INTRODUCTION TO RANDOM GRAPHS

ALAN FRIEZE and MICHAŁ KAROŃSKI

April 15, 2025

To Carol and Jola

Contents

Preface

Our purpose in writing this book is to provide a gentle introduction to a subject that is enjoying a surge in interest. We believe that the subject is fascinating in its own right, but the increase in interest can be attributed to several factors. One factor is the realization that networks are "everywhere". From social networks such as Facebook, the World Wide Web and the Internet to the complex interactions between proteins in the cells of our bodies, we face the challenge of understanding their structure and development. By and large natural networks grow in an unpredictable manner and this is often modeled by a random construction. Another factor is the realization by Computer Scientists that NP-hard problems are often easier to solve than their worst-case suggests and that an analysis of running times on random instances can be informative.

History

Random graphs were used by Erdős [330] to give a probabilistic construction of a graph with large girth and large chromatic number. It was only later that Erdős and Rényi began a systematic study of random graphs as objects of interest in their own right. Early on they defined the random graph $\mathbb{G}_{n,m}$ and founded the subject. Often neglected in this story is the contribution of Gilbert [433] who introduced the model $\mathbb{G}_{n,p}$, but clearly the credit for getting the subject off the ground goes to Erdős and Rényi. Their seminal series of papers [331], [333], [334], [335] and in particular [332], on the evolution of random graphs laid the groundwork for other mathematicians to become involved in studying properties of random graphs.

In the early eighties the subject was beginning to blossom and it received a boost from two sources. First was the publication of the landmark book of Béla Bollobás [155] on random graphs. Around the same time, the Discrete Mathematics group in Adam Mickiewicz University began a series of conferences in 1983. This series continues biennially to this day and is now a conference attracting more and more participants.

The next important event in the subject was the start of the journal Random Structures and Algorithms in 1990 followed by Combinatorics, Probability and

Computing a few years later. These journals provided a dedicated outlet for work in the area and are flourishing today.

Scope of the book

We have divided the book into four parts. Part one is devoted to giving a detailed description of the main properties of $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$. The aim is not to give best possible results, but instead to give some idea of the tools and techniques used in the subject, as well to display some of the basic results of the area. There is sufficient material in part one for a one semester course at the advanced undergraduate or beginning graduate level. Once one has finished the content of the first part, one is equipped to continue with material of the remainder of the book, as well as to tackle some of the advanced monographs such as Bollobás [155] and the more recent one by Janson, Łuczak and Ruciński [509].

Each chapter comes with a few exercises. Some are fairly simple and these are designed to give the reader practice with making some the estimations that are so prevalent in the subject. In addition each chapter ends with some notes that lead through references to some of the more advanced important results that have not been covered.

Part two deals with models of random graphs that naturally extend $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$. Part three deals with other models. Finally, in part four, we describe some of the main tools used in the area along with proofs of their validity.

Having read this book, the reader should be in a good position to pursue research in the area and we hope that this book will appeal to anyone interested in Combinatorics or Applied Probability or Theoretical Computer Science.

Acknowledgement

Several people have helped with the writing of this book and we would like to acknowledge their help. First there are the students who have sat in on courses based on early versions of this book and who helped to iron out the many typo's etc.

We would next like to thank the following people for reading parts of the book before final submission: Andrew Beveridge, Deepak Bal, Malgosia Bednarska, Patrick Bennett, Mindaugas Blozneliz, Antony Bonato, Boris Bukh, Fan Chung, Amin Coja-Oghlan, Colin Cooper, Andrzej Dudek, Asaf Ferber, Nikolas Fountoulakis, Catherine Greenhill, Dan Hefetz, Paul Horn, Hsien–Kuei Hwang, Tal Hershko, Jerzy Jaworski, Tony Johansson, Mihyun Kang, Michael Krivelevich, Tomasz Łuczak, Colin McDiarmid, Andrew McDowell, Hosam Mahmoud, Mike Molloy, Tobias Müller, Rajko Nenadov, Wesley Pegden, Huy Pham, Boris Pittel, Dan Poole, Pawel Prałat, Oliver Riordan, Andrzej Ruciński, Katarzyna Rybarczyk, Wojtek Samotij, Yilun Shang, Matas Šilekis, Greg Sorkin, Joel Spencer, Sam Spiro, Dudley Stark, Angelika Steger, Prasad Tetali, Andrew Thomason, Linnus Wästlund, Nick Wormald, Stephen Young.

Thanks also to Béla Bollobás for his advice on the structure of the book.

Conventions/Notation

Often in what follows, we will give an expression for a large positive integer. It might not be obvious that the expression is actually an integer. In which case, the reader can rest assured that he/she can round up or down and obtained any required property. We avoid this rounding for convenience and for notational purposes.

In addition we list the following notation:

Mathematical Relations

- $f(x) = O(g(x))$: $|f(x)| \leq K|g(x)|$ for some constant $K > 0$ and all $x \in \mathbf{R}$.
- $f(x) = \Theta(g(x))$: $f(n) = O(g(x))$ and $g(x) = O(f(x))$.
- $f(x) = \omega(g(x))$ if $g(x) = o(f(x))$.
- $f(x) = \Omega(g(x))$ if $f(x) \geq cg(x)$ for some positive constant $c.0$ and all $x \in \mathbf{R}$.
- $f(x) = o(g(x))$ as $x \to a$: $f(x)/g(x) \to 0$ as $x \to a$.
- $A \ll B$: $A/B \to 0$ as $n \to \infty$.
- $A \gg B$: $A/B \to \infty$ as $n \to \infty$.
- $A \approx B$: $A/B \to 1$ as some parameter converges to 0 or ∞ or another limit.
- $A \lesssim B$ or $B \gtrsim A$ if $A \leq (1+o(1))B$.
- $[n]$: This is $\{1,2,\ldots,n\}$. In general, if $a < b$ are positive integers, then $[a,b] = \{a,a+1,\ldots,b\}$.
- If S is a set and k is a non-negative integer then $\binom{S}{k}$ denotes the set of k-element subsets of S. In particular, $\binom{[n]}{k}$ dnotes the set of k-sets of $\{1,2,\ldots,n\}$. Furthermore, $\binom{S}{\leq k} = \bigcup_{j=0}^{k} \binom{S}{j}$.

Graph Notation

- $G = (V,E)$: $V = V(G)$ is the vertex set and $E = E(G)$ is the edge set.
- $e(G) = |E(G)|$ and for $S \subseteq V$ we have $e_G(S) = |\{e \in E : e \subseteq S\}|$.
- For $S,T \subseteq V, S \cap T = \emptyset$ we have $e_G(S:T) = \{e = \{x,y\} \in E : x \in S, y \in T\}$.
- $N(S) = N_G(S) = \{w \notin S : \exists v \in S \text{ such that } \{v,w\} \in E\}$ and $d_G(S) = |N_G(S)|$ for $S \subseteq V(G)$.
- $N_G(S,X) = N_G(S) \cap X$ for $X,S \subseteq V$.
- $\deg_S(x) = |\{y \in S : \{x,y\} \in E\}|$ for $x \in V, S \subseteq V$ and $\deg(v) = \deg_V(v)$.
- For sets $X,Y \subseteq V(G)$ we let $N_G(X,Y) = \{y \in Y : \exists x \in X, \{x,y\} \in E(G)\}$ and $e_G(X,Y) = |N_G(X,Y)|$.
- For a graph H, $aut(H)$ denotes the number of automorphisms of H.
- $\mathrm{dist}(\mathrm{v},\mathrm{w})$ denotes the graph distance between vertices v,w.
- The co-degree of vertices v,w of graph G is $N_G(v) \cap N_G(w)$.

Random Graph Models

- $[n]$: The set $\{1,2,\ldots,n\}$.
- $\mathscr{G}_{n,m}$: The family of all labeled graphs with vertex set $V = [n] = \{1,2,\ldots,n\}$ and exactly m edges.
- $\mathbb{G}_{n,m}$:A random graph chosen uniformly at random from $\mathscr{G}_{n,m}$.
- $E_{n,m} = E(\mathbb{G}_{n,m})$.
- $\mathbb{G}_{n,p}$: A random graph on vertex set $[n]$ where each possible edge occurs independently with probability p.
- $E_{n,p} = E(\mathbb{G}_{n,p})$.
- $\mathbb{G}_{n,m}^{\delta \geq k}$: $G_{n,m}$, conditioned on having minimum degree at least k.
- $\mathbb{G}_{n,n,p}$: A random bipartite graph with vertex set consisting of two disjoint copies of $[n]$ where each of the n^2 possible edges occurs independently with probability p.
- $\mathbb{G}_{n,r}$: A random r-regular graph on vertex set $[n]$.
- $\mathscr{G}_{n,\mathbf{d}}$: The set of graphs with vertex set $[n]$ and degree sequence $\mathbf{d} = (d_1,d_2,\ldots,d_n)$.

- $\mathbb{G}_{n,\mathbf{d}}$: A random graph chosen uniformly at random from $\mathscr{G}_{n,\mathbf{d}}$.
- $\mathbb{H}_{n,m;k}$: A random k-uniform hypergraph on vertex set $[n]$ and m edges of size k.
- $\mathbb{H}_{n,p;k}$: A random k-uniform hypergraph on vertex set $[n]$ where each of the $\binom{n}{k}$ possibles edge occurs independently with probability p.
- $\vec{\mathbb{G}}_{k-out}$: A random digraph on vertex set $[n]$ where each $v \in [n]$ independently chooses k random out-neighbors.
- $\mathbb{G}_{k-out}$: The graph obtained from $\vec{\mathbb{G}}_{k-out}$ by ignoring orientation and coalescing multiple edges.

Probability

- $\mathbb{P}(A)$: The probability of event A.
- $\mathbb{E}Z$: The expected value of random variable Z.
- $h(Z)$: The entropy of random variable Z.
- $Po(\lambda)$: A random variable with the Poisson distribution with mean λ.
- $N(0,1)$: A random variable with the normal distribution, mean 0 and variance 1.
- $\mathrm{Bin}(n,p)$: A random variable with the binomial distribution with parameters n, the number of trials and p, the probability of success.
- $EXP(\lambda)$: A random variable with the exponential distribution, mean λ i.e. $\mathbb{P}(EXP(\lambda) \geq x) = e^{-\lambda x}$. We sometimes say *rate* $1/\lambda$ in place of mean λ.
- w.h.p.: A sequence of events $\mathscr{A}_n, n = 1,2,\ldots,$ is said to occur *with high probability* (w.h.p.) if $\lim_{n\to\infty} \mathbb{P}(\mathscr{A}_n) = 1$.
- $\xrightarrow{d}$: We write $X_n \xrightarrow{d} X$ to say that a random variable X_n *converges in distribution* to a random variable X, as $n \to \infty$. Occasionally we write $X_n \xrightarrow{d} N(0,1)$ (resp. $X_n \xrightarrow{d} Po(\lambda)$) to mean that X has the corresponding normal (resp. Poisson) distribution.

Part I

Basic Models

Chapter 1

Random Graphs

Graph theory is a vast subject in which the goals are to relate various graph properties i.e. proving that Property A implies Property B for various properties A,B. In some sense, the goals of Random Graph theory are to prove results of the form "Property A almost always implies Property B". In many cases Property A could simply be "Graph G has m edges". A more interesting example would be the following: Property A is "G is an r-regular graph, $r \geq 3$" and Property B is "G is r-connected". This is proved in Chapter 9.

Before studying questions such as these, we will need to describe the basic models of a random graph.

1.1 Models and Relationships

The study of random graphs in their own right began in earnest with the seminal paper of Erdős and Rényi [332]. This paper was the first to exhibit the threshold phenomena that characterize the subject.

Let $\mathscr{G}_{n,m}$ be the family of all labeled graphs with vertex set $V = [n] = \{1, 2, \ldots, n\}$ and exactly m edges, $0 \leq m \leq \binom{n}{2}$. To every graph $G \in \mathscr{G}_{n,m}$, we assign a probability

$$\mathbb{P}(G) = \binom{\binom{n}{2}}{m}^{-1}.$$

Equivalently, we start with an empty graph on the set $[n]$, and insert m edges in such a way that all possible $\binom{\binom{n}{2}}{m}$ choices are equally likely. We denote such a random graph by $\mathbb{G}_{n,m} = ([n], E_{n,m})$ and call it a *uniform random graph*.

We now describe a similar model. Fix $0 \leq p \leq 1$. Then for $0 \leq m \leq \binom{n}{2}$, assign to each graph G with vertex set $[n]$ and m edges a probability

$$\mathbb{P}(G) = p^m (1-p)^{\binom{n}{2} - m},$$

where $0 \le p \le 1$. Equivalently, we start with an empty graph with vertex set $[n]$ and perform $\binom{n}{2}$ Bernoulli experiments inserting edges independently with probability p. We call such a random graph, a *binomial random graph* and denote it by $\mathbb{G}_{n,p} = ([n], E_{n,p})$. This was introduced by Gilbert [433]

As one may expect there is a close relationship between these two models of random graphs. We start with a simple observation.

Lemma 1.1. *A random graph $\mathbb{G}_{n,p}$, given that its number of edges is m, is equally likely to be one of the $\binom{\binom{n}{2}}{m}$ graphs that have m edges.*

Proof. Let G_0 be any labeled graph with m edges. Then since

$$\{\mathbb{G}_{n,p} = G_0\} \subseteq \{|E_{n,p}| = m\}$$

we have

$$\begin{aligned}
\mathbb{P}(\mathbb{G}_{n,p} = G_0 \mid |E_{n,p}| = m) &= \frac{\mathbb{P}(\mathbb{G}_{n,p} = G_0, |E_{n,p}| = m)}{\mathbb{P}(|E_{n,p}| = m)} \\
&= \frac{\mathbb{P}(\mathbb{G}_{n,p} = G_0)}{\mathbb{P}(|E_{n,p}| = m)} \\
&= \frac{p^m (1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}} \\
&= \binom{\binom{n}{2}}{m}^{-1}.
\end{aligned}$$

□

Thus $\mathbb{G}_{n,p}$ conditioned on the event $\{\mathbb{G}_{n,p}$ has m edges$\}$ is equal in distribution to $\mathbb{G}_{n,m}$, the graph chosen uniformly at random from all graphs with m edges. Obviously, the main difference between those two models of random graphs is that in $\mathbb{G}_{n,m}$ we choose its number of edges, while in the case of $\mathbb{G}_{n,p}$ the number of edges is the Binomial random variable with the parameters $\binom{n}{2}$ and p. Intuitively, for large n random graphs $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$ should behave in a similar fashion when the number of edges m in $\mathbb{G}_{n,m}$ equals or is "close" to the expected number of edges of $\mathbb{G}_{n,p}$, i.e., when

$$m = \binom{n}{2} p \approx \frac{n^2 p}{2}, \tag{1.1}$$

or, equivalently, when the edge probability in $\mathbb{G}_{n,p}$

$$p \approx \frac{2m}{n^2}. \tag{1.2}$$

Throughout the book, we will use the notation $f \approx g$ to indicate that $f = (1 + o(1))g$, where the $o(1)$ term will depend on some parameter going to 0 or ∞.

We next introduce a useful "coupling technique" that generates the random graph $\mathbb{G}_{n,p}$ in two independent steps. We will then describe a similar idea in relation to $\mathbb{G}_{n,m}$. Suppose that $p_1 < p$ and p_2 is defined by the equation

$$1 - p = (1 - p_1)(1 - p_2), \tag{1.3}$$

or, equivalently,

$$p = p_1 + p_2 - p_1 p_2.$$

Thus an edge is not included in $\mathbb{G}_{n,p}$ if it is not included in either of $\mathbb{G}_{n,p_1}$ or $\mathbb{G}_{n,p_2}$.

It follows that

$$\mathbb{G}_{n,p} = \mathbb{G}_{n,p_1} \cup \mathbb{G}_{n,p_2},$$

where the two graphs $\mathbb{G}_{n,p_1}, \mathbb{G}_{n,p_2}$ are independent. So when we write

$$\mathbb{G}_{n,p_1} \subseteq \mathbb{G}_{n,p},$$

we mean that the two graphs are *coupled* so that $\mathbb{G}_{n,p}$ is obtained from $\mathbb{G}_{n,p_1}$ by superimposing it with $\mathbb{G}_{n,p_2}$ and replacing eventual double edges by a single one.

We can also couple random graphs $\mathbb{G}_{n,m_1}$ and $\mathbb{G}_{n,m_2}$ where $m_2 \geq m_1$ via

$$\mathbb{G}_{n,m_2} = \mathbb{G}_{n,m_1} \cup \mathbb{H}.$$

Here $\mathbb{H}$ is the random graph on vertex set $[n]$ that has $m = m_2 - m_1$ edges chosen uniformly at random from $\binom{[n]}{2} \setminus E_{n,m_1}$.

Consider now a graph property $\mathscr{P}$ defined as a subset of the set of all labeled graphs on vertex set $[n]$, i.e., $\mathscr{P} \subseteq 2^{\binom{n}{2}}$. For example, all connected graphs (on n vertices), graphs with a Hamiltonian cycle, graphs containing a given subgraph, planar graphs, and graphs with a vertex of given degree form a specific "graph property".

We will state below two simple observations which show a general relationship between $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$ in the context of the probabilities of having a given graph property $\mathscr{P}$. The constant 10 in the next lemma is not best possible, but in the context of the usage of the lemma, any constant will suffice.

Lemma 1.2. *Let $\mathscr{P}$ be any graph property and $p = m/\binom{n}{2}$ where $m = m(n) \to \infty$, $\binom{n}{2} - m \to \infty$. Then, for large n,*

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq 10 m^{1/2}\, \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}).$$

Proof. By the law of total probability,

$$\begin{aligned}
\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) &= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P} \mid |E_{n,p}| = k)\,\mathbb{P}(|E_{n,p}| = k) \\
&= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P})\,\mathbb{P}(|E_{n,p}| = k) \qquad (1.4) \\
&\geq \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P})\,\mathbb{P}(|E_{n,p}| = m).
\end{aligned}$$

To justify (1.4), we write

$$\begin{aligned}
\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P} \mid |E_{n,p}| = k) &= \frac{\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P} \wedge |E_{n,p}| = k)}{\mathbb{P}(|E_{n,p}| = k)} \\
&= \sum_{\substack{G \in \mathscr{P} \\ |E(G)|=k}} \frac{p^k(1-p)^{N-k}}{\binom{N}{k} p^k (1-p)^{N-k}} \\
&= \sum_{\substack{G \in \mathscr{P} \\ |E(G)|=k}} \frac{1}{\binom{N}{k}} \\
&= \mathbb{P}(G_{n,k} \in \mathscr{P}).
\end{aligned}$$

Next recall that the number of edges $|E_{n,p}|$ of a random graph $\mathbb{G}_{n,p}$ is a random variable with the Binomial distribution with parameters $\binom{n}{2}$ and p. Applying Stirling's Formula:

$$k! = (1+o(1))\left(\frac{k}{e}\right)^k \sqrt{2\pi k}, \qquad (1.5)$$

and putting $N = \binom{n}{2}$, we get, after substituting (1.5) for the factorials in $\binom{N}{m}$,

$$\begin{aligned}
\mathbb{P}(|E_{n,p}| = m) &= \binom{N}{m} p^m (1-p)^{\binom{n}{2}-m} \\
&= (1+o(1))\,\frac{N^N\sqrt{2\pi N}\; p^m(1-p)^{N-m}}{m^m(N-m)^{N-m}\;2\pi\sqrt{m(N-m)}} \qquad (1.6) \\
&= (1+o(1))\sqrt{\frac{N}{2\pi m(N-m)}},
\end{aligned}$$

Hence

$$\mathbb{P}(|E_{n,p}| = m) \geq \frac{1}{10\sqrt{m}},$$

so

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq 10 m^{1/2}\,\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}).$$

□

We call a graph property $\mathscr{P}$ *monotone increasing* if $G \in \mathscr{P}$ implies $G+e \in \mathscr{P}$, i.e., adding an edge e to a graph G does not destroy the property. For example, connectivity and Hamiltonicity are monotone increasing properties. A monotone increasing property is *non-trivial* if the empty graph $\bar{K}_n \notin \mathscr{P}$ and the complete graph $K_n \in \mathscr{P}$.

A graph property is *monotone decreasing* if $G \in \mathscr{P}$ implies $G-e \in \mathscr{P}$, i.e., removing an edge from a graph does not destroy the property. Properties of a graph not being connected or being planar are examples of monotone decreasing graph properties. Obviously, a graph property $\mathscr{P}$ is monotone increasing if and only if its complement is monotone decreasing. Clearly not all graph properties are monotone. For example having at least half of the vertices having a given fixed degree d is not monotone.

From the coupling argument it follows that if $\mathscr{P}$ is a monotone increasing property then, whenever $p < p'$ or $m < m'$,

$$\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) \leq \mathbb{P}(\mathbb{G}_{n,p'} \in \mathscr{P}), \tag{1.7}$$

and

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq \mathbb{P}(\mathbb{G}_{n,m'} \in \mathscr{P}), \tag{1.8}$$

respectively.

For monotone increasing graph properties we can get a much better upper bound on $\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P})$, in terms of $\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P})$, than that given by Lemma 1.2.

Lemma 1.3. *Let $\mathscr{P}$ be a monotone increasing graph property and $p = \frac{m}{N}$. Then, for large n and $p = o(1)$ such that $Np, N(1-p)/(Np)^{1/2} \to \infty$,*

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq 3\,\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}).$$

Proof. Suppose $\mathscr{P}$ is monotone increasing and $p = \frac{m}{N}$, where $N = \binom{n}{2}$. Then

$$\begin{aligned}
\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) &= \sum_{k=0}^{N} \mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P})\,\mathbb{P}(|E_{n,p}| = k) \\
&\geq \sum_{k=m}^{N} \mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P})\,\mathbb{P}(|E_{n,p}| = k)
\end{aligned}$$

However, by the coupling property we know that for $k \geq m$,

$$\mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P}) \geq \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}).$$

The number of edges $|E_{n,p}|$ in $\mathbb{G}_{n,p}$ has the Binomial distribution with parameters N,p. Hence

$$\begin{aligned}\mathbb{P}(\mathbb{G}_{n,p}\in\mathscr{P})&\geq\mathbb{P}(\mathbb{G}_{n,m}\in\mathscr{P})\sum_{k=m}^{N}\mathbb{P}(|E_{n,p}|=k)\\&=\mathbb{P}(\mathbb{G}_{n,m}\in\mathscr{P})\sum_{k=m}^{N}u_k,\end{aligned}\tag{1.9}$$

where

$$u_k=\binom{N}{k}p^k(1-p)^{N-k}.$$

Now, using Stirling's formula,

$$u_m=(1+o(1))\frac{N^Np^m(1-p)^{N-m}}{m^m(N-m)^{N-m}(2\pi m)^{1/2}}=\frac{1+o(1)}{(2\pi m)^{1/2}}.$$

Furthermore, if $k=m+t$ where $0\leq t\leq m^{1/2}$ then

$$\frac{u_{k+1}}{u_k}=\frac{(N-k)p}{(k+1)(1-p)}=\frac{1-\frac{t}{N-m}}{1+\frac{t+1}{m}}\geq\exp\left\{-\frac{t}{N-m-t}-\frac{t+1}{m}\right\},$$

after using Lemma 27.1(a),(b) to obtain the inequality. and our assumptions on N,p to obtain the second.

It follows that for $0\leq t\leq m^{1/2}$,

$$u_{m+t}\geq\frac{1+o(1)}{(2\pi m)^{1/2}}\exp\left\{-\sum_{s=0}^{t-1}\left(\frac{s}{N-m-s}-\frac{s+1}{m}\right)\right\}\geq \frac{\exp\left\{-\frac{t^2}{2m}-o(1)\right\}}{(2\pi m)^{1/2}},$$

where we have used the fact that $m=o(N)$.

It follows that

$$\sum_{k=m}^{m+m^{1/2}}u_k\geq\frac{1-o(1)}{(2\pi)^{1/2}}\int_{x=0}^{1}e^{-x^2/2}dx\geq\frac{1}{3}$$

and the lemma follows from (1.9). □

Lemmas 1.2 and 1.3 are surprisingly applicable. In fact, since the $\mathbb{G}_{n,p}$ model is computationally easier to handle than $\mathbb{G}_{n,m}$, we will repeatedly use both lemmas to show that $\mathbb{P}(\mathbb{G}_{n,p}\in\mathscr{P})\to 0$ implies that $\mathbb{P}(\mathbb{G}_{n,m}\in\mathscr{P})\to 0$ when $n\to\infty$. In other situations we can use a stronger and more widely applicable result. The theorem below, which we state without proof, gives precise conditions for the asymptotic equivalence of random graphs $\mathbb{G}_{n,p}$ and $\mathbb{G}_{n,m}$. It is due to Łuczak [632].

Theorem 1.4. *Let* $0 \le p_0 \le 1$, $s(n) = n\sqrt{p(1-p)} \to \infty$, *and* $\omega(n) \to \infty$ *arbitrarily slowly as* $n \to \infty$.

(i) Suppose that $\mathscr{P}$ *is a graph property such that* $\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \to p_0$ *for all*

$$m \in \left[\binom{n}{2}p - \omega(n)s(n), \binom{n}{2}p + \omega(n)s(n)\right].$$

Then $\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) \to p_0$ *as* $n \to \infty$,

(ii) Let $p_- = p - \omega(n)s(n)/n^2$ *and* $p_+ = p + \omega(n)s(n)/n^2$ *Suppose that* $\mathscr{P}$ *is a monotone graph property such that* $\mathbb{P}(\mathbb{G}_{n,p_-} \in \mathscr{P}) \to p_0$ *and* $\mathbb{P}(\mathbb{G}_{n,p_+} \in \mathscr{P}) \to p_0$. *Then* $\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \to p_0$, *as* $n \to \infty$, *where* $m = \lfloor \binom{n}{2} p \rfloor$.

1.2 Thresholds and Sharp Thresholds

One of the most striking observations regarding the asymptotic properties of random graphs is the "abrupt" nature of the appearance and disappearance of certain graph properties. To be more precise in the description of this phenomenon, let us introduce *threshold functions* (or just *thresholds*) for monotone graph properties. We start by giving the formal definition of a threshold for a monotone increasing graph property $\mathscr{P}$.

Definition 1.5. A function $m^* = m^*(n)$ is a *threshold* for a monotone increasing property $\mathscr{P}$ in the random graph $\mathbb{G}_{n,m}$ if

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) = \begin{cases} 0 & \text{if } m/m^* \to 0, \\ 1 & \text{if } m/m^* \to \infty, \end{cases}$$

as $n \to \infty$.

A similar definition applies to the edge probability $p = p(n)$ in a random graph $\mathbb{G}_{n,p}$.

Definition 1.6. A function $p^* = p^*(n)$ is a *threshold* for a monotone increasing property $\mathscr{P}$ in the random graph $\mathbb{G}_{n,p}$ if

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) = \begin{cases} 0 & \text{if } p/p^* \to 0, \\ 1 & \text{if } p/p^* \to \infty, \end{cases}$$

as $n \to \infty$.

It is easy to see how to define thresholds for monotone decreasing graph properties and therefore we will leave this to the reader.

Notice also that the thresholds defined above are not unique since any function which differs from $m^*(n)$ (resp. $p^*(n)$) by a constant factor is also a threshold for $\mathscr{P}$.

A large body of the theory of random graphs is concerned with the search for thresholds for various properties, such as containing a path or cycle of a given length, or, in general, a copy of a given graph, or being connected or Hamiltonian, to name just a few. Therefore the next result is of special importance. It was proved by Bollobás and Thomason [178].

Theorem 1.7. *Every non-trivial monotone graph property has a threshold.*

Proof. Without loss of generality assume that $\mathscr{P}$ is a monotone increasing graph property. Given $0 < \varepsilon < 1$ we define $p(\varepsilon)$ by

$$\mathbb{P}(\mathbb{G}_{n,p(\varepsilon)} \in \mathscr{P}) = \varepsilon.$$

Note that $p(\varepsilon)$ exists because

$$\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) = \sum_{G\in\mathscr{P}} p^{|E(G)|}(1-p)^{N-|E(G|}$$

is a polynomial in p that increases from 0 to 1. This is not obvious from the expression, but it is obvious from the fact that $\mathscr{P}$ is monotone increasing and that increasing p increases the likelihood that $\mathbb{G}_{n,p} \in \mathscr{P}$.

We will show that $p^* = p(1/2)$ is a threshold for $\mathscr{P}$. Let $G_1, G_2, \ldots, G_k$ be independent copies of $\mathbb{G}_{n,p}$. The graph $G_1 \cup G_2 \cup \ldots \cup G_k$ is distributed as $\mathbb{G}_{n,1-(1-p)^k}$. Now $1-(1-p)^k \le kp$, and therefore by the coupling argument

$$\mathbb{G}_{n,1-(1-p)^k} \subseteq \mathbb{G}_{n,kp},$$

and so $\mathbb{G}_{n,kp} \notin \mathscr{P}$ implies $G_1, G_2, \ldots, G_k \notin \mathscr{P}$. Hence

$$\mathbb{P}(\mathbb{G}_{n,kp} \notin \mathscr{P}) \le [\mathbb{P}(\mathbb{G}_{n,p} \notin \mathscr{P})]^k.$$

Let ω be a function of n such that $\omega \to \infty$ arbitrarily slowly as $n \to \infty$, $\omega \ll \log\log n$. (We say that $f(n) \ll g(n)$ or $f(n) = o(g(n))$ if $f(n)/g(n) \to 0$ as $n \to \infty$. Of course in this case we can also write $g(n) \gg f(n)$.) Suppose also that $p = p^* = p(1/2)$ and $k = \omega$. Then

$$\mathbb{P}(\mathbb{G}_{n,\omega p^*} \notin \mathscr{P}) \leq 2^{-\omega} = o(1).$$

On the other hand for $p = p^*/\omega$,

$$\frac{1}{2} = \mathbb{P}(\mathbb{G}_{n,p^*} \notin \mathscr{P}) \leq \left[\mathbb{P}(\mathbb{G}_{n,p^*/\omega} \notin \mathscr{P})\right]^{\omega}.$$

So

$$\mathbb{P}(\mathbb{G}_{n,p^*/\omega} \notin \mathscr{P}) \geq 2^{-1/\omega} = 1 - o(1).$$

$\square$

In order to shorten many statements of theorems in the book we say that a sequence of events $\mathscr{E}_n$ occurs *with high probability* (w.h.p.) if

$$\lim_{n\to\infty} \mathbb{P}(\mathscr{E}_n) = 1.$$

Thus the statement that says p^* is a threshold for a property $\mathscr{P}$ in $\mathbb{G}_{n,p}$ is the same as saying that $\mathbb{G}_{n,p} \notin \mathscr{P}$ w.h.p. if $p \ll p^*$, while $\mathbb{G}_{n,p} \in \mathscr{P}$ w.h.p. if $p \gg p^*$.

In many situations we can observe that for some monotone graph properties more "subtle" thresholds hold. We call them "*sharp thresholds*". More precisely,

Definition 1.8. A function $m^* = m^*(n)$ is a *sharp threshold* for a monotone increasing property $\mathscr{P}$ in the random graph $\mathbb{G}_{n,m}$ if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) = \begin{cases} 0 & if \quad m/m^* \leq 1-\varepsilon \\ 1 & if \quad m/m^* \geq 1+\varepsilon. \end{cases}$$

A similar definition applies to the edge probability $p = p(n)$ in the random graph $\mathbb{G}_{n,p}$.

Definition 1.9. A function $p^* = p^*(n)$ is a *sharp threshold* for a monotone increasing property $\mathscr{P}$ in the random graph $\mathbb{G}_{n,p}$ if for every $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) = \begin{cases} 0 & if \quad p/p^* \leq 1-\varepsilon \\ 1 & if \quad p/p^* \geq 1+\varepsilon. \end{cases}$$

We will illustrate both types of threshold in a series of examples dealing with very simple graph properties. Our goal at the moment is to demonstrate basic

techniques to determine thresholds rather than to "discover" some "striking" facts about random graphs.
We will start with the random graph $\mathbb{G}_{n,p}$ and the property

$$\mathscr{P} = \{\text{all non-empty (non-edgeless) labeled graphs on } n \text{ vertices}\}.$$

This simple graph property is clearly monotone increasing and we will show below that $p^* = 1/n^2$ is a threshold for a random graph $\mathbb{G}_{n,p}$ of having at least one edge (being non-empty).

Lemma 1.10. *Let $\mathscr{P}$ be the property defined above, i.e., stating that $\mathbb{G}_{n,p}$ contains at least one edge. Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) = \begin{cases} 0 & \text{if } p \ll n^{-2} \\ 1 & \text{if } p \gg n^{-2}. \end{cases}$$

Proof. Let X be a random variable counting edges in $\mathbb{G}_{n,p}$. Since X has the Binomial distribution, then $\mathbb{E}X = \binom{n}{2}p$, and $\mathrm{Var}X = \binom{n}{2}p(1-p) = (1-p)\mathbb{E}X$.

A standard way to show the first part of the threshold statement, i.e. that w.h.p. a random graph $\mathbb{G}_{n,p}$ is empty when $p = o(n^{-2})$, is a very simple consequence of the Markov inequality, called the First Moment Method, see Lemma 26.2. It states that if X is a non-negative integer valued random variable, then

$$\mathbb{P}(X > 0) \le \mathbb{E}X.$$

Hence, in our case

$$\mathbb{P}(X > 0) \le \frac{n^2}{2} p \to 0$$

as $n \to \infty$, since $p \ll n^{-2}$.

On the other hand, if we want to show that $\mathbb{P}(X > 0) \to 1$ as $n \to \infty$ then we cannot use the First Moment Method and we should use the Second Moment Method, which is a simple consequence of the Chebyshev inequality, see Lemma 26.3. We will use the inequality to show *concentration around the mean*. By this we mean that w.h.p. $X \approx \mathbb{E}X$. The Chebyshev inequality states that if X is a non-negative integer valued random variable then

$$\mathbb{P}(X > 0) \ge 1 - \frac{\mathrm{Var}X}{(\mathbb{E}X)^2}.$$

Hence $\mathbb{P}(X > 0) \to 1$ as $n \to \infty$ whenever $\mathrm{Var}X/(\mathbb{E}X)^2 \to 0$ as $n \to \infty$. (For proofs of both of the above Lemmas see Section 26.1 of Chapter 26.)

Now, if $p \gg n^{-2}$ then $\mathbb{E}X \to \infty$ and therefore

$$\frac{\operatorname{Var} X}{(\mathbb{E}X)^2} = \frac{1-p}{\mathbb{E}X} \to 0$$

as $n \to \infty$, which shows that the second statement of Lemma 1.10 holds, and so $p^* = 1/n^2$ is a threshold for the property of $\mathbb{G}_{n,p}$ being non-empty. □

Let us now look at the degree of a fixed vertex in both models of random graphs. One immediately notices that if $\deg(v)$ denotes the degree of a fixed vertex in $\mathbb{G}_{n,p}$, then $\deg(v)$ is a binomially distributed random variable, with parameters $n-1$ and p, i.e., for $d = 0, 1, 2 \ldots, n-1$,

$$\mathbb{P}(\deg(v) = d) = \binom{n-1}{d} p^d (1-p)^{n-1-d},$$

while in $\mathbb{G}_{n,m}$ the distribution of $\deg(v)$ is Hypergeometric, i.e.,

$$\mathbb{P}(\deg(v) = d) = \frac{\binom{n-1}{d}\binom{\binom{n-1}{2}}{m-d}}{\binom{\binom{n}{2}}{m}}.$$

Consider the monotone decreasing graph property that a graph contains an isolated vertex, i.e. a vertex of degree zero:

$$\mathscr{P} = \{\text{all labeled graphs on } n \text{ vertices containing isolated vertices}\}.$$

We will show that $m^* = \frac{1}{2} n \log n$ is the sharp threshold function for the above property $\mathscr{P}$ in $\mathbb{G}_{n,m}$.

Lemma 1.11. *Let $\mathscr{P}$ be the property that a graph on n vertices contains at least one isolated vertex and let $m = \frac{1}{2}n(\log n + \omega(n))$. Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) = \begin{cases} 1 & \text{if } \omega(n) \to -\infty \\ 0 & \text{if } \omega(n) \to \infty. \end{cases}$$

Proof. To see that the second statement of Lemma 1.11 holds we use the First Moment Method. Namely, let $X_0 = X_{n,0}$ be the number of isolated vertices in the random graph $\mathbb{G}_{n,m}$. Then X_0 can be represented as the sum of indicator random variables

$$X_0 = \sum_{v \in V} I_v,$$

where

$$I_v = \begin{cases} 1 & \text{if } v \text{ is an isolated vertex in } \mathbb{G}_{n,m} \\ 0 & \text{otherwise.} \end{cases}$$

So

$$
\begin{aligned}
\mathbb{E} X_0 = \sum_{v \in V} \mathbb{E} I_v = n \frac{\binom{\binom{n-1}{2}}{m}}{\binom{\binom{n}{2}}{m}} = \\
n\left(\frac{n-2}{n}\right)^m \prod_{i=0}^{m-1}\left(1-\frac{4i}{n(n-1)(n-2)-2i(n-2)}\right) = \\
n\left(\frac{n-2}{n}\right)^m\left(1+O\left(\frac{(\log n)^2}{n}\right)\right), \quad (1.10)
\end{aligned}
$$

assuming that $\omega = o(\log n)$.
(For the product we use $1 \geq \prod_{i=0}^{m-1}(1-x_i) \geq 1-\sum_{i=0}^{m-1} x_i$ which is valid for all $0 \leq x_0, x_1, \ldots, x_{m-1} \leq 1$.)

Hence,

$$
\mathbb{E} X_0 \leq n\left(\frac{n-2}{n}\right)^m \leq n e^{-\frac{2m}{n}} = e^{-\omega},
$$

for $m = \frac{1}{2} n(\log n + \omega(n))$.
($1 + x \leq e^x$ is one of the basic inequalities stated in Lemma 27.1.)

So $\mathbb{E} X_0 \to 0$ when $\omega(n) \to \infty$ as $n \to \infty$ and the First Moment Method implies that $X_0 = 0$ w.h.p.

To show that Lemma 1.11 holds in the case when $\omega \to -\infty$ we first observe from (1.10) that in this case

$$
\begin{aligned}
\mathbb{E} X_0 &= (1-o(1)) n\left(\frac{n-2}{n}\right)^m \\
&\geq (1-o(1)) n \exp\left\{-\frac{2m}{n-2}\right\} \\
&\geq (1-o(1)) e^{-\omega} \to \infty, \quad (1.11)
\end{aligned}
$$

The second inequality in the above comes from Lemma 27.1(b), and we have once again assumed that $\omega = o(\log n)$ to justify the first equation.

We caution the reader that $\mathbb{E} X_0 \to \infty$ does not prove that $X_0 > 0$ w.h.p. In Chapter 5 we will see an example of a random variable X_H, where $\mathbb{E} X_H \to \infty$ and yet $X_H = 0$ w.h.p.

We will now use a stronger version of the Second Moment Method (for its proof see Section 26.1 of Chapter 26). It states that if X is a non-negative integer valued random variable then

$$
\mathbb{P}(X > 0) \geq \frac{(\mathbb{E} X)^2}{\mathbb{E} X^2} = 1 - \frac{\operatorname{Var} X}{\mathbb{E} X^2}. \quad (1.12)
$$

Notice that

$$\begin{aligned}
\mathbb{E}X_0^2 &= \mathbb{E}\left(\sum_{v\in V} I_v\right)^2 = \sum_{u,v\in V} \mathbb{E}(I_u I_v) \\
&= \sum_{u,v\in V} \mathbb{P}(I_u = 1, I_v = 1) \\
&= \sum_{u\neq v} \mathbb{P}(I_u = 1, I_v = 1) + \sum_{u=v} \mathbb{P}(I_u = 1, I_v = 1) \\
&= n(n-1)\frac{\binom{\binom{n-2}{2}}{m}}{\binom{\binom{n}{2}}{m}} + \mathbb{E}X_0 \\
&\leq n^2\left(\frac{n-2}{n}\right)^{2m} + \mathbb{E}X_0 \\
&= (1+o(1))(\mathbb{E}X_0)^2 + \mathbb{E}X_0.
\end{aligned}$$

The last equation follows from (1.10).

Hence, by (1.12),

$$\begin{aligned}
\mathbb{P}(X_0 > 0) &\geq \frac{(\mathbb{E}X_0)^2}{\mathbb{E}X_0^2} \\
&\geq \frac{(\mathbb{E}X_0)^2}{(1+o(1))((\mathbb{E}X_0)^2 + \mathbb{E}X_0)} \\
&= \frac{1}{(1+o(1)+(\mathbb{E}X_0)^{-1}} \\
&= 1 - o(1),
\end{aligned}$$

on using (1.11). Hence $\mathbb{P}(X_0 > 0) \to 1$ when $\omega(n) \to -\infty$ as $n \to \infty$, and so we can conclude that $m = m(n)$ is the sharp threshold for the property that $\mathbb{G}_{n,m}$ contains isolated vertices. □

For this simple random variable, we worked with $\mathbb{G}_{n,m}$. We will in general work with the more congenial independent model $\mathbb{G}_{n,p}$ and translate the results to $G_{n,m}$ if so desired.

For another simple example of the use of the second moment method, we will prove

Theorem 1.12. *If $m/n \to \infty$ then w.h.p. $\mathbb{G}_{n,m}$ contains at least one triangle.*

Proof. Because having a triangle is a monotone increasing property we can prove the result in $\mathbb{G}_{n,p}$ assuming that $np \to \infty$.

Assume first that $np = \omega \leq \log n$ where $\omega = \omega(n) \to \infty$ and let Z be the number of triangles in $\mathbb{G}_{n,p}$. Then

$$\mathbb{E}Z = \binom{n}{3} p^3 \geq (1 - o(1))\frac{\omega^3}{6} \to \infty.$$

We remind the reader that simply having $\mathbb{E}Z \to \infty$ is not sufficient to prove that $Z > 0$ w.h.p.

Next let $T_1, T_2, \ldots, T_M, M = \binom{n}{3}$ denote the triangles of K_n. Then

$$\begin{aligned}
\mathbb{E}Z^2 &= \sum_{i,j=1}^{M} \mathbb{P}(T_i, T_j \in \mathbb{G}_{n,p}) \\
&= \sum_{i=1}^{M} \mathbb{P}(T_i \in \mathbb{G}_{n,p}) \sum_{j=1}^{M} \mathbb{P}(T_j \in \mathbb{G}_{n,p} \mid T_i \in \mathbb{G}_{n,p}) && (1.13) \\
&= M\,\mathbb{P}(T_1 \in \mathbb{G}_{n,p}) \sum_{j=1}^{M} \mathbb{P}(T_j \in \mathbb{G}_{n,p} \mid T_1 \in \mathbb{G}_{n,p}) && (1.14) \\
&= \mathbb{E}Z \times \sum_{j=1}^{M} \mathbb{P}(T_j \in \mathbb{G}_{n,p} \mid T_1 \in \mathbb{G}_{n,p}).
\end{aligned}$$

Here (1.14) follows from (1.13) by symmetry.

Now suppose that T_j, T_1 share σ_j edges. Then

$$\begin{aligned}
&\sum_{j=1}^{M} \mathbb{P}(T_j \in \mathbb{G}_{n,p} \mid T_1 \in \mathbb{G}_{n,p}) \\
&= 1 + \sum_{j:\sigma_j=1} \mathbb{P}(T_j \in \mathbb{G}_{n,p} \mid T_1 \in \mathbb{G}_{n,p}) + \\
&\qquad\qquad \sum_{j:\sigma_j=0} \mathbb{P}(T_j \in \mathbb{G}_{n,p} \mid T_1 \in \mathbb{G}_{n,p}) \\
&= 1 + 3(n-3)p^2 + \left(\binom{n}{3} - 3n + 8\right) p^3 \\
&\leq 1 + \frac{3\omega^2}{n} + \mathbb{E}Z.
\end{aligned}$$

It follows that

$$\operatorname{Var} Z \leq (\mathbb{E}Z)\left(1 + \frac{3\omega^2}{n} + \mathbb{E}Z\right) - (\mathbb{E}Z)^2 \leq 2\,\mathbb{E}Z.$$

Applying the Chebyshev inequality we get

$$\mathbb{P}(Z = 0) \leq \mathbb{P}(|Z - \mathbb{E}Z| \geq \mathbb{E}Z) \leq \frac{\operatorname{Var} Z}{(\mathbb{E}Z)^2} \leq \frac{2}{\mathbb{E}Z} = o(1).$$

This proves the theorem for $p \leq \frac{\log n}{n}$. For larger p we can use (1.7). □

We can in fact use the second moment method to show that if $m/n \to \infty$ then w.h.p. $\mathbb{G}_{n,m}$ contains a copy of a k-cycle C_k for any fixed $k \geq 3$. See Theorem 5.3, see also Exercise 1.4.7.

1.3 Pseudo-Graphs

We sometimes use one of the two the following models that are related to $\mathbb{G}_{n,m}$ and have a little more independence. (We will use Model A in Section 7.3 and Model B in Section 6.4).

Model A: We let $\mathbf{x} = (x_1, x_2, \ldots, x_{2m})$ be chosen uniformly at random from $[n]^{2m}$.

Model B: We let $\mathbf{x} = (x_1, x_2, \ldots, x_{2m})$ be chosen uniformly at random from $\binom{[n]}{2}^m$.

The (multi-)graph $\mathbb{G}_{n,m}^{(X)}, X \in \{A,B\}$ has vertex set $[n]$ and edge set $E_m = \{\{x_{2i-1}, x_{2i}\} : 1 \leq i \leq m\}$. Basically, we are choosing edges with replacement. In Model A we allow loops and in Model B we do not. We get simple graphs from by removing loops and multiple edges to obtain graphs $\mathbb{G}_{n,m}^{(X*)}$ with m^* edges. It is not difficult to see that for $X \in \{A,B\}$ and conditional on the value of m^* that $\mathbb{G}_{n,m}^{(X*)}$ is distributed as $\mathbb{G}_{n,m^*}$, see Exercise (1.4.11).

More importantly, we have that for $G_1, G_2 \in \mathscr{G}_{n,m}$,

$$\mathbb{P}(\mathbb{G}_{n,m}^{(X)} = G_1 \mid \mathbb{G}_{n,m}^{(X)} \text{ is simple}) = \mathbb{P}(\mathbb{G}_{n,m}^{(X)} = G_2 \mid \mathbb{G}_{n,m}^{(X)} \text{ is simple}), \tag{1.15}$$

for $X = A, B$.

This is because for $i = 1, 2$,

$$\mathbb{P}(\mathbb{G}_{n,m}^{(A)} = G_i) = \frac{m!2^m}{n^{2m}} \text{ and } \mathbb{P}(\mathbb{G}_{n,m}^{(B)} = G_i) = \frac{m!2^m}{\binom{n}{2}^m 2^m}.$$

Indeed, we can permute the edges in $m!$ ways and permute the vertices within edges in 2^m ways without changing the underlying graph. This relies on $\mathbb{G}_{n,m}^{(X)}$ being simple.

Secondly, if $m = cn$ for a constant $c > 0$ then with $N = \binom{n}{2}$, and using Lemma 27.2,

$$\begin{aligned}\mathbb{P}(\mathbb{G}_{n,m}^{(X)} \text{ is simple}) \geq \binom{N}{m} \frac{m!2^m}{n^{2m}} \geq \\ (1 - o(1)) \frac{N^m}{m!} \exp\left\{ -\frac{m^2}{2N} - \frac{m^3}{6N^2} \right\} \frac{m!2^m}{n^{2m}} \\ = (1 - o(1)) e^{-(c^2+c)}. \end{aligned} \tag{1.16}$$

It follows that if $\mathscr{P}$ is some graph property then

$$\mathbb{P}(G_{n,m} \in \mathscr{P}) = \mathbb{P}(\mathbb{G}_{n,m}^{(X)} \in \mathscr{P} \mid \mathbb{G}_{n,m}^{(X)} \text{ is simple}) \leq$$

$$(1+o(1))e^{c^2+c}\,\mathbb{P}(\mathbb{G}_{n,m}^{(X)} \in \mathscr{P}). \quad (1.17)$$

Here we have used the inequality $\mathbb{P}(A \mid B) \leq \mathbb{P}(A)/\mathbb{P}(B)$ for events A, B.

We will use this model a couple of times and (1.17) shows that if $\mathbb{P}(\mathbb{G}_{n,m}^{(X)} \in \mathscr{P}) = o(1)$ then $\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) = o(1)$, for $m = O(n)$.

Model $\mathbb{G}_{n,m}^{(A)}$ was introduced independently by Bollobás and Frieze [166] and by Chvátal [229].

1.4 Exercises

We point out here that in the following exercises, we have not asked for best possible results. These exercises are for practise. You will need to use the inequalities from Section 27.1.

1.4.1 Suppose that $p = d/n$ where $d = o(n^{1/3})$. Show that w.h.p. $G_{n,p}$ has no copies of K_4.

1.4.2 Suppose that $p = d/n$ where $d > 1$. Show that w.h.p. $G_{n,p}$ contains an *induced* path of length $(\log n)^{1/2}$.

1.4.3 Suppose that $p = d/n$ where $d = O(1)$. Prove that w.h.p., in $G_{n,p}$, for all $S \subseteq [n], |S| \leq n/\log n$, we have $e(S) \leq 2|S|$, where $e(S)$ is the number of edges contained in S.

1.4.4 Suppose that $p = \log n/n$. Let a vertex of $G_{n,p}$ be small if its degree is less than $\log n/100$. Show that w.h.p. there is no edge of $G_{n,p}$ joining two small vertices.

1.4.5 Suppose that $p = d/n$ where d is constant. Prove that w.h.p., in $G_{n,p}$, no vertex belongs to more than one triangle.

1.4.6 Suppose that $p = d/n$ where d is constant. Prove that w.h.p. $G_{n,p}$ contains a vertex of degree exactly $\left\lceil (\log n)^{1/2} \right\rceil$.

1.4.7 Suppose that $k \geq 3$ is constant and that $np \to \infty$. Show that w.h.p. $\mathbb{G}_{n,p}$ contains a copy of the k-cycle, C_k.

1.4.8 Suppose that $0 < p < 1$ is constant. Show that w.h.p. $G_{n,p}$ has diameter two.

1.4.9 Let $f : [n] \to [n]$ be chosen uniformly at random from all n^n functions from $[n] \to [n]$. Let $X = \{j : \nexists i \text{ s.t. } f(i) = j\}$. Show that w.h.p. $|X| \approx e^{-1}n$.

1.4.10 Prove Theorem 1.4.

1.4.11 Show that conditional on the value of m^{X*} that $\mathbb{G}_{n,m}^{X*}$ is distributed as $\mathbb{G}_{n,m^*}$, where $X = A, B$.

1.5 Notes

Friedgut and Kalai [378] and Friedgut [379] and Bourgain [185] and Bourgain and Kalai [184] provide much greater insight into the notion of sharp thresholds. Friedgut [377] gives a survey of these aspects. For a graph property $\mathscr{A}$ let $\mu(p, \mathscr{A})$ be the probability that the random graph $\mathbb{G}_{n,p}$ has property $\mathscr{A}$. A threshold is *coarse* if it is not sharp. We can identify coarse thresholds with $p\frac{d\mu(p,\mathscr{A})}{dp} < C$ for some absolute constant $0 < C$. The main insight into coarse thresholds is that to exist, the occurrence of $\mathscr{A}$ can in the main be attributed to the existence of one of a bounded number of small subgraphs. For example, Theorem 2.1 of [377] states that there exists a function $K(C, \varepsilon)$ such that the following holds. Let $\mathscr{A}$ be a monotone property of graphs that is invariant under automorphism and assume that $p\frac{d\mu(p,\mathscr{A})}{dp} < C$ for some constant $0 < C$. Then for every $\varepsilon > 0$ there exists a finite list of graphs $G_1, G_2, \ldots, G_m$ all of which have no more than $K(\varepsilon, C)$ edges, such that if $\mathscr{B}$ is the family of graphs having one of these graphs as a subgraph then $\mu(p, \mathscr{A}\Delta\mathscr{B}) \leq \varepsilon$.

Chapter 2

Evolution

Here begins our story of the typical growth of a random graph. All the results up to Section 2.3 were first proved in a landmark paper by Erdős and Rényi [332]. The notion of the *evolution* of a random graph stems from a dynamic view of a *graph process*: viz. a sequence of graphs:

$$\mathbb{G}_0 = ([n], \emptyset), \mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_m, \ldots, \mathbb{G}_N = K_n.$$

where $\mathbb{G}_{m+1}$ is obtained from $\mathbb{G}_m$ by adding a random edge e_m. We see that there are $\binom{n}{2}!$ such sequences and $\mathbb{G}_m$ and $\mathbb{G}_{n,m}$ have the same distribution.
In process of the evolution of a random graph we consider properties possessed by $\mathbb{G}_m$ or $\mathbb{G}_{n,m}$ w.h.p., when $m = m(n)$ grows from 0 to $\binom{n}{2}$, while in the case of $\mathbb{G}_{n,p}$ we analyse its typical structure when $p = p(n)$ grows from 0 to 1 as $n \to \infty$.

In the current chapter we mainly explore how the typical component structure evolves as the number of edges m increases.

2.1 Sub-Critical Phase

The evolution of Erdős-Rényi type random graphs has clearly distinguishable phases. The first phase, at the beginning of the evolution, can be described as a period when a random graph is a collection of small components which are mostly trees. Indeed the first result in this section shows that a random graph $\mathbb{G}_{n,m}$ is w.h.p. a collection of tree-components as long as $m = o(n)$, or, equivalently, as long as $p = o(n^{-1})$ in $\mathbb{G}_{n,p}$. For clarity, all results presented in this chapter are stated in terms of $\mathbb{G}_{n,m}$. Due to the fact that computations are much easier for $\mathbb{G}_{n,p}$ we will first prove results in this model and then the results for $\mathbb{G}_{n,m}$ will follow by the equivalence established either in Lemmas 1.2 and 1.3 or in Theorem 1.4. We will also assume, throughout this chapter, that $\omega = \omega(n)$ is a function growing slowly with n, e.g. $\omega = \log\log n$ will suffice.

Theorem 2.1. *If $m \ll n$, then $\mathbb{G}_m$ is a forest w.h.p.*

Proof. Suppose $m = n/\omega$ and let $N = \binom{n}{2}$, so $p = m/N \leq 3/(\omega n)$. Let X be the number of cycles in $\mathbb{G}_{n,p}$. Then

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p^k \\
&\leq \sum_{k=3}^{n} \frac{n^k}{k!} \frac{(k-1)!}{2} p^k \\
&\leq \sum_{k=3}^{n} \frac{n^k}{2k} \frac{3^k}{\omega^k n^k} \\
&= O(\omega^{-3}) \to 0.
\end{aligned}$$

Therefore, by the First Moment Method, (see Lemma 26.2),

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is not a forest}) = \mathbb{P}(X \geq 1) \leq \mathbb{E}X = o(1),$$

which implies that

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is a forest}) \to 1 \text{ as } n \to \infty.$$

Notice that the property that a graph is a forest is monotone decreasing, so by Lemma 1.3

$$\mathbb{P}(\mathbb{G}_m \text{ is a forest}) \to 1 \text{ as } n \to \infty.$$

(Note that we have actually used Lemma 1.3 to show that $\mathbb{P}(\mathbb{G}_{n,p}$ is not a forest)=$o(1)$ implies that $\mathbb{P}(G_m$ is not a forest)=$o(1)$.) □

We will next examine the time during which the components of G_m are isolated vertices and single edges only, w.h.p.

Theorem 2.2. *If $m \ll n^{1/2}$ then $\mathbb{G}_m$ is the union of isolated vertices and edges w.h.p.*

Proof. Let $p = m/N$, $m = n^{1/2}/\omega$ and let X be the number of paths of length two in the random graph $\mathbb{G}_{n,p}$. By the First Moment Method,

$$\mathbb{P}(X > 0) \leq \mathbb{E}X = 3\binom{n}{3} p^2 \leq \frac{n^4}{2N^2\omega^2} \to 0,$$

as $n \to \infty$. Hence

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ contains a path of length two}) = o(1).$$

Notice that the property that a graph contains a path of a given length two is monotone increasing, so by Lemma 1.3,

$$\mathbb{P}(\mathbb{G}_m \text{ contains a path of length two}) = o(1),$$

and the theorem follows. □

Now we are ready to describe the next step in the evolution of $\mathbb{G}_m$.

Theorem 2.3. *If $m \gg n^{1/2}$, then $\mathbb{G}_m$ contains a path of length two w.h.p.*

Proof. Let $p = \frac{m}{N}, m = \omega n^{1/2}$ and X be the number of paths of length two in $\mathbb{G}_{n,p}$. Then

$$\mathbb{E}X = 3\binom{n}{3}p^2 \approx 2\omega^2 \to \infty,$$

as $n \to \infty$. This however does not imply that $X > 0$ w.h.p.! To show that $X > 0$ w.h.p. we will apply the Second Moment Method

Let $\mathscr{P}_2$ be the set of all paths of length two in the complete graph K_n, and let $\hat{X}$ be the number of *isolated* paths of length two in $\mathbb{G}_{n,p}$ i.e. paths that are also components of $\mathbb{G}_{n,p}$. We will show that w.h.p. $\mathbb{G}_{n,p}$ contains such an isolated path. Now,

$$\hat{X} = \sum_{P \in \mathscr{P}_2} I_{P \subseteq_i \mathbb{G}_{n,p}}.$$

We always use $I_{\mathscr{E}}$ to denote the indicator for an event $\mathscr{E}$. The notation $\subseteq_i$ indicates that P is contained in $\mathbb{G}_{n,p}$ as a component (i.e. P is isolated). Having a path of length two is a monotone increasing property. Therefore we can assume that $m = o(n)$ and so $np = o(1)$ and the result for larger m will follow from monotonicity and coupling. Then

$$\begin{aligned}\mathbb{E}\hat{X} &= 3\binom{n}{3}p^2(1-p)^{3(n-3)+1} \\ &\geq (1-o(1))\frac{n^3}{2}\frac{4\omega^2 n}{n^4}(1-3np) \to \infty,\end{aligned}$$

as $n \to \infty$.

In order to compute the second moment of the random variable $\hat{X}$ notice that,

$$\hat{X}^2 = \sum_{P \in \mathscr{P}_2} \sum_{Q \in \mathscr{P}_2} I_{P \subseteq_i \mathbb{G}_{n,p}} I_{Q \subseteq_i \mathbb{G}_{n,p}} = \sum\nolimits^*_{P,Q \in \mathscr{P}_2} I_{P \subseteq_i \mathbb{G}_{n,p}} I_{Q \subseteq_i \mathbb{G}_{n,p}},$$

where the last sum is taken over $P, Q \in \mathscr{P}_2$ such that either $P = Q$ or P and Q are vertex disjoint. The simplification that provides the last summation is precisely the reason that we introduce path-components (isolated paths). Now

$$\mathbb{E}\hat{X}^2 = \sum_P \left\{ \sum_Q \mathbb{P}(Q \subseteq_i \mathbb{G}_{n,p} | \, P \subseteq_i \mathbb{G}_{n,p}) \right\} \mathbb{P}(P \subseteq_i \mathbb{G}_{n,p}).$$

The expression inside the brackets is the same for all P and so

$$\mathbb{E}\hat{X}^2 = \mathbb{E}\hat{X} \left(1 + \sum_{Q \cap P_{(1,2,3)} = \emptyset} \mathbb{P}(Q \subseteq_i \mathbb{G}_{n,p} | \, P_{(1,2,3)} \subseteq_i \mathbb{G}_{n,p}) \right),$$

where $P_{\{1,2,3\}}$ denotes the path on vertex set $[3] = \{1,2,3\}$ with middle vertex 2. By conditioning on the event $P_{(1,2,3)} \subseteq_i \mathbb{G}_{n,p}$, i.e, assuming that $P_{(1,2,3)}$ is a component of $\mathbb{G}_{n,p}$, we see that all of the nine edges between Q and $P_{(1,2,3)}$ must be missing. Therefore

$$\mathbb{E}\hat{X}^2 \le \mathbb{E}\hat{X} \left(1 + 3\binom{n}{3} p^2 (1-p)^{3(n-6)+1} \right) \le \mathbb{E}\hat{X} \left(1 + (1-p)^{-9} \mathbb{E}\hat{X} \right).$$

So, by the Second Moment Method (see Lemma 26.5),

$$\mathbb{P}(\hat{X} > 0) \ge \frac{(\mathbb{E}\hat{X})^2}{\mathbb{E}\hat{X}^2} \ge \frac{(\mathbb{E}\hat{X})^2}{\mathbb{E}\hat{X}\left(1 + (1-p)^{-9}\mathbb{E}\hat{X}\right)}$$

$$= \frac{1}{(1-p)^{-9} + [\mathbb{E}\hat{X}]^{-1}} \to 1$$

as $n \to \infty$, since $p \to 0$ and $\mathbb{E}\hat{X} \to \infty$. Thus

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ contains an isolated path of length two}) \to 1,$$

which implies that $\mathbb{P}(\mathbb{G}_{n,p}$ contains a path of length two$) \to 1$. As the property of having a path of length two is monotone increasing it in turn implies that

$$\mathbb{P}(\mathbb{G}_m \text{ contains a path of length two}) \to 1$$

for $m \gg n^{1/2}$ and the theorem follows. □

From Theorems 2.2 and 2.3 we obtain the following corollary.

Corollary 2.4. *The function $m^*(n) = n^{1/2}$ is the threshold for the property that a random graph $\mathbb{G}_m$ contains a path of length two, i.e.,*

$$\mathbb{P}(\mathbb{G}_m \textit{ contains a path of length two}) = \begin{cases} o(1) & \textit{if } m \ll n^{1/2}. \\ 1 - o(1) & \textit{if } m \gg n^{1/2}. \end{cases}$$

As we keep adding edges, trees on more than three vertices start to appear. Note that isolated vertices, edges and paths of length two are also trees on one, two and three vertices, respectively. The next two theorems show how long we have to "wait" until trees with a given number of vertices appear w.h.p.

Theorem 2.5. *Fix $k \geq 3$. If $m \ll n^{\frac{k-2}{k-1}}$, then w.h.p. $\mathbb{G}_m$ contains no tree with k vertices.*

Proof. Let $m = n^{\frac{k-2}{k-1}}/\omega$ and then $p = \frac{m}{N} \approx \frac{2}{\omega n^{k/(k-1)}} \leq \frac{3}{\omega n^{k/(k-1)}}$. Let X_k denote the number of trees with k vertices in $\mathbb{G}_{n,p}$. Let $T_1, T_2, \ldots, T_M$ be an enumeration of the copies of k-vertex trees in K_n. Let

$$A_i = \{T_i \text{ occurs as a subgraph in } \mathbb{G}_{n,p}\}.$$

The probability that a tree T occurs in $\mathbb{G}_{n,p}$ is $p^{e(T)}$, where $e(T)$ is the number of edges of T. So,

$$\mathbb{E} X_k = \sum_{t=1}^{M} \mathbb{P}(A_t) = M p^{k-1}.$$

But $M = \binom{n}{k} k^{k-2}$ since one can choose a set of k vertices in $\binom{n}{k}$ ways and then by Cayley's formula choose a tree on these vertices in k^{k-2} ways. Hence

$$\mathbb{E} X_k = \binom{n}{k} k^{k-2} p^{k-1}. \tag{2.1}$$

Noting also that (see Lemma 27.1(c))

$$\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k,$$

we see that

$$\begin{aligned}
\mathbb{E} X_k &\leq \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{3}{\omega n^{k/(k-1)}}\right)^{k-1} \\
&= \frac{3^{k-1} e^k}{k^2 \omega^{k-1}} \to 0,
\end{aligned}$$

as $n \to \infty$, seeing as k is fixed.

Thus we see by the first moment method that,

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ contains a tree with } k \text{ vertices}) \to 0.$$

This property is monotone increasing and therefore

$$\mathbb{P}(\mathbb{G}_m \text{ contains a tree with } k \text{ vertices}) \to 0.$$

□

Let us check what happens if the number of edges in $\mathbb{G}_m$ is much larger than $n^{\frac{k-2}{k-1}}$.

Theorem 2.6. *Fix $k \geq 3$. If $m \gg n^{\frac{k-2}{k-1}}$, then w.h.p. $\mathbb{G}_m$ contains a copy of every fixed tree with k vertices.*

Proof. Let $p = \frac{m}{N}, m = \omega n^{\frac{k-2}{k-1}}$ where $\omega = o(\log n)$ and fix some tree T with k vertices. Denote by $\hat{X}_k$ the number of *isolated* copies of T (T-components) in $\mathbb{G}_{n,p}$. Let $aut(H)$ denote the number of automorphisms of a graph H. Note that there are $k!/aut(T)$ copies of T in the complete graph K_k. To see this choose a copy of T with vertex set $[k]$. There are $k!$ ways of mapping the vertices of T to the vertices of K_k. Each map f induces a copy of T and two maps f_1, f_2 induce the same copy iff $f_2 f_1^{-1}$ is an automorphism of T.
So,

$$\begin{aligned}
\mathbb{E}\hat{X}_k &= \binom{n}{k} \frac{k!}{aut(T)} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1} \qquad (2.2)\\
&= (1+o(1)) \frac{(2\omega)^{k-1}}{aut(T)} \to \infty.
\end{aligned}$$

In (2.2) we have approximated $\binom{n}{k} \leq \frac{n^k}{k!}$ and used the fact that $\omega = o(\log n)$ in order to show that $(1-p)^{k(n-k)+\binom{k}{2}-k+1} = 1-o(1)$.

Next let $\mathscr{T}$ be the set of copies of T in K_n and $T_{[k]}$ be a fixed copy of T on vertices $[k]$ of K_n. Then, arguing as in (2.3),

$$\begin{aligned}
\mathbb{E}(\hat{X}_k^2) &= \sum_{T_1, T_2 \in \mathscr{T}} \mathbb{P}(T_2 \subseteq_i \mathbb{G}_{n,p} | \, T_1 \subseteq_i \mathbb{G}_{n,p}) \mathbb{P}(T_1 \subseteq_i \mathbb{G}_{n,p}) \\
&= \mathbb{E}\hat{X}_k \left(1 + \sum_{\substack{T_2 \in \mathscr{T} \\ V(T_2) \cap [k] = \emptyset}} \mathbb{P}(T_2 \subseteq_i \mathbb{G}_{n,p} | \, T_{[k]} \subseteq_i \mathbb{G}_{n,p}) \right) \\
&\leq \mathbb{E}\hat{X}_k \left(1 + (1-p)^{-k^2} \mathbb{E} X_k \right).
\end{aligned}$$

Notice that the $(1-p)^{-k^2}$ factor comes from conditioning on the event $T_{[k]} \subseteq_i \mathbb{G}_{n,p}$ which forces the non-existence of fewer than k^2 edges.

Hence, by the Second Moment Method,

$$\mathbb{P}(\hat{X}_k > 0) \geq \frac{(\mathbb{E}\hat{X}_k)^2}{\mathbb{E}\hat{X}_k \left(1 + (1-p)^{-k^2} \mathbb{E}\hat{X}_k\right)} \to 1.$$

Then, by a similar reasoning to that in the proof of Theorem 2.3,

$$\mathbb{P}(\mathbb{G}_m \text{ contains a copy of } T) \to 1,$$

as $n \to \infty$. $\square$

Combining the two above theorems we arrive at the following conclusion.

Corollary 2.7. *The function $m^*(n) = n^{\frac{k-2}{k-1}}$ is the threshold for the property that a random graph $\mathbb{G}_m$ contains a tree with $k \geq 3$ vertices, i.e.,*

$$\mathbb{P}(\mathbb{G}_m \supseteq k\text{-vertex-tree}) = \begin{cases} o(1) & \text{if } m \ll n^{\frac{k-2}{k-1}} \\ 1 - o(1) & \text{if } m \gg n^{\frac{k-2}{k-1}} \end{cases}$$

In the next theorem we show that "on the threshold" for k vertex trees, i.e., if $m = cn^{\frac{k-2}{k-1}}$, where c is a constant, $c > 0$, the number of tree components of a given order asymptotically follows the Poisson distribution. This time we will formulate both the result and its proof in terms of $\mathbb{G}_m$.

Theorem 2.8. *If $m = cn^{\frac{k-2}{k-1}}$, where $c > 0$, and T is a fixed tree with $k \geq 3$ vertices, then*

$$\mathbb{P}(\mathbb{G}_m \text{ contains an isolated copy of tree } T) \to 1 - e^{-\lambda},$$

as $n \to \infty$, where $\lambda = \frac{(2c)^{k-1}}{aut(T)}$.

More precisely, the number of copies of T is asymptotically distributed as the Poisson distribution with expectation λ.

Proof. Let $T_1, T_2, \dots, T_M$ be an enumeration of the copies of some k vertex tree T in K_n.
Let

$$A_i = \{T_i \text{ occurs as a component in } \mathbb{G}_m\}.$$

Suppose $J \subseteq [M] = \{1, 2, \dots, M\}$ with $|J| = t$, where t is fixed. Let $A_J = \bigcap_{j \in J} A_j$. We have $\mathbb{P}(A_J) = 0$ if there are $i, j \in J$ such that T_i, T_j share a vertex. Suppose $T_i, i \in J$ are vertex disjoint. Then

$$\mathbb{P}(A_J) = \frac{\binom{\binom{n-kt}{2}}{m-(k-1)t}}{\binom{N}{m}}.$$

Note that in the numerator we count the number of ways of choosing m edges so that A_J occurs.
If, say, $t \leq \log n$, then

$$\binom{n-kt}{2} = N\left(1 - \frac{kt}{n}\right)\left(1 - \frac{kt}{n-1}\right) = N\left(1 - O\left(\frac{kt}{n}\right)\right),$$

and so

$$\frac{m^2}{\binom{n-kt}{2}} \to 0.$$

Then from Lemma 27.1(f),

$$\begin{aligned}\binom{\binom{n-kt}{2}}{m-(k-1)t} &= (1+o(1))\frac{\left(N\left(1-O\left(\frac{kt}{n}\right)\right)\right)^{m-(k-1)t}}{(m-(k-1)t)!}\\ &= (1+o(1))\frac{N^{m-(k-1)t}\left(1-O\left(\frac{mkt}{n}\right)\right)}{(m-(k-1)t)!}\\ &= (1+o(1))\frac{N^{m-(k-1)t}}{(m-(k-1)t)!}.\end{aligned}$$

Similarly, again by Lemma 27.1,

$$\binom{N}{m} = (1+o(1))\frac{N^m}{m!},$$

and so

$$\mathbb{P}(A_J) = (1+o(1))\frac{m!}{(m-(k-1)t)!}N^{-(k-1)t} = (1+o(1))\left(\frac{m}{N}\right)^{(k-1)t}.$$

Thus, if Z_T denotes the number of components of $\mathbb{G}_m$ that are copies of T, then,

$$\begin{aligned}\mathbb{E}\binom{Z_T}{t} &\approx \frac{1}{t!}\binom{n}{k,k,k,\ldots,k}\left(\frac{k!}{aut(T)}\right)^t\left(\frac{m}{N}\right)^{(k-1)t}\\ &\approx \frac{n^{kt}}{t!(k!)^t}\left(\frac{k!}{aut(T)}\right)^t\left(\frac{cn^{(k-2)/(k-1)}}{N}\right)^{(k-1)t}\\ &\approx \frac{\lambda^t}{t!},\end{aligned}$$

where

$$\lambda = \frac{(2c)^{k-1}}{aut(T)}.$$

So by Theorem 26.11 the number of copies of T-components is asymptotically distributed as the Poisson distribution with expectation λ given above, which combined with the statements of Theorem 2.1 and Corollary 2.7 proves the theorem. Note that Theorem 2.1 implies that w.h.p. there are no non-component copies of T. $\square$

We complete our presentation of the basic features of a random graph in its sub-critical phase of evolution with a description of the order of its largest component.

Theorem 2.9. *If* $m = \frac{1}{2}cn$, *where* $0 < c < 1$ *is a constant, then w.h.p. the order of the largest component of a random graph* $\mathbb{G}_m$ *is* $O(\log n)$.

The above theorem follows from the next three lemmas stated and proved in terms of $\mathbb{G}_{n,p}$ with $p = c/n,\ 0 < c < 1$. In fact the first of those three lemmas covers a little bit more than the case of $p = c/n,\ 0 < c < 1$.

Lemma 2.10. *If* $p \leq \frac{1}{n} - \frac{\omega}{n^{4/3}}$, *where* $\omega = \omega(n) \to \infty$, *then w.h.p. every component in* $\mathbb{G}_{n,p}$ *contains at most one cycle.*

Proof. Suppose that there is a pair of cycles that are in the same component. If such a pair exists then there is *minimal* pair C_1, C_2, i.e., either C_1 and C_2 are connected by a path (or meet at a vertex) or they form a cycle with a diagonal path (see Figure 2.1). Then in either case, $C_1 \cup C_2$ consists of a path P plus another two distinct edges, one from each endpoint of P joining it to another vertex in P. The number of such graphs on k labeled vertices can be bounded by $k^2 k!$.

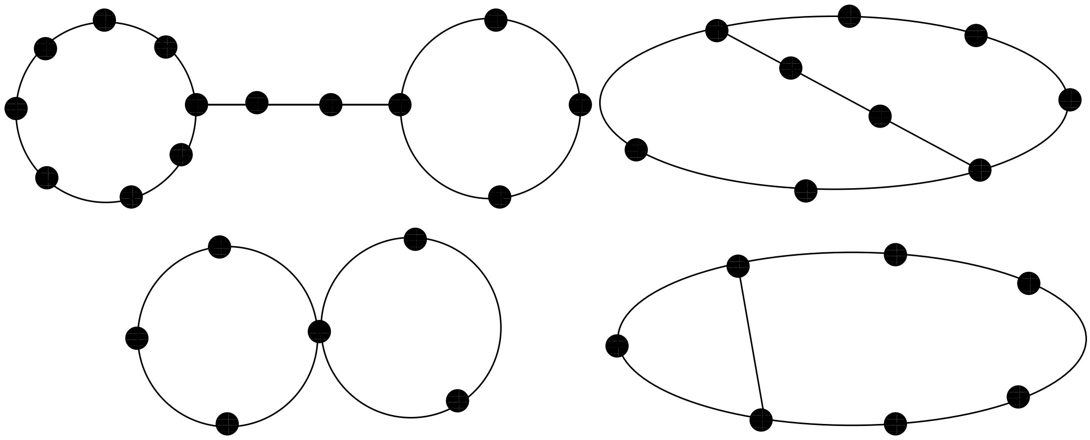


Figure 2.1: $C_1 \cup C_2$

Let X be the number of subgraphs of the above kind (shown in Figure 2.1) in the random graph $\mathbb{G}_{n,p}$. By the first moment method (see Lemma 26.2),

$$\begin{aligned}
\mathbb{P}(X>0) \leq \mathbb{E}X &\leq \sum_{k=4}^{n} \binom{n}{k} k^2 k! p^{k+1} \qquad (2.3)\\
&\leq \sum_{k=4}^{n} \frac{n^k}{k!} k^2 k! \frac{1}{n^{k+1}} \left(1-\frac{\omega}{n^{1/3}}\right)^{k+1}\\
&\leq \int_0^\infty \frac{x^2}{n} \exp\left(-\frac{\omega x}{n^{1/3}}\right) dx\\
&= \frac{2}{\omega^3}\\
&= o(1).
\end{aligned}$$

□

We remark for later use that if $p=c/n$, $0<c<1$ then (2.3) implies

$$\mathbb{P}(X>0) \leq \sum_{k=4}^{n} k^2 c^{k+1} n^{-1} = O(n^{-1}). \qquad (2.4)$$

Hence, in determining the order of the largest component we may concentrate our attention on unicyclic components and tree-components (isolated trees). However the number of vertices on unicyclic components tends to be rather small, as is shown in the next lemma.

Lemma 2.11. *If $p=c/n$, where $c \neq 1$ is a constant, then in $\mathbb{G}_{n,p}$ w.h.p. the number of vertices in components with exactly one cycle, is $O(\omega)$ for any growing function ω.*

Proof. Let X_k be the number of vertices on unicyclic components with k vertices. Then

$$\mathbb{E}X_k \leq \binom{n}{k} k^{k-2} \binom{k}{2} k p^k (1-p)^{k(n-k)+\binom{k}{2}-k}. \qquad (2.5)$$

The factor $k^{k-2}\binom{k}{2}$ in (2.5) is the number of choices for a tree plus an edge on k vertices in $[k]$. This bounds the number $C(k,k)$ of connected graphs on $[k]$ with k edges. This is off by a factor $O(k^{1/2})$ from the exact formula which is given below for completeness:

$$C(k,k) = \sum_{r=3}^{k} \binom{k}{r} \frac{(r-1)!}{2} r k^{k-r-1} \approx \sqrt{\frac{\pi}{8}} k^{k-1/2}. \qquad (2.6)$$

The remaining factor, other than $\binom{n}{k}$, in (2.5) is the probability that the k edges of the unicyclic component exist and that there are no other edges on $\mathbb{G}_{n,p}$ incident with the k chosen vertices.
Noting also that by Lemma 27.1(d),

$$\binom{n}{k} \le \frac{n^k}{k!} e^{-\frac{k(k-1)}{2n}}.$$

Assume next that $c < 1$ and then we get

$$\mathbb{E} X_k \le \frac{n^k}{k!} e^{-\frac{k(k-1)}{2n}} k^{k+1} \frac{c^k}{n^k} e^{-ck + \frac{ck(k-1)}{2n} + \frac{ck}{2n}} \tag{2.7}$$

$$\le \frac{e^k}{k^k} e^{-\frac{k(k-1)}{2n}} k^{k+1} c^k e^{-ck + \frac{k(k-1)}{2n} + \frac{c}{2}} \tag{2.8}$$

$$\le k \left(ce^{1-c}\right)^k e^{\frac{c}{2}}.$$

So,

$$\mathbb{E} \sum_{k=3}^{n} X_k \le \sum_{k=3}^{n} k \left(ce^{1-c}\right)^k e^{\frac{c}{2}} = O(1), \tag{2.9}$$

since $ce^{1-c} < 1$ for $c \neq 1$. By the Markov inequality, if $\omega = \omega(n) \to \infty$, (see Lemma 26.1)

$$\mathbb{P}\left(\sum_{k=3}^{n} X_k \ge \omega\right) = O\left(\frac{1}{\omega}\right) \to 0 \text{ as } n \to \infty,$$

and the Lemma follows for $c < 1$. If $c > 1$ then we cannot deduce (2.8) from (2.7). If however $k = o(n)$ then this does not matter, since then $e^{k^2/n} = e^{o(k)}$. Now we show in the proof of Theorem 2.14 below that when $c > 1$ there is w.h.p. a unique giant component of size $\Omega(n)$ and all other components are of size $O(\log n)$. This giant is not unicyclic. This enables us to complete the proof of this lemma for $c > 1$. $\square$

After proving the first two lemmas one can easily see that the only remaining candidate for the largest component of our random graph is an isolated tree.

Lemma 2.12. *Let $p = \frac{c}{n}$, where $c \neq 1$ is a constant, $\alpha = c - 1 - \log c$, and $\omega = \omega(n) \to \infty$, $\omega = o(\log\log n)$. Then*

(i) w.h.p. there exists an isolated tree of order

$$k_- = \frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right) - \omega,$$

(ii) w.h.p. there is no *isolated tree of order at least*

$$k_+ = \frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right) + \omega$$

Proof. Note that our assumption on c means that α is a positive constant.

Let X_k be the number of isolated trees of order k. Then

$$\mathbb{E}X_k = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1}. \tag{2.10}$$

To prove *(i)* suppose $k = O(\log n)$. Then $\binom{n}{k} \approx \frac{n^k}{k!}$ and by using Lemma 27.1(a),(b) and Stirling's approximation (1.5) for $k!$ we see that

$$\begin{aligned}
\mathbb{E}X_k &= (1+o(1))\frac{n}{c}\frac{k^{k-2}}{k!}(ce^{-c})^k \qquad\qquad (2.11)\\
&= \frac{(1+o(1))}{c\sqrt{2\pi}}\frac{n}{k^{5/2}}(ce^{1-c})^k\\
&= \frac{(1+o(1))}{c\sqrt{2\pi}}\frac{n}{k^{5/2}}e^{-\alpha k}, \qquad \text{for } k = O(\log n). \qquad (2.12)
\end{aligned}$$

Putting $k = k_-$ we see that

$$\mathbb{E}X_k = \frac{(1+o(1))}{c\sqrt{2\pi}}\frac{n}{k^{5/2}}\frac{e^{\alpha\omega}(\log n)^{5/2}}{n} \geq Ae^{\alpha\omega}, \tag{2.13}$$

for some constant $A > 0$.

We continue via the Second Moment Method, this time using the Chebyshev inequality as we will need a little extra precision for the proof of Theorem 2.14. Using essentially the same argument as for a fixed tree T of order k (see Theorem 2.6), we get

$$\mathbb{E}X_k^2 \leq \mathbb{E}X_k\left(1 + (1-p)^{-k^2}\,\mathbb{E}X_k\right).$$

So

$$\begin{aligned}
\operatorname{Var} X_k &\leq \mathbb{E}X_k + (\mathbb{E}X_k)^2\left((1-p)^{-k^2} - 1\right)\\
&\leq \mathbb{E}X_k + 2ck^2(\mathbb{E}X_k)^2/n, \qquad \text{for } k = O(\log n). \qquad (2.14)
\end{aligned}$$

Thus, by the Chebyshev inequality (see Lemma 26.3), we see that for any $\varepsilon > 0$,

$$\mathbb{P}\left(|X_k - \mathbb{E}X_k| \geq \varepsilon\,\mathbb{E}X_k\right) \leq \frac{1}{\varepsilon^2\,\mathbb{E}X_k} + \frac{2ck^2}{\varepsilon^2 n} = o(1). \tag{2.15}$$

Thus w.h.p. $X_k \geq Ae^{\alpha\omega/2}$ and this completes the proof of *(i)*.

For *(ii)* we go back to the formula (2.10) and write, for some new constant $A > 0$,

$$\begin{aligned}\mathbb{E}X_k &\leq \frac{A}{\sqrt{k}}\left(\frac{ne}{k}\right)^k k^{k-2}\left(1-\frac{k}{2n}\right)^{k-1}\left(\frac{c}{n}\right)^{k-1} e^{-ck+\frac{ck^2}{2n}} \\ &\leq \frac{2An}{\widehat{c}_k k^{5/2}}\left(\widehat{c}_k e^{1-\widehat{c}_k}\right)^k,\end{aligned}$$

where $\widehat{c}_k = c\left(1-\frac{k}{2n}\right)$.

In the case $c < 1$ we have $\widehat{c}_k e^{1-\widehat{c}_k} \leq ce^{1-c}$ and $\widehat{c}_k \approx c$ and so we can write

$$\begin{aligned}\sum_{k=k_+}^{n} \mathbb{E}X_k &\leq \frac{3An}{c}\sum_{k=k_+}^{n}\frac{\left(ce^{1-c}\right)^k}{k^{5/2}} \leq \frac{3An}{ck_+^{5/2}}\sum_{k=k_+}^{\infty} e^{-\alpha k} = \\ &= \frac{3Ane^{-\alpha k_+}}{ck_+^{5/2}(1-e^{-\alpha})} = \frac{(3A+o(1))\alpha^{5/2}e^{-\alpha\omega}}{c(1-e^{-\alpha})} = o(1). \quad (2.16)\end{aligned}$$

If $c > 1$ then for $k \leq \frac{n}{\log n}$ we use $\widehat{c}_k e^{1-\widehat{c}_k} = e^{-\alpha-O(1/\log n)}$ and for $k > \frac{n}{\log n}$ we use $c_k \geq c/2$ and $\widehat{c}_k e^{1-\widehat{c}_k} \leq 1$ and replace (2.16) by

$$\sum_{k=k_+}^{n} \mathbb{E}X_k \leq \frac{3An}{ck_+^{5/2}}\sum_{k=k_+}^{n/\log n} e^{-(\alpha+O(1/\log n))k} + \frac{6An}{c}\sum_{k=n/\log n}^{n}\frac{1}{k^{5/2}} = o(1).$$

□

Finally, applying Lemmas 2.11 and 2.12 we can prove the following useful identity: Suppose that $x = x(c)$ is given as

$$x = x(c) = \begin{cases} c & c \leq 1 \\ \text{The solution in } (0,1) \text{ to } xe^{-x} = ce^{-c} & c > 1 \end{cases}.$$

Note that xe^{-x} increases continuously as x increases from 0 to 1 and then decreases. This justifies the existence and uniqueness of x.

Lemma 2.13. *If $c > 0$, $c \neq 1$ is a constant, and $x = x(c)$ is defined above, then*

$$\frac{1}{x}\sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}\left(ce^{-c}\right)^k = 1.$$

Proof. Let $p = \frac{c}{n}$. Assume first that $c < 1$ and let X be the total number of vertices of $\mathbb{G}_{n,p}$ that lie in non-tree components. Let X_k be the number of tree-components of order k. Then,

$$n = \sum_{k=1}^{n} kX_k + X.$$

So,

$$n = \sum_{k=1}^{n} k\mathbb{E}X_k + \mathbb{E}X.$$

Now,

(i) by (2.4) and (2.9), $\mathbb{E}X = O(1)$,

(ii) by (2.11), if $k < k_+$ then

$$\mathbb{E}X_k = (1+o(1))\frac{n}{ck!}k^{k-2}\left(ce^{-c}\right)^k.$$

So, by Lemma 2.12,

$$\begin{aligned} n &= o(n) + \frac{n}{c}\sum_{k=1}^{k_+}\frac{k^{k-1}}{k!}\left(ce^{-c}\right)^k \\ &= o(n) + \frac{n}{c}\sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}\left(ce^{-c}\right)^k. \end{aligned}$$

Now divide through by n and let $n \to \infty$.

This proves the identity for the case $c < 1$. Suppose now that $c > 1$. Then, since x is a solution of equation $ce^{-c} = xe^{-x},\ 0 < x < 1$, we have

$$\sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}\left(ce^{-c}\right)^k = \sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}\left(xe^{-x}\right)^k = x,$$

by the first part of the proof (for $c < 1$). □

We note that in fact, Lemma 2.13 is also true for $c = 1$.

2.2 Super-Critical Phase

The structure of a random graph $\mathbb{G}_m$ changes dramatically when $m = \frac{1}{2}cn$ where $c > 1$ is a constant. We will give a precise characterisation of this phenomenon, presenting results in terms of $\mathbb{G}_m$ and proving them for $\mathbb{G}_{n,p}$ with $p = c/n,\ c > 1$.

Theorem 2.14. *If $m = cn/2$, $c > 1$, then w.h.p. $\mathbb{G}_m$ consists of a unique giant component, with $\left(1 - \frac{x}{c} + o(1)\right) n$ vertices and $\left(1 - \frac{x^2}{c^2} + o(1)\right) \frac{cn}{2}$ edges. Here $0 < x < 1$ is the solution of the equation $xe^{-x} = ce^{-c}$. The remaining components are of order at most $O(\log n)$.*

Proof. Suppose that Z_k is the number of components of order k in $\mathbb{G}_{n,p}$. Then, bounding the number of such components by the number of trees with k vertices that span a component, we get

$$\begin{aligned}
\mathbb{E} Z_k &\leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \qquad (2.17)\\
&\leq A \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck+ck^2/n}\\
&\leq \frac{An}{k^2} \left(ce^{1-c+ck/n}\right)^k
\end{aligned}$$

Now let $\beta_1 = \beta_1(c)$ be small enough so that

$$ce^{1-c+c\beta_1} < 1,$$

and let $\beta_0 = \beta_0(c)$ be large enough so that

$$\left(ce^{1-c+o(1)}\right)^{\beta_0 \log n} < \frac{1}{n^2}.$$

If we choose β_1 and β_0 as above then it follows that w.h.p. there is no component of order $k \in [\beta_0 \log n, \beta_1 n]$.
Our next task is to estimate the number of vertices on small components i.e. those of size at most $\beta_0 \log n$.

We first estimate the total number of vertices on small tree components, i.e., on isolated trees of order at most $\beta_0 \log n$.
Assume first that $1 \leq k \leq k_0$, where $k_0 = \frac{1}{2\alpha} \log n$, where α is from Lemma 2.12. It follows from (2.11) that

$$\begin{aligned}
\mathbb{E}\left(\sum_{k=1}^{k_0} kX_k\right) &\approx \frac{n}{c} \sum_{k=1}^{k_0} \frac{k^{k-1}}{k!} \left(ce^{-c}\right)^k\\
&\approx \frac{n}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} \left(ce^{-c}\right)^k,
\end{aligned}$$

using $k^{k-1}/k! < e^k$, and $ce^{-c} < e^{-1}$ for $c \neq 1$ to extend the summation from k_0 to infinity.

Putting $\varepsilon = 1/\log n$ and using (2.15) we see that the probability that any $X_k,\ 1 \le k \le k_0$, deviates from its mean by more than $1 \pm \varepsilon$ is at most

$$\sum_{k=1}^{k_0} \left[\frac{(\log n)^2}{n^{1/2-o(1)}} + O\left(\frac{(\log n)^4}{n} \right) \right] = o(1),$$

where the $n^{1/2-o(1)}$ term comes from putting $\omega \approx k_0/2$ in (2.13), which is allowed by (2.12), (2.14).
Thus, if $x = x(c),\ 0 < x < 1$ is the unique solution in $(0,1)$ of the equation $xe^{-x} = ce^{-c}$, then w.h.p.,

$$\begin{aligned} \sum_{k=1}^{k_0} kX_k &\approx \frac{n}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} \left(xe^{-x}\right)^k \\ &= \frac{nx}{c}, \end{aligned}$$

by Lemma 2.13.
Now consider $k_0 < k \le \beta_0 \log n$.

$$\begin{aligned} \mathbb{E}\left(\sum_{k=k_0+1}^{\beta_0 \log n} kX_k \right) &\le \frac{n}{c} \sum_{k=k_0+1}^{\beta_0 \log n} \left(ce^{1-c+ck/n} \right)^k \\ &= O\left(n(ce^{1-c})^{k_0} \right) \\ &= O\left(n^{1/2+o(1)} \right). \end{aligned}$$

So, by the Markov inequality (see Lemma 26.1), w.h.p.,

$$\sum_{k=k_0+1}^{\beta_0 \log n} kX_k = o(n).$$

Now consider the number Y_k of non-tree components with k vertices, $1 \le k \le \beta_0 \log n$.

$$\begin{aligned} \mathbb{E}\left(\sum_{k=1}^{\beta_0 \log n} kY_k \right) &\le \sum_{k=1}^{\beta_0 \log n} \binom{n}{k} k^{k-1} \binom{k}{2} \left(\frac{c}{n} \right)^k \left(1 - \frac{c}{n} \right)^{k(n-k)} \\ &\le \sum_{k=1}^{\beta_0 \log n} k \left(ce^{1-c+ck/n} \right)^k \\ &= O(1). \end{aligned}$$

So, again by the Markov inequality, w.h.p.,

$$\sum_{k=1}^{\beta_0 \log n} kY_k = o(n).$$

Summarising, we have proved so far that w.h.p. there are approximately $\frac{nx}{c}$ vertices on components of order k, where $1 \leq k \leq \beta_0 \log n$ and all the remaining *giant* components are of size at least $\beta_1 n$.
We complete the proof by showing the uniqueness of the giant component. Let

$$c_1 = c - \frac{\log n}{n} \text{ and } p_1 = \frac{c_1}{n}.$$

Define p_2 by

$$1 - p = (1 - p_1)(1 - p_2)$$

and note that $p_2 \geq \frac{\log n}{n^2}$. Then, see Section 1.2,

$$\mathbb{G}_{n,p} = \mathbb{G}_{n,p_1} \cup \mathbb{G}_{n,p_2}.$$

If $x_1 e^{-x_1} = c_1 e^{-c_1}$, then $x_1 \approx x$ and so, by our previous analysis, w.h.p., $\mathbb{G}_{n,p_1}$ has no components with number of vertices in the range $[\beta_0 \log n, \beta_1 n]$.
Suppose there are components $C_1, C_2, \ldots, C_l$ with $|C_i| > \beta_1 n$. Here $l \leq 1/\beta_1$.
Now we add edges of $\mathbb{G}_{n,p_2}$ to $\mathbb{G}_{n,p_1}$. Then

$$\begin{aligned}
\mathbb{P}\left(\exists i, j : \text{ no } \mathbb{G}_{n,p_2} \text{ edge joins } C_i \text{ with } C_j\right) &\leq \binom{l}{2}(1 - p_2)^{(\beta_1 n)^2} \\
&\leq l^2 e^{-\beta_1^2 \log n} \\
&= o(1).
\end{aligned}$$

So w.h.p. $\mathbb{G}_{n,p}$ has a unique component with more than $\beta_0 \log n$ vertices and it has $\approx \left(1 - \frac{x}{c}\right) n$ vertices.

We now consider the number of edges in the giant C_0. Now we switch to $G = \mathbb{G}_{n,m}$. Suppose that the edges of G are $e_1, e_2, \ldots, e_m$ in random order. We estimate the probability that $e = e_m = \{x, y\}$ is an edge of the giant. Let G_1 be the graph induced by $\{e_1, e_2, \ldots, e_{m-1}\}$. G_1 is distributed as $\mathbb{G}_{n,m-1}$ and so we know that w.h.p. G_1 has a unique giant C_1 and other components are of size $O(\log n)$. So the probability that e is an edge of the giant is $o(1)$ plus the probability that x or y is a vertex of C_1. Thus,

$$\begin{aligned}
\mathbb{P}\left(e \notin C_0 \mid |C_1| \approx n\left(1 - \frac{x}{c}\right)\right) &= \mathbb{P}\left(e \cap C_1 = \emptyset \mid |C_1| \approx n\left(1 - \frac{x}{c}\right)\right) \\
&= \left(1 - \frac{|C_1|}{n}\right)\left(1 - \frac{|C_1| + 1}{n}\right) \approx \left(\frac{x}{c}\right)^2. \quad (2.18)
\end{aligned}$$

It follows that the expected number of edges in the giant is as claimed. To prove concentration, it is simplest to use the Chebyshev inequality, see Lemma 26.3. So, now fix $i, j \leq m$ and let C_2 denote the unique giant component of $G_{n,m} - \{e_i, e_j\}$. Then, arguing as for (2.18),

$$\mathbb{P}(e_i, e_j \subseteq C_0) = \\ o(1) + \mathbb{P}(e_j \cap C_2 \neq \emptyset \mid e_i \cap C_2 \neq \emptyset)\,\mathbb{P}(e_i \cap C_2 \neq \emptyset) \\ = (1 + o(1))\,\mathbb{P}(e_i \subseteq C_0)\,\mathbb{P}(e_j \subseteq C_0).$$

In the $o(1)$ term, we hide the probability of the event

$$\{e_i \cap C_2 \neq \emptyset, e_j \cap C_2 \neq \emptyset, e_i \cap e_j \neq \emptyset\}$$

which has probability $o(1)$. We should double this $o(1)$ probability here to account for switching the roles of i, j.

The Chebyshev inequality can now be used to show that the number of edges is concentrated as claimed. □

We will see later, see Theorem 2.18, that w.h.p. each of the small components have at most one cycle.

From the above theorem and the results of previous sections we see that, when $m = cn/2$ and c passes the critical value equal to 1, the typical structure of a random graph changes from a scattered collection of small trees and unicyclic components to a coagulated lump of components (the giant component) that dominates the graph. This short period when the giant component emerges is called the *phase transition*. We will look at this fascinating period of the evolution more closely in Section 2.3.

We know that w.h.p. the giant component of $\mathbb{G}_{n,m}, m = cn/2, c > 1$ has $\approx \left(1 - \frac{x}{c}\right) n$ vertices and $\approx \left(1 - \frac{x^2}{c^2}\right) \frac{cn}{2}$ edges. So, if we look at the graph H induced by the vertices outside the giant, then w.h.p. H has $\approx n_1 = \frac{nx}{c}$ vertices and $\approx m_1 = xn_1/2$ edges. Thus we should expect H to resemble $\mathbb{G}_{n_1.m_1}$, which is sub-critical since $x < 1$. This can be made precise, but the intuition is clear.

Now increase m further and look on the outside of the giant component. The giant component subsequently consumes the small components not yet attached to it. When m is such that $m/n \to \infty$ then unicyclic components disappear and a random graph $\mathbb{G}_m$ achieves the structure described in the next theorem.

Theorem 2.15. *Let $\omega = \omega(n) \to \infty$ as $n \to \infty$ be some slowly growing function. If $m \geq \omega n$ but $m \leq n(\log n - \omega)/2$, then $\mathbb{G}_m$ is disconnected and all components, with the exception of the giant, are trees w.h.p.*

Tree-components of order k die out in the reverse order they were born, i.e., larger trees are "swallowed" by the giant earlier than smaller ones.

Cores

Given a positive integer k, the k-core of a graph $G=(V,E)$ is the largest set $S\subseteq V$ such that the minimum degree δ_S in the vertex induced subgraph $G[S]$ is at least k. This is unique because if $\delta_S\geq k$ and $\delta_T\geq k$ then $\delta_{S\cup T}\geq k$. Cores were first discussed by Bollobás [154]. It was shown by Łuczak [636] that for $k\geq 3$ either there is no k-core in $G_{n,p}$ or one of linear size, w.h.p. The precise size and first occurrence of k-cores for $k\geq 3$ was established in Pittel, Spencer and Wormald [742]. The 2-core, C_2 which is the set of vertices that lie on at least one cycle behaves differently to the other cores, $k\geq 3$. It grows gradually. We will need the following result in Section 17.2.

Lemma 2.16. *Suppose that $c>1$ and that $x<1$ is the solution to $xe^{-x}=ce^{-c}$. Then w.h.p. the 2-core C_2 of $G_{n,p}, p=c/n$ has $(1-x)\left(1-\frac{x}{c}+o(1)\right)n$ vertices and $\left(1-\frac{x}{c}+o(1)\right)^2\frac{cn}{2}$ edges.*

Proof. Fix $v\in[n]$. We estimate $\mathbb{P}(v\in C_2)$. Let C_1 denote the unique giant component of $G_1=\mathbb{G}_{n,p}-v$. Now G_1 is distributed as $\mathbb{G}_{n-1,p}$ and so C_1 exists w.h.p. To be in C_2, either (i) v has two neighbors in C_1 or (ii) v has two neighbors in some other component. Now because all components other than C_1 have size $O(\log n)$ w.h.p., we see that

$$\mathbb{P}((ii))=o(1)+n\binom{O(\log n)}{2}\left(\frac{c}{n}\right)^2=o(1).$$

Now w.h.p. $|C_1|\approx\left(1-\frac{x}{c}\right)n$ and it is independent of the edges incident with v and so

$$\begin{aligned}
&\mathbb{P}((i))=1-\mathbb{P}(0\text{ or }1\text{ neighbors in }C_1)=\\
&=o(1)+(1+o(1))\mathbb{E}\left(1-\left(\left(1-\frac{c}{n}\right)^{|C_1|}+|C_1|\left(1-\frac{c}{n}\right)^{|C_1|-1}\frac{c}{n}\right)\right) \qquad (2.19)\\
&=o(1)+1-(e^{-c+x}+(c-x)e^{-c+x})\\
&=o(1)+(1-x)\left(1-\frac{x}{c}\right),
\end{aligned}$$

where the last line follows from the fact that $e^{-c+x}=\frac{x}{c}$. Also, one has to be careful when estimating something like $\mathbb{E}\left(1-\frac{c}{n}\right)^{|C_1|}$. For this we note that Jensen's inequality implies that

$$\mathbb{E}\left(1-\frac{c}{n}\right)^{|C_1|}\geq\left(1-\frac{c}{n}\right)^{\mathbb{E}|C_1|}=e^{-c+x+o(1)}.$$

On the other hand, if $n_g=\left(1-\frac{x}{c}\right)n$,

$$
\begin{aligned}
\mathbb{E}\left(1-\frac{c}{n}\right)^{|C_1|} \leq & \\
& \mathbb{E}\left(\left(1-\frac{c}{n}\right)^{|C_1|} \middle| |C_1| \geq (1-o(1))n_g\right) \mathbb{P}\left(|C_1| \geq (1-o(1))n_g\right) \\
& \qquad\qquad + \mathbb{P}\left(|C_1| \leq (1-o(1))n_g\right) = e^{-c+x+o(1)}.
\end{aligned}
$$

It follows from (2.19) that $\mathbb{E}(|C_2|) \approx (1-x)\left(1-\frac{x}{c}\right)n$. To prove concentration of $|C_2|$, we can use the Chebyshev inequality as we did in the proof of Theorem 2.14 to prove concentration for the number of edges in the giant.

To estimate the expected number of edges in C_2, we proceed as in Theorem 2.14 and turn to $G = \mathbb{G}_{n,m}$ and estimate the probability that $e_1 \subseteq C_2$. If $G' = G \setminus e$ and C_1' is the giant of G' then e_1 is an edge of C_2 iff $e_1 \subseteq C_1'$ or e_1 is contained in a small component. This latter condition is unlikely. Thus,

$$
\mathbb{P}(e_1 \subseteq C_2) = o(1) + \mathbb{E}\left(\frac{|C_1'|}{n}\right)^2 = o(1) + \left(1-\frac{x}{c}\right)^2.
$$

The estimate for the expectation of the number of edges in the 2-core follows immediately and one can prove concentration using the Chebyshev inequality. □

2.3 Phase Transition

In the previous two sections we studied the asymptotic behavior of $\mathbb{G}_m$ (and $\mathbb{G}_{n,p}$) in the "sub-critical phase" when $m = cn/2, c < 1$ ($p = c/n, c < 1$), as well as in the "super-critical phase" when $m = cn/2, c > 1$ ($p = c/n, c > 1$) of its evolution.

We have learned that when $m = cn/2, c > 1$ our random graph consists w.h.p. of tree components and components with exactly one cycle (see Theorem 2.1 and Lemma 2.11). We call such components *simple* while components which are not simple, i.e. components with at least two cycles, will be called *complex*.

All components during the sub-critical phase are rather small, of order $\log n$, tree-components dominate the typical structure of $\mathbb{G}_m$, and there is no significant gap in the order of the first and the second largest component. This follows from Lemma 2.12. The proof of this lemma shows that w.h.p. there are many trees of height k_-. The situation changes when $m > n/2$, i.e., when we enter the super-critical phase and then w.h.p. $\mathbb{G}_m$ consists of a single giant complex component (of the order comparable to n), and some number of simple components, i.e., tree components and components with exactly one cycle (see Theorem 2.14). One can also observe a clear gap between the order of the largest component (the giant) and the second largest component which is of the order $O(\log n)$. This phenomenon of dramatic change of the typical structure of a random graph is called its *phase transition*.

A natural question arises as to what happens when $m/n \to 1/2$, either from below or above, as $n \to \infty$. It appears that one can establish, a so called, *scaling window* or *critical window* for the phase transition in which $\mathbb{G}_m$ is undergoing a rapid change in its typical structure. A characteristic feature of this period is that a random graph can w.h.p. consist of **more than one** complex component (recall: there are no complex components in the sub-critical phase and there is a unique complex component in the super-critical phase).

Erdős and Rényi [332] studied the size of the largest tree in the random graph $\mathbb{G}_{n,m}$ when $m = n/2$ and showed that it was likely to be around $n^{2/3}$. They called the transition from $O(\log n)$ through $\Theta(n^{2/3})$ to $\Omega(n)$ the "double jump". They did not study the regime $m = n/2 + o(n)$. Bollobás [153] opened the detailed study of this and Łuczak [634] continued this analysis. He established the precise size of the "scaling window" by removing a logarithmic factor from Bollobás's estimates. The component structure of $\mathbb{G}_{n,m}$ for $m = n/2 + o(n)$ is rather complicated and the proofs are technically challenging. We will begin by stating several results that give a an idea of the component structure in this range, referring the reader elsewhere for proofs: Chapter 5 of Janson, Łuczak and Ruciński [509]; Aldous [19]; Bollobás [153]; Janson [496]; Janson, Knuth, Łuczak and Pittel [513]; Łuczak [634], [635], [639]; Łuczak, Pittel and Wierman [642]. We will finish with a proof by Nachmias and Peres that when $p = 1/n$ the largest component is likely to have size of order $n^{2/3}$.

The first theorem is a refinement of Lemma 2.10.

Theorem 2.17. *Let $m = \frac{n}{2} - s$, where $s = s(n) \geq 0$.*

(a) The probability that $\mathbb{G}_{n,m}$ contains a complex component is at most $n^2/4s^3$.

(b) If $n^{2/3} \ll s \ll n$ then w.h.p. the largest component is a tree of size asymptotic to $\frac{n^2}{2s^2} \log \frac{s^3}{n}$.

The next theorem indicates when the phase in which we may have more than one complex component "ends", i.e., when a single giant component emerges.

Theorem 2.18. *Let $m = \frac{n}{2} + s$, where $s = s(n) \geq 0$. Then the probability that $\mathbb{G}_{n,m}$ contains more than one complex component is at most $6n^{2/9}/s^{1/3}$.*

□

For larger s, the next theorem gives a precise estimate of the size of the largest component for $s \gg n^{2/3}$. For $s > 0$ we let $\bar{s} > 0$ be defined by

$$\left(1 - \frac{2\bar{s}}{n}\right) \exp\left\{\frac{2\bar{s}}{n}\right\} = \left(1 + \frac{2s}{n}\right) \exp\left\{-\frac{2s}{n}\right\}.$$

Theorem 2.19. *Let $m = \frac{n}{2} + s$ where $s \gg n^{2/3}$. Then with probability at least $1 - 7n^{2/9}/s^{1/3}$,*

$$\left| L_1 - \frac{2(s+\bar{s})n}{n+2s} \right| \leq \frac{n^{2/3}}{5}$$

where L_1 is the size of the largest component in $\mathbb{G}_{n,m}$. In addition, the largest component is complex and all other components are either trees or unicyclic components.

To get a feel for this estimate of L_1 we remark that

$$\bar{s} = s - \frac{4s^2}{3n} + O\left(\frac{s^3}{n^2}\right).$$

The next theorem gives some information about ℓ-components inside the scaling window $m = n/2 + O(n^{2/3})$. An ℓ-component is one that has ℓ more edges than vertices. So trees are (-1)-components.

Theorem 2.20. *Let $m = \frac{n}{2} + O(n^{2/3})$ and let r_ℓ denote the number of ℓ-components in $\mathbb{G}_{n,m}$. For every $0 < \delta < 1$ there exists C_δ such that if n is sufficiently large, then with probability at least $1 - \delta$, $\sum_{\ell \geq 3} \ell r_\ell \leq C_\delta$ and the number of vertices on complex components is at most $C_\delta n^{2/3}$.*

One of the difficulties in analysing the phase transition stems from the need to estimate $C(k,\ell)$, which is the number of connected graphs with vertex set $[k]$ and ℓ edges. We need good estimates for use in first moment calculations. We have seen the values for $C(k,k-1)$ (Cayley's formula) and $C(k,k)$, see (2.6). For $\ell > 0$, things become more tricky. Wright [856], [857], [858] showed that $C_{k,k+\ell} \approx \gamma_\ell k^{k+(3\ell-1)/2}$ for $\ell = o(k^{1/3})$ where the Wright coefficients γ_ℓ satisfy an explicit recurrence and have been related to Brownian motion, see Aldous [19] and Spencer [811]. In a breakthrough paper, Bender, Canfield and McKay [92] gave an asymptotic formula valid for all k. Łuczak [633] in a beautiful argument simplified a large part of their argument, see Exercise (4.3.6). Bollobás [155] proved the useful simple estimate $C_{k,k+\ell} \leq c\ell^{-\ell/2} k^{k+(3\ell-1)/2}$ for some absolute constant $c > 0$. It is difficult to prove tight statements about $\mathbb{G}_{n,m}$ in the phase transition window without these estimates. Nevertheless, it is possible to see that the largest component should be of order $n^{2/3}$, using a nice argument from Nachmias and Peres. They have published a stronger version of this argument in [705].

Theorem 2.21. *Let $p = \frac{1}{n}$ and A be a large constant. Let Z be the size of the largest component in $\mathbb{G}_{n,p}$. Then*

$$(i) \quad \mathbb{P}\left(Z \leq \frac{1}{A} n^{2/3}\right) = O(A^{-1}),$$

$$(ii)\ \ \mathbb{P}\left(Z \geq An^{2/3}\right) = O(A^{-1}).$$

Proof. We will prove part (i) of the theorem first. This is a standard application of the first moment method, see for example Bollobás [155]. Let X_k be the number of tree components of order k and let $k \in \left[\frac{1}{A}n^{2/3}, An^{2/3}\right]$. Then, see also (2.10),

$$\mathbb{E}X_k = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1}.$$

But

$$\begin{aligned}(1-p)^{k(n-k)+\binom{k}{2}-k+1} &\approx (1-p)^{kn-k^2/2} \\ &= \exp\{(kn-k^2/2)\log(1-p)\} \\ &\approx \exp\left\{-\frac{kn-k^2/2}{n}\right\}.\end{aligned}$$

Hence, by the above and Lemma 27.2,

$$\mathbb{E}X_k \approx \frac{n}{\sqrt{2\pi}\,k^{5/2}} \exp\left\{-\frac{k^3}{6n^2}\right\}. \tag{2.20}$$

So if

$$X = \sum_{\frac{1}{A}n^{2/3}}^{An^{2/3}} X_k,$$

then

$$\begin{aligned}\mathbb{E}X &\approx \frac{1}{\sqrt{2\pi}} \int_{x=\frac{1}{A}}^{A} \frac{e^{-x^3/6}}{x^{5/2}} dx \\ &= \frac{4}{3\sqrt{\pi}} A^{3/2} + O(A^{1/2}).\end{aligned}$$

Arguing as in Lemma 2.12 we see that

$$\mathbb{E}X_k^2 \leq \mathbb{E}X_k + (1+o(1))(\mathbb{E}X_k)^2,$$

$$\mathbb{E}(X_k X_l) \leq (1+o(1))(\mathbb{E}X_k)(\mathbb{E}X_l), \ \ k \neq l.$$

It follows that

$$\mathbb{E}X^2 \leq \mathbb{E}X + (1+o(1))(\mathbb{E}X)^2.$$

Applying the second moment method, Lemma 26.6, we see that

$$\mathbb{P}(X > 0) \geq \frac{1}{(\mathbb{E}X)^{-1} + 1 + o(1)}$$

$$= 1 - O(A^{-1}),$$

which completes the proof of part (i).

To prove (ii) we first consider a breadth first search (BFS) starting from, say, vertex x. We construct a sequence of sets $S_1 = \{x\}, S_2, \ldots$, where

$$S_{i+1} = \{v \notin \bigcup_{j \le i} S_j : \exists w \in S_i \text{ such that } (v,w) \in E(\mathbb{G}_{n,p})\}.$$

We have

$$\begin{aligned}\mathbb{E}(|S_{i+1}|\ |S_i) &\le (n - |S_i|)\left(1 - (1-p)^{|S_i|}\right)\\ &\le (n - |S_i|)|S_i|p\\ &\le |S_i|.\end{aligned}$$

So

$$\mathbb{E}\,|S_{i+1}| \le \mathbb{E}\,|S_i| \le \cdots \le \mathbb{E}\,|S_1| = 1. \tag{2.21}$$

We prove next that

$$\pi_k = \mathbb{P}(S_k \ne \emptyset) \le \frac{4}{k}. \tag{2.22}$$

This is clearly true for $k \le 4$ and we obtain (2.22) by induction from

$$\pi_{k+1} \le \sum_{i=1}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} (1 - (1-\pi_k)^i). \tag{2.23}$$

To explain the above inequality note that we can couple the construction of $S_1, S_2, \ldots, S_k$ with a (branching) process where $T_1 = \{1\}$ and T_{k+1} is obtained from T_k as follows: each T_k independently spawns $\text{Bin}(n-1, p)$ individuals. Note that $|T_k|$ stochastically dominates $|S_k|$. This is because in the BFS process, each $w \in S_k$ gives rise to at most $\text{Bin}(n-1, p)$ new vertices. Inequality (2.23) follows, because $T_{k+1} \ne \emptyset$ implies that at least one of 1's children give rise to descendants at level k. Going back to (2.23) we get

$$\begin{aligned}\pi_{k+1} &\le 1 - (1-p)^{n-1} - (1 - p + p(1-\pi_k))^{n-1} + (1-p)^{n-1}\\ &= 1 - (1 - p\pi_k)^{n-1}\\ &\le 1 - 1 + (n-1)p\pi_k - \binom{n-1}{2} p^2 \pi_k^2 + \binom{n-1}{3} p^3 \pi_k^3\\ &\le \pi_k - \left(\frac{1}{2} + o(1)\right)\pi_k^2 + \left(\frac{1}{6} + o(1)\right)\pi_k^3\end{aligned}$$

$$
\begin{aligned}
&= \pi_k \left(1 - \pi_k \left(\left(\frac{1}{2} + o(1)\right) - \left(\frac{1}{6} + o(1)\right)\pi_k\right)\right) \\
&\le \pi_k \left(1 - \frac{1}{4}\pi_k\right).
\end{aligned}
$$

This expression increases for $0 \le \pi_k \le 1$ and immediately gives $\pi_5 \le 3/4 \le 4/5$. In general we have by induction that

$$\pi_{k+1} \le \frac{4}{k}\left(1 - \frac{1}{k}\right) \le \frac{4}{k+1},$$

completing the inductive proof of (2.22).

Let C_x be the component containing x and let $\rho_x = \max\{k : S_k \ne \emptyset\}$ in the BFS from x. Let

$$X = \left|\left\{x : |C_x| \ge n^{2/3}\right\}\right| \le X_1 + X_2,$$

where

$$
\begin{aligned}
X_1 &= \left|\left\{x : |C_x| \ge n^{2/3} \text{ and } \rho_x \le n^{1/3}\right\}\right|, \\
X_2 &= \left|\left\{x : \rho_x > n^{1/3}\right\}\right|.
\end{aligned}
$$

It follows from (2.22) that

$$\mathbb{P}(\rho_x > n^{1/3}) \le \frac{4}{n^{1/3}}$$

and so

$$\mathbb{E}X_2 \le 4n^{2/3}.$$

Furthermore,

$$
\begin{aligned}
&\mathbb{P}\left\{|C_x| \ge n^{2/3} \text{ and } \rho_x \le n^{1/3}\right\} \\
&\le \mathbb{P}\left(|S_1| + \ldots + |S_{n^{1/3}}| \ge n^{2/3}\right) \\
&\le \frac{\mathbb{E}(|S_1| + \ldots + |S_{n^{1/3}}|)}{n^{2/3}} \\
&\le \frac{1}{n^{1/3}},
\end{aligned}
$$

after using (2.21). So $\mathbb{E}X_1 \le n^{2/3}$ and $\mathbb{E}X \le 5n^{2/3}$.
Now let C_{max} denote the size of the largest component. Now

$$C_{max} \le |X| + n^{2/3}$$

where the addition of $n^{2/3}$ accounts for the case where $X = 0$.

So we have

$$\mathbb{E} C_{max} \leq 6n^{2/3}$$

and part (ii) of the theorem follows from the Markov inequality (see Lemma 26.1).

□

2.4 Exercises

2.4.1 Prove Theorem 2.15.

2.4.2 Show that if $p = \omega/n$ where $\omega = \omega(n) \to \infty$ then w.h.p. $\mathbb{G}_{n,p}$ contains no unicyclic components. (A component is *unicyclic* if it contains exactly one cycle i.e. is a tree plus one extra edge).

2.4.3 Prove Theorem 2.17.

2.4.4 Suppose that $m = cn/2$ where $c > 1$ is a constant. Let C_1 denote the giant component of $\mathbb{G}_{n,m}$, assuming that it exists. Suppose that C_1 has $n' \leq n$ vertices and $m' \leq m$ edges. Let G_1, G_2 be two connected graphs with n' vertices from $[n]$ and m' edges. Show that

$$\mathbb{P}(C_1 = G_1) = \mathbb{P}(C_1 = G_2).$$

(I.e. C_1 is a uniformly random connected graph with n' vertices and m' edges).

2.4.5 Suppose that Z is the length of the cycle in a randomly chosen connected unicyclic graph on vertex set $[n]$. Show that, where $N = \binom{n}{2}$,

$$\mathbb{E} Z = \frac{n^{n-2}(N-n+1)}{C(n,n)}.$$

2.4.6 Suppose that $c < 1$. Show that w.h.p. the length of the longest path in $\mathbb{G}_{n,p}$, $p = \frac{c}{n}$ is $\approx \frac{\log n}{\log 1/c}$.

2.4.7 Suppose that $c \neq 1$ is constant. Show that w.h.p. the number of edges in the largest component that is a path in $\mathbb{G}_{n,p}$, $p = \frac{c}{n}$ is $\approx \frac{\log n}{c - \log c}$.

2.4.8 Let $\mathbb{G}_{n,n,p}$ denote the random bipartite graph derived from the complete bipartite graph $K_{n,n}$ where each edge is included independently with probability p. Show that if $p = c/n$ where $c > 1$ is a constant then w.h.p. $\mathbb{G}_{n,n,p}$ has a unique giant component of size $\approx 2G(c)n$ where $G(c)$ is as in Theorem 2.14.

2.4.9 Consider the bipartite random graph $\mathbb{G}_{n,n,p=c/n}$, with constant $c > 1$. Define $0 < x < 1$ to be the solution to $xe^{-x} = ce^{-c}$. Prove that w.h.p. the 2-core of $\mathbb{G}_{n,n,p=c/n}$ has $\approx 2(1-x)\left(1-\frac{x}{c}\right)n$ vertices and $\approx c\left(1-\frac{x}{c}\right)^2 n$ edges.

2.4.10 Let $p = \frac{1+\varepsilon}{n}$. Show that if ε is a small positive constant then w.h.p. $\mathbb{G}_{n,p}$ contains a giant component of size $(2\varepsilon + O(\varepsilon^2))n$.

2.4.11 Let $m = \frac{n}{2} + s$, where $s = s(n) \geq 0$. Show that if $s \gg n^{2/3}$ then w.h.p. the random graph $\mathbb{G}_{n,m}$ contains exactly one complex component. (A component C is *complex* if it contains at least two distinct cycles. In terms of edges, C is complex iff it contains at last $|C|+1$ edges).

2.4.12 Let $m_k(n) = n(\log n + (k-1)\log\log n + \omega)/(2k)$, where $|\omega| \to \infty, |\omega| = o(\log n)$. Show that

$$\mathbb{P}(\mathbb{G}_{m_k} \not\supseteq k\text{-vertex-tree-component}) = \begin{cases} o(1) & \text{if } \omega \to -\infty \\ 1-o(1) & \text{if } \omega \to \infty \end{cases}.$$

2.4.13 Let $k \geq 3$ be fixed and let $p = \frac{c}{n}$. Show that if c is sufficiently large, then w.h.p. the k-core of $\mathbb{G}_{n,p}$ is non-empty.

2.4.14 Let $k \geq 3$ be fixed and let $p = \frac{c}{n}$. Show that there exists $\theta = \theta(c,k) > 0$ such that w.h.p. all vertex sets S with $|S| \leq \theta n$ contain fewer than $k|S|/2$ edges. Deduce that w.h.p. either the k-core of $\mathbb{G}_{n,p}$ is empty or it has size at least θn.

2.4.15 Suppose that $p = \frac{c}{n}$ where $c > 1$ is a constant. Show that w.h.p. the giant component of $\mathbb{G}_{n,p}$ is non-planar. (Hint: Assume that $c = 1+\varepsilon$ where ε is small. Remove a few vertices from the giant so that the girth is large. Now use Euler's formula).

2.4.16 Show that if $\omega = \omega(n) \to \infty$ then w.h.p. $\mathbb{G}_{n,p}$ has at most ω complex components.

2.4.17 Suppose that $np \to \infty$ and $3 \leq k = O(1)$. Show that $\mathbb{G}_{n,p}$ contains a k-cycle w.h.p.

2.4.18 Suppose that $p = c/n$ where $c > 1$ is constant and let $\beta = \beta(c)$ be the smallest root of the equation

$$\frac{1}{2}c\beta + (1-\beta)ce^{-c\beta} = \log\left(c(1-\beta)^{(\beta-1)/\beta}\right).$$

(a) Show that if $\omega/\log n \to \infty$ and $\omega \leq k \leq \beta n$ then w.h.p. $\mathbb{G}_{n,p}$ contains no *maximal* induced tree of size k.

(b) Show that w.h.p. $\mathbb{G}_{n,p}$ contains an induced tree of size $(\log n)^2$.

(c) Deduce that w.h.p. $\mathbb{G}_{n,p}$ contains an induced tree of size at least βn.

2.4.19 Show that if $c \neq 1$ and $xe^{-x} = ce^{-c}$ where $0 < x < 1$ then

$$\frac{1}{c}\sum_{k=1}^{\infty}\frac{k^{k-2}}{k!}(ce^{-c})^k = \begin{cases} 1-\frac{c}{2} & c < 1. \\ \frac{x}{c}\left(1-\frac{x}{2}\right) & c > 1. \end{cases}$$

2.4.20 Let $G_{N,M}^{\delta \geq k}$ denote a graph chosen uniformly at random from the set of graphs with vertex set $[N]$, M edges and minimum degree at least k. Let C_k denote the k core of $G_{n,m}$ (if it exists). Show that conditional on $|C_k| = N$ and $|E(C_k)| = M$ that the graph induced by C_k is distributed as $G_{N,M}^{\delta \geq k}$.

2.4.21 Let $p = c/n$. Run the Breadth First Search algorithm on $G_{n,p}$. Denote by S the set of vertices that have already been used and uncovered, Q the set of active vertices in the queue, and T the remaining vertices. Denote by $q(s)$ the size of Q at the time that $|S| = s$. Assume that the first vertex to enter Q belongs to the giant component. Then, finding the expected value of s for which $q(s) = 0$ again will give us the size of the giant. Use the Differential Equations method of Chapter 28 to obtain the size of the giant given in Theorem 2.14. (This idea was given to us in a private communication by Sahar Diskin and Michael Krivelevich.)

2.5 Notes

Phase transition

The paper by Łuczak, Pittel and Wierman [642] contains a great deal of information about the phase transition. In particular, [642] shows that if $m = n/2 + \lambda n^{2/3}$ then the probability that $\mathbb{G}_{n,m}$ is planar tends to a limit $p(\lambda)$, where $p(\lambda) \to 0$ as $\lambda \to \infty$. The landmark paper by Janson, Knuth, Łuczak and Pittel [513] gives the most detailed analysis to date of the events in the scaling window. Ambroggio and

Roberts [49], [50] discuss the probability of finding unusually large components in the scaling window and in critical percolation on random regular graphs. Ambroggio in [48] discusses the case of unusually small maximal components in the same models.

Outside of the critical window $\frac{n}{2} \pm O(n^{2/3})$ the size of the largest component is asymptotically determined. Theorem 2.17 describes $\mathbb{G}_{n,m}$ before reaching the window and on the other hand a unique "giant" component of size $\approx 4s$ begins to emerge at around $m = \frac{n}{2} + s$, for $s \gg n^{2/3}$. Ding, Kim, Lubetzky and Peres [296] give a useful model for the structure of this giant.

Achlioptas processes

Dimitris Achlipotas proposed the following variation on the basic graph process. Suppose that instead of adding a random edge e_i to add to $\mathbb{G}_{i-1}$ to create $\mathbb{G}_i$, one is given a choice of two random edges e_i, f_i and one chooses one of them to add. He asked whether it was possible to come up with a choice rule that would delay the occurrence of some graph property $\mathscr{P}$. As an initial challenge he asked whether it was possible to delay the production of a giant component beyond $n/2$. Bohman and Frieze [137] showed that this was possible by the use of a simple rule. Since that time this has grown into a large area of research. Kang, Perkins and Spencer [545] have given a more detailed analysis of the "Bohman-Frieze" process. Bohman and Kravitz [144] and in greater generality Spencer and Wormald [813] analyse "bounded size algorithms" in respect of avoiding giant components. Flaxman, Gamarnik and Sorkin [368] consider how to speed up the occurrence of a giant component. Riordan and Warnke [764] discuss the speed of transition at a critical point in an Achlioptas process.

The above papers concern component structure. Krivelevich, Loh and Sudakov [595] considered rules for avoiding specific subgraphs. Krivelevich, Lubetzky and Sudakov [596] discuss rules for speeding up Hamiltonicity.

Graph Minors

Fountoulakis, Kühn and Osthus [374] show that for every $\varepsilon > 0$ there exists C_ε such that if $np > C_\varepsilon$ and $p = o(1)$ then w.h.p. $\mathbb{G}_{n,p}$ contains a complete minor of size $(1 \pm \varepsilon)\left(\frac{n^2 p}{\log np}\right)$. This improves earlier results of Bollobás, Catlin and Erdős [159] and Krivelevich and Sudakov [603]. Ajtai, Komlós and Szemerédi [13] showed that if $np \geq 1 + \varepsilon$ and $np = o(n^{1/2})$ then w.h.p. $\mathbb{G}_{n,p}$ contains a toplogical clique of size almost as large as the maximum degree. If we know that $\mathbb{G}_{n,p}$ is non-planar w.h.p. then it makes sense to determine its *thickness*. This is the minimum number of planar graphs whose union is the whole graph. Cooper [243] showed that the thickness of $\mathbb{G}_{n,p}$ is strongly related to its *arboricity* and is asymptotic to $np/2$ for a large range of p.

Chapter 3

Vertex Degrees

In this chapter we study some typical properties of the degree sequence of a random graph. We begin by discussing the typical degrees in a sparse random graph i.e. one with $O(n)$ edges and prove some results on the asymptotic distribution of degrees. Next we look at the typical values of the minimum and maximum degrees in dense random graphs. We then describe a simple canonical labelling algorithm for the graph isomorphism problem on a dense random graph.

3.1 Degrees of Sparse Random Graphs

Recall that the degree of an individual vertex of $\mathbb{G}_{n,p}$ is a Binomial random variable with parameters $n-1$ and p. One should also notice that the degrees of different vertices are only mildly correlated.

We will first prove some simple but often useful properties of vertex degrees when $p = o(1)$. Let $X_0 = X_{n,0}$ be the number of isolated vertices in $\mathbb{G}_{n,p}$. In Lemma 1.11, we established the sharp threshold for "disappearance" of such vertices. Now we will be more precise and determine the asymptotic distribution of X_0 "below", "on" and "above" the threshold. Obviously,

$$\mathbb{E}X_0 = n(1-p)^{n-1},$$

and an easy computation shows that, as $n \to \infty$,

$$\mathbb{E}X_0 \to \begin{cases} \infty & \text{if } np - \log n \to -\infty \\ e^{-c} & \text{if } np - \log n \to c, \;\; c < \infty, \\ 0 & \text{if } np - \log n \to \infty \end{cases} \tag{3.1}$$

We denote by $Po(\lambda)$ a random variable with the Poisson distribution with parameter λ, while $N(0,1)$ denotes the random variable with the Standard Normal distribution. We write $X_n \xrightarrow{D} X$ to say that a random variable X_n *converges in distribution* to a random variable X, as $n \to \infty$.

The following theorem shows that the asymptotic distribution of X_0 passes through three phases: it starts in the Normal phase; next when isolated vertices are close to "dying out", it moves through a Poisson phase; it finally ends up at the distribution concentrated at 0.

Theorem 3.1. *Let X_0 be the random variable counting isolated vertices in a random graph $\mathbb{G}_{n,p}$. Then, as $n \to \infty$,*

(i) $\tilde{X}_0 = (X_0 - \mathbb{E}X_0)/(\mathrm{Var}X_0)^{1/2} \xrightarrow{D} \mathrm{N}(0,1)$, *if* $n^2p \to \infty$ *and* $np - \log n \to -\infty$,

(ii) $X_0 \xrightarrow{D} \mathrm{Po}(e^{-c})$, *if* $np - \log n \to c,\ \ c < \infty$,

(iii) $X_0 \xrightarrow{D} 0$, *if* $np - \log n \to \infty$.

Proof. For the proof of (i) we refer the reader to Chapter 6 of Janson, Łuczak and Ruciński [509] (or to [76] and [586]).

To prove (ii) one has to show that if $p = p(n)$ is such that $np - \log n \to c$, then

$$\lim_{n\to\infty} \mathbb{P}(X_0 = k) = \frac{e^{-ck}}{k!} e^{-e^{-c}}, \tag{3.2}$$

for $k = 0, 1, \dots$. Now,

$$X_0 = \sum_{v \in V} I_v,$$

where

$$I_v = \begin{cases} 1 & \text{if } v \text{ is an isolated vertex in } \mathbb{G}_{n,p} \\ 0 & \text{otherwise.} \end{cases}$$

So

$$\begin{aligned} \mathbb{E}X_0 &= \sum_{v \in V} \mathbb{E}I_v = n(1-p)^{n-1} \\ &= n\exp\{(n-1)\log(1-p)\} \\ &= n\exp\left\{-(n-1)\sum_{k=1}^{\infty} \frac{p^k}{k}\right\} \\ &= n\exp\left\{-(n-1)p + O(np^2)\right\} \\ &= n\exp\left\{-(\log n + c) + O\left(\frac{(\log n)^2}{n}\right)\right\} \end{aligned}$$

$$\approx e^{-c}. \tag{3.3}$$

The easiest way to show that (3.2) holds is to apply the Method of Moments (see Chapter 26). Briefly, since the distribution of the random variable X_0 is uniquely determined by its moments, it is enough to show, that either the kth factorial moment $\mathbb{E}X_0(X_0-1)\cdots(X_0-k+1)$ of X_0, or its binomial moment $\mathbb{E}\binom{X_0}{k}$, tend to the respective moments of the Poisson distribution, i.e., to either e^{-ck} or $e^{-ck}/k!$. We choose the binomial moments, and so let

$$B_k^{(n)} = \mathbb{E}\binom{X_0}{k},$$

then, for every non-negative integer k,

$$\begin{aligned} B_k^{(n)} &= \sum_{1\le i_1<i_2<\cdots<i_k\le n} \mathbb{P}(I_{v_{i_1}}=1, I_{v_{i_2}}=1,\ldots,I_{v_{i_k}}=1), \\ &= \binom{n}{k}(1-p)^{k(n-k)+\binom{k}{2}}. \end{aligned}$$

Hence

$$\lim_{n\to\infty} B_k^{(n)} = \frac{e^{-ck}}{k!},$$

and part (ii) of the theorem follows by Theorem 26.11, with $\lambda = e^{-c}$.

For part (iii), suppose that $np = \log n + \omega$ where $\omega \to \infty$. We repeat the calculation estimating $\mathbb{E}X_0$ and replace $\approx e^{-c}$ in (3.3) by $\le (1+o(1))e^{-\omega} \to 0$ and apply the first moment method. $\square$

From the above theorem we immediately see that if $np - \log n \to c$ then

$$\lim_{n\to\infty} \mathbb{P}(X_0 = 0) = e^{-e^{-c}}. \tag{3.4}$$

We next give a more general result describing the asymptotic distribution of the number $X_d = X_{n,d}$, $d \ge 1$ of vertices of any fixed degree d in a random graph.

Recall, that the degree of a vertex in $\mathbb{G}_{n,p}$ has the binomial distribution $\mathrm{Bin}(n-1,p)$. Hence,

$$\mathbb{E}X_d = n\binom{n-1}{d} p^d (1-p)^{n-1-d}. \tag{3.5}$$

Therefore, as $n \to \infty$,

$$\mathbb{E}X_d \to \begin{cases} 0 & \text{if } p \ll n^{-(d+1)/d}, \\ \lambda_1 & \text{if } p \approx cn^{-(d+1)/d},\ c<\infty, \\ \infty & \text{if } p \gg n^{-(d+1)/d)} \text{ but} \\ & \quad pn - \log n - d\log\log n \to -\infty, \\ \lambda_2 & \text{if } pn - \log n - d\log\log n \to c,\ c<\infty, \\ 0 & \text{if } pn - \log n - d\log\log n \to \infty, \end{cases} \tag{3.6}$$

where

$$\lambda_1 = \frac{c^d}{d!} \quad \text{and} \quad \lambda_2 = \frac{e^{-c}}{d!}. \tag{3.7}$$

The asymptotic behavior of the expectation of the random variable X_d suggests possible asymptotic distributions for X_d, for a given edge probability p.

Theorem 3.2. *Let $X_d = X_{n,d}$ be the number of vertices of degree d, $d \geq 1$, in $\mathbb{G}_{n,p}$ and let λ_1, λ_2 be given by (3.7). Then, as $n \to \infty$,*

(i) $X_d \xrightarrow{D} 0$ *if* $p \ll n^{-(d+1)/d}$,

(ii) $X_d \xrightarrow{D} \mathrm{Po}(\lambda_1)$ *if* $p \approx cn^{-(d+1)/d}$, $c < \infty$,

(iii) $\tilde{X}_d := (X_d - \mathbb{E} X_d)/(\mathrm{Var} X_d)^{1/2} \xrightarrow{D} \mathrm{N}(0,1)$ *if* $p \gg n^{-(d+1)/d}$, *but* $pn - \log n - d \log\log n \to -\infty$

(iv) $X_d \xrightarrow{D} \mathrm{Po}(\lambda_2)$ *if* $pn - \log n - d\log\log n \to c$, $-\infty < c < \infty$,

(v) $X_d \xrightarrow{D} 0$ *if* $pn - \log n - d\log\log n \to \infty$

Proof. The proofs of statements (i) and (v) are straightforward applications of the first moment method, while the proofs of (ii) and (iv) can be found in Chapter 3 of Bollobás [147] (see also Karoński and Ruciński [553] for estimates of the rate of convergence). The proof of (iii) can be found in [76].

□

The next theorem shows the concentration of X_d around its expectation when in $\mathbb{G}_{n,p}$ the edge probability $p = c/n$, i.e., when the average vertex degree is c.

Theorem 3.3. *Let $p = c/n$ where c is a constant. Let X_d denote the number of vertices of degree d in $\mathbb{G}_{n,p}$. Then, for $d = O(1)$, w.h.p.*

$$X_d \approx \frac{c^d e^{-c}}{d!} n.$$

Proof. Assume that vertices of $\mathbb{G}_{n,p}$ are labeled $1, 2, \ldots, n$. We first compute $\mathbb{E} X_d$. Thus,

$$\begin{aligned}\mathbb{E} X_d &= n \mathbb{P}(\deg(1) = d) = \\ &= n \binom{n-1}{d} \left(\frac{c}{n}\right)^d \left(1 - \frac{c}{n}\right)^{n-1-d}\end{aligned}$$

$$= n\frac{n^d}{d!}\left(1+O\left(\frac{d^2}{n}\right)\right)\left(\frac{c}{n}\right)^d \exp\left\{-(n-1-d)\left(\frac{c}{n}+O\left(\frac{1}{n^2}\right)\right)\right\}$$
$$= n\frac{c^d e^{-c}}{d!}\left(1+O\left(\frac{1}{n}\right)\right).$$

We now compute the second moment. For this we need to estimate

$$\begin{aligned}
&\mathbb{P}(\deg(1)=\deg(2)=d)\\
&=\frac{c}{n}\left(\binom{n-2}{d-1}\left(\frac{c}{n}\right)^{d-1}\left(1-\frac{c}{n}\right)^{n-1-d}\right)^2\\
&+\left(1-\frac{c}{n}\right)\left(\binom{n-2}{d}\left(\frac{c}{n}\right)^{d}\left(1-\frac{c}{n}\right)^{n-2-d}\right)^2\\
&=\mathbb{P}(\deg(1)=d)\,\mathbb{P}(\deg(2)=d)\left(1+O\left(\frac{1}{n}\right)\right).
\end{aligned}$$

The first line here accounts for the case where $\{1,2\}$ is an edge and the second line deals with the case where it is not.

Thus

$$\begin{aligned}
&\operatorname{Var} X_d =\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathbb{P}(\deg(i)=d,\deg(j)=d)-\mathbb{P}(\deg(1)=d)\,\mathbb{P}(\deg(2)=d)\right]\\
&\leq \sum_{i\neq j=1}^{n} O\left(\frac{1}{n}\right)+\mathbb{E}X_d \leq An,
\end{aligned}$$

for some constant $A=A(c)$.

Applying the Chebyshev inequality (Lemma 26.3), we obtain

$$\mathbb{P}(|X_d-\mathbb{E}X_d|\geq tn^{1/2})\leq \frac{A}{t^2},$$

which completes the proof. □

We conclude this section with a look at the asymptotic behavior of the maximum vertex degree, when a random graph is sparse.

Theorem 3.4. *Let $\Delta(\mathbb{G}_{n,p})$ ($\delta(\mathbb{G}_{n,p})$) denotes the maximum (minimum) degree of vertices of $\mathbb{G}_{n,p}$.*

(i) If $p=c/n$ for some constant $c>0$ then w.h.p.

$$\Delta(\mathbb{G}_{n,p})\approx \frac{\log n}{\log\log n}.$$

(ii) If $np = \omega \log n$ where $\omega \to \infty$, then w.h.p. $\delta(\mathbb{G}_{n,p}) \approx \Delta(\mathbb{G}_{n,p}) \approx np$.

Proof. (i) Let $d_\pm = \left\lceil \frac{\log n}{\log\log n \pm 2\log\log\log n} \right\rceil$. Then, if $d = d_-$,

$$\begin{aligned}
\mathbb{P}(\exists v : deg(v) \geq d) &\leq n \binom{n-1}{d} \left(\frac{c}{n}\right)^d \\
&\leq n \left(\frac{ce}{d}\right)^d \\
&= \exp\{\log n - d\log d + O(d)\} \qquad (3.8)
\end{aligned}$$

Let $\lambda = \frac{\log\log\log n}{\log\log n}$. Then

$$\begin{aligned}
d\log d &\geq \frac{\log n}{\log\log n} \cdot \frac{1}{1-2\lambda} \cdot (\log\log n - \log\log\log n + o(1)) \\
&= \frac{\log n}{\log\log n}(1 + 2\lambda + O(\lambda^2))(\log\log n - \log\log\log n + o(1)) \\
&= \frac{\log n}{\log\log n}(\log\log n + \log\log\log n + o(1)). \qquad (3.9)
\end{aligned}$$

Plugging this into (3.8) shows that $\Delta(\mathbb{G}_{n,p}) \leq d_-$ w.h.p.

Now let $d = d_+$ and let X_d be the number of vertices of degree d in $\mathbb{G}_{n,p}$. Then

$$\begin{aligned}
&\mathbb{E}(X_d) = n \binom{n-1}{d} \left(\frac{c}{n}\right)^d \left(1 - \frac{c}{n}\right)^{n-d-1} \\
&= \exp\{\log n - d\log d + O(d)\} \\
&= \exp\left\{\log n - \frac{\log n}{\log\log n}(\log\log n - \log\log\log n + o(1)) + O(d)\right\} \qquad (3.10) \\
&\to \infty.
\end{aligned}$$

Here (3.10) is obtained by using $-\lambda$ in place of λ in the argument for (3.9). Now, for vertices v, w, by the same argument as in the proof of Theorem 3.3, we have

$$\mathbb{P}(deg(v) = deg(w) = d) = (1 + o(1))\,\mathbb{P}(deg(v) = d)\,\mathbb{P}(deg(w) = d),$$

and the Chebyshev inequality implies that $X_d > 0$ w.h.p. This completes the proof of (i).

Statement (ii) is an easy consequence of the Chernoff bounds, Corollary 27.7. Let $\varepsilon = \omega^{-1/3}$. Then

$$\mathbb{P}(\exists v : |deg(v) - np| \geq \varepsilon np) \leq 2ne^{-\varepsilon^2 np/3} = 2n^{-\omega^{1/3}/3} = o(n^{-1}).$$

$\square$

3.2 Degrees of Dense Random Graphs

In this section we will concentrate on the case where edge probability p is constant and see how the degree sequence can be used to solve the graph isomorphism problem w.h.p. The main result deals with the maximum vertex degree in dense random graph and is instrumental in the solution of this problem.

Theorem 3.5. *Let $d_\pm = (n-1)p + (1\pm\varepsilon)\sqrt{2(n-1)pq\log n}$, where $q = 1-p$. If p is constant and $\varepsilon > 0$ is a small constant, then w.h.p.*

(i) $d_- \le \Delta(\mathbb{G}_{n,p}) \le d_+$.

(ii) There is a unique vertex of maximum degree.

Proof. We break the proof of Theorem 3.5 into two lemmas.

Lemma 3.6. *Let $d = (n-1)p + x\sqrt{(n-1)pq}$, p be constant, $x \le n^{1/3}(\log n)^2$, where $q = 1-p$. Then*

$$B_d = \binom{n-1}{d} p^d (1-p)^{n-1-d} = (1+o(1))\sqrt{\frac{1}{2\pi npq}}e^{-x^2/2}$$
$$= \textit{the probability that an individual vertex has degree } d.$$

Proof. Stirling's formula gives

$$B_d = (1+o(1))\sqrt{\frac{1}{2\pi npq}}\left(\left(\frac{(n-1)p}{d}\right)^{\frac{d}{n-1}}\left(\frac{(n-1)q}{n-1-d}\right)^{1-\frac{d}{n-1}}\right)^{n-1}. \tag{3.11}$$

Now

$$\left(\frac{d}{(n-1)p}\right)^{\frac{d}{n-1}} = \left(1+x\sqrt{\frac{q}{(n-1)p}}\right)^{\frac{d}{n-1}} =$$
$$= \exp\left\{\left(x\sqrt{\frac{q}{(n-1)p}} - \frac{x^2q}{2(n-1)p} + O\left(\frac{x^3}{n^{3/2}}\right)\right)\left(p + x\sqrt{\frac{pq}{n-1}}\right)\right\}$$
$$= \exp\left\{x\sqrt{\frac{pq}{n-1}} + \frac{x^2q}{2(n-1)} + O\left(\frac{x^3}{n^{3/2}}\right)\right\},$$

whereas

$$\left(\frac{n-1-d}{(n-1)q}\right)^{1-\frac{d}{n-1}} = \left(1-x\sqrt{\frac{p}{(n-1)q}}\right)^{1-\frac{d}{n-1}} =$$
$$= \exp\left\{-\left(x\sqrt{\frac{p}{(n-1)q}} + \frac{x^2p}{2(n-1)q} + O\left(\frac{x^3}{n^{3/2}}\right)\right)\left(q - x\sqrt{\frac{pq}{n-1}}\right)\right\}$$

$$= \exp\left\{-x\sqrt{\frac{pq}{n-1}} + \frac{x^2 p}{2(n-1)} + O\left(\frac{x^3}{n^{3/2}}\right)\right\},$$

So

$$\left(\frac{d}{(n-1)p}\right)^{\frac{d}{n-1}} \left(\frac{n-1-d}{(n-1)q}\right)^{1-\frac{d}{n-1}} = \exp\left\{\frac{x^2}{2(n-1)} + O\left(\frac{x^3}{n^{3/2}}\right)\right\},$$

and lemma follows from (3.11). □

The next lemma proves a strengthing of Theorem 3.5.

Lemma 3.7. *Let $\varepsilon = 1/10$, and p be constant and $q = 1-p$. If*

$$d_{\pm} = (n-1)p + (1\pm\varepsilon)\sqrt{2(n-1)pq\log n}.$$

then w.h.p.

(i) $\Delta(\mathbb{G}_{n,p}) \le d_+$,

(ii) There are $\Omega(n^{2\varepsilon(1-\varepsilon)})$ *vertices of degree at least* d_-,

(iii) $\nexists\, u \neq v$ *such that* $deg(u), deg(v) \ge d_-$ *and* $|deg(u) - deg(v)| \le 10$.

Proof. We first prove that as $x \to \infty$,

$$\frac{1}{x}e^{-x^2/2}\left(1-\frac{1}{x^2}\right) \le \int_x^\infty e^{-y^2/2}dy \le \frac{1}{x}e^{-x^2/2}. \tag{3.12}$$

To see this, notice

$$\begin{aligned}
\int_x^\infty e^{-y^2/2}dy &= -\int_x^\infty \frac{1}{y}\left(e^{-y^2/2}\right)' dy \\
&= -\left[\frac{1}{y}e^{-y^2/2}\right]_x^\infty - \int_x^\infty \frac{1}{y^2}e^{-y^2/2}dy \\
&= \frac{1}{x}e^{-x^2/2} + \left[\frac{1}{y^3}e^{-y^2/2}\right]_x^\infty + 3\int_x^\infty \frac{1}{y^4}e^{-y^2/2}dy \\
&= \frac{1}{x}e^{-x^2/2}\left(1-\frac{1}{x^2}\right) + O\left(\frac{1}{x^4}e^{-x^2/2}\right).
\end{aligned}$$

We can now prove statement (i).

Let X_d be the number of vertices of degree d. Then $\mathbb{E}X_d = nB_d$ and so Lemma 3.6 implies that

$$\mathbb{E}X_d = (1+o(1))\sqrt{\frac{n}{2\pi pq}}\exp\left\{-\frac{1}{2}\left(\frac{d-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\}$$

assuming that
$$d \le d_L = (n-1)p + (\log n)^2\sqrt{(n-1)pq}.$$

Also, if $d > (n-1)p$ then
$$\frac{B_{d+1}}{B_d} = \frac{(n-d-1)p}{(d+1)q} < 1$$
and so if $d \ge d_L$,
$$\mathbb{E}X_d \le \mathbb{E}X_{d_L} \le n\exp\{-\Omega(\log n)^4\}.$$
It follows that
$$\Delta(\mathbb{G}_{n,p}) \le d_L \quad w.h.p. \tag{3.13}$$
Now if $Y_d = X_d + X_{d+1} + \cdots + X_{d_L}$ for $d = d_\pm$ then
$$\begin{aligned}\mathbb{E}Y_d &\approx \sum_{l=d}^{d_L}\sqrt{\frac{n}{2\pi pq}}\exp\left\{-\frac{1}{2}\left(\frac{l-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\} \\ &\approx \sum_{l=d}^{\infty}\sqrt{\frac{n}{2\pi pq}}\exp\left\{-\frac{1}{2}\left(\frac{l-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\} \qquad (3.14) \\ &\approx \sqrt{\frac{n}{2\pi pq}}\int_{\lambda=d}^{\infty}\exp\left\{-\frac{1}{2}\left(\frac{\lambda-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\}d\lambda.\end{aligned}$$
The justification for (3.14) comes from
$$\sum_{l=d_L}^{\infty}\sqrt{\frac{n}{2\pi pq}}\exp\left\{-\frac{1}{2}\left(\frac{l-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\} = \\ = O(n)\sum_{x=(\log n)^2}^{\infty} e^{-x^2/2} = O(e^{-(\log n)^2/3}),$$
and
$$\sqrt{\frac{n}{2\pi pq}}\exp\left\{-\frac{1}{2}\left(\frac{d_+-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\} = n^{-O(1)}.$$

If $d = (n-1)p + x\sqrt{(n-1)pq}$ then, from (3.12) we have
$$\mathbb{E}Y_d \approx \sqrt{\frac{n}{2\pi pq}}\int_{\lambda=d}^{\infty}\exp\left\{-\frac{1}{2}\left(\frac{\lambda-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\}d\lambda$$

$$
\begin{aligned}
&= \sqrt{\frac{n}{2\pi pq}}\sqrt{(n-1)pq}\int_{y=x}^{\infty} e^{-y^2/2}dy \\
&\approx \frac{n}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2} \\
&\begin{cases} \le n^{-2\varepsilon(1+\varepsilon)} & d = d_+ \\ \ge n^{2\varepsilon(1-\varepsilon)} & d = d_- \end{cases}.
\end{aligned} \tag{3.15}
$$

Part (i) follows from (3.15).

When $d = d_-$ we see from (3.15) that $\mathbb{E} Y_d \to \infty$. We use the second moment method to show that $Y_{d_-} \neq 0$ w.h.p.

$$
\begin{aligned}
&\mathbb{E} Y_d(Y_d - 1) = n(n-1)\sum_{d \le d_1, d_2}^{d_L} \mathbb{P}(deg(1) = d_1, deg(2) = d_2) \\
&= n(n-1)\sum_{d \le d_1, d_2}^{d_L} (p\,\mathbb{P}(\hat{d}(1) = d_1 - 1, \hat{d}(2) = d_2 - 1) \\
&+ (1-p)\,\mathbb{P}(\hat{d}(1) = d_1, \hat{d}(2) = d_2)),
\end{aligned}
$$

where $\hat{d}(x)$ is the number of neighbors of x in $\{3,4,\ldots,n\}$. Note that $\hat{d}(1)$ and $\hat{d}(2)$ are independent, and

$$
\begin{aligned}
\frac{\mathbb{P}(\hat{d}(1) = d_1 - 1)}{\mathbb{P}(\hat{d}(1) = d_1)} &= \frac{\binom{n-2}{d_1-1}(1-p)}{\binom{n-2}{d_1}p} = \frac{d_1(1-p)}{(n-1-d_1)p} \\
&= 1 + \tilde{O}(n^{-1/2}).
\end{aligned}
$$

In $\tilde{O}$ we ignore polylog factors.

Hence

$$
\begin{aligned}
&\mathbb{E}(Y_d(Y_d - 1)) \\
&= n(n-1)\sum_{d \le d_1, d_2}^{d_L} \left[\mathbb{P}(\hat{d}(1) = d_1)\,\mathbb{P}(\hat{d}(2) = d_2)(1 + \tilde{O}(n^{-1/2}))\right] \\
&= n(n-1)\sum_{d \le d_1, d_2}^{d_L} \left[\mathbb{P}(deg(1) = d_1)\,\mathbb{P}(deg(2) = d_2)(1 + \tilde{O}(n^{-1/2}))\right] \\
&= \mathbb{E} Y_d(\mathbb{E} Y_d - 1)(1 + \tilde{O}(n^{-1/2})),
\end{aligned}
$$

since

$$
\frac{\mathbb{P}(\hat{d}(1) = d_1)}{\mathbb{P}(deg(1) = d_1)} = \frac{\binom{n-2}{d_1}}{\binom{n-1}{d_1}}(1-p)^{-1}
$$

$$= 1 + \tilde{O}(n^{-1/2}).$$

So, with $d = d_-$

$$\begin{aligned}
&\mathbb{P}\left(Y_d \le \frac{1}{2}\mathbb{E}Y_d\right) \\
&\le \frac{\mathbb{E}(Y_d(Y_d-1)) + \mathbb{E}Y_d - (\mathbb{E}Y_d)^2}{(\mathbb{E}Y_d)^2/4} \\
&= \tilde{O}\left(\frac{1}{n^{\varepsilon}}\right) \\
&= o(1).
\end{aligned}$$

This completes the proof of statement (ii). Finally,

$$\begin{aligned}
&\mathbb{P}(\neg(iii)) \le o(1) + \binom{n}{2} \sum_{d_1=d_-}^{d_L} \sum_{|d_2-d_1|\le 10} \mathbb{P}(deg(1) = d_1, deg(2) = d_2) \\
&= o(1) + \binom{n}{2} \sum_{d_1=d_-}^{d_L} \sum_{|d_2-d_1|\le 10} \Big[p\,\mathbb{P}(\hat{d}(1) = d_1 - 1)\,\mathbb{P}(\hat{d}(2) = d_2 - 1) \\
&+ (1-p)\,\mathbb{P}(\hat{d}(1) = d_1)\,\mathbb{P}(\hat{d}(2) = d_2)\Big],
\end{aligned}$$

Now

$$\begin{aligned}
&\sum_{d_1=d_-}^{d_L} \sum_{|d_2-d_1|\le 10} \mathbb{P}(\hat{d}(1) = d_1 - 1)\,\mathbb{P}(\hat{d}(2) = d_2 - 1) \\
&\le 21(1 + \tilde{O}(n^{-1/2})) \sum_{d_1=d_-}^{d_L} \left[\mathbb{P}(\hat{d}(1) = d_1 - 1)\right]^2,
\end{aligned}$$

and by Lemma 3.6 and by (3.12) we have with

$$x = \frac{d_- - (n-1)p}{\sqrt{(n-1)pq}} \approx (1-\varepsilon)\sqrt{2\log n},$$

$$\begin{aligned}
\sum_{d_1=d_-}^{d_L} \left[\mathbb{P}(\hat{d}(1) = d_1 - 1)\right]^2 &\approx \frac{1}{2\pi pqn} \int_{y=x}^{\infty} e^{-y^2} dy \\
&= \frac{1}{\sqrt{8\pi pqn}} \int_{z=x\sqrt{2}}^{\infty} e^{-z^2/2} dz \\
&\approx \frac{1}{\sqrt{8\pi pqn}} \frac{1}{x\sqrt{2}} n^{-2(1-\varepsilon)^2},
\end{aligned}$$

We get a similar bound for $\sum_{d_1=d_-}^{d_L} \sum_{|d_2-d_1|\leq 10} \left[\mathbb{P}(\hat{d}(1)=d_1\right]^2$. Thus

$$\begin{aligned}\mathbb{P}(\neg(iii)) &= o\left(n^{2-1-2(1-\varepsilon)^2}\right)\\ &= o(1).\end{aligned}$$

□

Application to graph isomorphism

In this section we describe a procedure for *canonically labelling* a graph G. It is taken from Babai, Erdős and Selkow [57]. If the procedure succeeds then it is possible to quickly tell whether $G \cong H$ for any other graph H. (Here $\cong$ stands for graph isomorphism).

Algorithm *LABEL*

Step 0: Input graph G and parameter L.

Step 1: Re-label the vertices of G so that they satisfy

$$d_G(v_1) \geq d_G(v_2) \geq \cdots \geq d_G(v_n).$$

If there exists $i < L$ such that $d_G(v_i) = d_G(v_{i+1})$, then **FAIL**.

Step 2: For $i > L$ let

$$X_i = \{j \in \{1,2,\ldots,L\} : \{v_i, v_j\} \in E(G)\}.$$

Re-label vertices $v_{L+1}, v_{L+2}, \ldots, v_n$ so that these sets satisfy

$$X_{L+1} \succ X_{L+2} \succ \cdots \succ X_n$$

where $\succ$ denotes lexicographic order.

If there exists $i < n$ such that $X_i = X_{i+1}$ then **FAIL**.

Suppose now that the above ordering/labelling procedure LABEL succeeds for G. Given an n vertex graph H, we run LABEL on H.

(i) If LABEL fails on H then $G \ncong H$.

(ii) Suppose that the ordering generated on $V(H)$ is $w_1, w_2, \ldots, w_n$. Then

$$G \cong H \Leftrightarrow v_i \to w_i \text{ is an isomorphism.}$$

It is straightforward to verify (i) and (ii).

Theorem 3.8. *Let p be a fixed constant, $q = 1 - p$, and let $\rho = p^2 + q^2$ and let $L = 3\log_{1/\rho} n$. Then w.h.p. LABEL succeeds on $\mathbb{G}_{n,p}$.*

Proof. Lemma 3.7 implies that Step 1 succeeds w.h.p. We must now show that w.h.p. $X_i \neq X_j$ for all $i \neq j > L$. There is a slight problem because the edges from $v_i, i > L$ to $v_j, j \leq L$ are conditioned by the fact that the latter vertices are those of highest degree.

Now fix i, j and let $\hat{G} = \mathbb{G}_{n,p} \setminus \{v_i, v_j\}$. It follows from Lemma 3.7 that if $i, j > L$ then w.h.p. the L largest degree vertices of $\hat{G}$ and $\mathbb{G}_{n,p}$ coincide. So, w.h.p., we can compute X_i, X_j with respect to $\hat{G}$ to create $\hat{X}_i, \hat{X}_j$, which are independent of the edges incident with v_i, v_j. It follows that if $i, j > L$ then $\hat{X}_i = X_i$ and $\hat{X}_j = X_j$and this avoids our conditioning problem. Denote by $N_{\hat{G}}(v)$ the set of the neighbors of vertex v in graph $\hat{G}$. Then

$$
\begin{aligned}
&\mathbb{P}(\text{Step 2 fails}) \\
&\leq o(1) + \mathbb{P}(\exists v_i, v_j : N_{\hat{G}}(v_i) \cap \{v_1, \ldots, v_L\} = N_{\hat{G}}(v_j) \cap \{v_1, \ldots, v_L\}) \\
&\leq o(1) + \binom{n}{2}(p^2 + q^2)^L \\
&= o(1).
\end{aligned}
$$

□

Corollary 3.9. *If $0 < p < 1$ is constant then w.h.p. $G_{n,p}$ has a unique automorphism, i.e. the identity automorphism.*

See Exercise 3.3.9.

Application to edge coloring

The *chromatic index* $\chi'(G)$ of a graph G is the minimum number of colors that can be used to color the edges of G so that if two edges share a vertex, then they have a different color. Vizing's theorem states that

$$\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1.$$

Also, if there is a unique vertex of maximum degree, then $\chi'(G) = \Delta(G)$. So, it follows from Theorem 3.5 (ii) that, for constant p, w.h.p. we have $\chi'(\mathbb{G}_{n,p}) = \Delta(\mathbb{G}_{n,p})$.

3.3 Exercises

3.3.1 Suppose that $m = dn/2$ where d is constant. Prove that the number of vertices of degree k in $\mathbb{G}_{n,m}$ is asymptotically equal to $\frac{d^k e^{-d}}{k!} n$ for any fixed positive integer k.

3.3.2 Suppose that $c > 1$ and that $x < 1$ is the solution to $xe^{-x} = ce^{-c}$. Show that if $c = O(1)$ is fixed then w.h.p. the giant component of $\mathbb{G}_{n,p}, p = \frac{c}{n}$ has $\approx \frac{c^k e^{-c}}{k!} \left(1 - \left(\frac{x}{c}\right)^k\right) n$ vertices of degree $k \geq 1$.

3.3.3 Suppose that $p \leq \frac{1+\varepsilon_n}{n}$ where $n^{1/4}\varepsilon_n \to 0$. Show that if Γ is the sub-graph of $\mathbb{G}_{n,p}$ induced by the 2-core C_2, then Γ has maximum degree at most three.

3.3.4 Let $p = \frac{\log n + d \log\log n + c}{n}$, $d \geq 1$. Using the method of moments, prove that the number of vertices of degree d in $\mathbb{G}_{n,p}$ is asymptotically Poisson with mean $\frac{e^{-c}}{d!}$.

3.3.5 Prove parts (i) and (v) of Theorem 3.2.

3.3.6 Show that if $0 < p < 1$ is constant then w.h.p. the minimum degree δ in $\mathbb{G}_{n,p}$ satisfies

$$|\delta - (n-1)q - \sqrt{2(n-1)pq\log n}| \leq \varepsilon\sqrt{2(n-1)pq\log n},$$

where $q = 1 - p$ and $\varepsilon = 1/10$.

3.3.7 Show that if $p = \frac{c\log n}{n}$ where $c > 1$ is a constant then w.h.p.the minimum degree in $\mathbb{G}_{n,p}$ is at least $\alpha_0 \log n$ where α_0 is the smallest root of $\alpha\log(ce/\alpha) = c - 1$.

3.3.8 Show that if $p = \frac{c\log n}{n}$ where $c > 1$ is a constant then w.h.p.the maximum degree in $\mathbb{G}_{n,p}$ is at most $\alpha_1 c \log n$ where α_1 is the largest root of $\alpha\log(ce/\alpha) = c - 1$.

3.3.9 Use the canonical labelling of Theorem 3.8 to show that w.h.p. $G_{n,1/2}$ has exactly one automprphism, the identity automorphism. (An automorphism of a graph $G = (V,E)$ is a map $\varphi : V \to V$ such that $\{x,y\} \in E$ if and only if $\{\varphi(x), \varphi(y)\} \in E$.)

3.4 Notes

For the more detailed account of the properties of the degree sequence of $\mathbb{G}_{n,p}$ the reader is referred to Chapter 3 of Bollobás [155].

Erdős and Rényi [331] and [333] were first to study the asymptotic distribution of the number X_d of vertices of degree d in relation with connectivity of a random graph. Bollobás [151] continued those investigations and provided detailed study of the distribution of X_d in $\mathbb{G}_{n,p}$ when $0 < \liminf np(n)/\log n \leq \limsup np(n)/\log n < \infty$. Palka [722] determined certain range of the edge probability p for which the number of vertices of a given degree of a random graph $\mathbb{G}_{n,p}$ has a Normal distribution. Barbour [73] and Karoński and Ruciński [553] studied the distribution of X_d using the Stein–Chen approach. A complete answer to the asymptotic Normality of X_d was given by Barbour, Karoński and Ruciński [76] (see also Kordecki [586]). Janson [502] extended those results and showed that random variables counting vertices of given degree are jointly normal, when $p \approx c/n$ in $\mathbb{G}_{n,p}$ and $m \approx cn$ in $\mathbb{G}_{n,m}$, where c is a constant.

Ivchenko [491] was the first to analyze the asymptotic behavior of the kth-largest and kth smallest element of the degree sequence of $\mathbb{G}_{n,p}$. In particular he analysed the span between the minimum and the maximum degree of sparse $\mathbb{G}_{n,p}$. Similar results were obtained independently by Bollobás [149] (see also Palka [723]). Bollobás [151] answered the question for what values of $p(n)$, $\mathbb{G}_{n,p}$ w.h.p. has a unique vertex of maximum degree (see Theorem 3.5).

Bollobás [146], for constant p, $0 < p < 1$, i.e., when $\mathbb{G}_{n,p}$ is dense, gave an estimate of the probability that maximum degree does not exceed $pn + O(\sqrt{n \log n})$. A more precise result was proved by Riordan and Selby [761] who showed that for constant p, the probability that the maximum degree of $\mathbb{G}_{n,p}$ does not exceed $pn + b\sqrt{np(1-p)}$, where b is fixed, is equal to $(c+o(1))^n$, for $c = c(b)$ the root of a certain equation. Surprisingly, $c(0) = 0.6102...$ is greater than $1/2$ and $c(b)$ is independent of p.

McKay and Wormald [673] proved that for a wide range of functions $p = p(n)$, the distribution of the degree sequence of $\mathbb{G}_{n,p}$ can be approximated by $\{(X_1,\ldots,X_n) | \sum X_i \text{ is even}\}$, where $X_1,\ldots,X_n$ are independent random variables each having the Binomial distribution $\text{Bin}(n-1,p')$, where p' is itself a random variable with a particular truncated normal distribution

Chapter 4

Connectivity

We first establish, rather precisely, the threshold for connectivity. We then view this property in terms of the graph process and show that w.h.p. the random graph becomes connected at precisely the time when the last isolated vertex joins the giant component. This "hitting time" result is the pre-cursor to several similar results. After this we deal with k-connectivity.

4.1 Connectivity

The first result of this chapter is from Erdős and Rényi [331].

Theorem 4.1. *Let* $m = \frac{1}{2}n(\log n + c_n)$. *Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_m \textit{ is connected}) = \begin{cases} 0 & \textit{if } c_n \to -\infty, \\ e^{-e^{-c}} & \textit{if } c_n \to c \ (\textit{constant}) \\ 1 & \textit{if } c_n \to \infty. \end{cases}$$

Proof. To prove the theorem we consider, as before, a random graph $\mathbb{G}_{n,p}$. It suffices to prove that, when $p = \frac{\log n + c}{n}$,

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is connected }) \to e^{-e^{-c}}.$$

and use Theorem 1.4 to translate to $\mathbb{G}_m$ and then use (1.7) and monotonicity for $c_n \to \pm\infty$.

Let $X_k = X_{k,n}$ be the number of components with k vertices in $\mathbb{G}_{n,p}$ and consider the complement of the event that $\mathbb{G}_{n,p}$ is connected. Then

$\mathbb{P}(\mathbb{G}_{n,p}$ is **not** connected)

$$= \mathbb{P}\left(\bigcup_{k=1}^{n/2}(\mathbb{G}_{n,p} \text{ has a component of order } k)\right) =$$

$$\mathbb{P}\left(\bigcup_{k=1}^{n/2}\{X_k > 0\}\right).$$

Note that X_1 counts here isolated vertices and therefore

$$\mathbb{P}(X_1 > 0) \leq \mathbb{P}(\mathbb{G}_{n,p} \text{ is } \textbf{not} \text{ connected }) \leq \mathbb{P}(X_1 > 0) + \sum_{k=2}^{n/2} \mathbb{P}(X_k > 0).$$

Now

$$\sum_{k=2}^{n/2} \mathbb{P}(X_k > 0) \leq \sum_{k=2}^{n/2} \mathbb{E} X_k \leq \sum_{k=2}^{n/2} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} = \sum_{k=2}^{n/2} u_k.$$

Now, for $2 \leq k \leq 10$,

$$\begin{aligned} u_k &\leq e^k n^k \left(\frac{\log n + c}{n}\right)^{k-1} e^{-k(n-10)\frac{\log n + c}{n}} \\ &\leq (1+o(1)) e^{k(1-c)} \left(\frac{\log n}{n}\right)^{k-1}, \end{aligned}$$

and for $k > 10$

$$\begin{aligned} u_k &\leq \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{\log n + c}{n}\right)^{k-1} e^{-k(\log n + c)/2} \\ &\leq n \left(\frac{e^{1-c/2+o(1)} \log n}{n^{1/2}}\right)^k. \end{aligned}$$

So

$$\begin{aligned} \sum_{k=2}^{n/2} u_k &\leq (1+o(1)) \frac{e^{-c} \log n}{n} + \sum_{k=10}^{n/2} n^{1+o(1)-k/2} \\ &= O\left(n^{o(1)-1}\right). \end{aligned}$$

It follows that

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is connected }) = \mathbb{P}(X_1 = 0) + o(1).$$

But we already know (Theorem 3.1) that for $p = (\log n + c)/n$ the number of isolated vertices in $\mathbb{G}_{n,p}$ has an asymptotically Poisson distribution and therefore, as in (3.4)

$$\lim_{n\to\infty} \mathbb{P}(X_1 = 0) = e^{-e^{-c}},$$

and so the theorem follows. □

It is possible to tweak the proof of Theorem 4.1 to give a more precise result stating that a random graph becomes connected exactly at the moment when the last isolated vertex disappears.

Theorem 4.2. *Consider the random graph process* $\{\mathbb{G}_m\}$. *Let*

$$m_1^* = \min\{m : \delta(\mathbb{G}_m) \geq 1\},$$

$$m_c^* = \min\{m : \mathbb{G}_m \text{ is connected}\}.$$

Then, w.h.p.,

$$m_1^* = m_c^*.$$

Proof. Let

$$m_\pm = \frac{1}{2} n \log n \pm \frac{1}{2} n \log\log n,$$

and

$$p_\pm = \frac{m_\pm}{N} \approx \frac{\log n \pm \log\log n}{n}.$$

We first show that w.h.p.

(i) G_{m_-} consists of a giant connected component plus a set V_1 of at most $2\log n$ isolated vertices,

(ii) G_{m_+} is connected.

Assume (i) and (ii). It follows that w.h.p.

$$m_- \leq m_1^* \leq m_c^* \leq m_+.$$

Since $\mathbb{G}_{m_-}$ consists of a connected component and a set of isolated vertices V_1, to create $\mathbb{G}_{m_+}$ we add $m_+ - m_-$ random edges. Note that $m_1^* = m_c^*$ if none of these edges are contained in V_1. Thus

$$\mathbb{P}(m_1^* < m_c^*) \leq o(1) + (m_+ - m_-)\frac{\frac{1}{2}|V_1|^2}{N - m_+}$$

$$\leq o(1) + \frac{2n((\log n)^2)\log\log n}{\frac{1}{2}n^2 - O(n\log n)}$$
$$= o(1).$$

Thus to prove the theorem, it is sufficient to verify (i) and (ii).

Let

$$p_- = \frac{m_-}{N} \approx \frac{\log n - \log\log n}{n},$$

and let X_1 be the number of isolated vertices in $\mathbb{G}_{n,p_-}$. Then

$$\begin{aligned}\mathbb{E}X_1 &= n(1-p_-)^{n-1}\\ &\approx ne^{-np_-}\\ &\approx \log n.\end{aligned}$$

Moreover

$$\begin{aligned}\mathbb{E}X_1^2 &= \mathbb{E}X_1 + n(n-1)(1-p_-)^{2n-3}\\ &\leq \mathbb{E}X_1 + (\mathbb{E}X_1)^2(1-p_-)^{-1}.\end{aligned}$$

So,

$$\operatorname{Var} X_1 \leq \mathbb{E}X_1 + 2(\mathbb{E}X_1)^2 p_-,$$

and

$$\begin{aligned}\mathbb{P}(X_1 \geq 2\log n) &= \mathbb{P}(|X_1 - \mathbb{E}X_1| \geq (1+o(1))\mathbb{E}X_1)\\ &\leq (1+o(1))\left(\frac{1}{\mathbb{E}X_1} + 2p_-\right)\\ &= o(1).\end{aligned}$$

Having at least $2\log n$ isolated vertices is a monotone property and so w.h.p. $\mathbb{G}_{m_-}$ has less then $2\log n$ isolated vertices.

To show that the rest of $\mathbb{G}_m$ is a single connected component we let X_k, $2 \leq k \leq n/2$ be the number of components with k vertices in $\mathbb{G}_{p_-}$. Repeating the calculations for p_- from the proof of Theorem 4.1, we have

$$\mathbb{E}\left(\sum_{k=2}^{n/2} X_k\right) = O\left(n^{o(1)-1}\right).$$

Let

$$\mathscr{E} = \{\exists \text{ component of order } 2 \leq k \leq n/2\}.$$

Then

$$\begin{aligned}\mathbb{P}(\mathbb{G}_{m_-} \in \mathscr{E}) &\leq O(\sqrt{n})\,\mathbb{P}(\mathbb{G}_{n,p_-} \in \mathscr{E}) \\ &= o(1),\end{aligned}$$

and this completes the proof of (i).

To prove (ii) (that G_{m_+} is connected w.h.p.) we note that (ii) follows from the fact that $\mathbb{G}_{n,p}$ is connected w.h.p. for $np - \log n \to \infty$ (see Theorem 4.1). By implication $\mathbb{G}_m$ is connected w.h.p. if $n\frac{m}{N} - \log n \to \infty$. But,

$$\begin{aligned}\frac{nm_+}{N} &= \frac{n(\frac{1}{2}n\log n + \frac{1}{2}n\log\log n)}{N} \\ &\approx \log n + \log\log n.\end{aligned}$$

□

4.2 k-connectivity

In this section we show that the threshold for the existence of vertices of degree k is also the threshold for the k-connectivity of a random graph. Recall that a graph G is k-connected if the removal of at most $k-1$ vertices of G does not disconnect it. In the light of the previous result it should be expected that a random graph becomes k-connected as soon as the last vertex of degree $k-1$ disappears. This is true and follows from the results of Erdős and Rényi [333]. Here is a weaker statement. The stronger statement is left as an exercise, Exercise 4.3.1.

Theorem 4.3. *Let* $m = \frac{1}{2}n\left(\log n + (k-1)\log\log n + c_n\right)$, $k = 1, 2, \ldots$. *Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_m \text{ is } k\text{-connected}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-\frac{e^{-c}}{(k-1)!}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

Proof. Let

$$p = \frac{\log n + (k-1)\log\log n + c}{n}.$$

We will prove that, in $\mathbb{G}_{n,p}$, with edge probability p above,

(i) the expected number of vertices of degree at most $k-2$ is $o(1)$,

(ii) the expected number of vertices of degree $k-1$ is, approximately $\frac{e^{-c}}{(k-1)!}$.

We have

$$
\begin{aligned}
&\mathbb{E}(\text{number of vertices of degree } t \le k-1)\\
&\qquad = n\binom{n-1}{t}p^t(1-p)^{n-1-t} \approx n\frac{n^t}{t!}\frac{(\log n)^t}{n^t}\frac{e^{-c}}{n(\log n)^{k-1}}
\end{aligned}
$$

and (i) and (ii) follow immediately.

The distribution of the number of vertices of degree $k-1$ is asymptotically Poisson, as may be verified by the method of moments. (See Exercise 3.3.4).

We now show that, if

$$
\mathscr{A}(S,T) = \left\{T \text{ is a component of } \mathbb{G}_{n,p} \setminus S\right\}
$$

then

$$
\mathbb{P}\left(\exists S,T,\ |S| < k,\ 2 \le |T| \le \frac{1}{2}(n-|S|) : \mathscr{A}(S,T)\right) = o(1).
$$

This implies that if $\delta(\mathbb{G}_{n,p}) \ge k$ then $\mathbb{G}_{n,p}$ is k-connected and Theorem 4.3 follows. $|T| \ge 2$ because if $T = \{v\}$ then v has degree less than k.

We can assume that S is minimal and then $N(T) = S$ and denote $s = |S|$, $t = |T|$. T is connected, and so it contains a tree with $t-1$ edges. Also each vertex of S is incident with an edge from S to T and so there are at least s edges between S and T. Thus, if $p = (1+o(1))\frac{\log n}{n}$ then

$$
\begin{aligned}
&\mathbb{P}(\exists S,T) \le o(1) +\\
&\sum_{s=1}^{k-1}\sum_{t=2}^{(n-s)/2}\binom{n}{s}\binom{n}{t}t^{t-2}p^{t-1}\binom{st}{s}p^s(1-p)^{t(n-s-t)}\\
&\le p^{-1}\sum_{s=1}^{k-1}\sum_{t=2}^{(n-s)/2}\left(\frac{ne}{s}\cdot(te)\cdot p\cdot e^{tp}\right)^s\left(ne\cdot p\cdot e^{-(n-t)p}\right)^t\\
&\le p^{-1}\sum_{s=1}^{k-1}\sum_{t=2}^{(n-s)/2}A^tB^s
\end{aligned}
\tag{4.1}
$$

where

$$
\begin{aligned}
A &= nepe^{-(n-t)p} = e^{1+o(1)}n^{-1+(t+o(t))/n}\log n\\
B &= ne^2tpe^{tp} = e^{2+o(1)}tn^{(t+o(t))/n}\log n.
\end{aligned}
$$

Now if $2 \le t \le \log n$ then $A = n^{-1+o(1)}$ and $B = O((\log n)^2)$. On the other hand, if $t > \log n$ then we can use $A \le n^{-1/3}$ and $B \le n^2$ to see that the sum in (4.1) is $o(1)$. $\square$

4.3 Exercises

4.3.1 Let $k = O(1)$ and let m_k^* be the hitting time for minimum degree at least k in the graph process. Let t_k^* be the hitting time for k-connectivity. Show that $m_k^* = t_k^*$ w.h.p.

4.3.2 Let $m = m_1^*$ be as in Theorem 4.2 and let $e_m = (u, v)$ where u has degree one. Let $0 < c < 1$ be a positive constant. Show that w.h.p. there is no triangle containing vertex v.

4.3.3 Let $m = m_1^*$ as in Theorem 4.2 and let $e_m = (u, v)$ where u has degree one. Let $0 < c < 1$ be a positive constant. Show that w.h.p. the degree of v in G_m is at least $c \log n$.

4.3.4 Suppose that $n \log n \ll m \leq n^{3/2}$ and let $d = 2m/n$. Let $S_i(v)$ be the set of vertices at distance i from vertex v. Show that w.h.p. $|S_i(v)| \geq (d/2)^i$ for all $v \in [n]$ and $1 \leq i \leq \frac{2\log n}{3\log d}$.

4.3.5 Suppose that $n \log n \ll m \leq n^{4/3-\varepsilon}$ and let $d = m/n$. Amend the proof of the previous question and show that w.h.p. there are at least $d/2$ internally vertex disjoint paths of length at most $\frac{4\log n}{3\log d}$ between any pair of vertices in $G_{n,m}$.

4.3.6 Suppose that $m \gg n \log n$ and let $d = m/n$. Suppose that we randomly color the edges of $G_{n,m}$ with q colors where $q \gg \frac{(\log n)^2}{(\log d)^2}$. Show that w.h.p. there is a *rainbow* path between every pair of vertices. (A path is rainbow if each of its edges has a different color).

4.3.7 Let $C_{k,k+\ell}$ denote the number of connected graphs with vertex set $[k]$ and $k+\ell$ edges where $\ell \to \infty$ with k and $\ell = o(k)$. Use the inequality

$$\binom{n}{k} C_{k,k+\ell} p^{k+\ell} (1-p)^{\binom{k}{2}-k-\ell+k(n-k)} \leq \frac{n}{k}$$

and a careful choice of p, n to prove (see Łuczak [633]) that

$$C_{k,k+\ell} \leq \sqrt{\frac{k^3}{\ell}} \left(\frac{e + O(\sqrt{\ell/k})}{12\ell} \right)^{\ell/2} k^{k+(3\ell-1)/2}.$$

4.3.8 Let $\mathbb{G}_{n,n,p}$ be the random bipartite graph with vertex bi-partition $V = (A, B)$, $A = [1, n], B = [n+1, 2n]$ in which each of the n^2 possible edges appears independently with probability p. Let $p = \frac{\log n + \omega}{n}$, where $\omega \to \infty$. Show that w.h.p. $\mathbb{G}_{n,n,p}$ is connected.

4.4 Notes

Disjoint paths

Being k-connected means that we can find disjoint paths between any two sets of vertices $A = \{a_1, a_2, \dots, a_k\}$ and $B = \{b_1, b_2, \dots, b_k\}$. In this statement there is no control over the endpoints of the paths i.e. we cannot specify a path from a_i to b_i for $i = 1, 2, \dots, k$. Specifying the endpoints leads to the notion of *linkedness*. Broder, Frieze, Suen and Upfal [200] proved that when we are above the connectivity threshold, we can w.h.p. link any two k-sets by edge disjoint paths, provided some natural restrictions apply. The result is optimal up to constants. Broder, Frieze, Suen and Upfal [199] considered the case of vertex disjoint paths. Frieze and Zhao [418] considered the edge disjoint path version in random regular graphs.

Rainbow Connection

The rainbow connection $rc(G)$ of a connected graph G is the minimum number of colors needed to color the edges of G so that there is a rainbow path between every pair of vertices. Caro, Lev, Roditty, Tuza and Yuster [209] proved that $p = \sqrt{\log n/n}$ is the sharp threshold for the property $rc(G) \leq 2$. This was sharpened to a hitting time result by Heckel and Riordan [475]. He and Liang [474] further studied the rainbow connection of random graphs. Specifically, they obtain a threshold for the property $rc(G) \leq d$ where d is constant. Frieze and Tsourakakis [417] studied the rainbow connection of $G = G(n, p)$ at the connectivity threshold $p = \frac{\log n + \omega}{n}$ where $\omega \to \infty$ and $\omega = o(\log n)$. They showed that w.h.p. $rc(G)$ is asymptotically equal to $\max\{diam(G), Z_1(G)\}$, where Z_1 is the number of vertices of degree one.

Chapter 5

Small Subgraphs

Graph theory is replete with theorems stating conditions for the existence of a subgraph H in a larger graph G. For example Turán's theorem [832] states that a graph with n vertices and more than $\left(1-\frac{1}{r}\right)\frac{n^2}{2}$ edges must contain a copy of K_{r+1}. In this chapter we see instead how many random edges are required to have a particular fixed size subgraph w.h.p. In addition, we will consider the distribution of the number of copies.

5.1 Thresholds

In this section we will look for a threshold for the appearance of any fixed graph H, with $v_H = |V(H)|$ vertices and $e_H = |E(H)|$ edges. The property that a random graph contains H as a subgraph is clearly monotone increasing. It is also transparent that ”denser” graphs appear in a random graph ”later” than ”sparser” ones. More precisely, denote by

$$d(H) = \frac{e_H}{v_H}, \tag{5.1}$$

the *density* of a graph H. Notice that $2d(H)$ is the average vertex degree in H. We begin with the analysis of the asymptotic behavior of the expected number of copies of H in the random graph $\mathbb{G}_{n,p}$.

Lemma 5.1. *Let X_H denote the number of copies of H in $\mathbb{G}_{n,p}$.*

$$\mathbb{E} X_H = \binom{n}{v_H} \frac{v_H!}{aut(H)} p^{e_H},$$

where $aut(H)$ is the number of automorphisms of H.

Proof. The complete graph on n vertices K_n contains $\binom{n}{v_H}a_H$ distinct copies of H, where a_H is the number of copies of H in K_{v_H}. Thus

$$\mathbb{E}X_H = \binom{n}{v_H} a_H p^{e_H},$$

and all we need to show is that

$$a_H \times aut(H) = v_H!.$$

Each permutation σ of $[v_H] = \{1,2,\ldots,v_H\}$ defines a unique copy of H as follows: A copy of H corresponds to a set of e_H edges of K_{v_H}. The copy H_σ corresponding to σ has edges $\{(x_{\sigma(i)}, y_{\sigma(i)}) :\ 1 \le i \le e_H\}$, where $\{(x_j, y_j) :\ 1 \le j \le e_H\}$ is some fixed copy of H in K_{v_H}. But $H_\sigma = H_{\tau\sigma}$ if and only if for each i there is j such that $(x_{\tau\sigma(i)}, y_{\tau\sigma(i)}) = (x_{\sigma(j)}, y_{\sigma(j)})$ i.e., if τ is an automorphism of H. □

Theorem 5.2. *Let H be a fixed graph with $e_H > 0$. Suppose $p = o\left(n^{-1/d(H)}\right)$. Then w.h.p. $\mathbb{G}_{n,p}$ contains no copies of H.*

Proof. Suppose that $p = \omega^{-1} n^{-1/d(H)}$ where $\omega = \omega(n) \to \infty$ as $n \to \infty$. Then

$$\mathbb{E}X_H = \binom{n}{v_H} \frac{v_H!}{aut(H)} p^{e_H} \le n^{v_H} \omega^{-e_H} n^{-e_H/d(H)} = \omega^{-e_H}.$$

Thus

$$\mathbb{P}(X_H > 0) \le \mathbb{E}X_H \to 0 \ \text{ as } \ n \to \infty.$$

□

From our previous experience one would expect that when $\mathbb{E}X_H \to \infty$ as $n \to \infty$ the random graph $\mathbb{G}_{n,p}$ would contain H as a subgraph w.h.p. Let us check whether such a phenomenon holds also in this case. So consider the case when $pn^{1/d(H)} \to \infty$, i.e. where $p = \omega n^{-1/d(H)}$ and $\omega = \omega(n) \to \infty$ as $n \to \infty$. Then for some constant $c_H > 0$

$$\mathbb{E}X_H \ge c_H n^{v_H} \omega^{e_H} n^{-e_H/d(H)} = c_H \omega^{e_H} \to \infty.$$

However, as we will see, this is not always enough for $\mathbb{G}_{n,p}$ to contain a copy of a given graph H w.h.p. To see this, consider the graph H given in Figure 5.1 below.

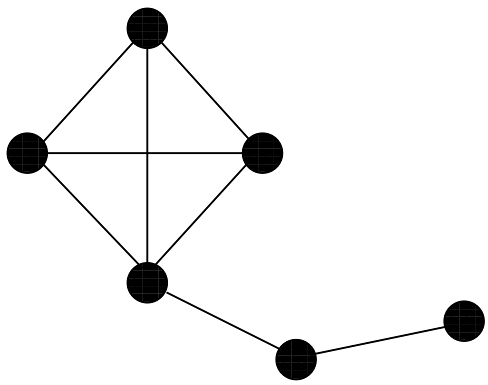


Figure 5.1: A Kite

Here $v_H = 6$ and $e_H = 8$. Let $p = n^{-5/7}$. Now $1/d(H) = 6/8 > 5/7$ and so

$$\mathbb{E}X_H \approx c_H n^{6-8\times 5/7} \to \infty.$$

On the other hand, if $\hat{H} = K_4$ then

$$\mathbb{E}X_{\hat{H}} \leq n^{4-6\times 5/7} \to 0,$$

and so w.h.p. there are no copies of $\hat{H}$ and hence no copies of H.

The reason for such "strange" behavior is quite simple. Our graph H is in fact *not balanced*, since its overall density is smaller than the density of one of its subgraphs, i.e., of $\hat{H} = K_4$. So we need to introduce another density characteristic of graphs, namely the maximum subgraph density defined as follows:

$$m(H) = \max\{d(K) : K \subseteq H\}. \tag{5.2}$$

A graph H is *balanced* if $m(H) = d(H)$. It is *strictly balanced* if $d(H) > d(K)$ for all proper subgraphs $K \subset H$.

Now we are ready to determine the threshold for the existence of a copy of H in $\mathbb{G}_{n,p}$. Erdős and Rényi [332] proved this result for balanced graphs. The threshold for any graph H was first found by Bollobás in [147] and an alternative, deterministic argument to derive the threshold was presented in [552]. A simple proof, given here, is due to Ruciński and Vince [777].

Theorem 5.3. *Let H be a fixed graph with $e_H > 0$. Then*

$$\lim_{n\to\infty} \mathbb{P}(H \subseteq \mathbb{G}_{n,p}) = \begin{cases} 0 & \text{if } pn^{1/m(H)} \to 0 \\ 1 & \text{if } pn^{1/m(H)} \to \infty. \end{cases}$$

Proof. Let $\omega = \omega(n) \to \infty$ as $n \to \infty$. The first statement follows from Theorem 5.2. Notice, that if we choose $\hat{H}$ to be a subgraph of H with $d(\hat{H}) = m(H)$ (such a subgraph always exists since we do not exclude $\hat{H} = H$), then $p = \omega^{-1} n^{-1/d(\hat{H})}$ implies that $\mathbb{E} X_{\hat{H}} \to 0$. Therefore, w.h.p. $\mathbb{G}_{n,p}$ contains no copies of $\hat{H}$, and so it does not contain H as well.

To prove the second statement we use the Second Moment Method. Suppose now that $p = \omega n^{-1/m(H)}$. Denote by $H_1, H_2, \ldots, H_t$ all copies of H in the complete graph on $\{1, 2, \ldots, n\}$. Note that

$$t = \binom{n}{v_H} \frac{v_H!}{aut(H)}, \tag{5.3}$$

where $aut(H)$ is the number of automorphisms of H. For $i = 1, 2, \ldots, t$ let

$$I_i = \begin{cases} 1 & \text{if } H_i \subseteq \mathbb{G}_{n,p}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $X_H = \sum_{i=1}^t I_i$. Then

$$\begin{aligned} \operatorname{Var} X_H &= \sum_{i=1}^t \sum_{j=1}^t \operatorname{Cov}(I_i, I_j) = \sum_{i=1}^t \sum_{j=1}^t (\mathbb{E}(I_i I_j) - (\mathbb{E} I_i)(\mathbb{E} I_j)) \\ &= \sum_{i=1}^t \sum_{j=1}^t (\mathbb{P}(I_i = 1, I_j = 1) - \mathbb{P}(I_i = 1)\mathbb{P}(I_j = 1)) \\ &= \sum_{i=1}^t \sum_{j=1}^t \left(\mathbb{P}(I_i = 1, I_j = 1) - p^{2e_H}\right). \end{aligned}$$

Observe that random variables I_i and I_j are independent iff H_i and H_j are edge disjoint. In this case $\operatorname{Cov}(I_i, I_j) = 0$ and such terms vanish from the above summation. Therefore we consider only pairs (H_i, H_j) with $H_i \cap H_j = K$, for some graph K with $e_K > 0$. So,

$$\begin{aligned} \operatorname{Var} X_H &= O\Big(\sum_{K \subseteq H, e_K > 0} n^{2v_H - v_K} \left(p^{2e_H - e_K} - p^{2e_H}\right)\Big) \\ &= O\Big(n^{2v_H} p^{2e_H} \sum_{K \subseteq H, e_K > 0} n^{-v_K} p^{-e_K}\Big). \end{aligned}$$

On the other hand,

$$\mathbb{E} X_H = \binom{n}{v_H} \frac{v_H!}{aut(H)} p^{e_H} = \Omega\left(n^{v_H} p^{e_H}\right),$$

thus, by Lemma 26.4,

$$\begin{aligned}
\mathbb{P}(X_H = 0) &\leq \frac{\operatorname{Var} X_H}{(\mathbb{E} X_H)^2} = O\left(\sum_{K \subseteq H, e_K > 0} n^{-v_K} p^{-e_K}\right) \\
&= O\left(\sum_{K \subseteq H, e_K > 0} \left(\frac{1}{\omega n^{1/d(K) - 1/m(H)}}\right)^{e_K}\right) \\
&= o(1).
\end{aligned}$$

Hence w.h.p., the random graph $\mathbb{G}_{n,p}$ contains a copy of the subgraph H when $pn^{1/m(H)} \to \infty$. □

5.2 Asymptotic Distributions

We will now study the asymptotic distribution of the number X_H of copies of a fixed graph H in $\mathbb{G}_{n,p}$. We start at the threshold, so we assume that $np^{m(H)} \to c,\ c > 0$, where $m(H)$ denotes as before, the maximum subgraph density of H. Now, if H is not balanced, i.e., its maximum subgraph density exceeds the density of H, then $\mathbb{E} X_H \to \infty$ as $n \to \infty$, and one can show that there is a sequence of numbers a_n, increasing with n, such that the asymptotic distribution of X_H/a_n coincides with the distribution of a random variable counting the number of copies of a subgraph K of H for which $m(H) = d(K)$. Note that K is itself a balanced graph. However the asymptotic distribution of balanced graphs on the threshold, although computable, cannot be given in a closed form. The situation changes dramatically if we assume that the graph H whose copies in $\mathbb{G}_{n,p}$ we want to count is *strictly balanced*, i.e., when for every proper subgraph K of H, $d(K) < d(H) = m(H)$.

The following result is due to Bollobás [147], and Karoński and Ruciński [551].

Theorem 5.4. *If H is a strictly balanced graph and $np^{m(H)} \to c$, $c > 0$, then $X_H \xrightarrow{D} \operatorname{Po}(\lambda)$, as $n \to \infty$, where $\lambda = c^{v_H}/aut(H)$.*

Proof. Denote, as before, by $H_1, H_2, \ldots, H_t$ all copies of H in the complete graph on $\{1, 2, \ldots, n\}$. For $i = 1, 2, \ldots, t$, let

$$I_{H_i} = \begin{cases} 1 & \text{if } H_i \subseteq \mathbb{G}_{n,p} \\ 0 & \text{otherwise} \end{cases}$$

Then $X_H = \sum_{i=1}^{t} I_{H_i}$ and the kth factorial moment of X_H, $k = 1, 2 \ldots$,

$$\mathbb{E}(X_H)_k = \mathbb{E}[X_H(X_H - 1) \cdots (X_H - k + 1)],$$

can be written as

$$\begin{aligned}\mathbb{E}(X_H)_k &= \sum_{i_1,i_2,\ldots,i_k} \mathbb{P}(I_{H_{i_1}} = 1, I_{H_{i_2}} = 1, \ldots, I_{H_{i_k}} = 1) \\ &= D_k + \overline{D}_k,\end{aligned}$$

where the summation is taken over all k-element sequences of distinct indices i_j from $\{1,2,\ldots,t\}$, while D_k and $\overline{D}_k$ denote the partial sums taken over all (ordered) k tuples of copies of H which are, respectively, pairwise vertex disjoint (D_k) and not all pairwise vertex disjoint ($\overline{D}_k$). Now, observe that

$$\begin{aligned}D_k &= \sum_{i_1,i_2,\ldots,i_k} \mathbb{P}(I_{H_{i_1}} = 1)\,\mathbb{P}(I_{H_{i_2}} = 1) \cdots \mathbb{P}(I_{H_{i_k}} = 1) \\ &= \binom{n}{v_H, v_H, \ldots, v_H} (a_H p^{e_H})^k \\ &\approx (\mathbb{E} X_H)^k.\end{aligned}$$

So assuming that $np^{d(H)} = np^{m(H)} \to c$ as $n \to \infty$,

$$D_k \approx \left(\frac{c^{v_H}}{aut(H)}\right)^k. \tag{5.4}$$

On the other hand we will show that

$$\overline{D}_k \to 0 \text{ as } n \to \infty. \tag{5.5}$$

Consider the family $\mathscr{F}_k$ of all (mutually non-isomorphic) graphs obtained by taking unions of k not all pairwise vertex disjoint copies of the graph H. Suppose $F \in \mathscr{F}_k$ has v_F vertices ($v_H \le v_F \le kv_H - 1$) and e_F edges, and let $d(F) = e_F/v_F$ be its density. To prove that (5.5) holds we need the following Lemma.

Lemma 5.5. *If $F \in \mathscr{F}_k$ then $d(F) > m(H)$.*

Proof. Define

$$f_F = m(H)v_F - e_F. \tag{5.6}$$

We will show (by induction on $k \ge 2$) that $f_F < 0$ for all $F \in \mathscr{F}_k$. First note that $f_H = 0$ and that $f_K > 0$ for every proper subgraph K of H, since H is strictly balanced. Notice also that the function f is modular, i.e., for any two graphs F_1 and F_2,

$$f_{F_1 \cup F_2} = f_{F_1} + f_{F_2} - f_{F_1 \cap F_2}. \tag{5.7}$$

Assume that the copies of H composing F are numbered in such a way that $H_{i_1} \cap H_{i_2} \neq \emptyset$. If $F = H_{i_1} \cup H_{i_2}$ then (5.6) and $f_{H_1} = f_{H_2} = 0$ implies

$$f_{H_{i_1} \cup H_{i_2}} = -f_{H_{i_1} \cap H_{i_2}} < 0.$$

For arbitrary $k \geq 3$, let $F' = \bigcup_{j=1}^{k-1} H_{i_j}$ and $K = F' \cap H_{i_k}$. Then by the inductive assumption we have $f_{F'} < 0$ while $f_K \geq 0$ since K is a subgraph of H (in extreme cases K can be H itself or an empty graph). Therefore

$$f_F = f_{F'} + f_{H_{i_k}} - f_K = f_{F'} - f_K < 0,$$

which completes the induction and implies that $d(F) > m(H)$. □

Let C_F be the number of sequences $H_{i_1}, H_{i_2}, \ldots, H_{i_k}$ of k distinct copies of H, such that

$$V\Big(\bigcup_{j=1}^{k} H_{i_j}\Big) = \{1, 2, \ldots, v_F\} \quad \text{and} \quad \bigcup_{j=1}^{k} H_{i_j} \cong F.$$

Then, by Lemma 5.5,

$$\begin{aligned}\overline{D}_k &= \sum_{F \in \mathscr{F}_k} \binom{n}{v_F} C_F \; p^{e_F} = O(n^{v_F} p^{e_F}) \\ &= O\left(\left(np^{d(F)}\right)^{v(F)}\right) = o(1),\end{aligned}$$

and so (5.5) holds.
Summarizing,

$$\mathbb{E}(X_H)_k \approx \left(\frac{c^{v_H}}{aut(H)}\right)^k,$$

and the theorem follows by the Method of Moments (see Theorem 26.11). □

The following theorem describes the asymptotic behavior of the number of copies of a graph H in $\mathbb{G}_{n,p}$ past the threshold for the existence of a copy of H. It holds regardless of whether or not H is balanced or strictly balanced (see Ruciński [776]).

Theorem 5.6. *Let H be a fixed (not-empty) graph. If $np^{m(H)} \to \infty$ and $n^2(1-p) \to \infty$, then $(X_H - \mathbb{E}X_H)/(\operatorname{Var} X_H)^{1/2} \xrightarrow{D} \mathrm{N}(0,1)$, as $n \to \infty$*

Proof. The proof is by the Method of Moments (see Lemma 26.7 and Corollary 26.8). Here, instead of the original proof from [776], we shall reproduce its more compact version, presented in [509].
As in the previous theorem, denote by $H_1, H_2, \ldots, H_t$ all copies of H in the complete graph on $\{1, 2, \ldots, n\}$, where t is given by (5.3). For $i = 1, 2, \ldots, t$, let

$$I_{H_i} = \begin{cases} 1 & \text{if } H_i \subseteq \mathbb{G}_{n,p} \\ 0 & \text{otherwise} \end{cases}$$

Then $X_H = \sum_{i=1}^{t} I_{H_i}$ and thus for every $k = 1, 2, ...,$

$$\mathbb{E}(X_H - \mathbb{E}X_H)^k = \sum_{H_1,\dots,H_k} \mathbb{E}((I_{H_1} - \mathbb{E}I_{H_1}) \cdots (I_{H_k} - \mathbb{E}I_{H_k})), \tag{5.8}$$

where the summation is over all k-tuples $H_1, \dots, H_k$ of copies of H in K_n.
Denote each term of the sum from (5.8) by $T(H_1, \dots, H_k)$, i.e.,

$$T(H_1, \dots, H_k) = \mathbb{E}((I_{H_1} - \mathbb{E}I_{H_1}) \cdots (I_{H_k} - \mathbb{E}I_{H_k})).$$

For each $T(H_1, \dots, H_k)$ define a graph $L(H_1, \dots, H_k)$ with vertex $\{1, \dots, k\}$ and an edge $\{i, j\}$ whenever H_i and H_j have at least one edge in common. An edge implies that respective indicator random variables I_{H_i} and I_{H_j} are not independent and we call graph $L(H_1, \dots, H_k)$ a dependency graph for the variables $I_{H_1}, \dots I_{H_k}$. Next, we group the terms of (5.8) according to the structure of the graph $L(H_1, \dots, H_k)$. So,

$$\mathbb{E}(X_H - \mathbb{E}X_H)^k = \sum_{L(H_1,\dots,H_k)} T(H_1, \dots, H_k) = \sum_{(1)} + \sum_{(2)}, \tag{5.9}$$

where the summation in $\sum_{(1)}$ is taken over all graphs L with an even number of vertices $k = 2m$ and with exactly $k/2$ edges forming a perfect matching, i.e., $k/2$ disjoint edges, while $\sum_{(2)}$ takes care of all the remaining graphs L, for k odd or even.

In the first step, we estimate $\sum_{(1)}$. One can easily check that in this case

$$T(H_1, \dots, H_k) = (\mathrm{Var} X_H)^{k/2}(1 + O(1/n)). \tag{5.10}$$

Furthermore, since there are

$$\frac{(2m)!}{2^m m!} = (2m-1)(2m-3)\cdots 3 \cdot 1 = (2m-1)!! = (k-1)!!$$

such graphs, so

$$\sum_{(1)} = (k-1)!!(\mathrm{Var} X_H)^{k/2}(1 + O(1/n)). \tag{5.11}$$

To estimate $\sum_{(2)}$, first notice that all terms corresponding to graphs L with an isolated vertex vanish. Indeed, if a vertex j is isolated, it means that $I_{H_j} - \mathbb{E}I_{H_j}$ is independent from the product $\prod_{i \neq j}(I_{H_i} - \mathbb{E}I_{H_i})$, and so $T(H_1, \dots, H_k) = 0$.

Also notice that, in all the remaining cases, the dependency graph L, for any k odd or even, has less than $k/2$ components since each component has to have at least two and some component has at least three vertices.

Denote the number of components of L by $c(L)$ and without loss of generality, reorder the vertices of L in such a way that vertices of the first component are labeled $\{1,\ldots,r_1\}$, vertices of the second component are labeled $\{r_1+1,\ldots,r_2\}$, and vertices of the last one, respectively, by $\{r_{c(L)-1}+1,\ldots,r_{c(L)}=k\}$. Moreover the relabeling is such that if $j \notin \{1,r_1+1,r_2+1,\ldots,r_{c(L)-1}+1\}$, then L contains an edge $\{i,j\}$ with $i<j$.

Consider $T(H_1,\ldots,H_k)$ with such an L and let $H^{(j)}=\bigcup_{i=1}^{j} H_i$. Let F_j be, the possibly empty, subgraph of H which corresponds to $H^{(j-1)} \cap H_j$ under isomorphism $H_j \cong H$. Note that, by our relabeling assumption, when $j \in \{1,r_1+1,r_2+1\ldots,r_{c(L)-1}+1\}$, the number of edges $e(F_j)=0$.
If the edge probability $p \leq 1/2$, we estimate $T(H_1,\ldots,H_k)$ by taking absolute values

$$|T(H_1,\ldots,H_k)| \leq \mathbb{E}((I_{H_1}+\mathbb{E}I_{H_1})\cdots(I_{H_k}+\mathbb{E}I_{H_k})).$$

The product can be expanded into 2^k terms, where the largest expectation is of the product $I_{H_1}\cdots I_{H_k}$, so

$$|T(H_1,\ldots,H_k)| \leq 2^k\,\mathbb{E}(I_{H_1}\cdots I_{H_k}) = O\left(p^{e(H^{(k)})}\right). \tag{5.12}$$

In the case of $1/2<p\leq 1$ we estimate $T(H_1,\ldots,H_k)$ by taking one factor from each component only, i.e.,

$$|T(H_1,\ldots,H_k)| \leq \mathbb{E}\prod_{i=1}^{c(L)}|I_{H_{r_i}}-\mathbb{E}I_{H_{r_i}}|.$$

These factors are independent, and each has the expected value

$$\mathbb{E}\,|I_{H_r}-\mathbb{E}I_{H_r}| = 2p^{e(H)}(1-p^{e(H)}) \leq 2(1-p^{e(H)}) \leq 2e(H)(1-p).$$

Hence, when $1/2<p\leq 1$,

$$T(H_1,\ldots,H_k) = O\left((1-p)^{c(L)}\right). \tag{5.13}$$

Combining (5.12) and (5.13), by introducing redundant factors in each bound, we get that for all $0\leq p\leq 1$, that

$$\begin{aligned} T(H_1,\ldots,H_k) &= O\left(p^{e(H^{(k)})}(1-p)^{c(L)}\right) \\ &= O\left((1-p)^{c(L)}p^{ke(H)-\sum_1^k e(F_i)}\right), \end{aligned} \tag{5.14}$$

since $e(H^{(k)}) = ke(H)-\sum_1^k e(F_i)$. Similarly, the number of vertices $v(H^{(k)}) = kv(H)-\sum_1^k v(F_i)$, so there are at most $O\left(n^{kv(H)-\sum_1^k v(F_i)}\right)$ possible choices of $H_1,\ldots,H_k$

yielding L and a particular sequence $F_1,\dots,F_k$. So, fixing L and $F_1,\dots,F_k$ generates a contribution to $\sum_{(2)}$ of the order

$$\begin{aligned}
&O\left(n^{kv(H)-\sum_1^k v(F_i)}(1-p)^{c(L)}p^{ke(H)-\sum_1^k e(F_i)}\right)\\
&=O\left(\left(n^{v(H)}p^{e(H)}\right)^k(1-p)^{c(L)}\prod_{i=1}^k\left(n^{v(F_i)}p^{e(F_i)}\right)^{-1}\right). \qquad (5.15)
\end{aligned}$$

To estimate $\prod_{i=1}^k\left(n^{v(F_i)}p^{e(F_i)}\right)^{-1}$, notice that $c(L)$ of the F_i's have no edges, so $n^{v(F_i)}p^{e(F_i)}=n^{v(F_i)}\geq 1$, while $k-c(L)$ others have $e(F_i)\geq 1$ and thus

$$n^{v(F_i)}p^{e(F_i)}\geq \mathbb{E}X_{F_i}\geq \min\{\mathbb{E}X_G : G\subseteq H,\ e(G)>0\}=\Phi_H.$$

Thus

$$\begin{aligned}
&\left(n^{v(H)}p^{e(H)}\right)^k(1-p)^{c(L)}\prod_{i=1}^k\left(n^{v(F_i)}p^{e(F_i)}\right)^{-1}\\
&\leq \left(n^{v(H)}p^{e(H)}\right)^k(1-p)^{c(L)}\Phi_H^{c(l)-k}=\Theta\left((\mathbb{E}X_H)^k(1-p)^{c(L)}\Phi_H^{c(l)-k}\right)\\
&=\Theta\left((\operatorname{Var}X_H)^{k/2}\left((1-p)\Phi_H\right)^{c(L)-k/2)}\right). \qquad (5.16)
\end{aligned}$$

Indeed, if H' and H'' are copies of H in the complete graph K_n then

$$\operatorname{Var}X_H=\sum_{H',H''}Cov(I_{H'},I_{H''})=\sum_{E(H')\cap E(H'')\neq\emptyset}\left(\mathbb{E}(I_{H'}I_{H''})-\mathbb{E}(I_{H'})\,\mathbb{E}(I_{H''})\right),$$

since indicator random variables $I_{H'}$ and $I_{H''}$ are independent if H' and H'' do not share an edge. Moreover, noticing that for each $G\subseteq H$ there are $\Theta(n^{v(G)}n^{2(v(H)-v(G))})=\Theta(n^{2v(H)-v(G)})$, we get

$$\begin{aligned}
\operatorname{Var}X_H&=\Theta\left(\sum_{G\subseteq H,e(G)>0}n^{2v(H)-v(G)}\left(p^{2e(H)-e(G)}-p^{2e(H)}\right)\right)\\
&=\Theta\left((1-p)\sum_{G\subseteq H,e(G)>0}n^{2v(H)-v(G)}p^{2e(H)-e(G)}\right)\\
&=\Theta\left((1-p)\max_{G\subseteq H,e(G)>0}\frac{(\mathbb{E}X_H)^2}{\mathbb{E}X_G}\right)\\
&=\Theta\left((1-p)\frac{(\mathbb{E}X_H)^2}{\Phi_H}\right),
\end{aligned}$$

and (5.16) follows.

Now recall that $c(L) < k/2$ and $m(H)$ denotes maximum subgraph density (see (5.2))) and notice that, when the edge probability p is such that $np^{m(H)} \to \infty$ and $n^2(1-p) \to \infty$, then $(1-p)\Phi_H \to \infty$.

Indeed, one can observe that the condition $np^{m(H)} \to \infty$ implies that $\Phi_H \to \infty$ and so $(1-p)\Phi_H \to \infty$ provided $p \to 0$. On the other hand, when p is a constant, or $p \to 1$, then $\Phi_H = \Theta(n^2)$, and thus $(1-p)\Phi_H = \Theta(n^2(1-p)) \to \infty$.
So, by (5.16), (5.15) and (5.14) it follows that for fixed L and $F_1, \ldots, F_k$, each term contributes $o((\text{Var} H_H)^{k/2})$ to $\sum_{(2)}$. Since there are finitely many possible sequences $F_1, \ldots, F_k$, therefore

$$\sum_{(2)} = o((\text{Var} X_H)^{k/2}). \tag{5.17}$$

Finally, merging (5.11) and (5.17) and taking $a_n^2 = \text{Var} X_H$ in Corollary 26.8 , we arrive at the thesis. □

5.3 Exercises

5.3.1 Draw a graph which is: (a) balanced but not strictly balanced, (b) unbalanced.

5.3.2 Are the small graphs listed below, balanced or unbalanced: (a) a tree, (b) a cycle, (c) a complete graph, (d) a regular graph, (d) the Petersen graph, (e) a graph composed of a complete graph on 4 vertices and a triangle, sharing exactly one vertex.

5.3.3 Determine (directly, not from the statement of Theorem 5.3) thresholds $\hat{p}$ for $\mathbb{G}_{n,p} \supseteq G$, for graphs listed in exercise (ii). Do the same for the thresholds of G in $\mathbb{G}_{n,m}$.

5.3.4 For a graph G a *balanced extension* of G is a graph F, such that $G \subseteq F$ and $m(F) = d(F) = m(G)$. Applying the result of Győri, Rothschild and Ruciński [458] that every graph has a balanced extension, deduce Bollobás's result (Theorem 5.3) from that of Erdős and Rényi (threshold for balanced graphs).

5.3.5 Let F be a graph obtained by taking a union of triangles such that not every pair of them is vertex-disjoint, Show (by induction) that $e_F > v_F$.

5.3.6 Let f_F be a graph function defined as

$$f_F = a\, v_F + b\, e_F,$$

where a,b are constants, while v_F and e_F denote, respectively, the number of vertices and edges of a graph F. Show that the function f_F is modular.

5.3.7 Determine (directly, using exercise (v)) when the random variable counting the number of copies of a triangle in $\mathbb{G}_{n,p}$ has asymptotically the Poisson distribution.

5.3.8 Let X_e be the number of isolated edges (edge-components) in $\mathbb{G}_{n,p}$ and let

$$\omega(n) = 2pn - \log n - \log\log n.$$

Prove that

$$\mathbb{P}(X_e > 0) \to \begin{cases} 0 & \text{if } p \ll n^{-2} \quad \text{or} \quad \omega(n) \to \infty \\ 1 & \text{if } p \gg n^{-2} \quad \text{and} \quad \omega(n) \to \infty. \end{cases}$$

5.3.9 Determine when the random variable X_e defined in exercise (vii) has asymptotically the Poisson distribution.

5.3.10 Use Janson's inequality, Theorem 27.13, to prove (5.18) below.

5.3.11 Check that (5.10) holds.

5.3.12 In the proof of Theorem 5.6 show that the condition $np^{m(H)} \to \infty$ is equivalent to $\Phi_H \to \infty$, as well as that $\Phi_H = \Theta(n^2)$, when p is a constant.

5.3.13 Prove that the conditions of Theorem 5.6 are also necessary.

5.4 Notes

Distributional Questions

In 1982 Barbour [73] adapted the Stein–Chen technique for obtaining estimates of the rate of convergence to the Poisson and the normal distribution (see Section 26.3 or [74]) to random graphs. The method was next applied by Karoński and Ruciński [553] to prove the convergence results for semi-induced graph properties of random graphs.

Barbour, Karoński and Ruciński [76] used the original Stein's method for normal approximation to prove a general central limit theorem for the wide class of decomposable random variables. Their result is illustrated by a variety of applications to random graphs. For example, one can deduce from it the asymptotic distribution of the number of k-vertex tree-components in $\mathbb{G}_{n,p}$, as well as of the number of vertices of fixed degree d in $\mathbb{G}_{n,p}$ (in fact, Theorem 3.2 is a direct consequence of the last result).

Barbour, Janson, Karoński and Ruciński [75] studied the number X_k of maximal complete subgraphs (cliques) of a given fixed size $k \geq 2$ in the random graph $\mathbb{G}_{n,p}$. They show that if the edge probability $p = p(n)$ is such that the $\mathbb{E}X_k$ tends to a finite constant λ as $n \to \infty$, then X_k tends in distribution to the Poisson random variable with the expectation λ. When its expectation tends to infinity, X_k converges in distribution to a random variable which is normally distributed. Poisson convergence was proved using Stein–Chen method, while for the proof of the normal part, different methods for different ranges of p were used such as the first projection method or martingale limit theorem (for details of these methods see Chapter 6 of Janson, Łuczak and Ruciński [509]).

Svante Janson in an a sequence of papers [492],[493], [494], [497] (see also [510]) developed or accommodated various methods to establish asymptotic normality of various numerical random graph characteristics. In particular, in [493] he established the normal convergence by higher semi-invariants of sums of dependent random variables with direct applications to random graphs. In [494] he proved a functional limit theorem for subgraph count statistics in random graphs (see also [510]).

In 1997 Janson [492] answered the question posed by Paul Erdős: What is the length Y_n of the first cycle appearing in the random graph process $\mathbb{G}_m$? He proved that

$$\lim_{n\to\infty} \mathbb{P}(Y_n = j) = \frac{1}{2}\int_0^1 t^{j-1} e^{t/2+t^2/4}\sqrt{1-t}\, dt, \text{ for every } j \geq 3.$$

Tails of Subgraph Counts in $\mathbb{G}_{n,p}$.

Often one needs exponentially small bounds for the probability that X_H deviates from its expectation. In 1990 Janson [495] showed that for fixed $\varepsilon \in (0,1]$,

$$\mathbb{P}(X_H \leq (1-\varepsilon)\mathbb{E}X_H) = \exp\{-\Theta(\Phi_H)\}, \tag{5.18}$$

where $\Phi_H = \min_{K \subseteq H : e_K > 0} n^{v_K} p^{e_K}$.

The upper tail $\mathbb{P}(X_H \geq (1+\varepsilon)\mathbb{E}X_H)$ proved to be much more elusive. To simplify the results, let us assume that ε is fixed, and p is above the existence threshold, that is, $p \gg n^{-1/m(H)}$, but small enough to make sure that $(1+\varepsilon)\mathbb{E}X_H$ is at most the number of copies of H in K_n.

Given a graph G, let Δ_G be the maximum degree of G and α^*_G the *fractional independence number* of G, defined as the maximum of $\sum_{v\in V(G)} w(v)$ over all functions $w : V(G) \to [0,1]$ satisfying $w(u) + w(v) \leq 1$ for every $uv \in E(G)$.

In 2004, Janson, Oleszkiewicz and Ruciński [507] proved that

$$\exp\{-O(M_H \log(1/p))\} \leq \mathbb{P}(X_H \geq (1+\varepsilon)\mathbb{E}X_H) \leq \exp\{-\Omega(M_H)\}, \tag{5.19}$$

where the implicit constants in (5.19) may depend on ε, and

$$M_H = \begin{cases} \min_{K\subseteq H}(n^{v_K}p^{e_K})^{1/\alpha^*_K}, & \text{if } n^{-1/m(H)} \leq p \leq n^{-1/\Delta_H}, \\ n^2 p^{\Delta_H}, & \text{if } p \geq n^{-1/\Delta_H}. \end{cases}$$

For example, if H is k-regular, then $M_H = n^2p^k$ for every p.

The logarithms of the upper and lower bounds in (5.19) differ by a multiplicative factor $\log(1/p)$. In 2011, DeMarco and Kahn formulated the following plausible conjecture (stated in [282] for $\varepsilon = 1$).

Conjecture: For any H and $\varepsilon > 0$,

$$\mathbb{P}(X_H \geq (1+\varepsilon)\mathbb{E}X_H) = \exp(-\Theta(\min\{\Phi_H, M_H \log(1/p)\})). \tag{5.20}$$

A careful look reveals that, when $\Delta_H \geq 2$, the minimum in (5.20) is only attained by Φ_H in a tiny range above the existence threshold (when $p \leq n^{-1/m(H)}(\log n)^{a_H}$ for some $a_H > 0$). In 2018, Šileikis and Warnke [806] found counterexample-graphs (all balanced but not strictly balanced) which violate (5.20) close to the threshold, and conjectured that (5.20) should hold under the stronger assumption $p \geq n^{-1/m_H+\delta}$.

DeMarco and Kahn [282] proved (5.20) for cliques $H = K_k$, $k = 3,4,\ldots$. Adamczak and Wolff [7] proved a polynomial concentration inequality which confirms (5.20) for any cycle $H = C_k$, $k = 3,4,\ldots$ and $p \geq n^{-\frac{k-2}{2(k-1)}}$. Moreover, Lubetzky and Zhao [631], via a large deviations framework of Chatterjee and Dembo [213], showed that (5.20) holds for any H and $p \geq n^{-\alpha}$ for a sufficiently small constant $\alpha > 0$. For more recent developments see [237], where it is shown that one can take $\alpha > 1/6\Delta_H$.

Chapter 6

Spanning Subgraphs

The previous chapter dealt with the existence of small subgraphs of a fixed size. In this chapter we concern ourselves with the existence of large subgraphs, most notably perfect matchings and Hamilton Cycles. The celebrated theorems of Hall and Tutte give necessary and sufficient conditions for a bipartite and arbitrary graph respectively to contain a perfect matching. Hall's theorem in particular can be used to establish that the threshold for having a perfect matching in a random bipartite graph can be identified with that of having no isolated vertices.

For general graphs we view a perfect matching as half a Hamilton cycle and prove thresholds for the existence of perfect matchings and Hamilton cycles in a similar way.

Having dealt with perfect matchings and Hamilton cycles, we turn our attention to long paths in sparse random graphs, i.e. in those where we expect a linear number of edges. We then analyse a simple greedy matching algorithm using differential equations.

We then consider random subgraphs of some fixed graph G, as opposed to random subgraphs of K_n. We give sufficient conditions for the existence of long paths and cycles.

We finally consider the existence of arbitrary spanning subgraphs H where we bound the maximum degree $\Delta(H)$.

6.1 Perfect Matchings

Before we move to the problem of the existence of a perfect matching, i.e., a collection of independent edges covering all of the vertices of a graph, in our main object of study, the random graph $\mathbb{G}_{n,p}$, we will analyse the same problem in a random bipartite graph. This problem is much simpler than the respective one for $\mathbb{G}_{n,p}$, but provides a general approach to finding a perfect matching in a random graph.

Bipartite Graphs

Let $\mathbb{G}_{n,n,p}$ be the random bipartite graph with vertex bi-partition $V = (A,B)$, $A = [1,n], B = [n+1,2n]$ in which each of the n^2 possible edges appears independently with probability p. The following theorem was first proved by Erdős and Rényi [334].

Theorem 6.1. *Let $\omega = \omega(n)$, $c > 0$ be a constant, and $p = \frac{\log n + \omega}{n}$. Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,n,p} \text{ has a perfect matching}) = \begin{cases} 0 & \text{if } \omega \to -\infty \\ e^{-2e^{-c}} & \text{if } \omega \to c \\ 1 & \text{if } \omega \to \infty. \end{cases}$$

Moreover,

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,n,p} \text{ has a perfect matching}) = \lim_{n\to\infty} \mathbb{P}(\delta(\mathbb{G}_{n,n,p}) \geq 1).$$

Proof. We will use Hall's condition for the existence of a perfect matching in a bipartite graph. It states that a bipartite graph contains a perfect matching if and only if the following condition is satisfied:

$$\forall S \subseteq A,\ |N(S)| \geq |S|, \tag{6.1}$$

where for a set of vertices S, $N(S)$ denotes the set of neighbors of S. We refer to S as a *witness*.

Now we can restrict our attention to minimal witnesses that satisfy $S \subseteq A, T \subseteq B$ satisfying (a) $|S| = |T| + 1$ and (b) each vertex in T has at least 2 neighbors in S and (c) $|S| \leq n/2$. Take a pair S,T with $|S| + |T|$ as small as possible. If the minimum degree $\delta \geq 1$ then $|S| \geq 2$.

(i) If $|S| > |T| + 1$, we can remove $|S| - |T| - 1$ vertices from $|S|$ – contradiction.

(ii) Suppose $\exists w \in T$ such that w has less than 2 neighbors in S. Remove w and its (unique) neighbor in $|S|$ – contradiction.

It is convenient to replace (6.1) by

$$\forall S \subseteq A,\ |S| \leq \frac{n}{2},\ \ |N(S)| \geq |S|, \tag{6.2}$$

$$\forall T \subseteq B,\ |T| \leq \frac{n}{2},\ \ |N(T)| \geq |T|. \tag{6.3}$$

This is because if we have a minimal witness S with $|S| > n/2$ and $|N(S)| < |S|$ then $T = B \setminus N(S)$ will violate (6.3).

It follows that

$$\begin{aligned}
&\mathbb{P}(\exists v : v \text{ is isolated}) \leq \mathbb{P}(\nexists \text{ a perfect matching}) \\
&\leq \mathbb{P}(\exists v : v \text{ is isolated}) + 2\,\mathbb{P}(\exists S \subseteq A, T \subseteq B, 2 \leq k = |S| \leq n/2, \\
&|T| = k-1, N(S) \subseteq T \text{and } e(S:T) \geq 2k-2).
\end{aligned}$$

Here $e(S:T)$ denotes the number of edges between S and T, and $e(S:T)$ can be assumed to be at least $2k-2$, because of (b) above.

Suppose now that $p = \frac{\log n + c}{n}$ for some constant c. Then let Y denote the number of sets S and T not satisfying the conditions (6.2), (6.3). Then

$$\begin{aligned}
\mathbb{E}Y &\leq 2\sum_{k=2}^{n/2} \binom{n}{k}\binom{n}{k-1}\binom{k(k-1)}{2k-2} p^{2k-2}(1-p)^{k(n-k)} \\
&\leq 2\sum_{k=2}^{n/2} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{k-1}\right)^{k-1} \left(\frac{ke(\log n + c)}{2n}\right)^{2k-2} e^{-npk(1-k/n)} \\
&\leq \sum_{k=2}^{n/2} n\left(\frac{e^{O(1)} n^{k/n} (\log n)^2}{n^{1-1/k}}\right)^k \\
&= \sum_{k=2}^{n/2} u_k.
\end{aligned}$$

Case 1: $\ 2 \leq k \leq n^{3/4}$.

$$u_k = n((e^{O(1)} n^{-1} \log n)^2)^k.$$

So

$$\sum_{k=2}^{n^{3/4}} u_k = O\left(\frac{1}{n^{1/2-o(1)}}\right).$$

Case 2: $\ n^{3/4} < k \leq n/2$.

$$u_k \leq n^{1-k(1/2-o(1))}$$

So

$$\sum_{n^{3/4}}^{n/2} u_k = O\left(n^{-n^{3/4}/3}\right).$$

So
$$\mathbb{P}(\nexists \text{ a perfect matching}) = \mathbb{P}(\exists \text{ isolated vertex}) + o(1).$$
Let X_0 denote the number of isolated vertices in $\mathbb{G}_{n,n,p}$. Then
$$\mathbb{E}X_0 = 2n(1-p)^n \approx 2e^{-c}.$$
It follows in fact via inclusion-exclusion or the method of moments that we have
$$\mathbb{P}(X_0 = 0) \approx e^{-2e^{-c}}$$
To prove the case for $|\omega| \to \infty$ we can use monotonicity and (1.7) and the fact that $e^{-e^{-2c}} \to 0$ if $c \to -\infty$ and $e^{-e^{-2c}} \to 1$ if $c \to \infty$. □

Non-Bipartite Graphs

We now consider $\mathbb{G}_{n,p}$. We could try to replace Hall's theorem by Tutte's theorem. A proof along these lines was given by Erdős and Rényi [335]. We can however get away with a simpler approach based on simple expansion properties of $\mathbb{G}_{n,p}$. The proof here can be traced back to Bollobás and Frieze [166].

Theorem 6.2. *Let $\omega = \omega(n)$, $c > 0$ be a constant, and let $p = \frac{\log n + c_n}{n}$. Then*
$$\lim_{\substack{n\to\infty \\ n \text{ even}}} \mathbb{P}(\mathbb{G}_{n,p} \text{ has a perfect matching}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty. \end{cases}$$
Moreover,
$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \text{ has a perfect matching}) = \lim_{n\to\infty} \mathbb{P}(\delta(\mathbb{G}_{n,p}) \ge 1).$$

Proof. We will for convenience only consider the case where $c_n = \omega \to \infty$ and $\omega = o(\log n)$. If $c_n \to -\infty$ then there are isolated vertices, w.h.p. and our proof can easily be modified to handle the case $c_n \to c$.

Our combinatorial tool that replaces Tutte's theorem is the following: We say that a matching M *isolates* a vertex v if no edge of M contains v.

For a graph G we let
$$\mu(G) = \max\{|M| : M \text{ is a matching in } G\}. \tag{6.4}$$
Let $G = (V,E)$ be a graph without a perfect matching i.e. $\mu(G) < \lfloor |V|/2 \rfloor$. Fix $v \in V$ and suppose that M is a maximum matching that isolates v. Let $S_0(v,M) = \{u \neq v : M \text{ isolates } u\}$. If $u \in S_0(v,M)$ and $e = \{x,y\} \in M$ and $f = \{u,x\} \in E$

then *flipping* e,f replaces M by $M' = M + f - e$. Here e is *flipped-out*. Note that $y \in S_0(v,M')$.

Now fix a maximum matching M that isolates v and let

$$A(v,M) = \bigcup_{M'} S_0(v,M')$$

where we take the union over M' obtained from M by a sequence of flips.

Lemma 6.3. *Let G be a graph without a perfect matching and let M be a maximum matching and v be a vertex isolated by M. Then $|N_G(A(v,M))| < |A(v,M)|$.*

Proof. Suppose that $x \in N_G(A(v,M))$ and that $f = \{u,x\} \in E$ where $u \in A(v,M)$. Now there exists y such that $e = \{x,y\} \in M$, else $x \in S_0(M) \subseteq A(v,M)$. We claim that $y \in A(v,M)$ and this will prove the lemma. Since then, every neighbor of $A(v,M)$ is the neighbor via an edge of M.

Suppose that $y \notin A(v,M)$. Let M' be a maximum matching that (i) isolates u and (ii) is obtainable from M by a sequence of flips. Now $e \in M'$ because if e has been flipped out then either x or y is placed in $A(v,M)$. But then we can do another flip with M', e and the edge $f = \{u,x\}$, placing $y \in A(v,M)$, contradiction. $\square$

We now change notation and write $A(v)$ in place of $A(v,M)$, understanding that there is some maximum matching that isolates v. Note that if $u \in A(v)$ then there is some maximum matching that isolates u and so $A(u)$ is well-defined. Furthermore, it always that case that if v is isolated by some maximum matching and $u \in A(v)$ then $\mu(G + \{u,v\}) = \mu(G) + 1$.

Now let

$$p = \frac{\log n + \theta \log\log n + \omega}{n}$$

where $\theta \geq 0$ is a fixed integer and $\omega \to \infty$ and $\omega = o(\log\log n)$.

We have introduced θ so that we can use some of the following results for the Hamilton cycle problem.

We write

$$\mathbb{G}_{n,p} = \mathbb{G}_{n,p_1} \cup \mathbb{G}_{n,p_2},$$

where

$$p_1 = \frac{\log n + \theta \log\log n + \omega/2}{n}$$

and

$$1 - p = (1 - p_1)(1 - p_2) \text{ so that } p_2 \approx \frac{\omega}{2n}.$$

Note that Theorem 4.3 implies:

$$\text{The minimum degree in } \mathbb{G}_{n,p_1} \text{ is at least } \theta + 1 \textit{ w.h.p.} \tag{6.5}$$

We consider a process where we add the edges of $\mathbb{G}_{n,p_2}$ one at a time to $\mathbb{G}_{n,p_1}$. We want to argue that if the current graph does not have a perfect matching then there is a good chance that adding such an edge $\{x,y\}$ will increase the size of a largest matching. This will happen if $y \in A(x)$. If we know that w.h.p. every set S for which $|N_{\mathbb{G}_{n,p_1}}(S)| < |S|$ satisfies $|S| \geq \alpha n$ for some constant $\alpha > 0$, then

$$\mathbb{P}(y \in A(x)) \geq \frac{\binom{\alpha n}{2} - i}{\binom{n}{2}} \geq \frac{\alpha^2}{2}, \tag{6.6}$$

provided $i = O(n)$.

This is because the edges we add will be uniformly random and there will be at least $\binom{\alpha n}{2}$ edges $\{x,y\}$ where $y \in A(x)$. Here given an initial x we can include edges $\{x',y'\}$ where $x' \in A(x)$ and $y' \in A(x')$. We have subtracted i to account for not re-using edges in $f_1, f_2, \ldots, f_{i-1}$.

In the light of this we now argue that sets S, with $|N_{\mathbb{G}_{n,p_1}}(S)| < (1+\theta)|S|$ are w.h.p. of size $\Omega(n)$.

Lemma 6.4. *Let $M = 100(\theta+7)$. W.h.p. $S \subseteq [n], |S| \leq \frac{n}{2e(\theta+5)M}$ implies $|N(S)| \geq (\theta+1)|S|$, where $N(S) = N_{\mathbb{G}_{n,p_1}}(S)$.*

Proof. Let a vertex of graph $G_1 = \mathbb{G}_{n,p_1}$ be large if its degree is at least $\lambda = \frac{\log n}{100}$, and small otherwise. Denote by *LARGE* and *SMALL*, the set of large and small vertices in G_1, respectively.

***Claim* 1.** *W.h.p. if $v,w \in SMALL$ then $dist(v,w) \geq 5$.*

Proof. If v,w are small and connected by a short path P, then v,w will have few neighbors outside P and conditional on P existing, v having few neighbors outside P is independent of w having few neighbors outside P. Hence,

$$\begin{aligned}
&\mathbb{P}(\exists v,w \in SMALL \text{ in } \mathbb{G}_{n,p_1} \text{ such that } dist(v,w) < 5) \\
&\leq \binom{n}{2}\left(\sum_{l=0}^{3} n^l p_1^{l+1}\right)\left(\sum_{k=0}^{\lambda}\binom{n}{k} p_1^k (1-p_1)^{n-k-5}\right)^2 \\
&\leq n(\log n)^4 \left(\sum_{k=0}^{\lambda} \frac{(\log n)^k}{k!} \cdot \frac{(\log n)^{(\theta+1)/100} \cdot e^{-\omega/2}}{n \log n}\right)^2 \\
&\leq 2n(\log n)^4 \left(\frac{(\log n)^{\lambda}}{\lambda!} \cdot \frac{(\log n)^{(\theta+1)/100} \cdot e^{-\omega/2}}{n \log n}\right)^2
\end{aligned} \tag{6.7}$$

$$
\begin{aligned}
&= O\left(\frac{(\log n)^{O(1)}}{n}(100e)^{\frac{2\log n}{100}}\right)\\
&= O(n^{-3/4})\\
&= o(1).
\end{aligned}
$$

The bound in (6.7) holds since $\lambda! \geq \left(\frac{\lambda}{e}\right)^{\lambda}$ and $\frac{u_{k+1}}{u_k} > 100$ for $k \leq \lambda$, where

$$
u_k = \frac{(\log n)^k}{k!} \cdot \frac{(\log n)^{(\theta+1)/100} \cdot e^{-\omega/2}}{n \log n}.
$$

□

***Claim* 2.** *W.h.p.* $\mathbb{G}_{n,p_1}$ *does not have a 4-cycle containing a small vertex.*

Proof.

$$
\begin{aligned}
&\mathbb{P}(\exists \text{ a 4-cycle containing a small vertex })\\
&\leq 4n^4 p_1^4 \sum_{k=0}^{(\log n)/100} \binom{n-4}{k} p_1^k (1-p_1)^{n-4-k}\\
&\leq n^{-3/4}(\log n)^4\\
&= o(1).
\end{aligned}
$$

□

***Claim* 3.** *W.h.p. in* $\mathbb{G}_{n,p_1}$ *for every* $S \subseteq [n], |S| \leq \frac{n}{2eM}, e(S) < \frac{|S|\log n}{M}$.

Proof.

$$
\begin{aligned}
&\mathbb{P}\left(\exists |S| \leq \frac{n}{2eM} \text{ and } e(S) \geq \frac{|S|\log n}{M}\right)\\
&\leq \sum_{s=\log n/M}^{n/2eM} \binom{n}{s}\binom{\binom{s}{2}}{s\log n/M} p_1^{s\log n/M}\\
&\leq \sum_{s=\log n/M}^{n/2eM} \left(\frac{ne}{s}\left(\frac{Me^{1+o(1)}s}{2n}\right)^{\log n/M}\right)^s\\
&\leq \sum_{s=\log n/M}^{n/2eM} \left(\left(\frac{s}{n}\right)^{-1+\log n/M} \cdot (Me^{1+o(1)})^{\log n/M}\right)^s\\
&= o(1).
\end{aligned}
$$

□

***Claim* 4.** *Let M be as in Claim 3. Then, w.h.p. in $\mathbb{G}_{n,p_1}$, if $S \subseteq LARGE, |S| \leq \frac{n}{2e(\theta+5)M}$ then $|N(S)| \geq (\theta+4)|S|$.*

Proof. Let $T = N(S), s = |S|, t = |T|$. Then we have

$$e(S \cup T) \geq e(S,T) \geq \frac{|S|\log n}{100} - 2e(S) \geq \frac{|S|\log n}{100} - \frac{2|S|\log n}{M}.$$

Then if $|T| \leq (\theta+4)|S|$ we have $|S \cup T| \leq (\theta+5)|S| \leq \frac{n}{2eM}$ and

$$e(S \cup T) \geq \frac{|S \cup T|}{\theta+5}\left(\frac{1}{100} - \frac{2}{M}\right)\log n = \frac{|S \cup T|\log n}{M}.$$

This contradicts Claim 3.

We can now complete the proof of Lemma 6.4. Let $|S| \leq \frac{n}{2e(\theta+5)M}$ and assume that $\mathbb{G}_{n,p_1}$ has minimum degree at least $\theta+1$.

Let $S_1 = S \cap SMALL$ and $S_2 = S \setminus S_1$. Then

$$\begin{aligned}
&|N(S)| \\
&\geq |N(S_1)| + |N(S_2)| - |N(S_1) \cap S_2| - |N(S_2) \cap S_1| - |N(S_1) \cap N(S_2)| \\
&\geq |N(S_1)| + |N(S_2)| - |S_2| - |N(S_2) \cap S_1| - |N(S_1) \cap N(S_2)|.
\end{aligned}$$

But Claim 1 and Claim 2 and minimum degree at least $\theta+1$ imply that

$$|N(S_1)| \geq (\theta+1)|S_1|,\ |N(S_2) \cap S_1| \leq \min\{|S_1|, |S_2|\}, |N(S_1) \cap N(S_2)| \leq |S_2|.$$

So, from this and Claim 4 we obtain

$$|N(S)| \geq (\theta+1)|S_1| + (\theta+4)|S_2| - 3|S_2| = (\theta+1)|S|.$$

□

We now go back to the proof of Theorem 6.2 for the case $c = \omega \to \infty$. Let the edges of $\mathbb{G}_{n,p_2}$ be $\{f_1, f_2, \ldots, f_s\}$ in random order, where $s \approx \omega n/4$. Let $\mathbb{G}_0 = \mathbb{G}_{n,p_1}$ and $\mathbb{G}_i = \mathbb{G}_{n,p_1} + \{f_1, f_2, \ldots, f_i\}$ for $i \geq 1$. It follows from Lemmas 6.3 and 6.4 that with $\mu(G)$ as in (6.4), and if $\mu(\mathbb{G}_i) < n/2$ then, assuming $\mathbb{G}_{n,p_1}$ has the expansion claimed in Lemma 6.4, with $\theta = 0$ and $\alpha = \frac{1}{10eM}$,

$$\mathbb{P}(\mu(\mathbb{G}_{i+1}) \geq \mu(\mathbb{G}_i) + 1 \mid f_1, f_2, \ldots, f_i) \geq \frac{\alpha^2}{2}, \tag{6.8}$$

see (6.6).

It follows that

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ does not have a perfect matching}) \leq \\ o(1) + \mathbb{P}(\mathrm{Bin}(s, \alpha^2/2) < n/2) = o(1).$$

We have used the notion of dominance, see Section 27.9 in order to use the binomial distribution in the above inequality. □

6.2 Hamilton Cycles

This was a difficult question left open in [332]. A breakthrough came with the result of Pósa [746]. The precise theorem given below can be credited to Komlós and Szemerédi [579], Bollobás [154] and Ajtai, Komlós and Szemerédi [15].

Theorem 6.5. *Let* $p = \frac{\log n + \log\log n + c_n}{n}$. *Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \text{ has a Hamilton cycle}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

Moreover,

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \text{ has a Hamilton cycle }) = \lim_{n\to\infty} \mathbb{P}(\delta(\mathbb{G}_{n,p}) \geq 2).$$

Proof. We will first give a proof of the first statement under the assumption that $c_n = \omega \to \infty$ where $\omega = o(\log\log n)$. The proof of the second statement is postponed to Section 6.3. Under this assumption, we have $\delta(G_{n,p}) \geq 2$ w.h.p., see Theorem 4.3. The result for larger p follows by monotonicity.

We now set up the main tool, viz. Pósa's Lemma. Let P be a path with end points a, b, as in Figure 6.1. Suppose that b does not have a neighbor outside of P.

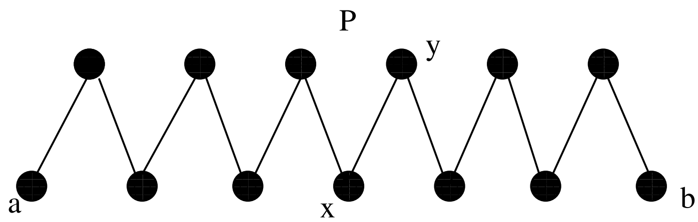


Figure 6.1: The path P

Notice that the P' below in Figure 6.2 is a path of the same length as P, obtained by a *rotation* with vertex a as the *fixed endpoint*. To be precise, suppose that $P = (a, \dots, x, y, y', \dots, b', b)$ and $\{b, x\}$ is an edge where x is an interior vertex of P. The path $P' = (a, \dots, x, b, b', \dots, y', y)$ is said to be obtained from P by a rotation.

Now let $END = END(P)$ denote the set of vertices v such that there exists a path P_v from a to v such that P_v is obtained from P by a sequence of rotations with vertex a fixed as in Figure 6.3.

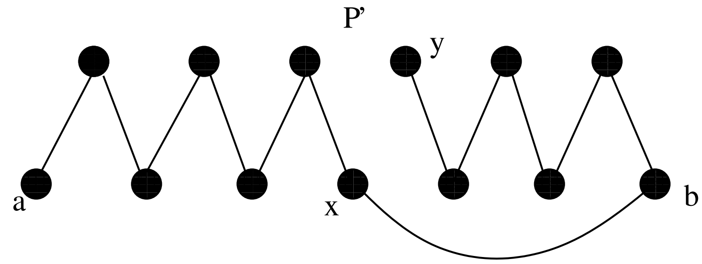


Figure 6.2: The path P' obtained after a single rotation

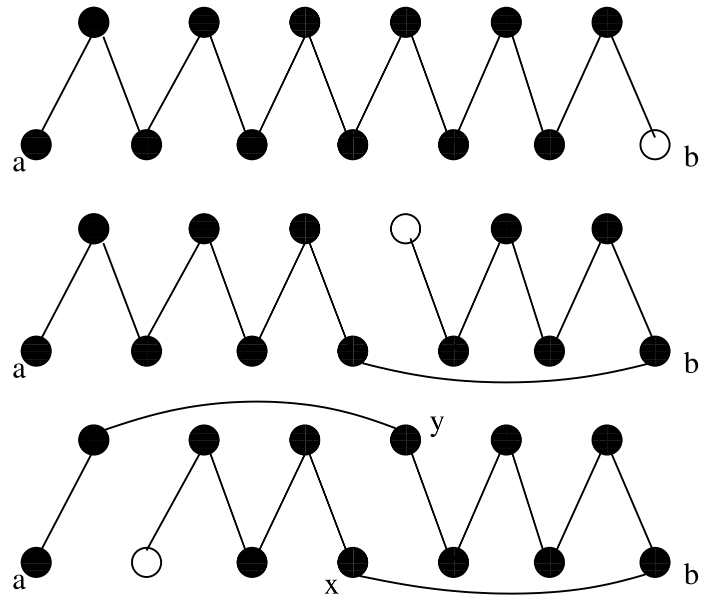


Figure 6.3: A sequence of rotations

Here the set *END* consists of all the white vertices on the path drawn below in Figure 6.4.

Lemma 6.6. *If $v \in P \setminus END$ and v is adjacent to $w \in END$ then there exists $x \in END$ such that the edge $\{v,x\} \in P$.*

Proof. Suppose to the contrary that x,y are the neighbors of v on P and that $v,x,y \notin END$ and that v is adjacent to $w \in END$. Consider the path P_w. Let $\{r,t\}$ be the neighbors of v on P_w. Now $\{r,t\} = \{x,y\}$ because if a rotation deleted $\{v,y\}$ say then v or y becomes an endpoint. But then after a further rotation from P_w we see that $x \in END$ or $y \in END$.

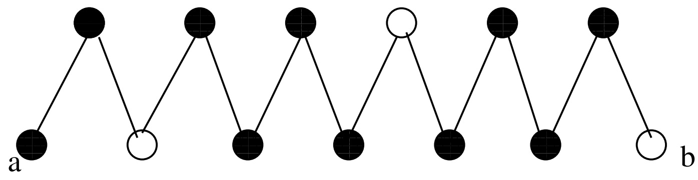


Figure 6.4: The set END

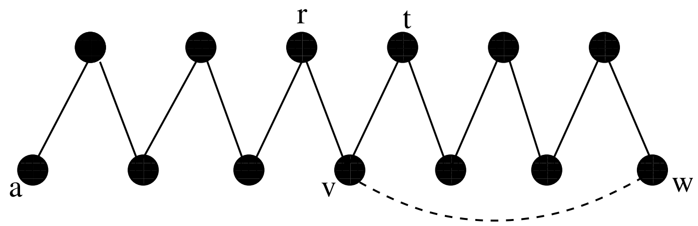


Figure 6.5: One of r,t will become an endpoint after a rotation

□

Corollary 6.7.

$$|N(END)| < 2|END|.$$

□

It follows from Lemma 6.4 with $\theta = 1$ that w.h.p. we have

$$|END| \geq \alpha n \text{ where } \alpha = \frac{1}{12eM}. \tag{6.9}$$

We now consider the following algorithm that searches for a Hamilton cycle in a connected graph G. The probability p_1 is above the connectivity threshold and so $\mathbb{G}_{n,p_1}$ is connected w.h.p. Our algorithm will proceed in stages. At the beginning of Stage k we will have a path of length k in G and we will try to grow it by one vertex in order to reach Stage $k+1$. In Stage $n-1$, our aim is simply to create a Hamilton cycle, given a Hamilton path. We start the whole procedure with an arbitrary path of G.

Algorithm Pósa:

(a) Let P be our path at the beginning of Stage k. Let its endpoints be x_0, y_0. If x_0 or y_0 have neighbors outside P then we can simply extend P to include one of these neighbors and move to stage $k+1$.

(b) Failing this, we do a sequence of rotations with x_0 as the fixed vertex until one of two things happens: (i) We produce a path Q with an endpoint y that has a neighbor outside of Q. In this case we extend Q and proceed to stage $k+1$. (ii) No sequence of rotations leads to Case (i). In this case let END denote the set of endpoints of the paths produced. If $y \in END$ then P_y denotes a path with endpoints x_0, y that is obtained from P by a sequence of rotations.

(c) If we are in Case (bii) then for each $y \in END$ we let $END(y)$ denote the set of vertices z such that there exists a longest path Q_z from y to z such that Q_z is obtained from P_y by a sequence of rotations with vertex y fixed. Repeating the argument above in (b) for each $y \in END$, we either extend a path and begin Stage $k+1$ or we go to (d).

(d) Suppose now that we do not reach Stage $k+1$ by an extension and that we have constructed the sets END and $END(y)$ for all $y \in END$. Suppose that G contains an edge (y,z) where $z \in END(y)$. Such an edge would imply the existence of a cycle $C = (z, Q_y, z)$. If this is not a Hamilton cycle then connectivity implies that there exist $u \in C$ and $v \notin C$ such that u, v are joined by an edge. Let w be a neighbor of u on C and let P' be the path obtained from C by deleting the edge (u,w). This creates a path of length $k+1$ viz. the path w, P', v, and we can move to Stage $k+1$.

A pair z, y where $z \in END(y)$ is called a *booster* in the sense that if we added this edge to $\mathbb{G}_{n,p_1}$ then it would either (i) make the graph Hamiltonian or (ii) make the current path longer. We argue now that $\mathbb{G}_{n,p_2}$ can be used to "boost" P to a Hamilton cycle, if necessary.

We observe now that when $G = \mathbb{G}_{n,p_1}$, $|END| \geq \alpha n$ w.h.p., see (6.9). Also, $|END(y)| \geq \alpha n$ for all $y \in END$. So we will have $\Omega(n^2)$ boosters.

For a graph G let $\lambda(G)$ denote the length of a longest path in G, when G is not Hamiltonian and let $\lambda(G) = n$ when G is Hamiltonian. Let the edges of $\mathbb{G}_{n,p_2}$ be $\{f_1, f_2, \ldots, f_s\}$ in random order, where $s \approx \omega n/4$. Let $\mathbb{G}_0 = \mathbb{G}_{n,p_1}$ and $\mathbb{G}_i = \mathbb{G}_{n,p_1} + \{f_1, f_2, \ldots, f_i\}$ for $i \geq 1$. It follows from Lemmas 6.3 and 6.4 that if $\lambda(\mathbb{G}_i) < n$ then, assuming $\mathbb{G}_{n,p_1}$ has the expansion claimed in Lemma 6.4,

$$\mathbb{P}(\lambda(\mathbb{G}_{i+1}) \geq \lambda(\mathbb{G}_i) + 1 \mid f_1, f_2, \ldots, f_i) \geq \frac{\alpha^2}{2}, \tag{6.10}$$

see (6.6), replacing $A(v)$ by $END(v)$.

It follows that

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is not Hamiltonian}) \leq o(1) + \mathbb{P}(\mathrm{Bin}(s, \alpha^2/2) < n) = o(1). \tag{6.11}$$

6.3 Long Paths and Cycles in Sparse Random Graphs

In this section we study the length of the longest path and cycle in $\mathbb{G}_{n,p}$ when $p = c/n$ where $c = O(\log n)$, most importantly for c is a large constant. We have seen in Chapter 1 that under these conditions, $\mathbb{G}_{n,p}$ will w.h.p. have isolated vertices and so it will not be Hamiltonian. We can however show that it contains a cycle of length $\Omega(n)$ w.h.p.

The question of the existence of a long path/cycle was posed by Erdős and Rényi in [332]. The first positive answer to this question was given by Ajtai, Komlós and Szemerédi [14] and by de la Vega [838]. The proof we give here is due to Krivelevich, Lee and Sudakov. It is subsumed by the more general results of [594].

Theorem 6.8. *Let $p = c/n$ where c is sufficiently large but $c = O(\log n)$. Then w.h.p.*

(a) $\mathbb{G}_{n,p}$ has a path of length at least $\left(1 - \frac{6\log c}{c}\right) n$.

(b) $\mathbb{G}_{n,p}$ has a cycle of length at least $\left(1 - \frac{12\log c}{c}\right) n$.

Proof. We prove this theorem by analysing simple properties of Depth First Search (DFS). This is a well known algorithm for exploring the vertices of a component of a graph. We can describe the progress of this algorithm using three sets: U is the set of *unexplored* vertices that have not yet been reached by the search. D is the set of *dead* vertices. These have been fully explored and no longer take part in the process. $A = \{a_1, a_2, \ldots, a_r\}$ is the set of active vertices and they form a path from a_1 to a_r. We start the algorithm by choosing a vertex v from which to start the process. Then we let

$$A = \{v\} \text{ and } D = \emptyset \text{ and } U = [n] \setminus \{v\} \text{ and } r = 1.$$

We now describe how these sets change during one step of the algorithm.

Step (a) If there is an edge $\{a_r, w\}$ for some $w \in U$ then we choose one such w and extend the path defined by A to include w.

$$a_{r+1} \leftarrow w; A \leftarrow A \cup \{w\}; U \leftarrow U \setminus \{w\}; r \leftarrow r+1.$$

We now repeat Step (a).
If there is no such w then we do Step (b):

Step (b) We have now completely explored a_r.

$$D \leftarrow D \cup \{a_r\}; A \leftarrow A \setminus \{a_r\}; r \leftarrow r-1.$$

If $r \geq 1$ we go to Step (a). Otherwise, if $U = \emptyset$ at this point then we terminate the algorithm. If $U \neq \emptyset$ then we choose some $v \in U$ to re-start the process with $r = 1$. We then go to Step (a).

We make the following simple observations:

- A step of the algorithm increases $|D|$ by one or decreases $|U|$ by one and so at some stage we must have $|D| = |U| = s$ for some positive integer s.

- There are no edges between D and U because we only add a_r to D when there are no edges from a_r to U edges and U does not increase from this point on.

Thus at some stage we have two disjoint sets D, U of size s with no edges between them and a path of length $|A| - 1 = n - 2s - 1$. This plus the following claim implies that $\mathbb{G}_{n,p}$ has a path P of length at least $\left(1 - \frac{6\log c}{c}\right) n$ w.h.p. Note that if c is large then

$$\alpha > \frac{3\log c}{c} \text{ implies } c > \frac{2}{\alpha} \log\left(\frac{e}{\alpha}\right).$$

Claim 5. *Let $0 < \alpha < 1$ be a positive constant. If $p = c/n$ and $c > \frac{2}{\alpha}\log\left(\frac{e}{\alpha}\right)$ then w.h.p. in $\mathbb{G}_{n,p}$, every pair of disjoint sets S_1, S_2 of size at least $\alpha n - 1$ are joined by at least one edge.*

Proof. The probability that there exist sets S_1, S_2 of size (at least) $\alpha n - 1$ with no joining edge is at most

$$\binom{n}{\alpha n - 1}^2 (1-p)^{(\alpha n - 1)^2} \leq \left(\frac{e^{2+o(1)}}{\alpha^2} e^{-c\alpha}\right)^{\alpha n - 1} = o(1).$$

□

To complete the proof of the theorem, we apply the above lemma to the vertices S_1, S_2 on the two sub-paths P_1, P_2 of length $\frac{3\log c}{c} n$ at each end of P. There will w.h.p. be an edge joining S_1, S_2, creating the cycle of the claimed length. □

Krivelevich and Sudakov [604] used DFS to give simple proofs of good bounds on the size of the largest component in $\mathbb{G}_{n,p}$ for $p = \frac{1+\varepsilon}{n}$ where ε is a small constant. Exercises 6.7.19, 6.7.20 and 6.7.21 elaborate on their results.

Completing the proof of Theorem 6.5

We need to prove part (b). So we let $1-p=(1-p_1)(1-p_2)$ where $p_2=\frac{1}{n\log\log n}$. Then we apply Theorem 6.8(a) to argue that w.h.p. G_{n,p_1} has a path of length $n\left(1-O\left(\frac{\log\log n}{\log n}\right)\right)$.

Now, conditional on G_{n,p_1} having minimum degree at least two, the proof of the statement of Lemma 6.4 goes through without change for $\theta=1$ i.e. $S\subseteq [n], |S|\leq \frac{n}{10000}$ implies $|N(S)|\geq 2|S|$. We can then use use the extension-rotation argument that we used to prove Theorem 6.5(c). This time we only need to close $O\left(\frac{n\log\log n}{\log n}\right)$ cycles and we have $\Omega\left(\frac{n}{\log\log n}\right)$ edges. Thus (6.11) is replaced by

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is not Hamiltonian} \mid \delta(G_{n,p_1})\geq 2)\leq \\ o(1)+\mathbb{P}\left(\text{Bin}\left(\frac{c_1 n}{\log\log n},10^{-8}\right)<\frac{c_2 n\log\log n}{\log n}\right)=o(1),$$

for some hidden constants c_1,c_2.

6.4 Greedy Matching Algorithm

In this section we see how we can use differential equations to analyse the performance of a greedy algorithm for finding a large matching in a random graph. Finding a large matching is a standard problem in Combinatorial Optimisation. The first polynomial time algorithm to solve this problem was devised by Edmonds in 1965 and runs in time $O(|V|^4)$ [326]. Over the years, many improvements have been made. Currently the fastest such algorithm is that of Micali and Vazirani which dates back to 1980. Its running time is $O(|E|\sqrt{|V|})$ [680]. These algorithms are rather complicated and there is a natural interest in the performance of simpler heuristic algorithms which should find large, but not necessarily maximum matchings. One well studied class of heuristics goes under the general title of the GREEDY heuristic.

The following simple greedy algorithm proceeds as follows: Beginning with graph $G=(V,E)$ we choose a random edge $e=\{u,v\}\in E$ and place it in a set M. We then delete u,v and their incident edges from G and repeat. In the following, we analyse the size of the matching M produced by this algorithm.

Algorithm *GREEDY*

begin

  $M\leftarrow\emptyset$;

  while $E(G)\neq\emptyset$ **do**

  begin

A: Randomly choose $e = \{x,y\} \in E$
$G \leftarrow G \setminus \{x,y\}$;
$M \leftarrow M \cup \{e\}$

end;
Output M
end

($G \setminus \{x,y\}$ is the graph obtained from G by deleting the vertices x,y and all incident edges.)

We will study this algorithm in the context of the pseudo-graph model $\mathbb{G}_{n,m}^{(B)}$ of Section 1.3 and apply (1.17) to bring the results back to $\mathbb{G}_{n,m}$. We will argue next that if at some stage G has ν vertices and μ edges then G is equally likely to be any pseudo-graph with these parameters.

We will use the method of *deferred decisions*, a term coined in Knuth, Motwani and Pittel [573]. In this scenario, we do not expose the edges of the pseudo-graph until we actually need to. So, as a thought experiment, think that initially there are m boxes, each containing a uniformly random ordered pair of distinct integers x,y from $[n]$. Until the box is opened, the contents are unknown except for their distribution. Observe that opening box A and observing its contents tells us nothing more about the contents of box B. This would not be the case if as in $\mathbb{G}_{n,m}$ we insisted that no two boxes had the same contents.

Remark 6.9. A step of GREEDY involves choosing the first unopened box at random to expose its contents x,y.

After this, the contents of the remaining boxes will of course remain uniformly random. The algorithm will then ask for each box with x or y to be opened. Other boxes will remain unopened and all we will learn is that their contents do not contain x or y and so they are still uniform over the remaining possible edges.

We need the following

Lemma 6.10. *Suppose that $m = cn$ for some constant $c > 0$. Then w.h.p. the maximum degree in $\mathbb{G}_{n,m}^{(B)}$ is at most* $\log n$.

Proof. The degree of a vertex is distributed as $\mathrm{Bin}(m, 2/n)$. So, if Δ denotes the maximum degree in $\mathbb{G}_{n,m}^{(B)}$, then with $\ell = \log n$,

$$\mathbb{P}(\Delta \geq \ell) \leq n \binom{m}{\ell} \left(\frac{2}{n}\right)^{\ell} \leq n \left(\frac{2ce}{\ell}\right)^{\ell} = o(1).$$

□

Now let $X(t) = (\nu(t), \mu(t)), t = 1, 2, \ldots$, denote the number of vertices and edges in the graph at the end of the tth iteration of GREEDY. Also, let $G_t = (V_t, E_t) = G$ at this point and let $G'_t = (V_t, E_t \setminus e)$ where e is a uniform random edge of E_t. Thus $\nu(1) = n, \mu(1) = m$ and $G_1 = \mathbb{G}^{(B)}_{n,m}$. Now $\nu(t+1) = \nu(t) - 2$ and so $\nu(t) = n - 2t + 2$. Let $d_t(\cdot)$ denote degree in G'_t and let $\theta_t(x, y)$ denote the number of copies of the edge $\{x, y\}$ in G_t, excluding e. Then we have

$$\mathbb{E}(\mu(t+1) \mid G_t) = \mu(t) - (d_t(x) + d_t(y) - 1) + \theta_t(x, y)).$$

Taking expectations over G_t we have

$$\mathbb{E}(\mu(t+1)) = \mathbb{E}(\mu(t)) - \mathbb{E}(d_t(x)) - \mathbb{E}(d_t(y)) + 1 + \mathbb{E}(\theta_t(x, y)).$$

Now we will see momentarily that $\mathbb{E}(d_t(x)^2) = O(1)$. And,

$$\begin{aligned}
&\mathbb{E}(d_t(x) + d_t(y) \mid G_t) \\
&= \sum_{i=1}^{\nu(t)} \sum_{j \neq i=1}^{\nu(t)} \frac{d_t(i) d_t(j)}{2\mu(t)(2\mu(t)-1)} (d_t(i) + d_t(j) \\
&= \frac{1}{2\mu(t)(2\mu(t)-1)} \left(\sum_{i=1}^{\nu(t)} d_t(i) \sum_{j \neq i} d_j(t)^2 + \sum_{i=1}^{\nu(t)} d_t(i)^2 \sum_{j \neq i} d_j(t) \right) \\
&= \frac{1}{2\mu(t)(2\mu(t)-1)} \left(2\mu(t) \left(\sum_{j=1}^{\nu(t)} d_j(t)^2 - O(1) \right) + (2\mu(t) - O(1)) \sum_{i=1}^{\nu(t)} d_t(i)^2 \right) \\
&= \frac{1}{\mu(t)} \sum_{i=1}^{\nu(t)} d_t(i)^2 + O\left(\frac{1}{\mu(t)} \right).
\end{aligned}$$

In the model $G^B_{n,m}$,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\mu(t)} \sum_{i=1}^{\nu(t)} d_t(i)^2 \middle| \mu(t) \right) &= \frac{\nu(t)}{\mu(t)} \sum_{k=0}^{\mu(t)} k^2 \binom{\mu(t)}{k} \left(\frac{2}{\nu(t)} \right)^k \left(1 - \frac{2}{\nu(t)} \right)^{\mu(t)-k} \\
&= 2\left(1 - \frac{2}{\nu(t)} + \frac{2\mu(t)}{\nu(t)} \right).
\end{aligned}$$

So,

$$\mathbb{E}(\mu(t+1)) = \mathbb{E}(\mu(t)) - \frac{4\,\mathbb{E}(\mu(t))}{n-2t} - 1 + O\left(\frac{1}{n-2t} \right). \tag{6.12}$$

Here we use Remark 6.9 to argue that $\mathbb{E}\,\theta_t(x, y)) = O(1/(n-2t))$.

This *suggests* that w.h.p. $\mu(t) \approx nz(t/n)$ where $z(0) = c$ and $z(\tau)$ is the solution to the differential equation

$$\frac{dz}{d\tau} = -\frac{4z(\tau)}{1-2\tau} - 1.$$

This is easy to solve and gives

$$z(\tau) = \left(c + \frac{1}{2}\right)(1-2\tau)^2 - \frac{1-2\tau}{2}.$$

The smallest root of $z(\tau) = 0$ is $\tau = \frac{c}{2c+1}$. This suggests the following theorem.

Theorem 6.11. *W.h.p., running GREEDY on $\mathbb{G}_{n,cn}$ finds a matching of size $\frac{c+o(1)}{2c+1}n$.*

Proof. We will replace $\mathbb{G}_{n,m}$ by $\mathbb{G}_{n,m}^{(B)}$ and consider the random sequence $\mu(t)$, $t = 1, 2, \ldots$. The number of edges in the matching found by GREEDY equals one less than the first value of t for which $\mu(t) = 0$. We show that w.h.p. $\mu(t) > 0$ if and only if $t \leq \frac{c+o(1)}{2c+1}n$. We will use Theorem 28.1 of Chapter 28.

In our set up for the theorem we let

$$f(\tau, x) = -\frac{4x}{1-2\tau} - 1.$$
$$D = \left\{(\tau, x) : -\frac{1}{n} < t < \mathscr{T}_D = \frac{c}{2c+1}, 0 < x < \frac{1}{2}\right\}.$$

We let $X(t) = \mu(t)$ for the statement of the theorem. Then we have to check the conditions:

(P1) $|\mu(t)| \leq cn, \ \forall t < T_D = \mathscr{T}_D n.$

(P2) $|\mu(t+1) - \mu(t)| \leq 2\log n, \ \forall t < T_D.$

(P3) $|\mathbb{E}(\mu(t+1) - \mu(t)|H_t, \mathscr{E}) - f(t/n, X(t)/n)| \leq \frac{A}{n}, \forall t < T_D.$
Here $\mathscr{E} = \{\Delta \leq \log n\}$ and this is needed for (P2).

(P4) $f(t,x)$ is continuous and satisfies a Lipschitz condition
$|f(t,x) - f(t',x')| \leq L\|(t,x) - (t',x')\|_\infty$ where $L = 10(2c+1)^2$,
for $(t,x), (t',x') \in D \cap \{(t,x) : t \geq 0\}$

Here $f(t,x) = -1 - \frac{4x}{1-2t}$ and we can justify L of P4 as follows:

$$|f(t,x) - f(t',x')|$$
$$= \left|\frac{4x}{1-2t} - \frac{4x'}{1-2t'}\right|$$

$$\leq \left|\frac{4(x-x')}{(1-2t)(1-2t')}\right| + \left|\frac{8x'(t-t')}{(1-2t)(1-2t')}\right| + \left|\frac{8't(x-x')}{(1-2t)(1-2t')}\right|$$
$$\leq 10(2c+1)^2.$$

Now let $\beta = n^{1/5}$ and $\lambda = n^{-1/20}$ and $\sigma = \mathscr{T}_D - 10\lambda$ and apply the theorem. This shows that w.h.p. $\mu(t) = nz(t/n) + O(n^{19/20})$ for $t \leq \sigma n$. $\square$

The result in Theorem 6.11 is taken from Dyer, Frieze and Pittel [324], where a central limit theorem is proven for the size of the matching produced by GREEDY.

The use of differential equations to approximate the trajectory of a stochastic process is quite natural and is often very useful. It is however not always best practise to try and use an "off the shelf" theorem like Theorem 28.1 in order to get a best result. It is hard to design a general theorem that can deal optimally with terms that are $o(n)$.

6.5 Random Subgraphs of Graphs with Large Minimum Degree

Here we prove an extension of Theorem 6.8. The setting is this. We have a sequence of graphs G_k with minimum degree at least k, where $k \to \infty$. We construct a random subgraph G_p of $G = G_k$ by including each edge of G, independently with probability p. Thus if $G = K_n$, G_p is $G_{n,p}$. The theorem we prove was first proved by Krivelevich, Lee and Sudakov [594]. The argument we present here is due to Riordan [763].

In the following we abbreviate $(G_k)_p$ to G_p where the parameter k is to be understood.

Theorem 6.12. *Let G_k be a sequence of graphs with minimum degree at least k where $k \to \infty$. Let p be such that $pk \to \infty$ as $k \to \infty$. Then w.h.p. G_p contains a cycle of length at least $(1-o(1))k$.*

Proof. We will assume that G has n vertices. We let T denote the *forest* produced by depth first search. We also let D, U, A be as in the proof of Theorem 6.8. Let v be a vertex of the rooted forest T. There is a unique *vertical* path from v to the root of its component. We write $\mathscr{A}(v)$ for the set of ancestors of v, i.e., vertices (excluding v) on this path. We write $\mathscr{D}(v)$ for the set of descendants of v, again excluding v. Thus $w \in \mathscr{D}(v)$ if and only if $v \in \mathscr{A}(w)$. The distance $d(u,v)$ between two vertices u and v on a common vertical path is just their graph distance along this path. We write $\mathscr{A}_i(v)$ and $\mathscr{D}_i(v)$ for the set of ancestors/descendants of v at distance exactly i, and $\mathscr{A}_{\leq i}(v), \mathscr{D}_{\leq i}(v)$ for those at distance at most i. By the depth of a vertex we mean its distance from the root. The height of a vertex v is $\max\{i : \mathscr{D}_i(v) \neq \emptyset\}$. Let R denote the set of edges of G that are not tested for inclusion in G_p during the exploration.

Lemma 6.13. *Every edge e of R joins two vertices on some vertical path in T.*

Proof. Let $e = \{u,v\}$ and suppose that u is placed in D before v. When u is placed in D, v cannot be in U, else $\{u,v\}$ would have been tested. Also, v cannot be in D by our choice of u. Therefore at this time $v \in A$ and there is a vertical path from v to u. □

Lemma 6.14. *With high probability, at most $2n/p = o(kn)$ edges are tested during the depth first search exploration.*

Proof. Each time an edge is tested, the test succeeds (the edge is found to be present) with probability p. The Chernoff bound implies that the probability that more than $2n/p$ tests are made but fewer than n succeed is $o(1)$. But every successful test contributes an edge to the forest T, so w.h.p. at most n tests are successful. □

From now on let us fix an arbitrary (small) constant $0 < \varepsilon < 1/10$. We call a vertex v *full* if it is incident with at least $(1-\varepsilon)k$ edges in R.

Lemma 6.15. *With high probability, all but $o(n)$ vertices of T_k are full.*

Proof. Since G has minimum degree at least k, each $v \in V(G) = V(T)$ that is not full is incident with at least εk tested edges. If for some constant $c > 0$ there are at least cn such vertices, then there are at least $c\varepsilon kn/2$ tested edges. But the probability of this is $o(1)$ by Lemma 6.14. □

Let us call a vertex v *rich* if $|\mathscr{D}(v)| \geq \varepsilon k$, and *poor* otherwise. In the next two lemmas, (T_k) is a sequence of rooted forests with n vertices. We suppress the dependence on k in notation.

Lemma 6.16. *Suppose that $T = T_k$ contains $o(n)$ poor vertices. Then, for any constant C, all but $o(n)$ vertices of T are at height at least Ck.*

Proof. For each rich vertex v, let $P(v)$ be a set of $\lceil \varepsilon k \rceil$ descendants of v, obtained by choosing vertices of $\mathscr{D}(v)$ one-by-one starting with those furthest from v. For every $w \in P(v)$ we have $\mathscr{D}(w) \subseteq P(v)$, so $|\mathscr{D}(w)| < \varepsilon k$, i.e., w is poor. Consider the set S_1 of ordered pairs (v,w) with v rich and $w \in P(v)$. Each of the $n - o(n)$ rich vertices appears in at least εk pairs, so $|S_1| \geq (1-o(1))\varepsilon kn$.

For any vertex w we have $|\mathscr{A}_{\leq i}(w)| \leq i$, since there is only one ancestor at each distance, until we hit the root. Since $(v,w) \in S_1$ implies that w is poor and $v \in \mathscr{A}(w)$, and there are only $o(n)$ poor vertices, at most $o(Ckn) = o(kn)$ pairs $(v,w) \in S_1$ satisfy $d(v,w) \leq Ck$. Thus $S_1' = \{(v,w) \in S_1 : d(v,w) > Ck\}$ satisfies

$|S_1'| \geq (1-o(1))\varepsilon kn$. Since each vertex v is the first vertex of at most $\lceil \varepsilon k \rceil \approx \varepsilon k$ pairs in $S_1 \supseteq S_1'$, it follows that $n - o(n)$ vertices v appear in pairs $(v,w) \in S_1'$. Since any such v has height at least Ck, the proof is complete. □

Let us call a vertex v *light* if $|\mathscr{D}_{\leq (1-5\varepsilon)k}(v)| \leq (1-4\varepsilon)k$, and *heavy* otherwise. Let H denote the set of heavy vertices in T.

Lemma 6.17. *Suppose that $T = T_k$ contains $o(n)$ poor vertices, and let $X \subseteq V(T)$ with $|X| = o(n)$. Then, for k large enough, T contains a vertical path P of length at least $\varepsilon^{-2}k$ containing at most $\varepsilon^2 k$ vertices in $X \cup H$.*

Proof. Let S_2 be the set of pairs (u,v) where u is an ancestor of v and $0 < d(u,v) \leq (1-5\varepsilon)k$. Since a vertex has at most one ancestor at any given distance, we have $|S_2| \leq (1-5\varepsilon)kn$. On the other hand, by Lemma 6.16 all but $o(n)$ vertices u are at height at least k and so appear in at least $(1-5\varepsilon)k$ pairs $(u,v) \in S_2$. It follows that only $o(n)$ vertices u are in more than $(1-4\varepsilon)k$ such pairs, i.e., $|H| = o(n)$.

Let S_3 denote the set of pairs (u,v) where $v \in X \cup H$, u is an ancestor of v, and $d(u,v) \leq \varepsilon^{-2}k$. Since a given v can only appear in $\varepsilon^{-2}k$ pairs $(u,v) \in S_3$, we see that $|S_3| \leq \varepsilon^{-2}k|X \cup H| = o(kn)$. Hence only $o(n)$ vertices u appear in more than $\varepsilon^2 k$ pairs $(u,v) \in S_3$.

By Lemma 6.16, all but $o(n)$ vertices are at height at least $\varepsilon^{-2}k$. Let u be such a vertex appearing in at most $\varepsilon^2 k$ pairs $(u,v) \in S_3$, and let P be the vertical path from u to some $v \in \mathscr{D}_{\varepsilon^{-2}k}(u)$. Then P has the required properties. □

Proof of Theorem 6.12

Fix $\varepsilon > 0$. It suffices to show that w.h.p. G_p contains a cycle of length at least $(1-5\varepsilon)k$, say. Explore G_p by depth-first search as described above. We condition on the result of the exploration, noting that the edges of R are still present independently with probability p. By Lemma 6.13, $\{u,v\} \in R$ implies that u is either an ancestor or a descendant of v. By Lemma 6.15, we may assume that all but $o(n)$ vertices are full.

Suppose that

$$|\{u : \{u,v\} \in R, d(u,v) \geq (1-5\varepsilon)k\}| \geq \varepsilon k. \tag{6.13}$$

for some vertex v. Then, since $\varepsilon kp \to \infty$, testing the relevant edges $\{u,v\}$ one-by-one, w.h.p we find one present in G_p, forming, together with T, the required long cycle. On the other hand, suppose that (6.13) fails for every v. Suppose that some vertex v is full but poor. Since v has at most εk descendants, there are at least $(1-2\varepsilon)k$ pairs $\{u,v\} \in R$ with $u \in \mathscr{A}(v)$. Since v has only one ancestor at each distance, it follows that (6.13) holds for v, a contradiction.

We have shown that we can assume that no poor vertex is full. Hence there are $o(n)$ poor vertices, and we may apply Lemma 6.17, with X the set of vertices that are not full. Let P be the path whose existence is guaranteed by the lemma, and let Z be the set of vertices on P that are full and light, so $|V(P)\setminus Z| \le \varepsilon^2 k$. For any $v \in Z$, since v is full, there are at least $(1-\varepsilon)k$ vertices $u \in \mathscr{A}(v) \cup \mathscr{D}(v)$ with $\{u,v\} \in R$. Since (6.13) does not hold, at least $(1-2\varepsilon)k$ of these vertices satisfy $d(u,v) \le (1-5\varepsilon)k$. Since v is light, in turn at least $2\varepsilon k$ of these u must be in $\mathscr{A}(v)$. Recalling that a vertex has at most one ancestor at each distance, we find a set $R(v)$ of at least εk vertices $u \in \mathscr{A}(v)$ with $\{u,v\} \in R$ and $\varepsilon k \le d(u,v) \le (1-5\varepsilon)k \le k$.

It is now easy to find a (very) long cycle w.h.p. Recall that $Z \subseteq V(P)$ with $|V(P)\setminus Z| \le \varepsilon^2 k$. Thinking of P as oriented upwards towards the root, let v_0 be the lowest vertex in Z. Since $|R(v_0)| \ge \varepsilon k$ and $kp \to \infty$, w.h.p. there is an edge $\{u_0, v_0\}$ in G_p with $u_0 \in R(v_0)$. Let v_1 be the first vertex below u_0 along P with $v_1 \in Z$. Note that we go up at least εk steps from v_0 to u_0 and down at most $1+|V(P)\setminus Z| \le 2\varepsilon^2 k$ from u_0 to v_1, so v_1 is above v_0. Again w.h.p. there is an edge $\{u_1, v_1\}$ in G_p with $u_1 \in R(v_1)$, and so at least εk steps above v_1. Continue downwards from u_1 to the first $v_2 \in Z$, and so on. Since $\varepsilon^{-1} = O(1)$, w.h.p. we may continue in this way to find overlapping chords $\{u_i, v_i\}$ for $0 \le i \le \lfloor 2\varepsilon^{-1} \rfloor$, say. (Note that we remain within P as each upwards step has length at most k.) These chords combine with P to give a cycle of length at least $(1-2\varepsilon^{-1} \times 2\varepsilon^2)k = (1-4\varepsilon)k$, as shown in Figure 6.6.

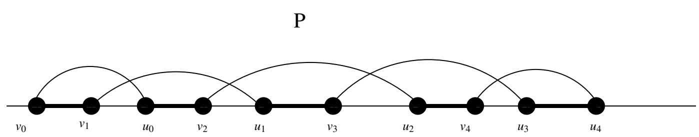


Figure 6.6: The path P, with the root off to the right. Each chord $\{v_i, u_i\}$ has length at least εk (and at most k); from u_i to v_{i+1} is at most $2\varepsilon^2 k$ steps back along P. The chords and the thick part of P form a cycle.

6.6 Spanning Subgraphs

Consider a fixed sequence $H^{(d)}$ of graphs where $n = |V(H^{(d)})| \to \infty$. In particular, we consider a sequence Q_d of d-dimensional cubes where $n = 2^d$ and a sequence of 2-dimensional lattices L_d of order $n = d^2$. We ask when $\mathbb{G}_{n,p}$ or $\mathbb{G}_{n,m}$ contains a copy of $H = H^{(d)}$ w.h.p.

We give a condition that can be proved in quite an elegant and easy way. This proof is from Alon and Füredi [31].

Theorem 6.18. *Let H be fixed sequence of graphs with $n = |V(H)| \to \infty$ and* *maximum degree Δ, where $(\Delta^2+1)^2 < n$. If*

$$p^\Delta > \frac{10\log\lfloor n/(\Delta^2+1)\rfloor}{\lfloor n/(\Delta^2+1)\rfloor}, \tag{6.14}$$

then $\mathbb{G}_{n,p}$ contains an isomorphic copy of H w.h.p.

Proof. To prove this we first apply the Hajnal-Szemerédi Theorem to the square H^2 of our graph H.

Recall that we square a graph if we add an edge between any two vertices of our original graph which are at distance at most two. The Hajnal-Szemerédi Theorem states that every graph with n vertices and maximum vertex degree at most d is $d+1$-colorable with all color classes of size $\lfloor n/(d+1)\rfloor$ or $\lceil n/(d+1)\rceil$, i.e, the $(d+1)$-coloring is equitable.

Since the maximum degree of H^2 is at most Δ^2, there exists an equitable Δ^2+1-coloring of H^2 which induces a partition of the vertex set of H, say $U = U(H)$, into Δ^2+1 pairwise disjoint subsets $U_1, U_2, \ldots, U_{\Delta^2+1}$, so that each U_k is an independent set in H^2 and the cardinality of each subset is either $\lfloor n/(\Delta^2+1)\rfloor$ or $\lceil n/(\Delta^2+1)\rceil$.

Next, partition the set V of vertices of the random graph $\mathbb{G}_{n,p}$ into pairwise disjoint sets $V_1, V_2, \ldots, V_{\Delta^2+1}$, so that $|U_k| = |V_k|$ for $k = 1, 2, \ldots, \Delta^2+1$.

We define a one-to-one function $f : U \mapsto V$, which maps each U_k onto V_k resulting in a mapping of H into an isomorphic copy of H in $\mathbb{G}_{n,p}$. In the first step, choose an arbitrary mapping of U_1 onto V_1. Now U_1 is an independent subset of H and so $\mathbb{G}_{n,p}[V_1]$ trivially contains a copy of $H[U_1]$. Assume, by induction, that we have already defined

$$f : U_1 \cup U_2 \cup \ldots \cup U_k \mapsto V_1 \cup V_2 \cup \ldots \cup V_k,$$

and that f maps the induced subgraph of H on $U_1 \cup U_2 \cup \ldots \cup U_k$ into a copy of it in $V_1 \cup V_2 \cup \ldots \cup V_k$. Now, define f on U_{k+1}, using the following construction. Suppose first that $U_{k+1} = \{u_1, u_2, \ldots, u_m\}$ and $V_{k+1} = \{v_1, v_2, \ldots, v_m\}$ where $m \in \{\lfloor n/(\Delta^2+1)\rfloor, \lceil n/(\Delta^2+1)\rceil\}$.

Next, construct a random bipartite graph $G^{(k)}_{m,m,p^*}$ with a vertex set $V = (X,Y)$, where $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$ and connect x_i and y_j with an edge if and only if in $\mathbb{G}_{n,p}$ the vertex v_j is joined by an edge to all vertices $f(u)$, where u is a neighbor of u_i in H which belongs to $U_1 \cup U_2 \cup \ldots \cup U_k$. Hence, we join x_i with y_j if and only if we can define $f(u_i) = v_j$.

Note that for each i and j, the edge probability $p^* \geq p^\Delta$ and that edges of $G^{(k)}_{m,m,p^*}$ are independent of each other, since they depend on pairwise disjoint sets of edges of $\mathbb{G}_{n,p}$. This follows from the fact that U_{k+1} is independent in H^2. Assuming that the condition (6.14) holds and that $(\Delta^2+1)^2 < n$, then by Theorem 6.1, the random graph $G^{(k)}_{m,m,p^*}$ has a perfect matching w.h.p. Moreover, we can conclude that the probability that there is no perfect matching in $G^{(k)}_{m,m,p^*}$ is at most $\frac{1}{(\Delta^2+1)n}$. It is here that we have used the extra factor 10 in the RHS of (6.14). We use a perfect matching in $G^{(k)}(m,m,p^*)$ to define f, assuming that if x_i and y_j are matched then $f(u_i) = v_j$. To define our mapping $f : U \mapsto V$ we have to find perfect matchings in all $G^{(k)}(m,m,p^*), k = 1,2,\ldots,\Delta^2+1$. The probability that we can succeed in this is at least $1-1/n$. This implies that $\mathbb{G}_{n,p}$ contains an isomorphic copy of H w.h.p. $\square$

Corollary 6.19. *Let $n = 2^d$ and suppose that $d \to \infty$ and $p \geq \frac{1}{2} + o_d(1)$, where $o_d(1)$ is a function that tends to zero as $d \to \infty$. Then w.h.p. $\mathbb{G}_{n,p}$ contains a copy of a d-dimensional cube Q_d.*

Corollary 6.20. *Let $n = d^2$ and $p \gg \left(\frac{\log n}{n}\right)^{1/4}$, where $\omega(n), d \to \infty$. Then w.h.p. $\mathbb{G}_{n,p}$ contains a copy of the 2-dimensional lattice L_d.*

6.7 Exercises

6.7.1 Consider the bipartite graph process $\Gamma_m, m = 0,1,2,\ldots,n^2$ where we add the n^2 edges in $A \times B$ in random order, one by one. Show that w.h.p. the hitting time for Γ_m to have a perfect matching is identical with the hitting time for minimum degree at least one.

6.7.2 Show that

$$\lim_{\substack{n\to\infty\\ n \text{ odd}}} \mathbb{P}(\mathbb{G}_{n,p} \text{ has a near perfect matching}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

A *near pefect matching* is one of size $\lfloor n/2 \rfloor$.

6.7.3 Show that if $p = \frac{\log n + (k-1)\log\log n + \omega}{n}$ where $k = O(1)$ and $\omega \to \infty$ then w.h.p. $G_{n,n,p}$ contains a k-regular spanning subgraph.

6.7.4 Consider the random bipartite graph G with bi-partition A, B where $|A| = |B| = n$. Each vertex $a \in A$ independently chooses $\lceil 2\log n \rceil$ random neighbors in B. Show that w.h.p. G contains a perfect matching.

6.7.5 Show that if $p = \frac{\log n + (k-1)\log\log n + \omega}{n}$ where $k = O(1)$ and $\omega \to \infty$ then w.h.p. $G_{n,p}$ contains $\lfloor k/2 \rfloor$ edge disjoint Hamilton cycles. If k is odd, show that in addition there is an edge disjoint matching of size $\lfloor n/2 \rfloor$. (Hint: Use Lemma 6.4 to argue that after "peeling off" a few Hamilton cycles, we can still use the arguments of Sections 6.1, 6.2).

6.7.6 Let m_k^* denote the first time that G_m has minimum degree at least k. Show that w.h.p. in the graph process (i) $G_{m_1^*}$ contains a perfect matching and (ii) $G_{m_2^*}$ contains a Hamilton cycle.

6.7.7 Show that if $p = \frac{\log n + \log\log n + \omega}{n}$ where $\omega \to \infty$ then w.h.p.$G_{n,n,p}$ contains a Hamilton cycle. (Hint: Start with a 2-regular spanning subgraph from (ii). Delete an edge from a cycle. Argue that rotations will always produce paths beginning and ending at different sides of the partition. Proceed more or less as in Section 6.2).

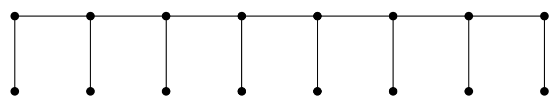
6.7.8 Show that if $p = \frac{\log n + \log\log n + \omega}{n}$ where n is even and $\omega \to \infty$ then w.h.p. $G_{n,p}$ contains a pair of vertex disjoint $n/2$-cycles. (Hint: Randomly partition $[n]$ into two sets of size $n/2$. Then move some vertices between parts to make the minimum degree at least two in both parts).

6.7.9 Show that if three divides n and $np^2 \gg \log n$ then w.h.p. $G_{n,p}$ contains $n/3$ vertex disjoint triangles. (Hint: Randomly partition $[n]$ into three sets A, B, C of size $n/3$. Choose a perfect matching M between A and B and then match C into M).

6.7.10 Let $G = (X, Y, E)$ be an arbitrary bipartite graph where the bi-partition X, Y satisfies $|X| = |Y| = n$. Suppose that G has minimum degree at least $3n/4$. Let $p = \frac{K\log n}{n}$ where K is a large constant. Show that w.h.p. G_p contains a perfect matching.

6.7.11 Let $p = (1+\varepsilon)\frac{\log n}{n}$ for some fixed $\varepsilon > 0$. Prove that w.h.p. $G_{n,p}$ is Hamilton connected i.e. every pair of vertices are the endpoints of a Hamilton path.

6.7.12 Show that if $p = \frac{(1+\varepsilon)\log n}{n}$ for $\varepsilon > 0$ constant, then w.h.p. $G_{n,p}$ contains a copy of a caterpillar on n vertices. The diagram below is the case $n = 16$.



6.7.13 Show that for any fixed $\varepsilon > 0$ there exists c_ε such that if $c \geq c_\varepsilon$ then $\mathbb{G}_{n,p}$ contains a cycle of length $(1-\varepsilon)n$ with probability $1-e^{-c\varepsilon^2 n/10}$.

6.7.14 Let $p = (1+\varepsilon)\frac{\log n}{n}$ for some fixed $\varepsilon > 0$. Prove that w.h.p. $G_{n,p}$ is pancyclic i.e. it contains a cycle of length k for every $3 \leq k \leq n$.
(See Cooper and Frieze [248] and Cooper [242], [244]).

6.7.15 Show that if p is constant then

$$\mathbb{P}(\mathbb{G}_{n,p} \text{ is not Hamiltonian}) = O(e^{-\Omega(np)}).$$

6.7.16 Let T be a tree on n vertices and maximum degree less than $c_1 \log n$. Suppose that T has at least $c_2 n$ leaves. Show that there exists $K = K(c_1, c_2)$ such that if $p \geq \frac{K \log n}{n}$ then $G_{n,p}$ contains a copy of T w.h.p.

6.7.17 Let $p = \frac{1000}{n}$ and $G = G_{n,p}$. Show that w.h.p. any red-blue coloring of the edges of G contains a mono-chromatic path of length $\frac{n}{1000}$. (Hint: Apply the argument of Section 6.3 to both the red and blue sub-graphs of G to show that if there is no long monochromatic path then there is a pair of large sets S, T such that no edge joins S, T.)
This question is taken from Dudek and Pralat [312]

6.7.18 Suppose that $p = n^{-\alpha}$ for some constant $\alpha > 0$. Show that if $\alpha > \frac{1}{3}$ then w.h.p. $\mathbb{G}_{n,p}$ does not contain a maximal spanning planar subgraph i.e. a planar subgraph with $3n-6$ edges. Show that if $\alpha < \frac{1}{3}$ then it contains one w.h.p. (see Bollobás and Frieze [167]).

6.7.19 Show that the hitting time for the existence of k edge-disjoint spanning trees coincides w.h.p. with the hitting time for minimum degree k, for $k = O(1)$.
(See Palmer and Spencer [724]).

6.7.20 Let $p = \frac{c}{n}$ where $c > 1$ is constant. Consider the greedy algorithm for constructing a large independent set I: choose a random vertex v and put v into I. Then delete v and all of its neighbors. Repeat until there are no vertices left. Use the differential equation method (see Section 6.4) and show that w.h.p. this algorithm chooses an independent set of size at least $\frac{\log c}{c} n$.

6.7.21 Consider the modified greedy matching algorithm where you first choose a random vertex x and then choose a random edge $\{x, y\}$ incident with x. Show that applied to $\mathbb{G}_{n,m}$, with $m = cn$, that w.h.p. it produces a matching of size $\left(\frac{1}{2} + o(1) - \frac{\log(2-e^{-2c})}{4c}\right) n$.

6.7.22 Let $X_1, X_2, \ldots, N = \binom{n}{2}$ be a sequence of independent Bernouilli random variables with common probability p. Let $\varepsilon > 0$ be sufficiently small. (See [604]).

(a) Let $p = \frac{1-\varepsilon}{n}$ and let $k = \frac{7\log n}{\varepsilon^2}$. Show that w.h.p. there is no interval I of length kn in $[N]$ in which at least k of the variables take the value 1.

(b) Let $p = \frac{1+\varepsilon}{n}$ and let $N_0 = \frac{\varepsilon n^2}{2}$. Show that w.h.p.

$$\left| \sum_{i=1}^{N_0} X_i - \frac{\varepsilon(1+\varepsilon)n}{2} \right| \leq n^{2/3}.$$

6.7.23 Use the result of Exercise 6.7.21(a) to show that if $p = \frac{1-\varepsilon}{n}$ then w.h.p. the maximum component size in $\mathbb{G}_{n,p}$ is at most $\frac{7\log n}{\varepsilon^2}$.

6.7.24 Use the result of Exercise 6.7.21(b) to show that if $p = \frac{1+\varepsilon}{n}$ then w.h.p $\mathbb{G}_{n,p}$ contains a path of length at least $\frac{\varepsilon^2 n}{5}$.

6.8 Notes

Hamilton cycles

Multiple Hamilton cycles

There are several results pertaining to the number of distinct Hamilton cycles in $G_{n,m}$. Cooper and Frieze [247] showed that in the graph process $G_{m_2^*}$ contains $(\log n)^{n-o(n)}$ distinct Hamilton cycles w.h.p. This number was improved by Glebov and Krivelevich [435] to $n!p^n e^{o(n)}$ for $G_{n,p}$ and $\left(\frac{\log n}{e}\right)^n e^{o(n)}$ at time m_2^*. McDiarmid [664] showed that for Hamilton cycles, perfect matchings, spanning trees the *expected* number was much higher. This comes from the fact that although there is a small probability that m_2^* is of order n^2, most of the expectation comes from here. (m_k^* is defined in Exercise 6.7.5).

Bollobás and Frieze [166] (see Exercise 6.7.4) showed that in the graph process, $G_{m_k^*}$ contains $\lfloor k/2 \rfloor$ edge disjoint Hamilton cycles plus another edge disjoint matching of size $\lfloor n/2 \rfloor$ if n is odd. We call this property $\mathscr{A}_k$. This was the case $k = O(1)$. The more difficult case of the occurrence of $\mathscr{A}_k$ at m_k^*, where $k \to \infty$ was verified in two papers, Krivelevich and Samotij [601] and Knox, Kühn and Osthus [574].

Conditioning on minimum degree

Suppose that instead of taking enough edges to make the minimum degree in $\mathbb{G}_{n,m}$ two very likely, we instead condition on having minimum degree at least two. Let $G_{n,m}^{\delta\geq k}$ denote $G_{n,m}$ conditioned on having minimum degree at least $k = O(1)$. Bollobás, Fenner and Frieze [164] proved that if

$$m = \frac{n}{2}\left(\frac{\log n}{k+1} + k\log\log n + \omega(n)\right)$$

then $G_{n,m}^{\delta\geq k}$ has $\mathscr{A}_k$ w.h.p.

Bollobás, Cooper, Fenner and Frieze [161] prove that w.h.p. $G_{n,cn}^{\delta\geq k}$ has property $\mathscr{A}_{k-1}$ w.h.p. provided $3 \leq k = O(1)$ and $c \geq (k+1)^3$. For $k = 3$, Frieze [391] showed that $G_{n,cn}^{\delta\geq 3}$ is Hamiltonian w.h.p. for $c \geq 10$.

The k-core of a random graphs is distributed like $G_{\nu,\mu}^{\delta\geq k}$ for some (random) ν,μ. Krivelevich, Lubetzky and Sudakov [598] prove that when a k-core first appears, $k \geq 15$, w.h.p. it has $\lfloor (k-3)/2 \rfloor$ edge disjoint Hamilton cycles.

Algorithms for finding Hamilton cycles

Gurevich and Shelah [457] and Thomason [831] gave linear expected time algorithms for finding a Hamilton cycle in a sufficiently dense random graph i.e. $G_{n,m}$ with $m \gg n^{5/3}$ in the Thomason paper. Bollobás, Fenner and Frieze [163] gave an $O(n^{3+o(1)})$ time algorithm that w.h.p. finds a Hamilton cycle in the graph $G_{m_2^*}$. Frieze and Haber [393] gave an $O(n^{1+o(1)})$ time algorithm for finding a Hamilton cycle in $G_{n,cn}^{\delta\geq 3}$ for c sufficiently large.

Long cycles

A sequence of improvements, Bollobás [150]; Bollobás, Fenner and Frieze [165] to Theorem 6.8 in the sense of replacing $O(\log c/c)$ by something smaller led finally to Frieze [384]. He showed that w.h.p. there is a cycle of length $n(1 - ce^{-c}(1+\varepsilon_c))$ where $\varepsilon_c \to 0$ with c. Up to the value of ε_c this is best possible.

Glebov, Naves and Sudakov [436] prove the following generalisation of (part of) Theorem 6.5. They prove that if a graph G has minimum degree at least k and $p \geq \frac{\log k + \log\log k + \omega_k(1)}{k}$ then w.h.p. G_p has a cycle of length at least $k+1$.

Spanning Subgraphs

Riordan [760] used a second moment calculation to prove the existence of a certain (sequence of) spanning subgraphs $H = H^{(i)}$ in $\mathbb{G}_{n,p}$. Suppose that we denote the number of vertices in a graph H by $|H|$ and the number of edges by $e(H)$. Suppose that $|H| = n$. For $k \in [n]$ we let $e_H(k) = \max\{e(F) : F \subseteq H, |F| = k\}$ and $\gamma = \max_{3 \le k \le n} \frac{e_H(k)}{k-2}$. Riordan proved that if the following conditions hold, then $\mathbb{G}_{n,p}$ contains a copy of H w.h.p.: (i) $e(H) \ge n$, (ii) $Np, (1-p)n^{1/2} \to \infty$, (iii) $np^{\gamma}/\Delta(H)^4 \to \infty$.

This for example replaces the $\frac{1}{2}$ in Corollary 6.19 by $\frac{1}{4}$.

Spanning trees

Gao, Pérez-Giménez and Sato [424] considered the existence of k edge disjoint spanning trees in $\mathbb{G}_{n,p}$. Using a characterisation of Nash-Williams [708] they were able to show that w.h.p. one can find $\min\left\{\delta, \frac{m}{n-1}\right\}$ edge disjoint spanning trees. Here δ denotes the minimum degree and m denotes the number of edges.

When it comes to spanning trees of a fixed structure, Kahn conjectured that the threshold for the existence of any fixed bounded degree tree T, in terms of number of edges, is $O(n \log n)$. For example, a *comb* consists of a path P of length $n^{1/2}$ with each $v \in P$ being one endpoint of a path P_v of the same length. The paths P_v, P_w being vertex disjoint for $v \neq w$. Hefetz, Krivelevich and Szabó [477] proved this for a restricted class of trees i.e. those with a linear number of leaves or with an induced path of length $\Omega(n)$. Kahn, Lubetzky and Wormald [534], [535] verified the conjecture for combs. Montgomery [687], [688] sharpened the result for combs, replacing $m = Cn \log n$ by $m = (1+\varepsilon)n \log n$ and proved that any tree can be found w.h.p. when $m = O(\Delta n (\log n)^5)$, where Δ is the maximum degree of T. More recently, Montgomery [690] improved the upper bound on m to the optimal, $m = O(\Delta n (\log n))$.

Large Matchings

Karp and Sipser [558] analysed a greedy algorithm for finding a large matching in the random graph $\mathbb{G}_{n,p}, p = c/n$ where $c > 0$ is a constant. It has a much better performance than the algorithm described in Section 6.4. It follows from their work that if $\mu(G)$ denotes the size of the largest matching in G then w.h.p.

$$\frac{\mu(\mathbb{G}_{n,p})}{n} \approx 1 - \frac{\gamma^* + \gamma_* + \gamma^* \gamma_*}{2c}$$

where γ_* is the smallest root of $x = c \exp\{-ce^{-x}\}$ and $\gamma^* = ce^{-\gamma_*}$.

Later, Aronson, Frieze and Pittel [52] tightened their analysis. This led to the consideration of the size of the largest matching in $G^{\delta \ge 2}_{n,m=cn}$. Frieze and Pittel [413] showed that w.h.p. this graph contains a matching of size $n/2 - Z$ where Z is a random variable with bounded expectation. Frieze [389] proved that in the bipartite analogue of this problem, a perfect matching exists w.h.p. Building on this work, Chebolu, Frieze and Melsted [215] showed how to find an exact maximum sized matching in $\mathbb{G}_{n,m}, m = cn$ in $O(n)$ expected time.

H-factors

By an H-factor of a graph G, we mean a collection of vertex disjoint copies of a fixed graph H that together cover all the vertices of G. Some early results on the existence of H-factors in random graphs are given in Alon and Yuster [38] and Ruciński [778]. For the case of when H is a tree, Łuczak and Ruciński [644] found the precise threshold. For general H, there is a recent breakthrough paper of Johansson, Kahn and Vu [529] that gives the threshold for strictly balanced H and good estimates in general. See Gerke and McDowell [423] for some further results.

Chapter 7

Extreme Characteristics

This chapter is devoted to the extremes of certain graph parameters. We look first at the diameter of random graphs i.e. the extreme value of the shortest distance between a pair of vertices. Then we look at the size of the largest independent set and the the related value of the chromatic number. We decribe an important recent result on "interpolation" that proves certain limits exist. We end the chapter with the likely values of the first and second eigenvalues of a random graph.

7.1 Diameter

In this section we will first discuss the threshold for $\mathbb{G}_{n,p}$ to have diameter d, when $d \geq 2$ is a constant. The diameter of a connected graph G is the maximum over distinct vertices v,w of $dist(v,w)$ where $dist(v,w)$ is the minimum number of edges in a path from v to w. The theorem below was proved independently by Burtin [203], [204] and by Bollobás [148]. The proof we give is due to Spencer [810].

Theorem 7.1. *Let $d \geq 2$ be a fixed positive integer. Suppose that $c > 0$ and*

$$p^d n^{d-1} = \log(n^2/c).$$

Then

$$\lim_{n\to\infty} \mathbb{P}(diam(\mathbb{G}_{n,p}) = k) = \begin{cases} e^{-c/2} & \text{if } k = d \\ 1 - e^{-c/2} & \text{if } k = d+1. \end{cases}$$

Proof. (a): w.h.p. diam$(G) \geq d$.

Fix $v \in V$ and let

$$N_k(v) = \{w : dist(v,w) = k\}. \tag{7.1}$$

It follows from Theorem 3.4 that w.h.p. for $0 \leq k < d$,

$$|N_k(v)| \leq \Delta^k \approx (np)^k \approx (n \log n)^{k/d} = o(n). \tag{7.2}$$

(b) w.h.p. $\text{diam}(G) \leq d+1$

Fix $v, w \in [n]$. Then for $1 \leq k < d$, define the event

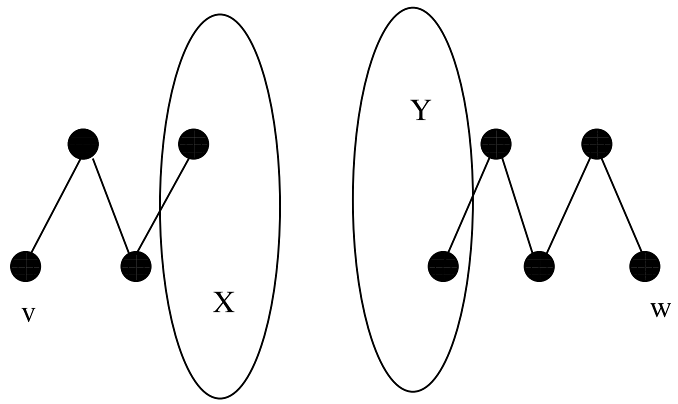
$$\mathscr{F}_k = \left\{ |N_k(v)| \in I_k = \left[\left(\frac{np}{2}\right)^k, (2np)^k \right] \right\}.$$

Then for $k \leq \lceil d/2 \rceil$ we have

$$\begin{aligned}
&\mathbb{P}(\neg \mathscr{F}_k \,|\mathscr{F}_1, \ldots, \mathscr{F}_{k-1}) = \\
&= \mathbb{P}\left(\text{Bin}\left(n - \sum_{i=0}^{k-1} |N_i(v)|, 1-(1-p)^{|N_{k-1}(v)|} \right) \notin I_k \right) \\
&\leq \mathbb{P}\left(\text{Bin}\left(n - o(n), \frac{3}{4}\left(\frac{np}{2}\right)^{k-1} p \right) \leq \left(\frac{np}{2}\right)^k \right) \\
&\qquad\qquad\qquad + \mathbb{P}\left(\text{Bin}\left(n - o(n), \frac{5}{4}(2np)^{k-1} p \right) \geq (2np)^k \right) \\
&\leq \exp\left\{ -\Omega\left((np)^k \right) \right\} \\
&= O(n^{-3}).
\end{aligned}$$

So with probability $1 - O(n^{-3})$,

$$|N_{\lfloor d/2 \rfloor}(v)| \geq \left(\frac{np}{2}\right)^{\lfloor d/2 \rfloor} \text{ and } |N_{\lceil d/2 \rceil}(w)| \geq \left(\frac{np}{2}\right)^{\lceil d/2 \rceil}.$$



If $X = N_{\lfloor d/2 \rfloor}(v)$ and $Y = N_{\lceil d/2 \rceil}(w)$ then, either

$$X \cap Y \neq \emptyset \text{ and } dist(v,w) \leq \lfloor d/2 \rfloor + \lceil d/2 \rceil = d,$$

or since the edges between X and Y are unconditioned by our construction,

$$\begin{aligned}\mathbb{P}(\nexists \text{ an } X:Y \text{ edge }) &\leq (1-p)^{\left(\frac{np}{2}\right)^d} \leq \exp\left\{-\left(\frac{np}{2}\right)^d p\right\} \\ &\leq \exp\{-(2-o(1))np\log n\} = o(n^{-3}).\end{aligned}$$

So

$$\mathbb{P}(\exists v,w : dist(v,w) > d+1) = o(n^{-1}).$$

We now consider the probability that d or $d+1$ is the diameter. We will use Janson's inequality, see Section 27.6. More precisely, we will use the earlier inequality, Corollary 27.14, from Janson, Łuczak and Ruciński [508].
We will first use this to estimate the probability of the following event: Let $v \neq w \in [n]$ and let

$$\mathscr{A}_{v,w} = \{v,w \text{ are not joined by a path of length } d\}.$$

For $\mathbf{x} = x_1, x_2, \ldots, x_{d-1}$ let

$$\mathscr{B}_{v,\mathbf{x},w} = \{(v, x_1, x_2, \ldots, x_{d-1}, w) \text{ is a path in } \mathbb{G}_{n,p}\}.$$

Let

$$Z = \sum_{\mathbf{x}} Z_{\mathbf{x}},$$

where

$$Z_{\mathbf{x}} = \begin{cases} 1 & \text{if } \mathscr{B}_{v,\mathbf{x},w} \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Janson's inequality allows us to estimate the probability that $Z = 0$, which is precisely the probability of $\mathscr{A}_{v,w}$.

Now

$$\mu = \mathbb{E}Z = (n-2)(n-3)\cdots(n-d)p^d = \log\left(\frac{n^2}{c}\right)\left(1 + O\left(\frac{1}{n}\right)\right).$$

Let $\mathbf{x} = x_1, x_2, \cdots, x_{d-1}$, $\mathbf{y} = y_1, y_2, \ldots, y_{d-1}$ and

$$\Delta = \sum_{\substack{\mathbf{x},\mathbf{y}:\mathbf{x}\neq\mathbf{y} \\ v,\mathbf{x},w \text{ and } v,\mathbf{y},w \\ \text{share an edge}}} \mathbb{P}(\mathscr{B}_{\mathbf{x}} \cap \mathscr{B}_{\mathbf{y}})$$

$$\leq \sum_{t=1}^{d-1} \binom{d}{t} n^{2(d-1)-t} p^{2d-t}, \qquad t \text{ is the number of shared edges,}$$

$$= O\left(\sum_{t=1}^{d-1} n^{2(d-1)-t-\frac{d-1}{d}(2d-t)} (\log n)^{\frac{2d-t}{d}}\right)$$

$$= O\left(\sum_{t=1}^{d-1} n^{-t/d+o(1)}\right)$$

$$= o(1).$$

Applying Corollary 27.14, $\mathbb{P}(Z=0) \leq e^{-\mu+\Delta}$, we get

$$\mathbb{P}(Z=0) \leq \frac{c+o(1)}{n^2}.$$

On the other hand the FKG inequality (see Section 27.3) implies that

$$\mathbb{P}(Z=0) \geq \left(1-p^d\right)^{(n-2)(n-3)\cdots(n-d)} = \frac{c+o(1)}{n^2}.$$

So

$$\mathbb{P}(\mathscr{A}_{v,w}) = \mathbb{P}(Z=0) = \frac{c+o(1)}{n^2}.$$

So

$$\mathbb{E}(\#v,w : \mathscr{A}_{v,w} \text{ occurs}) = \frac{c+o(1)}{2}$$

and we should expect that

$$\mathbb{P}(\nexists\, v,w : \mathscr{A}_{v,w} \text{ occurs}) \approx e^{-c/2}. \tag{7.3}$$

Indeed if we choose $v_1, w_1, v_2, w_2, \ldots, v_k, w_k$, k constant, we will find that

$$\mathbb{P}\left(\mathscr{A}_{v_1,w_1}, \mathscr{A}_{v_2,w_2}, \ldots, \mathscr{A}_{v_k,w_k}\right) \approx \left(\frac{c}{n^2}\right)^k \tag{7.4}$$

and (7.3) follows from the method of moments.

The proof of (7.3) is just a more involved version of the proof of the special case $k=1$ that we have just completed. We now let

$$\mathscr{B}_{\mathbf{x}} = \bigcup_{i=1}^{k} \mathscr{B}_{v_i,\mathbf{x},w_i}$$

and re-define

$$Z = \sum_{\mathbf{x}} Z_{\mathbf{x}},$$

where now

$$Z_{\mathbf{x}} = \begin{cases} 1 & \text{if } \mathscr{B}_{\mathbf{x}} \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Then we have $\{Z = 0\}$ is equivalent to $\bigcap_{i=1}^{k} \mathscr{A}_{v_i,w_i}$.

Now,

$$\mathbb{E} Z = k(n-2)(n-3)\cdots(n-d)p^d = k \log\left(\frac{n^2}{c}\right)\left(1 + O\left(\frac{1}{n}\right)\right).$$

We need to show that the corresponding $\Delta = o(1)$. But,

$$\begin{aligned} \Delta &\leq \sum_{r,s} \sum_{\substack{\mathbf{x},\mathbf{y}:\mathbf{x}\neq\mathbf{y} \\ v_r,\mathbf{x},w_r \text{ and } v_s,\mathbf{y},w_s \\ \text{share an edge}}} \mathbb{P}(\mathscr{B}_{v_r,\mathbf{x},w_r} \cap \mathscr{B}_{v_s,\mathbf{y},w_s}) \\ &\leq k^2 \sum_{t=1}^{d-1} \binom{d}{t} n^{2(d-1)-t} p^{2d-t} \\ &= o(1). \end{aligned}$$

This shows that

$$\mathbb{P}(Z = 0) \leq e^{-k\log(n^2/c + o(1)} = \left(\frac{c + o(1)}{n^2}\right)^k.$$

On the other hand, the FKG inequality (see Section 27.3) shows that

$$\mathbb{P}\left(\mathscr{A}_{v_1,w_1}, \mathscr{A}_{v_2,w_2}, \ldots, \mathscr{A}_{v_k,w_k}\right) \geq \prod_{i=1}^{k} \mathbb{P}\left(\mathscr{A}_{v_i,w_i}\right).$$

This verifies (7.4) and completes the proof of Theorem 7.1. □

We turn next to a sparser case and prove a somewhat weaker result.

Theorem 7.2. *Suppose that $p = \frac{\omega \log n}{n}$ where $\omega \to \infty$. Then*

$$diam(\mathbb{G}_{n,p}) \approx \frac{\log n}{\log np} \qquad w.h.p.$$

Proof. Fix $v \in [n]$ and let $N_i = N_i(v)$ be as in (7.1). Let $N_{\leq k} = \bigcup_{i \leq k} N_i$. Using the proof of Theorem 3.4(b) we see that we can assume that

$$(1 - \omega^{-1/3})np \leq \deg(x) \leq (1 + \omega^{-1/3})np \qquad \text{for all } x \in [n]. \tag{7.5}$$

It follows that if $\gamma = \omega^{-1/3}$ and

$$k_0 = \frac{\log n - \log 3}{\log np + \gamma} \approx \frac{\log n}{\log np}$$

then w.h.p.

$$|N_{\leq k_0}| \leq \sum_{k \leq k_0} ((1+\gamma)np)^k \leq 2((1+\gamma)np)^{k_0} = \frac{2n}{3+o(1)}$$

and so the diameter of $\mathbb{G}_{n,p}$ is at least $(1-o(1))\frac{\log n}{\log np}$.

We can assume that $np = n^{o(1)}$ as larger p are dealt with in Theorem 7.1. Now fix $v, w \in [n]$ and let N_i be as in the previous paragraph. Now consider a Breadth First Search (BFS) that constructs $N_1, N_2, \dots, N_{k_1}$ where

$$k_1 = \frac{3\log n}{5\log np}.$$

It follows that if (7.5) holds then for $k \leq k_1$ we have

$$|N_{i \leq k}| \leq n^{3/4} \text{ and } |N_k|p \leq n^{-1/5}. \tag{7.6}$$

Observe now that the edges from N_i to $[n] \setminus N_{\leq i}$ are unconditioned by the BFS up to layer k and so for $x \in [n] \setminus N_{\leq k}$,

$$\mathbb{P}(x \in N_{k+1} \mid N_{\leq k}) = 1 - (1-p)^{|N_k|} \geq |N_k|p(1 - |N_k|p) \geq \rho_k = |N_k|p(1 - n^{-1/5}).$$

The events $x \in N_{k+1}$ are independent and so $|N_{k+1}|$ stochastically dominates the binomial $\mathrm{Bin}(n - n^{3/4}, \rho_k)$. Assume inductively that $|N_k| \geq (1-\gamma)^k (np)^k$ for some $k \geq 1$. This is true w.h.p. for $k = 1$ by (7.5). Let $\mathscr{A}_k$ be the event that (7.6) holds. It follows that

$$\mathbb{E}(|N_{k+1}| \mid \mathscr{A}_k) \geq np|N_k|(1 - O(n^{-1/5})).$$

It then follows from the Chernoff bounds (Theorem 27.6) that

$$\mathbb{P}(|N_{k+1}| \leq ((1-\gamma)np)^{k+1} \leq \exp\left\{-\frac{\gamma^2}{4}|N_k|np\right\} = o(n^{-anyconstant}).$$

There is a small point to be made about conditioning here. We can condition on (7.5) holding and then argue that this only multiplies small probabilities by $1 + o(1)$ if we use $\mathbb{P}(A \mid B) \leq \mathbb{P}(A)/\mathbb{P}(B)$.

It follows that if

$$k_2 = \frac{\log n}{2(\log np + \log(1-\gamma)} \approx \frac{\log n}{2\log np}$$

then w.h.p. we have

$$|N_{k_2}| \geq n^{1/2}.$$

Analogously, if we do BFS from w to create $N'_k, i = 1,2,\dots,k_2$ then $|N'_{k_2}| \geq n^{1/2}$. If $N_{\leq k_2} \cap N'_{\leq k_2} \neq \emptyset$ then $dist(v,w) \leq 2k_2$ and we are done. Otherwise, we observe that the edges $N_{k_2} : N'_{k_2}$ between N_{k_2} and N'_{k_2} are unconditioned (except for (7.5)) and so

$$\mathbb{P}(N_{k_2} : N'_{k_2} = \emptyset) \leq (1-p)^{n^{1/2} \times n^{1/2}} \leq n^{-\omega}.$$

If $N_{k_2} : N'_{k_2} \neq \emptyset$ then $dist(v,w) \leq 2k_2 + 1$ and we are done. Note that given (7.5), all other unlikely events have probability $O(n^{-anyconstant})$ of occurring and so we can inflate these latter probabilities by n^2 to account for all choices of v,w. This completes the proof of Theorem 7.2. □

7.2 Largest Independent Sets

Let $\alpha(G)$ denote the size of the largest independent set in a graph G.

Dense case

The following theorem was first proved by Matula [655].

Theorem 7.3. *Suppose* $0 < p < 1$ *is a constant and* $b = \frac{1}{1-p}$. *Then w.h.p.*

$$\alpha(\mathbb{G}_{n,p}) \approx 2\log_b n.$$

Proof. Let X_k be the number of independent sets of order k.
(i) Let

$$k = \lceil 2\log_b n \rceil$$

Then,

$$\begin{aligned}
\mathbb{E}X_k &= \binom{n}{k}(1-p)^{\binom{k}{2}} \\
&\leq \left(\frac{ne}{k(1-p)^{1/2}}(1-p)^{k/2}\right)^k \\
&\leq \left(\frac{e}{k(1-p)^{1/2}}\right)^k \\
&= o(1).
\end{aligned}$$

(ii) Let now

$$k = \lfloor 2\log_b n - 5\log_b \log n \rfloor.$$

Let
$$\overline{\Delta} = \sum_{\substack{i,j \\ S_i \sim S_j}} \mathbb{P}(S_i, S_j \text{ are independent in } \mathbb{G}_{n,p}),$$
where $S_1, S_2, \dots, S_{\binom{n}{k}}$ are all the k-subsets of $[n]$ and $S_i \sim S_j$ iff $|S_i \cap S_j| \geq 2$. By Janson's inequality, see Theorem 27.13,
$$\mathbb{P}(X_k = 0) \leq \exp\left\{ -\frac{(\mathbb{E}X_k)^2}{2\overline{\Delta}} \right\}.$$
Here we apply the inequality in the context of X_k being the number of k-cliques in the complement of $G_{n,p}$. The set $[N]$ will be the edges of the complete graph and the sets D_i will the edges of the k-cliques. Now
$$\begin{aligned}
\frac{\overline{\Delta}}{(\mathbb{E}X_k)^2} &= \frac{\binom{n}{k}(1-p)^{\binom{k}{2}} \sum_{j=2}^{k} \binom{n-k}{k-j}\binom{k}{j}(1-p)^{\binom{k}{2}-\binom{j}{2}}}{\left(\binom{n}{k}(1-p)^{\binom{k}{2}}\right)^2} \\
&= \sum_{j=2}^{k} \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} (1-p)^{-\binom{j}{2}} \\
&= \sum_{j=2}^{k} u_j.
\end{aligned}$$
Notice that for $j \geq 2$,
$$\begin{aligned}
\frac{u_{j+1}}{u_j} &= \frac{k-j}{n-2k+j+1} \frac{k-j}{j+1} (1-p)^{-j} \\
&\leq \left(1 + O\left(\frac{\log_b n}{n}\right)\right) \frac{k^2(1-p)^{-j}}{n(j+1)}.
\end{aligned}$$
Therefore,
$$\begin{aligned}
\frac{u_j}{u_2} &\leq (1+o(1)) \left(\frac{k^2}{n}\right)^{j-2} \frac{2(1-p)^{-(j-2)(j+1)/2}}{j!} \\
&\leq (1+o(1)) \left(\frac{2k^2 e}{nj}(1-p)^{-\frac{j+1}{2}}\right)^{j-2} \leq 1.
\end{aligned}$$
So
$$\frac{(\mathbb{E}X_k)^2}{\overline{\Delta}} \geq \frac{1}{ku_2} \geq \frac{n^2(1-p)}{k^5}.$$

Therefore

$$\mathbb{P}(X_k = 0) \le e^{-\Omega(n^2/(\log n)^5)}. \tag{7.7}$$

$\square$

Matula used the Chebyshev inequality and so he was not able to prove an exponential bound like (7.7). This will be important when we come to discuss the chromatic number.

Sparse Case

We now consider the case where $p = d/n$ and d is a large constant. Frieze [388] proved

Theorem 7.4. *Let $\varepsilon > 0$ be a fixed constant. Then for $d \ge d(\varepsilon)$ we have that w.h.p.*

$$\left|\alpha(\mathbb{G}_{n,p})) - \frac{2n}{d}(\log d - \log\log d - \log 2 + 1)\right| \le \frac{\varepsilon n}{d}.$$

Dani and Moore [276] have given an even sharper result.

In this section we will prove that if $p = d/n$ and d is sufficiently large then w.h.p.

$$\left|\alpha(\mathbb{G}_{n,p}) - \frac{2\log d}{d} n\right| \le \frac{\varepsilon \log d}{d} n. \tag{7.8}$$

This will follow from the following. Let X_k be as defined in the previous section. Let

$$k_0 = \frac{(2-\varepsilon/8)\log d}{d} n \text{ and } k_1 = \frac{(2+\varepsilon/8)\log d}{d} n.$$

Then,

$$\mathbb{P}\left(\left|\alpha(\mathbb{G}_{n,p}) - \mathbb{E}(\alpha(\mathbb{G}_{n,p}))\right| \ge \frac{\varepsilon \log d}{8d} n\right) \le \exp\left\{-\Omega\left(\frac{(\log d)^2}{d^2}\right) n\right\}. \tag{7.9}$$

$$\mathbb{P}(\alpha(\mathbb{G}_{n,p}) \ge k_1) = \mathbb{P}(X_{k_1} > 0) \le \exp\left\{-\Omega\left(\frac{(\log d)^2}{d}\right) n\right\}. \tag{7.10}$$

$$\mathbb{P}(\alpha(\mathbb{G}_{n,p}) \ge k_0) = \mathbb{P}(X_{k_0} > 0) \ge \exp\left\{-O\left(\frac{(\log d)^{3/2}}{d^2}\right) n\right\}. \tag{7.11}$$

Let us see how (7.8) follows from these three inequalities. Indeed, (7.9) and (7.11) imply that

$$\mathbb{E}(\alpha(\mathbb{G}_{n,p})) \ge k_0 - \frac{\varepsilon \log d}{8d} n. \tag{7.12}$$

Furthermore (7.9) and (7.10) imply that

$$\mathbb{E}(\alpha(\mathbb{G}_{n,p})) \leq k_1 + \frac{\varepsilon \log d}{8d} n. \tag{7.13}$$

It follows from (7.12) and (7.13) that

$$|k_0 - \mathbb{E}(\alpha(\mathbb{G}_{n,p}))| \leq \frac{\varepsilon \log d}{2d} n.$$

We obtain (7.8) by applying (7.9) once more.

Proof of (7.9)**:** This follows directly from the Azuma-Hoeffding inequality – see Section 27.7, in particular Lemma 27.17. If $Z = \alpha(\mathbb{G}_{n,p})$ then we write $Z = Z(Y_2, Y_3, \ldots, Y_n)$ where Y_i is the set of edges between vertex i and vertices $[i-1]$ for $i \geq 2$. $Y_2, Y_3, \ldots, Y_n$ are independent and changing a single Y_i can change Z by at most one. Therefore, for any $t > 0$ we have

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq t) \leq \exp\left\{-\frac{t^2}{2n-2}\right\}.$$

Setting $t = \frac{\varepsilon \log d}{8d} n$ yields (7.9).

Proof of (7.10)**:** The first moment method gives

$$\begin{aligned}
\Pr(X_{k_1} > 0) &\leq \binom{n}{k_1}\left(1 - \frac{d}{n}\right)^{\binom{k_1}{2}} \leq \left(\frac{ne}{k_1} \cdot \left(1 - \frac{d}{n}\right)^{(k_1-1)/2}\right)^{k_1} \\
&\leq \left(\frac{de}{2\log d} \cdot d^{-(1+\varepsilon/5)}\right)^{k_1} = \exp\left\{-\Omega\left(\frac{(\log d)^2}{d}\right) n\right\}.
\end{aligned}$$

Proof of (7.11)**:** Now, after using Lemma 27.1(g),

$$\begin{aligned}
\frac{1}{\mathbb{P}(X_{k_0} > 0)} &\leq \frac{\mathbb{E}(X_{k_0}^2)}{\mathbb{E}(X_{k_0})^2} = \sum_{j=0}^{k_0} \frac{\binom{n-k_0}{k_0-j}\binom{k_0}{j}}{\binom{n}{k_0}} (1-p)^{-\binom{j}{2}} \\
&\leq \sum_{j=0}^{k_0} \left(\frac{k_0 e}{j} \cdot \exp\left\{\frac{jd}{2n} + O\left(\frac{jd^2}{n^2}\right)\right\}\right)^j \times \\
&\qquad \left(\frac{k_0}{n}\right)^j \left(\frac{n-k_0}{n-j}\right)^{k_0-j} \qquad (7.14) \\
&\leq \sum_{j=0}^{k_0} \left(\frac{k_0 e}{j} \cdot \frac{k_0}{n} \cdot \exp\left\{\frac{jd}{2n} + O\left(\frac{jd^2}{n^2}\right)\right\}\right)^j \times
\end{aligned}$$

$$
\begin{aligned}
&\exp\left\{-\frac{(k_0-j)^2}{n-j}\right\} \\
&\leq_b \sum_{j=0}^{k_0}\left(\frac{k_0 e}{j}\cdot\frac{k_0}{n}\cdot\exp\left\{\frac{jd}{2n}+\frac{2k_0}{n}\right\}\right)^j \times \exp\left\{-\frac{k_0^2}{n}\right\} \\
&= \sum_{j=0}^{k_0} v_j. \qquad (7.15)
\end{aligned}
$$

(The notation $A \leq_b B$ is shorthand for $A = O(B)$ when the latter is considered to be ugly looking).

We observe first that $(A/x)^x \leq e^{A/e}$ for $A > 0$ implies that

$$\left(\frac{k_0 e}{j}\cdot\frac{k_0}{n}\right)^j \times \exp\left\{-\frac{k_0^2}{n}\right\} \leq 1.$$

So,

$$
\begin{aligned}
j \leq j_0 = \frac{(\log d)^{3/4}}{d^{3/2}} n \Longrightarrow v_j &\leq \exp\left\{\frac{j^2 d}{2n}+\frac{2jk_0}{n}\right\} \\
&= \exp\left\{O\left(\frac{(\log d)^{3/2}}{d^2}\right) n\right\}. \qquad (7.16)
\end{aligned}
$$

Now put

$$j = \frac{\alpha \log d}{d} n \text{ where } \frac{1}{d^{1/2}(\log d)^{1/4}} < \alpha < 2 - \frac{\varepsilon}{4}.$$

Then

$$
\begin{aligned}
\frac{k_0 e}{j}\cdot\frac{k_0}{n}\cdot\exp\left\{\frac{jd}{2n}+\frac{2k_0}{n}\right\} &\leq \frac{4e\log d}{\alpha d}\cdot\exp\left\{\frac{\alpha\log d}{2}+\frac{4\log d}{d}\right\} \\
&= \frac{4e\log d}{\alpha d^{1-\alpha/2}}\exp\left\{\frac{4\log d}{d}\right\} \\
&< 1.
\end{aligned}
$$

To see this note that if $f(\alpha) = \alpha d^{1-\alpha/2}$ then f increases between $d^{-1/2}$ and $2/\log d$ after which it decreases. Then note that

$$\min\left\{f(d^{-1/2}), f(2-\varepsilon)\right\} > 4e\exp\left\{\frac{4\log d}{d}\right\}.$$

Thus $v_j < 1$ for $j \geq j_0$ and (7.11) follows from (7.16). □

7.3 Interpolation

The following theorem is taken from Bayati, Gamarnik and Tetali [82]. Note that it is not implied by Theorem 7.4. This paper proves a number of other results of a similar flavor for other parameters. It is an important paper in that it verifies some very natural conjectures about some graph parameters, that have not been susceptible to proof until now.

Theorem 7.5. *There exists a function $H(d)$ such that*

$$\lim_{n\to\infty} \frac{\mathbb{E}(\alpha(G_{n,\lfloor dn \rfloor}))}{n} = H(d).$$

Proof. For this proof we use the model $\mathbb{G}^{(A)}_{n,m}$ of Section 1.3. This is proper since we we know that w.h.p.

$$|\alpha(\mathbb{G}^{(A)}_{n,m}) - \alpha(\mathbb{G}_{n,m})| \leq ||E(\mathbb{G}^{(A)}_{n,m})| - m| \leq \log n.$$

We will prove that for every $1 \leq n_1, n_2 \leq n-1$ such that $n_1 + n_2 = n$,

$$\mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n,\lfloor dn \rfloor})) \geq \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n_1,m_1})) + \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n_2,m_2})) \tag{7.17}$$

where $m_i = \mathrm{Bin}(\lfloor dn \rfloor, n_i/n), i = 1,2$.

Assume (7.17). We have $\mathbb{E}(|m_j - \lfloor dn_j \rfloor|) = O(n^{1/2})$. This and (7.17) and the fact that adding/deleting one edge changes α by at most one implies that

$$\mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n,\lfloor dn \rfloor})) \geq \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n_1,\lfloor dn_1 \rfloor})) + \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n_2,\lfloor dn_2 \rfloor})) - O(n^{1/2}). \tag{7.18}$$

Thus the sequence $u_n = \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n,\lfloor dn \rfloor}))$ satisfies the conditions of Lemma 7.6 below and the proof of Theorem 7.5 follows.

Proof of (7.17): We begin by constructing a sequence of graphs interpolating between $\mathbb{G}^{(A)}_{n,\lfloor dn \rfloor}$ and a disjoint union of $\mathbb{G}^{(A)}_{n_1,m_1}$ and $\mathbb{G}^{(A)}_{n_2,m_2}$. Given n, n_1, n_2 such that $n_1 + n_2 = n$ and any $0 \leq r \leq m = \lfloor dn \rfloor$, let $\mathbb{G}(n,m,r)$ be the random (pseudo-)graph on vertex set $[n]$ obtained as follows. It contains precisely m edges. The first r edges $e_1, e_2, \ldots, e_r$ are selected randomly from $[n]^2$. The remaining $m - r$ edges $e_{r+1}, \ldots, e_m$ are generated as follows. For each $j = r+1, \ldots, m$, with probability n_j/n, e_j is selected randomly from $M_1 = [n_1]^2$ and with probability n_2/n, e_j is selected randomly from $M_2 = [n_1+1, n]^2$. Observe that when $r = m$ we have $\mathbb{G}(n,m,r) = \mathbb{G}^{(A)}(n,m)$ and when $r = 0$ it is the disjoint union of $\mathbb{G}^{(A)}_{n_1,m_1}$ and $\mathbb{G}^{(A)}_{n_2,m_2}$ where $m_j = \mathrm{Bin}(m, n_j/n)$ for $j = 1,2$. We will show next that

$$\mathbb{E}(\alpha(\mathbb{G}(n,m,r))) \geq \mathbb{E}(\alpha(\mathbb{G}(n,m,r-1))) \text{ for } r = 1, \ldots, m. \tag{7.19}$$

It will follow immediately that

$$\mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n,m})) = \mathbb{E}(\alpha(\mathbb{G}(n,m,m))) \geq \\ \mathbb{E}(\alpha(\mathbb{G}(n,m,0))) = \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n_1,m_1})) + \mathbb{E}(\alpha(\mathbb{G}^{(A)}_{n_2,m_2}))$$

which is (7.17).

Proof of (7.19)**:** Observe that $\mathbb{G}(n,m,r-1)$ is obtained from $\mathbb{G}(n,m,r)$ by deleting the random edge e_r and then adding an edge from M_1 or M_2. Let $\mathbb{G}_0$ be the graph obtained after deleting e_r, but before adding its replacement. Remember that

$$\mathbb{G}(n,m,r) = \mathbb{G}_0 + e_r.$$

We will show something stronger than (7.19) viz. that

$$\mathbb{E}(\alpha(\mathbb{G}(n,m,r)) \mid \mathbb{G}_0) \geq \mathbb{E}(\alpha(\mathbb{G}(n,m,r-1)) \mid \mathbb{G}_0) \text{ for } r = 1,\dots,m. \tag{7.20}$$

Now let $O^* \subseteq [n]$ be the set of vertices that belong to every largest independent set in $\mathbb{G}_0$. Then for $e_r = (x,y)$, $\alpha(\mathbb{G}_0 + e) = \alpha(\mathbb{G}_0) - 1$ if $x, y \in O^*$ and $\alpha(\mathbb{G}_0 + e) = \alpha(\mathbb{G}_0)$ if $x \notin O^*$ or $y \notin O^*$. Because e_r is randomly chosen, we have

$$\mathbb{E}(\alpha(\mathbb{G}_0 + e_r) \mid \mathbb{G}_0) - \mathbb{E}(\alpha(\mathbb{G}_0)) = -\left(\frac{|O^*|}{n}\right)^2.$$

By a similar argument

$$\begin{aligned}
&\mathbb{E}(\alpha(\mathbb{G}(n,m,r-1) \mid \mathbb{G}_0) - \alpha(\mathbb{G}_0) \\
&= -\frac{n_1}{n}\left(\frac{|O^* \cap M_1|}{n_1}\right)^2 - \frac{n_2}{n}\left(\frac{|O^* \cap M_2|}{n_2}\right)^2 \\
&\leq -\left(\frac{n_1}{n}\frac{|O^* \cap M_1|}{n_1} + \frac{n_2}{n}\frac{|O^* \cap M_2|}{n_2}\right)^2 \\
&= -\left(\frac{|O^*|}{n}\right)^2 \\
&= \mathbb{E}(\alpha(\mathbb{G}_0 + e_r) \mid \mathbb{G}_0) - \mathbb{E}(\alpha(\mathbb{G}_0)),
\end{aligned}$$

completing the proof of (7.20). □

The proof of the following lemma is left as an exercise.

Lemma 7.6. *Given $\gamma \in (0,1)$, suppose that the non-negative sequence $u_n, n \geq 1$ satisfies*

$$u_n \geq u_{n_1} + u_{n_2} - O(n^\gamma)$$

for every n_1, n_2 such that $n_1 + n_2 = n$. Then $\lim_{n\to\infty} \frac{u_n}{n}$ exists.

7.4 Chromatic Number

Let $\chi(G)$ denote the chromatic number of a graph G, i.e., the smallest number of colors with which one can properly color the vertices of G. A coloring is proper if no two adjacent vertices have the same color.

Dense Graphs

We will first describe the asymptotic behavior of the chromatic number of dense random graphs. The following theorem is a major result, due to Bollobás [156]. The upper bound without the 2 in the denominator follows directly from Theorem 7.3. An intermediate result giving 3/2 instead of 2 was already proved by Matula [656].

Theorem 7.7. *Suppose* $0 < p < 1$ *is a constant and* $b = \frac{1}{1-p}$. *Then w.h.p.*

$$\chi(\mathbb{G}_{n,p}) \approx \frac{n}{2\log_b n}.$$

Proof. (i) By Theorem 7.3

$$\chi(\mathbb{G}_{n,p}) \geq \frac{n}{\alpha(\mathbb{G}_{n,p})} \approx \frac{n}{2\log_b n}.$$

(ii) Let $\nu = \frac{n}{(\log_b n)^2}$ and $k_0 = 2\log_b n - 4\log_b \log_b n$. It follows from (7.7) that

$$\begin{aligned} &\mathbb{P}(\exists S : |S| \geq \nu, S \text{ does not contain an independent set of order } \geq k_0) \\ &\leq \binom{n}{\nu} \exp\left\{-\Omega\left(\frac{\nu^2}{(\log n)^5}\right)\right\} \qquad (7.21) \\ &= o(1). \end{aligned}$$

So assume that every set of order at least ν contains an independent set of order at least k_0. We repeatedly choose an independent set of order k_0 among the set of uncolored vertices. Give each vertex in this set a new color. Repeat until the number of uncolored vertices is at most ν. Give each remaining uncolored vertex its own color. The number of colors used is at most

$$\frac{n}{k_0} + \nu \approx \frac{n}{2\log_b n}.$$

□

It should be noted that Bollobás did not have the Janson inequality available to him and he had to make a clever choice of random variable for use with the Azuma-Hoeffding inequality. His choice was the maximum size of a family of edge independent independent sets. Łuczak [637] proved the corresponding result to Theorem 7.7 in the case where $np \to 0$.

Concentration

Theorem 7.8. *Suppose* $0 < p < 1$ *is a constant. Then*

$$\mathbb{P}(|\chi(\mathbb{G}_{n,p}) - \mathbb{E}\,\chi(\mathbb{G}_{n,p})| \geq t) \leq 2\exp\left\{-\frac{t^2}{2n}\right\}$$

Proof. Write

$$\chi = Z(Y_1, Y_2, \dots, Y_n) \tag{7.22}$$

where

$$Y_j = \{(i,j) \in E(\mathbb{G}_{n,p}) : i < j\}.$$

Then

$$|Z(Y_1, Y_2, \dots, Y_n) - Z(Y_1, Y_2, \dots, \hat{Y}_i, \dots, Y_n)| \leq 1$$

and the theorem follows from the Azuma-Hoeffding inequality, see Section 27.7, in particular Lemma 27.17. □

Greedy Coloring Algorithm

We show below that a simple greedy algorithm performs very efficiently. It uses twice as many colors as it "should" in the light of Theorem 7.7. This algorithm is discussed in Bollobás and Erdős [162] and by Grimmett and McDiarmid [453]. It starts by greedily choosing an independent set C_1 and at the same time giving its vertices color 1. C_1 is removed and then we greedily choose an independent set C_2 and give its vertices color 2 and so on, until all vertices have been colored.

Algorithm *GREEDY*

- k is the current color.
- A is the current set of vertices that might get color k in the current round.
- U is the current set of uncolored vertices.

begin
  $k \longleftarrow 0$, $A \longleftarrow [n]$, $U \longleftarrow [n]$, $C_k \longleftarrow \emptyset$.
  while $U \neq \emptyset$ **do**
  $k \longleftarrow k+1$ $A \longleftarrow U$
  while $A \neq \emptyset$
    begin
    Choose $v \in A$ and put it into C_k
    $U \longleftarrow U \setminus \{v\}$
    $A \longleftarrow A \setminus (\{v\} \cup N(v))$
    end
end

Theorem 7.9. *Suppose* $0 < p < 1$ *is a constant and* $b = \frac{1}{1-p}$. *Then w.h.p. algorithm GREEDY uses approximately* $n/\log_b n$ *colors to color the vertices of* $\mathbb{G}_{n,p}$.

Proof. At the start of an iteration the edges inside U are un-examined. Suppose that

$$|U| \geq \nu = \frac{n}{(\log_b n)^2}.$$

We show that approximately $\log_b n$ vertices get color k i.e. at the end of round k, $|C_k| \approx \log_b n$.
Each iteration chooses a *maximal* independent set from the remaining uncolored vertices. Let $k_0 = \log_b n - 5\log_b \log_b n$. Then

$$\begin{aligned}
&\mathbb{P}(\exists\, T : |T| \leq k_0, T \text{ is maximally independent in U}) \\
&\leq \sum_{t=1}^{k_0} \binom{n}{t} (1-p)^{\binom{t}{2}} \left(1-(1-p)^t\right)^{\nu - t} \leq \sum_{t=1}^{k_0} (ne)^t e^{-(\nu - k_0)(1-p)^t} \leq \\
&\qquad k_0\,(ne)^{k_0}\, e^{-\frac{1}{2}(\log_b n)^3} \leq e^{-\frac{1}{3}(\log_b n)^3}.
\end{aligned}$$

So the probability that we fail to use at least k_0 colors while $|U| \geq \nu$ is at most

$$ne^{-\frac{1}{3}(\log_b \nu)^3} = o(1).$$

So w.h.p. GREEDY uses at most

$$\frac{n}{k_0} + \nu \approx \frac{n}{\log_b n} \text{ colors.}$$

We now put a lower bound on the number of colors used by GREEDY. Let

$$k_1 = \log_b n + 2\log_b \log_b n.$$

Consider one round. Let $U_0 = U$ and suppose $u_1, u_2, \ldots \in C_k$ and $U_{i+1} = U_i \setminus (\{u_i\}) \cup N(u_i))$.
Then

$$\mathbb{E}(\ |U_{i+1}|\ |U_i\ ) \leq |U_i\ |(1-p),$$

and so, for $i = 1, 2, ..$

$$\mathbb{E}\,|U_i| \leq n(1-p)^i.$$

So

$$\mathbb{P}(k_1 \text{ vertices colored in one round}) \leq \frac{1}{(\log_b n)^2},$$

and

$$\mathbb{P}(2k_1 \text{ vertices colored in one round}) \leq \frac{1}{n}.$$

So let

$$\delta_i = \begin{cases} 1 & \text{if at most } k_1 \text{ vertices are colored in round } i \\ 0 & \text{otherwise} \end{cases}$$

We see that

$$\mathbb{P}(\delta_i = 1 | \delta_1, \delta_2, \ldots, \delta_{i-1}) \geq 1 - \frac{1}{(\log_b n)^2}.$$

So the number of rounds that color more than k_1 vertices is stochastically dominated by a binomial with mean $n/(\log_b n)^2$. The Chernoff bounds imply that w.h.p. the number of rounds that color more than k_1 vertices is less than $2n/(\log_b n)^2$. Strictly speaking we need to use Lemma 27.24 to justify the use of the Chernoff bounds. Because no round colors more than $2k_1$ vertices we see that w.h.p. GREEDY uses at least

$$\frac{n - 2k_1 \times 2n/(\log_b n)^2}{k_1} \approx \frac{n}{\log_b n} \text{ colors.}$$

$\square$

Sparse Graphs

We now consider the case of sparse random graphs. We first state an important conjecture about the chromatic number.

Conjecture: Let $k \geq 3$ be a fixed positive integer. Then there exists $d_k > 0$ such that if ε is an arbitrary positive constant and $p = \frac{d}{n}$ then w.h.p. (i) $\chi(\mathbb{G}_{n,p}) \leq k$ for $d \leq d_k - \varepsilon$ and (ii) $\chi(\mathbb{G}_{n,p}) \geq k+1$ for $d \geq d_k + \varepsilon$.

In the absence of a proof of this conjecture, we present the following result due to Łuczak [638]. It should be noted that Shamir and Spencer [799] had already proved six point concentration.

Theorem 7.10. *If $p < n^{-5/6-\delta}, \delta > 0$, then the chromatic number of $\mathbb{G}_{n,p}$ is w.h.p. two point concentrated.*

Proof. To prove this theorem we need three lemmas.

Lemma 7.11.

(a) *Let $0 < \delta < 1/10$, $0 \le p < 1$ and $d = np$. Then w.h.p. each subgraph H of $\mathbb{G}_{n,p}$ on less than $nd^{-3(1+2\delta)}$ vertices has less than $(3/2-\delta)|H|$ edges.*

(b) *Let $0 < \delta < 1.0001$ and let $0 \le p \le \delta/n$. Then w.h.p. each subgraph H of $\mathbb{G}_{n,p}$ has less than $3|H|/2$ edges.*

The above lemma can be proved easily by the first moment method, see Exercise 7.6.6. Note also that Lemma 7.11 implies that each subgraph H satisfying the conditions of the lemma has minimum degree less than three, and thus is 3-colorable, due to the following simple observation (see Bollobás [157] Theorem V.1)

Lemma 7.12. *Let $k = \max_{H \subseteq G} \delta(H)$, where the maximum is taken over all induced subgraphs of G. Then $\chi(G) \le k+1$.*

Proof. This is an easy exercise in Graph Theory. We proceed by induction on $|V(G)|$. We choose a vertex of minimum degree v, color $G-v$ inductively and then color v. □

The next lemma is an immediate consequence of the Azuma-Hoeffding inequality, see Section 27.7, in particular Lemma 27.17.

Lemma 7.13. *Let $k = k(n)$ be such that*

$$\mathbb{P}(\chi(\mathbb{G}_{n,p}) \le k) > \frac{1}{\log\log n}. \tag{7.23}$$

Then w.h.p. all but at most $n^{1/2}\log n$ vertices of $\mathbb{G}_{n,p}$ can be properly colored using k colors.

Proof. Let Z be the maximum number of vertices in $\mathbb{G}_{n,p}$ that can be properly colored with k colors. Write $Z = Z(Y_1, Y_2, \ldots, Y_n)$ as in (7.22). Then we have

$$\mathbb{P}(Z = n) > \frac{1}{\log\log n} \text{ and } \mathbb{P}(|Z - \mathbb{E}Z)| \ge t) \le 2\exp\left\{-\frac{t^2}{2n}\right\}. \tag{7.24}$$

Putting $t = \frac{1}{2}n^{1/2}\log n$ into (7.24) shows that $\mathbb{E}Z \geq n - t$ and the lemma follows after applying the concentration inequality in (7.24) once again. $\square$

Now we are ready to present Łuczak's ingenious argument to prove Theorem 7.10. Note first that when p is such that $np \to 0$ as $n \to \infty$, then by Theorem 2.1 $\mathbb{G}_{n,p}$ is a forest w.h.p. and so its chromatic number is either 1 or 2. Furthermore, for $1/\log n < d < 1.0001$ the random graph $\mathbb{G}_{n,p}$ w.h.p. contains at least one edge and no subgraph with minimal degree larger than two (see Lemma 7.11), which implies that $\chi(\mathbb{G}_{n,p})$ is equal to 2 or 3 (see Lemma 7.12). Now let us assume that the edge probability p is such that $1.0001 < d = np < n^{1/6-\delta}$. Observe that in this range of p the random graph $\mathbb{G}_{n,p}$ w.h.p. contains an odd cycle, so $\chi(\mathbb{G}_{n,p}) \geq 3$.
Let k be as in Lemma 7.13 and let U_0 be a set of size at most $u_0 = n^{1/2}\log n$ such that $[n] \setminus U_0$ can be properly colored with k colors. Let us construct a nested sequence of subsets of vertices $U_0 \subseteq U_1 \subseteq \ldots \subseteq U_m$ of $\mathbb{G}_{n,p}$, where we define $U_{i+1} = U_i \cup \{v,w\}$, where $v,w \notin U_i$ are connected by an edge and both v and w have a neighbor in U_i. The construction stops at $i = m$ if such a pair $\{v,w\}$ does not exist.
Notice that m can not exceed $m_0 = n^{1/2}\log n$, since if $m > m_0$ then a subgraph of $\mathbb{G}_{n,p}$ induced by vertices of U_{m_0} would have

$$|U_{m_0}| = u_0 + 2m_0 \leq 3n^{1/2}\log n < nd^{-3(1+2\delta)}$$

vertices and at least $3m_0 \geq (3/2 - \delta)|U_{m_0}|$ edges, contradicting the statement of Lemma 7.11.
As a result, the construction produces a set U_m in $\mathbb{G}_{n,p}$, such that its size is smaller than $nd^{-3(1+2\delta)}$ and, moreover, all neighbors $N(U_m)$ of U_m form an independent set, thus "isolating" U_m from the "outside world".
Now, the coloring of the vertices of $\mathbb{G}_{n,p}$ is an easy task. Namely, by Lemma 7.13, we can color the vertices of $\mathbb{G}_{n,p}$ outside the set $U_m \cup N(U_m)$ with k colors. Then we can color the vertices from $N(U_m)$ with color $k+1$, and finally, due to Lemmas 7.11 and 7.12, the subgraph induced by U_m is 3-colorable and we can color U_m with any three of the first k colors. $\square$

7.5 Eigenvalues

Separation of first and remaining eigenvalues

The following theorem is a weaker version of a theorem of Füredi and Komlós [421], which was itself a strengthening of a result of Juhász [533]. See also Coja–Oghlan [232] and Vu [843]. In their papers, $2\omega\log n$ is replaced by $2 + o(1)$ and this is best possible.

Theorem 7.14. *Suppose that $\omega \to \infty, \omega = o(\log n)$ and $\omega^3(\log n)^2 \leq np \leq n - \omega^3(\log n)^2$. Let A denote the adjacency matrix of $\mathbb{G}_{n,p}$. Let the eigenvalues of A be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then w.h.p.*

(i) $\lambda_1 \approx np$

(ii) $|\lambda_i| \leq 2\omega \log n \sqrt{np(1-p)}$ *for* $2 \leq i \leq n$.

The proof of the above theorem is based on the following lemma.

In the following $|x|$ denotes the Euclidean norm of $x \in \mathbf{R}$.

Lemma 7.15. *Let J be the all 1's matrix and $M = pJ - A$. Then w.h.p.*

$$\|M\| \leq 2\omega \log n \sqrt{np(1-p)}$$

where

$$\|M\| = \max_{|x|=1} |Mx| = \max\left\{|\lambda_1(M)|, |\lambda_n(M)|\right\}.$$

We first show that the lemma implies the theorem. Let $\mathbf{e}$ denote the all 1's vector. Suppose that $|\xi| = 1$ and $\xi \perp \mathbf{e}$. Then $J\xi = 0$ and

$$|A\xi| = |M\xi| \leq \|M\| \leq 2\omega \log n \sqrt{np(1-p)}.$$

Now let $|x| = 1$ and let $x = \alpha u + \beta y$ where $u = \frac{1}{\sqrt{n}}\mathbf{e}$ and $y \perp \mathbf{e}$ and $|y| = 1$. Then

$$|Ax| \leq |\alpha||Au| + |\beta||Ay|.$$

We have, writing $A = pJ + M$, that

$$\begin{aligned}
|Au| = \frac{1}{\sqrt{n}}|A\mathbf{e}| &\leq \frac{1}{\sqrt{n}}(np|\mathbf{e}| + \|M\||\mathbf{e}|) \\
&\leq np + 2\omega \log n \sqrt{np(1-p)}
\end{aligned}$$

$$|Ay| \leq 2\omega \log n \sqrt{np(1-p)}$$

Thus

$$\begin{aligned}
|Ax| &\leq |\alpha| np + (|\alpha| + |\beta|) 2\omega \log n \sqrt{np(1-p)} \\
&\leq np + 3\omega \log n \sqrt{np(1-p)}.
\end{aligned}$$

This implies that $\lambda_1 \leq (1 + o(1))np$.
But

$$|Au| \geq |(A+M)u| - |Mu|$$

$$
\begin{aligned}
&= |pJu| - |Mu| \\
&\geq np - 2\omega \log n \sqrt{np(1-p)},
\end{aligned}
$$

implying $\lambda_1 \geq (1+o(1))np$, which completes the proof of (i).
Now

$$
\begin{aligned}
\lambda_2 &= \min_{\eta} \max_{0 \neq \xi \perp \eta} \frac{|A\xi|}{|\xi|} \\
&\leq \max_{0 \neq \xi \perp u} \frac{|A\xi|}{|\xi|} \\
&\leq \max_{0 \neq \xi \perp u} \frac{|M\xi|}{|\xi|} \\
&\leq 2\omega \log n \sqrt{np(1-p)} \\
\lambda_n &= \min_{|\xi|=1} \xi^T A\xi \geq \min_{|\xi|=1} \xi^T A\xi - p\xi^T J\xi \\
&= \min_{|\xi|=1} -\xi^T M\xi \geq -\|M\| \geq -2\omega \log n \sqrt{np(1-p)}.
\end{aligned}
$$

This completes the proof of (ii).

Proof of Lemma 7.15:

As in previously mentioned papers, we use the trace method of Wigner [853]. Putting $\hat{M} = M - pI_n$ we see that

$$\|M\| \leq \|\hat{M}\| + \|pI_n\| = \|\hat{M}\| + p$$

and so we bound $\|\hat{M}\|$.
Letting m_{ij} denote the (i,j)th entry of $\hat{M}$ we have

(i) $\mathbb{E}\, m_{ij} = 0$

(ii) $\mathrm{Var}\, m_{ij} \leq p(1-p) = \sigma^2$

(iii) $m_{ij}, m_{i'j'}$ are independent, unless $(i',j') = (j,i)$, in which case they are identical.

Now let $k \geq 2$ be an even integer.

$$
\begin{aligned}
Trace(\hat{M}^k) &= \sum_{i=1}^{n} \lambda_i(\hat{M}^k) \\
&\geq \max\left\{\lambda_1(\hat{M}^k), \lambda_n(\hat{M}^k)\right\} \\
&= \|\hat{M}^k\|.
\end{aligned}
$$

We estimate

$$\|\hat{M}\| \leq Trace(\hat{M}^k)^{1/k},$$

where $k = \omega \log n$.

Now,

$$\mathbb{E}(Trace(\hat{M}^k)) = \sum_{i_0,i_1,\dots,i_{k-1}\in[n]} \mathbb{E}(m_{i_0i_1}m_{i_1i_2}\cdots m_{i_{k-2}i_{k-1}}m_{i_{k-1}i_0}).$$

Recall that the i,jth entry of $\hat{M}^k$ is the sum over all products $m_{i,i_1}m_{i_1,i_2}\cdots m_{i_{k-1}j}$.

Continuing, we therefore have

$$\mathbb{E}\,\|\hat{M}\|^k \leq \sum_{\rho=2}^{k} \mathbb{E}_{n,k,\rho}$$

where

$$\mathbb{E}_{n,k,\rho} = \sum_{\substack{i_0,i_1,\dots,i_{k-1}\in[n]\\ |\{i_0,i_1,i_2,\dots,i_{k-1}\}|=\rho}} \left|\mathbb{E}\left(\prod_{j=0}^{k-1} m_{i_ji_{j+1}}\right)\right|.$$

Note that as $m_{ii} = 0$ by construction of $\hat{M}$ we have that $\mathbb{E}_{n,k,1} = 0$

Each sequence $\underline{i} = i_0,i_1,\dots,i_{k-1},i_0$ corresponds to a walk $W(\underline{i})$ on the graph K_n with n loops added. Note that

$$\mathbb{E}\left(\prod_{j=0}^{k-1} m_{i_ji_{j+1}}\right) = 0 \tag{7.25}$$

if the walk $W(\underline{i})$ contains an edge that is crossed exactly once, by condition (i). On the other hand, $|m_{ij}| \leq 1$ and so by conditions (ii), (iii),

$$\left|\mathbb{E}\left(\prod_{j=0}^{k-1} m_{i_ji_{j+1}}\right)\right| \leq \sigma^{2(\rho-1)}$$

if each edge of $W(\underline{i})$ is crossed at least twice and if $|\{i_0,i_1,\dots,i_{k-1}\}| = \rho$.

Let $R_{k,\rho}$ denote the number of (k,ρ) walks i.e closed walks of length k that visit ρ distinct vertices and do not cross any edge exactly once. We use the following trivial estimates:

(i) $\rho > \frac{k}{2}+1$ implies $R_{k,\rho} = 0$. (ρ this large will invoke (7.25)).

(ii) $\rho \leq \frac{k}{2}+1$ implies $R_{k,\rho} \leq n^\rho k^k$,

where n^ρ bounds from above the the number of choices of ρ distinct vertices, while k^k bounds the number of walks of length k.
We have

$$\mathbb{E}\,\|\hat{M}\|^k \le \sum_{\rho=2}^{\frac{1}{2}k+1} R_{k,\rho}\sigma^{2(\rho-1)} \le \sum_{\rho=2}^{\frac{1}{2}k+1} n^\rho k^k \sigma^{2(\rho-1)} \le 2n^{\frac{1}{2}k+1}k^k\sigma^k.$$

Therefore,

$$\mathbb{P}\left(\|\hat{M}\| \ge 2k\sigma n^{\frac{1}{2}}\right) = \mathbb{P}\left(\|\hat{M}\|^k \ge \left(2k\sigma n^{\frac{1}{2}}\right)^k\right) \le \frac{\mathbb{E}\,\|\hat{M}\|^k}{\left(2k\sigma n^{\frac{1}{2}}\right)^k}$$

$$\le \frac{2n^{\frac{1}{2}k+1}k^k\sigma^k}{\left(2k\sigma n^{\frac{1}{2}}\right)^k} = \left(\frac{(2n)^{1/k}}{2}\right)^k = \left(\frac{1}{2}+o(1)\right)^k = o(1).$$

It follows that w.h.p. $\|\hat{M}\| \le 2\sigma\omega(\log n)n^{1/2} \le 2\omega\log n\sqrt{np(1-p)}$ and completes the proof of Theorem 7.14. □

Concentration of eigenvalues

We show here how one can use Talagrand's inequality, Theorem 27.18, to show that the eigenvalues of random matrices are highly concentrated around their median values. The result is from Alon, Krivelevich and Vu [35].

Theorem 7.16. *Let A be an $n \times n$ random symmetric matrix with independent entries $a_{i,j} = a_{j,i}$, $1 \le i \le j \le n$ with absolute value at most one. Let its eigenvalues be $\lambda_1(A) \ge \lambda_2(A) \ge \cdots \ge \lambda_n(A)$. Suppose that $1 \le s \le n$. Let μ_s denote the median value of $\lambda_s(A)$ i.e. $\mu_s = \inf_\mu \{\mathbb{P}(\lambda_s(A) \le \mu) \ge 1/2\}$. Then for any $t \ge 0$ we have*

$$\mathbb{P}(|\lambda_s(A) - \mu_s| \ge t) \le 4e^{-t^2/32s^2}.$$

The same estimate holds for the probability that $\lambda_{n-s+1}(A)$ deviates from its median by more than t.

Proof. We will use Talagrand's inequality, Theorem 27.18. We let $m = \binom{n+1}{2}$ and let $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m$ where for each $1 \le k \le m$ we have $\Omega_k = \left\{a_{i,j}\right\}$ for some $i \le j$. Fix a positive integer s and let M,t be real numbers. Let $\mathscr{A}$ be the set of matrices A for which $\lambda_s(A) \le M$ and let $\mathscr{B}$ be the set of matrices for which $\lambda_s(B) \ge M + t$. When applying Theorem 27.18 it is convenient to view A as an m-vector.

Fix $B \in \mathscr{B}$ and let $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(s)}$ be an orthonormal set of eigenvectors for the s largest eigenvalues of B. Let $\mathbf{v}^{(k)} = (v_1^{(k)}, v_2^{(k)}, \ldots, v_n^{(k)})$,

$$\alpha_{i,i} = \sum_{k=1}^{s} (v_i^{(k)})^2 \quad \text{for } 1 \leq i \leq n$$

and

$$\alpha_{i,j} = 2\sqrt{\sum_{k=1}^{s} (v_i^{(k)})^2}\sqrt{\sum_{k=1}^{s} (v_j^{(k)})^2} \quad \text{for } 1 \leq i < j \leq n.$$

Lemma 7.17.

$$\sum_{1 \leq i \leq j \leq n} \alpha_{i,j}^2 \leq 2s^2.$$

Proof.

$$\begin{aligned}\sum_{1 \leq i \leq j \leq n} \alpha_{i,j}^2 &= \sum_{i=1}^{n} \left(\sum_{k=1}^{s} (v_i^{(k)})^2 \right)^2 + 4 \sum_{1 \leq i < j \leq n} \left(\sum_{k=1}^{s} (v_i^{(k)})^2 \sum_{k=1}^{s} (v_j^{(k)})^2 \right) \\ &\leq 2 \left(\sum_{i=1}^{n} \sum_{k=1}^{s} (v_i^{(k)})^2 \right)^2 = 2 \left(\sum_{k=1}^{s} \sum_{i=1}^{n} (v_i^{(k)})^2 \right)^2 = 2s^2,\end{aligned}$$

where we have used the fact that each $\mathbf{v}^{(k)}$ is a unit vector.

□

Lemma 7.18. *For every $A = (a_{i,j}) \in \mathscr{A}$ and $B = (b_{i,j}) \in \mathscr{B}$,*

$$\sum_{1 \leq i \leq j \leq n: a_{i,j} \neq b_{i,j}} \alpha_{i,j} \geq t/2.$$

Fix $A \in \mathscr{A}$. Let $\mathbf{u} = \sum_{k=1}^{s} c_k \mathbf{v}^{(k)}$ be a unit vector in the span S of the vectors $\mathbf{v}^{(k)}, k = 1, 2, \ldots, s$ which is orthogonal to the eigenvectors of the $(s-1)$ largest eigenvalues of A. Recall that $\mathbf{v}^{(k)}, k = 1, 2, \ldots, s$ are eigenvectors of B. Then $\sum_{k=1}^{s} c_k^2 = 1$ and $\mathbf{u}^t A \mathbf{u} \leq \lambda_s(A) \leq M$, whereas $\mathbf{u}^t B \mathbf{u} \geq \min_{\mathbf{v} \in S} \mathbf{v}^t B \mathbf{v} = \lambda_s(B) \geq M + t$. Recall that all entries of A and B are bounded in absolute value by 1, implying that $|b_{i,j} - a_{i,j}| \leq 2$ for all $1 \leq i, j \leq n$. It follows that if X is the set of ordered pairs (i, j) for which $a_{i,j} \neq b_{i,j}$ then

$$\begin{aligned}t \leq \mathbf{u}^t (B - A) \mathbf{u} &= \sum_{(i,j) \in X} (b_{i,j} - a_{i,j}) \left(\sum_{k=1}^{s} c_k v_i^{(k)} \right)^t \sum_{k=1}^{s} c_k v_j^{(k)} \\ &\leq 2 \sum_{(i,j) \in X} \left| \sum_{k=1}^{s} c_k v_i^{(k)} \right| \left| \sum_{k=1}^{s} c_k v_j^{(k)} \right|\end{aligned}$$

$$\leq 2 \sum_{(i,j)\in X} \left(\sqrt{\sum_{k=1}^{s} c_k^2} \sqrt{\sum_{k=1}^{s} \left(v_i^{(k)}\right)^2} \right) \left(\sqrt{\sum_{k=1}^{s} c_k^2} \sqrt{\sum_{k=1}^{s} \left(v_j^{(k)}\right)^2} \right)$$
$$= 2 \sum_{(i,j)\in X} \alpha_{i,j}$$

as claimed. (We obtained the third inequality by use of the Cauchy-Schwarz inequality).

□

By the above two lemmas, and by Theorem 27.18 for every M and every $t > 0$

$$\mathbb{P}(\lambda_s(A) \leq M)\,\mathbb{P}(\lambda_s(B) \geq M+t) \leq e^{-t^2/(32s^2)}. \tag{7.26}$$

If M is the median of $\lambda_s(A)$ then $\mathbb{P}(\lambda_s(A) \leq M) \geq 1/2$, by definition, implying that

$$\mathbb{P}(\lambda_s(A) \geq M+t) \leq 2e^{-t^2/(32s^2)}.$$

Similarly, by applying (7.26) with $M+t$ being the median of $\lambda_s(A)$ we conclude that

$$\mathbb{P}(\lambda_s(A) \leq M-t) \leq 2e^{-t^2/(32s^2)}.$$

This completes the proof of Theorem 7.16 for $\lambda_s(A)$. The proof for λ_{n-s+1} follows by applying the theorem to s and $-A$. □

7.6 Exercises

7.6.1 Let $p = d/n$ where d is a positive constant. Let S be the set of vertices of degree at least $\frac{2\log n}{3\log\log n}$. Show that w.h.p., S is an independent set.

7.6.2 Let $p = d/n$ where d is a large positive constant. Use the first moment method to show that w.h.p.

$$\alpha(G_{n,p}) \leq \frac{2n}{d}(\log d - \log\log d - \log 2 + 1 + \varepsilon)$$

for any positive constant ε.

7.6.3 Complete the proof of Theorem 7.4.
Let $m = d/(\log d)^2$ and partition $[n]$ into $n_0 = \frac{n}{m}$ sets $S_1, S_2, \ldots, S_{n_0}$ of size m. Let $\beta(G)$ be the maximum size of an independent set S that satisfies $|S \cap S_i| \leq 1$ for $i = 1, 2, \ldots, n_0$. Use the proof idea of Theorem 7.4 to show that w.h.p.

$$\beta(\mathbb{G}_{n,p}) \geq k_{-\varepsilon} = \frac{2n}{d}(\log d - \log\log d - \log 2 + 1 - \varepsilon).$$

7.6.4 Prove Theorem 7.4 using Talagrand's inequality, Theorem 27.22. (Hint: Let $A = \left\{\alpha(\mathbb{G}_{n,p}) \leq k_{-\varepsilon} - 1\right\}$).

7.6.5 Prove Lemma 7.6.

7.6.6 Prove Lemma 7.11.

7.6.7 Prove that if $\omega = \omega(n) \to \infty$ then there exists an interval I of length $\omega n^{1/2}/\log n$ such that w.h.p. $\chi(G_{n,1/2}) \in I$. (See Scott [797]).

7.6.8 A *topological clique* of size s is a graph obtained from the complete graph K_s by subdividing edges. Let $tc(G)$ denote the size of the largest topological clique contained in a graph G. Prove that w.h.p. $tc(G_{n,1/2}) = \Theta(n^{1/2})$.

7.6.9 Suppose that H is obtained from $G_{n,1/2}$ by planting a clique C of size $m = n^{1/2}\log n$ inside it. describe a polynomial time algorithm that w.h.p. finds C. (Think that an adversary adds the clique without telling you where it is).

7.6.10 Show that if $d > 2k\log k$ for a positive integer $k \geq 2$ then w.h.p. $G(n, d/n)$ is not k-colorable. (Hint:Consider the expected number of proper k-coloring's).

7.6.11 Let $p = K\log n/n$ for some large constant $K > 0$. Show that w.h.p. the diameter of $\mathbb{G}_{n,p}$ is $\Theta(\log n/\log\log n)$.

7.6.12 Suppose that $1+\varepsilon \leq np = o(\log n)$, where $\varepsilon > 0$ is constant. Show that given $A > 0$, there exists $B = B(A)$ such that

$$\mathbb{P}\left(diam(K) \geq B\frac{\log n}{\log np}\right) \leq n^{-A},$$

where K is the giant component of $\mathbb{G}_{n,p}$.

7.6.13 Let $p = d/n$ for some constant $d > 0$. Let A be the adjacency matrix of $G_{n,p}$. Show that w.h.p. $\lambda_1(A) \approx \Delta^{1/2}$ where Δ is the maximum degree in $G_{n,p}$. (Hint: the maximum eigenvalue of the adjacency matrix of $K_{1,m}$ is $m^{1/2}$).

7.6.14 A proper *2-tone* k-coloring of a graph $G = (V,E)$ is an assignment of pairs of colors $C_v \subseteq [k], |C_v| = 2$ such that (i) $|C_v \cap C_w| < d(v,w)$ where $d(v,w)$ is the graph distance from v to w. If $\chi_2(G)$ denotes the minimum k for which there exists a 2-tone coloring of G, show that w.h.p. $\chi_2(\mathbb{G}_{n,p}) \approx 2\chi(\mathbb{G}_{n,p})$. (This question is taken from [60]).

7.6.15 The *set chromatic number* $\chi_s(G)$ of a graph $G=(V,E)$ is defined as follows: Let C denote a set of colors. Color each $v \in V$ with a color $f(v) \in C$. Let $C_v = \{f(w) : \{v,w\} \in G\}$. The coloring is proper if $C_v \neq C_w$ whenever $\{v,w\} \in E$. χ_s is the minimum size of C in a proper coloring of G. Prove that if $0<p<1$ is constant then w.h.p. $\chi_s(\mathbb{G}_{n,p}) \approx r\log_2 n$ where $r=\frac{2}{\log_2 1/s}$ and $s=\min\left\{q^{2\ell}+(1-q^\ell)^2 : \ell=1,2,\ldots\right\}$ where $q=1-p$. (This question is taken from Dudek, Mitsche and Pralat [311]).

7.7 Notes

Chromatic number

There has been a lot of progress in determining the chromatic number of sparse random graphs. Alon and Krivelevich [32] extended the result in [638] to the range $p \leq n^{-1/2-\delta}$. A breakthrough came when Achlioptas and Naor [6] identified the two possible values for $np=d$ where $d=O(1)$: Let k_d be the smallest integer k such that $d<2k\log k$. Then w.h.p. $\chi(\mathbb{G}_{n,p}) \in \{k_d, k_d+1\}$. This implies that d_k, the (conjectured) threshold for a random graph to have chromatic number at most k, satisfies $d_k \geq 2k\log k - 2\log_k - 2 + o_k(1)$ where $o_k(1) \to 0$ as $k \to \infty$. Coja–Oghlan, Panagiotou and Steger [234] extended the result of [6] to $np \leq n^{1/4-\varepsilon}$, although here the guaranteed range is three values. More recently, Coja–Oghlan and Vilenchik [235] proved the following. Let $d_{k,cond} = 2k\log k - \log k - 2\log 2$. Then w.h.p. $d_k \geq d_{k,cond} - o_k(1)$. On the other hand Coja–Oghlan [233] proved that $d_k \leq d_{k,cond} + (2\log 2 - 1) + o_k(1)$.

It follows from Chapter 2 that the chromatic number of $\mathbb{G}_{n,p}, p \leq 1/n$ is w.h.p. at most 3. Achlioptas and Moore [4] proved that in fact $\chi(\mathbb{G}_{n,p}) \leq 3$ w.h.p. for $p \leq 4.03/n$. Now a graph G is s-colorable iff it has a homomorphism $\varphi : G \to K_s$. (A *homomorphism* from G to H is a mapping $\varphi : V(G) \to V(H)$ such that if $\{u,v\} \in E(G)$ then $(\varphi(u),\varphi(v)) \in E(H)$). It is therefore of interest in the context of coloring, to consider homomorphisms from $\mathbb{G}_{n,p}$ to other graphs. Frieze and Pegden [411] show that for any $\ell > 1$ there is an $\varepsilon > 0$ such that with high probability, $G_{n,\frac{1+\varepsilon}{n}}$ either has odd-girth $< 2\ell+1$ or has a homomorphism to the odd cycle $C_{2\ell+1}$. They also showed that w.h.p. there is no homomorphism from $\mathbb{G}_{n,p}, p=4/n$ to C_5. Previously, Hatami [470] has shown that w.h.p. there is no homomorphism from a random cubic graph to C_7.

Alon and Sudakov [37] considered how many edges one must add to $\mathbb{G}_{n,p}$ in order to significantly increase the chromatic number. They show that if $n^{-1/3+\delta} \leq p \leq 1/2$ for some fixed $\delta > 0$ then w.h.p. for every set E of $\frac{2^{-12}\varepsilon^2 n^2}{(\log_b(np))^2}$ edges, the chromatic number of $\mathbb{G}_{n,p} \cup E$ is still at most $\frac{(1+\varepsilon)n}{2\log_b(np)}$.

Let L_k be an arbitrary function that assigns to each vertex of G a list of k colors. We say that G is **L_k-list-colorable** if there exists a proper coloring of the vertices such that every vertex is colored with a color from its own list. A graph is **k-choosable**, if for every such function L_k, G is L_k-list-colorable. The minimum k for which a graph is k-choosable is called the **list chromatic number**, or the **choice number**, and denoted by $\chi_L(G)$. The study of the choice number of $\mathbb{G}_{n,p}$ was initiated in [26], where Alon proved that w.h.p., the choice number of $\mathbb{G}_{n,1/2}$ is $o(n)$. Kahn then showed (see [27]) that w.h.p. the choice number of $\mathbb{G}_{n,1/2}$ equals $(1+o(1))\chi(\mathbb{G}_{n,1/2})$. In [589], Krivelevich showed that this holds for $p \gg n^{-1/4}$, and Krivelevich, Sudakov, Vu, and Wormald [607] improved this to $p \gg n^{-1/3}$. On the other hand, Alon, Krivelevich, Sudakov [33] and Vu [842] showed that for any value of p satisfying $2 < np \leq n/2$, the choice number is $\Theta(np/\log(np))$. Krivelevich and Vu [608] generalized this to hypergraphs; they also improved the leading constants and showed that the choice number for $C/n \leq p \leq 0.9$ (where C is a sufficiently large constant) is at most a multiplicative factor of $2+o(1)$ away from the chromatic number, the best known factor for $p \leq n^{-1/3}$.

Algorithmic questions

We have seen that the Greedy algorithm applied to $\mathbb{G}_{n,p}$ generally produces a coloring that uses roughly twice the minimum number of colors needed. Note also that the analysis of Theorem 7.9, when $k = 1$, implies that a simple greedy algorithm for finding a large independent set produces one of roughly half the maximum size. In spite of much effort neither of these two results have been significantly improved. We mention some negative results. Jerrum [526] showed that the *Metropolis* algorithm was unlikely to do very well in finding an independent set that was significantly larger than GREEDY. Other earlier negative results include: Chvátal [228], who showed that for a significant set of densities, a large class of algorithms will w.h.p. take exponential time to find the size of the largest independent set and McDiarmid [660] who carried out a similar analysis for the chromatic number.

Frieze, Mitsche, Pérez-Giménez and Pralat [409] study list coloring in an on-line setting and show that for a wide range of p, one can asymptotically match the best known constants of the off-line case. Moreover, if $pn \geq \log^{\omega} n$, then they get the same multiplicative factor of $2+o(1)$.

Randomly Coloring random graphs

A substantial amount of research in Theoretical Computer Science has been associated with the question of random sampling from complex distributions. Of relevance here is the following: Let G be a graph and k be a positive integer. Then let $\Omega_k(G)$ be the set of proper k-coloring's of the vertices of G. There has been a good deal of work on the problem of efficiently choosing a (near) random member of $\Omega_k(G)$. For example, Vigoda [840] has described an algorithm that produces a (near) random sample in polynomial time provided $k > 11\Delta(G)/6$. When it comes to $\mathbb{G}_{n,p}$, Dyer, Flaxman, Frieze and Vigoda [321] showed that if $p = d/n, d = O(1)$ then w.h.p. one can sample a random coloring if $k = O(\log\log n) = o(\Delta)$. The bound on k was reduced to $k = O(d^{O(1)})$ by Mossell and Sly [697] and then to $k = O(d)$ by Efthymiou [327].

Diameter of sparse random graphs

The diameter of the giant component of $\mathbb{G}_{n,p}$, $p = \lambda/n, \lambda > 1$ was considered by Fernholz and Ramachandran [359] and by Riordan and Wormald [765]. In particular, [765] proves that w.h.p. the diameter is $\frac{\log n}{\log \lambda} + 2\frac{\log n}{\log 1/\lambda^*} + W$ where $\lambda^* < 1$ and $\lambda^* e^{-\lambda^*} = \lambda e^{-\lambda}$ and $W = O_p(1)$ i.e. is bounded in probability for $\lambda = O(1)$ and $O(1)$ for $\lambda \to \infty$. In addition, when $\lambda = 1 + \varepsilon$ where $\varepsilon^3 n \to \infty$ i.e. the case of the emerging giant, [765] shows that w.h.p. the diameter is $\frac{\log \varepsilon^3 n}{\log \lambda} + 2\frac{\log \varepsilon^3 n}{\log 1/\lambda^*} + W$ where $W = O_p(1/\varepsilon)$. If $\lambda = 1 - \varepsilon$ where $\varepsilon^3 n \to \infty$ i.e. the sub-critical case, then Łuczak [640] showed that w.h.p. the diameter is $\frac{\log(2\varepsilon^3 n) + O_p(1)}{-\log \lambda}$.

Part II

Basic Model Extensions

Chapter 8

Inhomogeneous Graphs

Thus far we have concentrated on the properties of the random graphs $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$. We first consider a generalisation of $\mathbb{G}_{n,p}$ where the probability of edge (i,j) is p_{ij} is not the same for all pairs i,j. We call this the *generalized binomial graph* . Our main result on this model concerns the probability that it is connected. For this model we concentrate on its degree sequence and the existence of a giant component. After this we move onto a special case of this model, viz. the *expected degree model*. Here p_{ij} is proportional to $w_i w_j$ for weights w_i. In this model, we prove results about the size of the largest components. We finally consider another special case of the generalized binomial graph, viz. the *Kronecker random graph*.

8.1 Generalized Binomial Graph

Consider the following natural generalisation of the binomial random graph $\mathbb{G}_{n,p}$, first considered by Kovalenko [587].
Let $V = \{1,2,\ldots,n\}$ be the vertex set. The random graph $\mathbb{G}_{n,\mathbf{P}}$ has vertex set V and two vertices i and j from V, $i \neq j$, are joined by an edge with probability $p_{ij} = p_{ij}(n)$, independently of all other edges. Denote by

$$\mathbf{P} = \left[p_{ij}\right]$$

the symmetric $n \times n$ matrix of edge probabilities, where $p_{ii} = 0$. Put $q_{ij} = 1 - p_{ij}$ and for $i,k \in \{1,2,\ldots,n\}$ define

$$Q_i = \prod_{j=1}^{n} q_{ij},\ \lambda_n = \sum_{i=1}^{n} Q_i.$$

Note that Q_i is the probability that vertex i is isolated and λ_n is the expected number of isolated vertices. Next let

$$R_{ik} = \min_{1 \le j_1 < j_2 < \cdots < j_k \le n} q_{ij_1} \cdots q_{ij_k}.$$

Suppose that the edge probabilities p_{ij} are chosen in such a way that the following conditions are simultaneously satisfied as $n \to \infty$:

$$\max_{1 \le i \le n} Q_i \to 0, \tag{8.1}$$

$$\lim_{n\to\infty} \lambda_n = \lambda = constant, \tag{8.2}$$

and

$$\lim_{n\to\infty} \sum_{k=1}^{n/2} \frac{1}{k!} \left(\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} \right)^k = e^{\lambda} - 1. \tag{8.3}$$

The next two theorems are due to Kovalenko [587].
We will first give the asymptotic distribution of the number of isolated vertices in $\mathbb{G}_{n,\mathbf{P}}$, assuming that the above three conditions are satisfied. The next theorem is a generalisation of the corresponding result for the classical model $\mathbb{G}_{n,p}$ (see Theorem 3.1(ii)).

Theorem 8.1. *Let X_0 denote the number of isolated vertices in the random graph $\mathbb{G}_{n,\mathbf{P}}$. If conditions* (8.1) (8.2) *and* (8.3) *hold, then*

$$\lim_{n\to\infty} \mathbb{P}(X_0 = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

for $k = 0, 1, \ldots$, i.e., the number of isolated vertices is asymptotically Poisson distributed with mean λ.

Proof. Let

$$X_{ij} = \begin{cases} 1 & \text{with prob. } p_{ij} \\ 0 & \text{with prob. } q_{ij} = 1 - p_{ij}. \end{cases}$$

Denote by X_i, for $i = 1, 2, \ldots n$, the indicator of the event that vertex i is isolated in $\mathbb{G}_{n,\mathbf{P}}$. To show that X_0 converges in distribution to the Poisson random variable with mean λ one has to show (see Theorem 26.11) that for any natural number k

$$\mathbb{E}\left(\sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} X_{i_1} X_{i_2} \cdots X_{i_k} \right) \to \frac{\lambda^k}{k!} \tag{8.4}$$

as $n \to \infty$. But

$$\mathbb{E}\left(X_{i_1} X_{i_2} \cdots X_{i_k}\right) = \prod_{r=1}^{k} \mathbb{P}\left(X_{i_r} = 1 | X_{i_1} = \ldots = X_{i_{r-1}} = 1\right), \tag{8.5}$$

where in the case of $r = 1$ we condition on the sure event.

Since the LHS of (8.4) is the sum of $\mathbb{E}\left(X_{i_1}X_{i_2}\cdots X_{i_k}\right)$ over all $i_1 < \cdots < i_k$, we need to find matching upper and lower bounds for this expectation. Now $\mathbb{P}\left(X_{i_r} = 1 | X_{i_1} = \ldots = X_{i_{r-1}} = 1\right)$ is the unconditional probability that i_r is not adjacent to any vertex $j \neq i_1, \ldots, i_{r-1}$ and so

$$\mathbb{P}\left(X_{i_r} = 1 | X_{i_1} = \ldots = X_{i_{r-1}} = 1\right) = \frac{\prod_{j=1}^{n} q_{i_r j}}{\prod_{s=1}^{r-1} q_{i_r i_s}}.$$

Hence

$$Q_{i_r} \leq P\left(X_{i_r} = 1 | X_{i_1} = \ldots = X_{i_{r-1}} = 1\right) \leq \frac{Q_{i_r}}{R_{i_r, r-1}} \leq \frac{Q_{i_r}}{R_{i_r k}}.$$

It follows from (8.5) that

$$Q_{i_1} \cdots Q_{i_k} \leq \mathbb{E}\left(X_{i_1} \cdots X_{i_k}\right) \leq \frac{Q_{i_1}}{R_{i_1 k}} \cdots \frac{Q_{i_k}}{R_{i_k k}}. \tag{8.6}$$

Applying conditions (8.1) and (8.2) we get that

$$\begin{aligned}
\sum_{1 \leq i_1 < \cdots < i_k \leq n} Q_{i_1} \cdots Q_{i_k} = \frac{1}{k!} \sum_{1 \leq i_1 \neq \cdots \neq i_r \leq n} Q_{i_1} \cdots Q_{i_k} \geq \\
\frac{1}{k!} \sum_{1 \leq i_1, \ldots, i_k \leq n} Q_{i_1} \cdots Q_{i_k} - \frac{k}{k!} \sum_{i=1}^{n} Q_i^2 \left(\sum_{1 \leq i_1, \ldots, i_{k-2} \leq n} Q_{i_1} \cdots Q_{i_{k-2}} \right) \\
\geq \frac{\lambda_n^k}{k!} - (\max_i Q_i) \lambda_n^{k-1} = \frac{\lambda_n^k}{k!} - (\max_i Q_i) \lambda_n^{k-1} \to \frac{\lambda^k}{k!},
\end{aligned} \tag{8.7}$$

as $n \to \infty$.
Now,

$$\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} \geq \lambda_n = \sum_{i=1}^{n} Q_i,$$

and if $\limsup_{n\to\infty} \sum_{i=1}^{n} \frac{Q_i}{R_{ik}} > \lambda$ then $\limsup_{n\to\infty} \sum_{k=1}^{n/2} \frac{1}{k!} \left(\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} \right)^k > e^{\lambda} - 1$, which contradicts (8.3). It follows that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{Q_i}{R_{ik}} = \lambda.$$

Therefore

$$\sum_{1 \leq i_1 < \ldots < i_k \leq n} Q_{i_1} \cdots Q_{i_k} \leq \frac{1}{k!} \left(\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} \right)^k \to \frac{\lambda^k}{k!}.$$

as $n \to \infty$.
Combining this with (8.7) gives us (8.4) and completes the proof of Theorem 8.1. $\square$

One can check that the conditions of the theorem are satisfied when

$$p_{ij} = \frac{\log n + x_{ij}}{n},$$

where x_{ij}'s are uniformly bounded by a constant.

The next theorem shows that under certain circumstances, the random graph $\mathbb{G}_{n,\mathbf{P}}$ behaves in a similar way to $\mathbb{G}_{n,p}$ at the connectivity threshold.

Theorem 8.2. *If the conditions* (8.1)*,* (8.2) *and* (8.3) *hold, then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,\mathbf{P}} \textit{ is connected}) = e^{-\lambda}.$$

Proof. To prove the this we will show that if (8.1), (8.2) and (8.3) are satisfied then w.h.p. $\mathbb{G}_{n,\mathbf{P}}$ consists of $X_0 + 1$ connected components, i.e., $\mathbb{G}_{n,\mathbf{P}}$ consists of a single giant component plus components that are isolated vertices only. This, together with Theorem 8.1, implies the conclusion of Theorem 8.2.
Let $U \subseteq V$ be a subset of the vertex set V. We say that U is *closed* if $X_{ij} = 0$ for every i and j, where $i \in U$ and $j \in V \setminus U$. Furthermore, a closed set U is called *simple* if either U or $V \setminus U$ consists of isolated vertices only. Denote the number of non-empty closed sets in $\mathbb{G}_{n,\mathbf{P}}$ by Y_1 and the number of non-empty simple sets by Y. Clearly $Y_1 \geq Y$.
We will prove first that

$$\liminf_{n\to\infty} \mathbb{E}Y \geq 2e^{\lambda} - 1. \tag{8.8}$$

Denote the set of isolated vertices in $\mathbb{G}_{n,\mathbf{P}}$ by J. If $V \setminus J$ is not empty then $Y = 2^{X_0+1} - 1$ (the number of non-empty subsets of J plus the number of their complements, plus V itself). If $V \setminus J = \emptyset$ then $Y = 2^n - 1$. Now, by Theorem 8.1, for every fixed $k = 0, 1, \ldots,$

$$\lim_{n\to\infty} \mathbb{P}(Y = 2^{k+1} - 1) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Observe that for any $\ell \geq 0$,

$$\mathbb{E}Y \geq \sum_{k=0}^{\ell} (2^{k+1} - 1)\,\mathbb{P}(Y = 2^{k+1} - 1)$$

and hence

$$\liminf_{n\to\infty} \mathbb{E}Y \geq \sum_{k=0}^{\ell}(2^{k+1}-1)\frac{\lambda^k e^{-\lambda}}{k!}.$$

So,

$$\liminf_{n\to\infty} \mathbb{E}Y \geq \lim_{\ell\to\infty}\sum_{k=0}^{\ell}(2^{k+1}-1)\frac{\lambda^k e^{-\lambda}}{k!} = 2e^{\lambda}-1$$

which completes the proof of (8.8).

We will show next that

$$\limsup_{n\to\infty} \mathbb{E}Y_1 \leq 2e^{\lambda}-1. \tag{8.9}$$

To prove (8.9) denote by Z_k the number of closed sets of order k in $\mathbb{G}_{n,\mathbf{P}}$ so that $Y_1 = \sum_{k=1}^{n} Z_k$. Note that

$$Z_k = \sum_{i_1<\ldots<i_k} Z_{i_1\ldots i_k},$$

where $Z_{i_1,\ldots i_k}$ indicates whether set $I_k = \{i_1 \ldots i_k\}$ is closed. Then

$$\mathbb{E}Z_{i_1,\ldots i_k} = \mathbb{P}(X_{ij}=0, i\in I_k, j\notin I_k) = \prod_{i\in I_k, j\notin I_k} q_{ij}.$$

Consider first the case when $k \leq n/2$. Then

$$\prod_{i\in I_k, j\notin I_k} q_{ij} = \frac{\prod_{i\in I_k, 1\leq j\leq n} q_{ij}}{\prod_{i\in I_k, j\in I_k} q_{ij}} = \prod_{i\in I_k}\frac{Q_i}{\prod_{j\in I_k} q_{ij}} \leq \prod_{i\in I_k}\frac{Q_i}{R_{ik}}.$$

Hence

$$\mathbb{E}Z_k \leq \sum_{i_1<\ldots<i_k}\prod_{i\in I_k}\frac{Q_i}{R_{ik}} \leq \frac{1}{k!}\left(\sum_{i=1}^{n}\frac{Q_i}{R_{ik}}\right)^k.$$

Now, (8.3) implies that

$$\limsup_{n\to\infty}\sum_{k=1}^{n/2}\mathbb{E}Z_k \leq e^{\lambda}-1.$$

To complete the estimation of $\mathbb{E}Z_k$ (and thus for $\mathbb{E}Y_1$) consider the case when $k > n/2$. For convenience let us switch k with $n-k$, i.e, consider $\mathbb{E}Z_{n-k}$, when $0 \leq k < n/2$. Notice that $\mathbb{E}Z_n = 1$ since V is closed. So for $1 \leq k < n/2$

$$\mathbb{E}Z_{n-k} = \sum_{i_1<\ldots<i_k}\prod_{i\in I_k, j\notin I_k} q_{ij}.$$

But $q_{ij} = q_{ji}$ so, for such k, $\mathbb{E} Z_{n-k} = \mathbb{E} Z_k$. This gives

$$\limsup_{n\to\infty} \mathbb{E} Y_1 \le 2(e^{\lambda} - 1) + 1,$$

where the +1 comes from $Z_n = 1$ This completes the proof of (8.9).

Now,

$$\mathbb{P}(Y_1 > Y) = \mathbb{P}(Y_1 - Y \ge 1) \le \mathbb{E}(Y_1 - Y).$$

Estimates (8.8) and (8.9) imply that

$$\limsup_{n\to\infty} \mathbb{E}(Y_1 - Y) \le 0,$$

which in turn leads to the conclusion that

$$\lim_{n\to\infty} \mathbb{P}(Y_1 > Y) = 0.$$

i.e., asymptotically, the probability that there is a closed set that is not simple, tends to zero as $n \to \infty$. It is easy to check that $X_0 < n$ w.h.p. and therefore $Y = 2^{X_0+1} - 1$ w.h.p. and so w.h.p. $Y_1 = 2^{X_0+1} - 1$. If $\mathbb{G}_{n,\mathbf{P}}$ has more than $X_0 + 1$ connected components then the graph after removal of all isolated vertices would contain at least one closed set, i.e., the number of closed sets would be at least 2^{X_0+1}. But the probability of such an event tends to zero and the theorem follows. $\square$

We finish this section by presenting a sufficient condition for $\mathbb{G}_{n,\mathbf{P}}$ to be connected w.h.p. as proven by Alon [28].

Theorem 8.3. *For every positive constant b there exists a constant $c = c(b) > 0$ so that if, for every non-trivial $S \subset V$,*

$$\sum_{i \in S, j \in V \setminus S} p_{ij} \ge c \log n,$$

then probability that $\mathbb{G}_{n,\mathbf{P}}$ is connected is at least $1 - n^{-b}$.

Proof. In fact Alon's result is much stronger. He considers a random subgraph $\mathbb{G}_{p_e}$ of a multi-graph G on n vertices, obtained by deleting each edge e independently with probability $1 - p_e$. The random graph $\mathbb{G}_{n,\mathbf{P}}$ is a special case of $\mathbb{G}_{p_e}$ when G is the complete graph K_n. Therefore, following in his footsteps, we will prove that Theorem 8.3 holds for $\mathbb{G}_{p_e}$ and thus for $\mathbb{G}_{n,\mathbf{P}}$.

So, let $G = (V, E)$ be a loopless undirected multigraph on n vertices, with probability p_e, $0 \le p_e \le 1$ assigned to every edge $e \in E$ and suppose that for any non-trivial $S \subset V$ the expectation of the number E_S of edges in a cut $(S, V \setminus S)$ of $\mathbb{G}_{p_e}$ satisfies

$$\mathbb{E} E_S = \sum_{e \in (S, V \setminus S)} p_e \geq c \log n. \tag{8.10}$$

Create a new graph $G' = (V, E')$ from G by replacing each edge e by $k = c \log n$ parallel copies with the same endpoints and giving each copy e' of e a probability $p'_{e'} = p_e/k$.

Observe that for $S \subset V$

$$\mathbb{E} E'_S = \sum_{e' \in (S, V \setminus S)} p'_{e'} = \mathbb{E} E_S.$$

Moreover, for every edge e of G, the probability that no copy e' of e survives in a random subgraph $\mathbb{G}'_{p'}$ is $(1 - p_e/k)^k \geq 1 - p_e$ and hence the probability that $\mathbb{G}_{p_e}$ is connected exceeds the probability of $\mathbb{G}'_{p'_e}$ being connected, and so in order to prove the theorem it suffice to prove that

$$\mathbb{P}(\mathbb{G}'_{p'_e} \text{ is connected}) \geq 1 - n^{-b}. \tag{8.11}$$

To prove this, let $E'_1 \cup E'_2 \cup \ldots \cup E'_k$ be a partition of the set E' of the edges of G', such that each E'_i consists of a single copy of each edge of G. For $i = 0, 1, \ldots, k$ define $\mathbb{G}'_i$ as follows. $\mathbb{G}'_0$ is the subgraph of G' that has no edges, and for all $i \geq 1$, $\mathbb{G}'_i$ is the random subgraph of G' obtained from $\mathbb{G}'_{i-1}$ by adding to it each edge $e' \in E'_i$ independently, with probability $p'_{e'}$.

Let C_i be the number of connected components of $\mathbb{G}'_i$. Then we have $C_0 = n$ and we have $\mathbb{G}'_k \equiv \mathbb{G}'_{p'_e}$. Let us call the stage i, $1 \leq i \leq k$, *successful* if either $C_{i-1} = 1$ (i.e., $\mathbb{G}'_{i-1}$ is connected) or if $C_i < 0.9 C_{i-1}$. We will prove that

$$\mathbb{P}(C_{i-1} = 1 \text{ or } C_i < 0.9 C_{i-1} | \mathbb{G}'_{i-1}) \geq \frac{1}{2}. \tag{8.12}$$

To see that (8.12) holds, note first that if $\mathbb{G}'_{i-1}$ is connected then there is nothing to prove. Otherwise let $\mathbb{H}_i = (U, F)$ be the graph obtained from $\mathbb{G}'_{i-1}$ by (i) contracting every connected component of $\mathbb{G}'_{i-1}$ to a single vertex and (ii) adding to it each edge $e' \in E'_i$ independently, with probability $p'_{e'}$ and throwing away loops. Note that since for every nontrivial S, $\mathbb{E} E'_S \geq k$, we have that for every vertex $u \in U$ (connected component of $\mathbb{G}'_{i-1}$),

$$\sum_{u \in e' \in F} p'_{e'} = \sum_{e \in U : U^c} \frac{p_e}{k} \geq 1.$$

Moreover, the probability that a fixed vertex $u \in U$ is isolated in $\mathbb{H}_i$ is

$$\prod_{u \in e' \in F} (1 - p'_{e'}) \leq \exp\left\{ - \sum_{u \in e' \in F} p'_{e'} \right\} \leq e^{-1}.$$

Hence the expected number of isolated vertices of $\mathbb{H}_i$ does not exceed $|U|e^{-1}$. Therefore, by the Markov inequality, it is at most $2|U|e^{-1}$ with probability at least 1/2. But in this case the number of connected components of $\mathbb{H}_i$ is at most

$$2|U|e^{-1}+\frac{1}{2}(|U|-2|U|e^{-1})=\left(\frac{1}{2}+e^{-1}\right)|U|<0.9|U|,$$

and so (8.12) follows. Observe that if $C_k>1$ then the total number of successful stages is strictly less than $\log n/\log 0.9<10\log n$. However, by (8.12), the probability of this event is at most the probability that a Binomial random variable with parameters k and $1/2$ will attain a value at most $10\log n$. It follows from (27.22) that if $k=c\log n=(20+t)\log n$ then the probability that $C_k>1$ (i.e., that $\mathbb{G}'_{p'_e}$ is disconnected) is at most $n^{-t^2/4c}$. This completes the proof of (8.11) and the theorem follows. □

8.2 Expected Degree Model

In this section we will consider a special case of Kovalenko's generalized binomial model, introduced by Chung and Lu in [220], where edge probabilities p_{ij} depend on weights assigned to vertices. This was also meant as a model for "Real World networks", see Chapter 18.

Let $V=\{1,2,\ldots,n\}$ and let w_i be the *weight* of vertex i. Now insert edges between vertices $i,j\in V$ independently with probability p_{ij} defined as

$$p_{ij}=\frac{w_iw_j}{W} \text{ where } W=\sum_{k=1}^n w_k.$$

We assume that $\max_i w_i^2<W$ so that $p_{ij}\le 1$. The resulting graph is denoted as $\mathbb{G}_{n,\mathbf{P^w}}$. Note that putting $w_i=np$ for $i\in[n]$ yields the random graph $\mathbb{G}_{n,p}$.

Notice that loops are allowed here but we will ignore them in what follows. Moreover, for vertex $i\in V$ its expected degree is

$$\sum_j \frac{w_iw_j}{W}=w_i.$$

Denote the average vertex weight by $\overline{w}$ (average expected vertex degree) i.e.,

$$\overline{w}=\frac{W}{n},$$

while, for any subset U of a vertex set V define the *volume* of U as

$$\mathrm{w}(U)=\sum_{k\in U} w_k.$$

Chung and Lu in [220] and [222] proved the following results summarized in the next theorem.

Theorem 8.4. *The random graph $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$ with a given expected degree sequence has a unique giant component w.h.p. if the average expected degree is strictly greater than one (i.e., $\overline{w} > 1$). Moreover, if $\overline{w} > 1$ then w.h.p. the giant component has volume*

$$\lambda_0 W + O\left(\sqrt{n}(\log n)^{3.5}\right),$$

where λ_0 is the unique nonzero root of the following equation

$$\sum_{i=1}^{n} w_i e^{-w_i \lambda} = (1-\lambda) \sum_{i=1}^{n} w_i,$$

Furthermore w.h.p., the second-largest component has size at most

$$(1+o(1)\mu(\overline{w}) \log n,$$

where

$$\mu(\overline{w}) = \begin{cases} 1/(\overline{w} - 1 - \log \overline{w}) & \text{if } 1 < \overline{w} < 2, \\ 1/(1 + \log \overline{w} - \log 4) & \text{if } \overline{w} > 4/e. \end{cases}$$

Here we will prove a weaker and restricted version of the above theorem. In the current context, a giant component is one with volume $\Omega(W)$.

Theorem 8.5. *If the average expected degree $\overline{w} > 4$, then a random graph $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$ w.h.p. has a unique giant component and its volume is at least*

$$\left(1 - \frac{2}{\sqrt{e\overline{w}}}\right) W$$

while the second-largest component w.h.p. has the size at most

$$(1+o(1)) \frac{\log n}{1 + \log \overline{w} - \log 4}.$$

The proof is based on a key lemma given below, proved under stronger conditions on $\overline{w}$ than in fact Theorem 8.5 requires.

Lemma 8.6. *For any positive $\varepsilon < 1$ and $\overline{w} > \frac{4}{e(1-\varepsilon)^2}$ w.h.p. every connected component in the random graph $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$ either has volume at least εW or has at most $\frac{\log n}{1+\log \overline{w} - \log 4 + 2\log(1-\varepsilon)}$ vertices.*

Proof. We first estimate the probability of the existence of a connected component with k vertices (component of *size* k) in the random graph $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$. Let $S \subseteq V$ and suppose that vertices from $S = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$ have respective weights $w_{i_1}, w_{i_2}, \ldots, w_{i_k}$. If the set S induces a connected subgraph of $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$ than it contains at least one spanning tree T. The probability of such event equals

$$\mathbb{P}(T) = \prod_{\{v_{i_j}, v_{i_l}\} \in E(T)} w_{i_j} w_{i_l} \rho,$$

where

$$\rho := \frac{1}{W} = \frac{1}{n\overline{w}}.$$

So, the probability that S induces a connected subgraph of our random graph can be bounded from above by

$$\sum_T \mathbb{P}(T) = \sum_T \prod_{\{v_{i_j}, v_{i_l}\} \in E(T)} w_{i_j} w_{i_l} \rho,$$

where T ranges over all spanning trees on S.
By the matrix-tree theorem (see West [852]) the above sum equals the determinant of any $k-1$ by $k-1$ principal sub-matrix of $(D-A)\rho$, where A is defined as

$$A = \begin{pmatrix} 0 & w_{i_1} w_{i_2} & \cdots & w_{i_1} w_{i_k} \\ w_{i_2} w_{i_1} & 0 & \cdots & w_{i_2} w_{i_k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i_k} w_{i_1} & w_{i_k} w_{i_2} & \cdots & 0 \end{pmatrix},$$

while D is the diagonal matrix

$$D = \operatorname{diag}\left(w_{i_1}(W - w_{i_1}), \ldots, w_{i_k}(W - w_{i_k})\right).$$

(To evaluate the determinant of the first principal co-factor of $D-A$, delete row and column k of $D-A$; Take out a factor $w_{i_1} w_{i_2} \cdots w_{i_{k-1}}$; Add the last $k-2$ rows to row 1; Row 1 is now $(w_{i_k}, w_{i_k}, \ldots, w_{i_k})$, so we can take out a factor w_{i_k}; Now subtract column 1 from the remaining columns to get a $(k-1) \times (k-1)$ upper triangular matrix with diagonal equal to $\operatorname{diag}(1, \mathrm{w}(S), \mathrm{w}(S), \ldots, \mathrm{w}(S))$).

It follows that

$$\sum_T \mathbb{P}(T) = w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{w}(S)^{k-2} \rho^{k-1}. \tag{8.13}$$

To show that this subgraph is in fact a component one has to multiply by the probability that there is no edge leaving S in $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$. Obviously, this probability equals $\prod_{v_i \in S, v_j \notin S} (1 - w_i w_j \rho)$ and can be bounded from above

$$\prod_{v_i \in S, v_j \in V \setminus S} (1 - w_i w_j \rho) \le e^{-\rho \mathrm{w}(S)(W - \mathrm{w}(S))}. \tag{8.14}$$

Let X_k be the number of components of size k in $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$. Then, using bounds from (8.13) and (8.14) we get

$$\mathbb{E}X_k \le \sum_S \mathrm{w}(S)^{k-2}\rho^{k-1}e^{-\rho\mathrm{w}(S)(W-\mathrm{w}(S))}\prod_{i\in S}w_i,$$

where the sum ranges over all $S \subseteq V$, $|S| = k$. Now, we focus our attention on k-vertex components whose volume is at most εW. We call such components *small* or ε-*small*. So, if Y_k is the number of small components of size k in $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$ then

$$\mathbb{E}Y_k \le \sum_{\text{small } S} \mathrm{w}(S)^{k-2}\rho^{k-1}e^{-\mathrm{w}(S)(1-\varepsilon)}\prod_{i\in S}w_i = f(k). \tag{8.15}$$

Now, using the arithmetic-geometric mean inequality, we have

$$f(k) \le \sum_{\text{small } S}\left(\frac{\mathrm{w}(S)}{k}\right)^k \mathrm{w}(S)^{k-2}\rho^{k-1}e^{-\mathrm{w}(S)(1-\varepsilon)}.$$

The function $x^{2k-2}e^{-x(1-\varepsilon)}$ achieves its maximum at $x = (2k-2)/(1-\varepsilon)$. Therefore

$$\begin{aligned}
f(k) &\le \binom{n}{k}\frac{\rho^{k-1}}{k^k}\left(\frac{2k-2}{1-\varepsilon}\right)^{2k-2}e^{-(2k-2)} \\
&\le \left(\frac{ne}{k}\right)^k\frac{\rho^{k-1}}{k^k}\left(\frac{2k-2}{1-\varepsilon}\right)^{2k-2}e^{-(2k-2)} \\
&\le \frac{(n\rho)^k}{4\rho(k-1)^2}\left(\frac{2}{1-\varepsilon}\right)^{2k}e^{-k} \\
&= \frac{1}{4\rho(k-1)^2}\left(\frac{4}{e\overline{w}(1-\varepsilon)^2}\right)^k \\
&= \frac{e^{-ak}}{4\rho(k-1)^2},
\end{aligned}$$

where

$$a = 1 + \log\overline{w} - \log 4 + 2\log(1-\varepsilon) > 0$$

under the assumption of Lemma 8.6.

Let $k_0 = \frac{\log n}{a}$. When k satisfies $k_0 < k < 2k_0$ we have

$$f(k) \le \frac{1}{4n\rho(k-1)^2} = o\left(\frac{1}{\log n}\right),$$

while, when $\frac{2\log n}{a} \leq k \leq n$, we have

$$f(k) \leq \frac{1}{4n^2\rho(k-1)^2} = o\left(\frac{1}{n\log n}\right).$$

So, the probability that there exists an ε-small component of size exceeding k_0 is at most

$$\sum_{k>k_0} f(k) \leq \frac{\log n}{a} \times o\left(\frac{1}{\log n}\right) + n \times o\left(\frac{1}{n\log n}\right) = o(1).$$

This completes the proof of Lemma 8.6. □

To prove Theorem 8.5 assume that for some fixed $\delta > 0$ we have

$$\overline{w} = 4 + \delta = \frac{4}{e(1-\varepsilon)^2} \text{ where } \varepsilon = 1 - \frac{2}{(e\overline{w})^{1/2}} \tag{8.16}$$

and suppose that $w_1 \geq w_2 \geq \cdots \geq w_n$. We show next that there exists $i_0 \geq n^{1/3}$ such that

$$w_{i_0} \geq \sqrt{\frac{\left(1+\frac{\delta}{8}\right)W}{i_0}}. \tag{8.17}$$

Suppose the contrary, i.e., for all $i \geq n^{1/3}$,

$$w_i < \sqrt{\frac{\left(1+\frac{\delta}{8}\right)W}{i}}.$$

Then

$$\begin{aligned} W &\leq n^{1/3}W^{1/2} + \sum_{i=n^{1/3}}^{n} \sqrt{\frac{\left(1+\frac{\delta}{8}\right)W}{i}} \\ &\leq n^{1/3}W^{1/2} + 2\sqrt{\left(1+\frac{\delta}{8}\right)Wn}. \end{aligned}$$

Hence

$$W^{1/2} \leq n^{1/3} + 2\left(1+\frac{\delta}{8}\right)n^{1/2}.$$

This is a contradiction since for our choice of $\overline{w}$

$$W = n\overline{w} \geq 4(1+\delta)n.$$

We have therefore verified the existence of i_0 satisfying (8.17).

Now consider the subgraph G of $\mathbb{G}_{n,\mathbf{P^w}}$ on the first i_0 vertices. The probability that there is an edge between vertices v_i and v_j, for any $i, j \leq i_0$, is at least

$$w_i w_j \rho \geq w_{i_0}^2 \rho \geq \frac{1+\frac{\delta}{8}}{i_0}.$$

So the asymptotic behavior of G can be approximated by a random graph $\mathbb{G}_{n,p}$ with $n = i_0$ and $p > 1/i_0$. So, w.h.p. G has a component of size $\Theta(i_0) = \Omega(n^{1/3})$. Applying Lemma 8.6 with ε as in (8.16) we see that any component with size $\gg \log n$ has volume at least εW.

Finally, consider the volume of a giant component. Suppose first that there exists a giant component of volume cW which is ε-small i.e. $c \leq \varepsilon$. By Lemma 8.6, the size of the giant component is then at most $\frac{\log n}{2\log 2}$. Hence, there must be at least one vertex with weight w greater than or equal to the average

$$w \geq \frac{2cW \log 2}{\log n}.$$

But it implies that $w^2 \gg W$, which contradicts the general assumption that all $p_{ij} < 1$.

We now prove uniqueness in the same way that we proved the uniqueness of the giant component in $G_{n,p}$. Choose $\eta > 0$ such that $\overline{w}(1-\eta) > 4$. Then define $w_i' = (1-\eta)w_i$ and decompose

$$\mathbb{G}_{n,\mathbf{P^w}} = G_1 \cup G_2$$

where the edge probability in G_1 is $p_{ij}' = \frac{w_i' w_j'}{(1-\eta)W}$ and the edge probability in G_2 is p_{ij}'' where $1 - \frac{w_i w_j}{W} = (1-p_{i,j}')(1-p_{ij}'')$. Simple algebra gives $p_{ij}'' \geq \frac{\eta w_i w_j}{W}$. It follows from the previous analysis that G_1 contains between one and $1/\varepsilon$ giant components. Let C_1, C_2 be two such components. The probability that there is no G_2 edge between them is at most

$$\prod_{\substack{i \in C_1 \\ j \in C_2}} \left(1 - \frac{\eta w_i w_j}{W}\right) \leq \exp\left\{-\frac{\eta w(C_1) w(C_2)}{W}\right\} \leq e^{-\eta W} = o(1).$$

As $1/\varepsilon < 4$, this completes the proof of Theorem 8.5. $\square$

To add to the picture of the asymptotic behavior of the random graph $\mathbb{G}_{n,\mathbf{P^w}}$ we will present one more result from [220]. Denote by $\overline{w^2}$ the expected second-order average degree, i.e.,

$$\overline{w^2} = \sum_j \frac{w_j^2}{W}.$$

Notice that

$$\overline{w^2} = \frac{\sum_j w_j^2}{W} \geq \frac{W}{n} = \overline{w}.$$

Chung and Lu [220] proved the following.

Theorem 8.7. *If the average expected square degree* $\overline{w^2} < 1$ *then, with probability at least* $1 - \frac{\overline{w}\left(\overline{w^2}\right)^2}{C^2\left(1-\overline{w^2}\right)}$, *all components of* $\mathbb{G}_{n,\mathbf{P^w}}$ *have volume at most* $C\sqrt{n}$.

Proof. Let

$$x = \mathbb{P}(\exists S : w(S) \geq Cn^{1/2}\text{and } S \text{ is a component}).$$

Randomly choose two vertices u and v from V, each with probability proportional to its weight. Then, for each vertex, the probability that it is in a set S with $w(S) \geq C\sqrt{n}$ is at least $C\sqrt{n}\rho$. Hence the probability that both vertices are in the same component is at least

$$x(C\sqrt{n}\rho)^2 = C^2 xn\rho^2. \tag{8.18}$$

On the other hand, for any two fixed vertices, say u and v, the probability $P_k(u,v)$ of u and v being connected via a path of length $k+1$ can be bounded from above as follows

$$P_k(u,v) \leq \sum_{i_1,i_2,\dots,i_k} (w_u w_{i_1}\rho)(w_{i_1} w_{i_2}\rho)\cdots(w_{i_k} w_v\rho) \leq w_u w_v \rho (\overline{w^2})^k.$$

So the probability that u and v belong to the same component is at most

$$\sum_{k=0}^{n} P_k(u,v) \leq \sum_{k=0}^{\infty} w_u w_v \rho (\overline{w^2})^k = \frac{w_u w_v \rho}{1-\overline{w^2}}.$$

Recall that the probabilities of u and v being chosen from V are $w_u\rho$ and $w_v\rho$, respectively. so the probability that a random pair of vertices are in the same component is at most

$$\sum_{u,v} w_u\rho\ w_v\rho\ \frac{w_u w_v \rho}{1-\overline{w^2}} = \frac{\left(\overline{w^2}\right)^2 \rho}{1-\overline{w^2}}.$$

Combining this with (8.18) we have

$$C^2 xn\rho^2 \leq \frac{\left(\overline{w^2}\right)^2 \rho}{1-\overline{w^2}},$$

which implies

$$x \leq \frac{\overline{w}\left(\overline{w^2}\right)^2}{C^2\left(1-\overline{w^2}\right)},$$

and Theorem 8.7 follows. □

8.3 Kronecker Graphs

Kronecker random graphs were introduced by Leskovec, Chakrabarti, Kleinberg and Faloutsos [622] (see also [621]). It is meant as a model of "Real World networks", see Chapter 18. Here we consider a special case of this model of an inhomogeneous random graph. To construct it we begin with a *seed* matrix

$$\mathbf{P} = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix},$$

where $0 < \alpha, \beta, \gamma < 1$, and let $\mathbf{P}^{[k]}$ be the kth Kronecker power of $\mathbf{P}$. Here $\mathbf{P}^{[k]}$ is obtained from $\mathbf{P}^{[k-1]}$ as in the diagram below:

$$\mathbf{P}^{[k]} = \begin{bmatrix} \alpha\mathbf{P}^{[k-1]} & \beta\mathbf{P}^{[k-1]} \\ \beta\mathbf{P}^{[k-1]} & \gamma\mathbf{P}^{[k-1]} \end{bmatrix}$$

and so for example

$$\mathbf{P}^{[2]} = \begin{bmatrix} \alpha^2 & \alpha\beta & \beta\alpha & \beta^2 \\ \alpha\beta & \alpha\gamma & \beta^2 & \beta\gamma \\ \beta\alpha & \beta^2 & \gamma\alpha & \gamma\beta \\ \beta^2 & \beta\gamma & \gamma\beta & \gamma^2 \end{bmatrix}.$$

Note that $\mathbf{P}^{[k]}$ is symmetric and has size $2^k \times 2^k$.

We define a *Kronecker random graph* as a copy of $\mathbb{G}_{n,\mathbf{P}^{[k]}}$ for some $k \geq 1$ and $n = 2^k$. Thus each vertex is a binary string of length k, and between any two such vertices (strings) $\mathbf{u}, \mathbf{v}$ we put an edge independently with probability

$$p_{\mathbf{u},\mathbf{v}} = \alpha^{\mathbf{uv}}\gamma^{(\mathbf{1}-\mathbf{u})(\mathbf{1}-\mathbf{v})}\beta^{k-\mathbf{uv}-(\mathbf{1}-\mathbf{u})(\mathbf{1}-\mathbf{v})},$$

or equivalently

$$p_{\mathbf{uv}} = \alpha^i\beta^j\gamma^{k-i-j},$$

where i is the number of positions t such that $u_t = v_t = 1$, j is the number of t where $u_t \neq v_t$ and hence $k-i-j$ is the number of t that $u_t = v_t = 0$. We observe that when $\alpha = \beta = \gamma$ then $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ becomes $\mathbb{G}_{n,p}$ with $n = 2^k$ and $p = \alpha^k$.

Connectivity

We will first examine, following Mahdian and Xu [647], conditions under which is $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ connected w.h.p.

Theorem 8.8. *Suppose that $\alpha \geq \beta \geq \gamma$. The random graph $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ is connected w.h.p. (for $k \to \infty$) if and only if either (i) $\beta + \gamma > 1$ or (ii) $\alpha = \beta = 1, \gamma = 0$.*

Proof. We show first that $\beta + \gamma \geq 1$ is a necessary condition. Denote by $\mathbf{0}$ the vertex with all 0's. Then the expected degree of vertex $\mathbf{0}$ is

$$\sum_{\mathbf{v}} p_{\mathbf{0v}} = \sum_{j=0}^{k} \binom{k}{j} \beta^j \gamma^{k-j} = (\beta + \gamma)^k = o(1), \quad \text{when } \beta + \gamma < 1.$$

Thus in this case vertex $\mathbf{0}$ is isolated w.h.p.

Moreover, if $\beta + \gamma = 1$ and $0 < \beta < 1$ then $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ cannot be connected w.h.p. since the probability that vertex $\mathbf{0}$ is isolated is bounded away from 0. Indeed, $0 < \beta < 1$ implies that $\beta^j \gamma^{k-j} \leq \zeta < 1, 0 \leq j \leq k$ for some absolute constant ζ. Thus, using Lemma 27.1(b),

$$\begin{aligned}\mathbb{P}(\mathbf{0} \text{ is isolated}) = \prod_{\mathbf{v}} (1 - p_{\mathbf{0v}}) &\geq \prod_{j=0}^{k} \left(1 - \beta^j \gamma^{k-j}\right)^{\binom{k}{j}} \\ &\geq \exp\left\{-\sum_{j=0}^{k} \frac{\binom{k}{j} \beta^j \gamma^{k-j}}{1 - \zeta}\right\} = e^{-1/\zeta}.\end{aligned}$$

Now when $\alpha = \beta = 1, \gamma = 0$, the vertex with all 1's has degree $n - 1$ with probability one and so $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ will be connected w.h.p. in this case.

It remains to show that the condition $\beta + \gamma > 1$ is also sufficient. To show that $\beta + \gamma > 1$ implies connectivity we will apply Theorem 8.3. Notice that the expected degree of vertex $\mathbf{0}$, excluding its self-loop, given that β and γ are constants independent of k and $\beta + \gamma > 1$, is

$$(\beta + \gamma)^k - \gamma^k \geq 2c \log n,$$

for some constant $c > 0$, which can be as large as needed.

Therefore the cut $(\mathbf{0}, V \setminus \{\mathbf{0}\})$ has weight at least $2c \log n$. Remove vertex $\mathbf{0}$ and consider any cut $(S, V \setminus S)$. Then at least one side of the cut gets at least half of the weight of vertex $\mathbf{0}$. Without loss of generality assume that it is S, i.e.,

$$\sum_{\mathbf{u} \in S} p_{\mathbf{0u}} \geq c \log n.$$

Take any vertices $\mathbf{u},\mathbf{v}$ and note that $p_{\mathbf{uv}} \geq p_{\mathbf{u0}}$ because we have assumed that $\alpha \geq \beta \geq \gamma$. Therefore

$$\sum_{\mathbf{u}\in S}\sum_{\mathbf{v}\in V\setminus S} p_{\mathbf{uv}} \geq \sum_{\mathbf{u}\in S} p_{\mathbf{u0}} > c\log n,$$

and so the claim follows by Theorem 8.3. $\square$

To add to the picture of the structure of $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ when $\beta+\gamma>1$ we state (without proof) the following result on the diameter of $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$.

Theorem 8.9. *If $\beta+\gamma>1$ then w.h.p. $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ has constant diameter.*

Giant Component

We now consider when $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ has a giant component (see Horn and Radcliffe [489]).

Theorem 8.10. *$\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ has a giant component of order $\Theta(n)$ w.h.p., if and only if $(\alpha+\beta)(\beta+\gamma)>1$.*

Proof. We prove a weaker version of the Theorem 8.10, assuming that for $\alpha \geq \beta \geq \gamma$ as in [647]. For the proof of the more general case, see [489].

We will show first that the above condition is necessary. We prove that if

$$(\alpha+\beta)(\beta+\gamma) \leq 1,$$

then w.h.p. $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ has $n-o(n)$ isolated vertices. First let

$$(\alpha+\beta)(\beta+\gamma) = 1-\varepsilon,\ \varepsilon>0.$$

First consider those vertices with weight (counted as the number of 1's in their label) less than $k/2+k^{2/3}$ and let $X_{\mathbf{u}}$ be the degree of a vertex $\mathbf{u}$ with weight l where $l=0,\ldots,k$. It is easily observed that

$$\mathbb{E}X_{\mathbf{u}} = (\alpha+\beta)^l(\beta+\gamma)^{k-l}. \tag{8.19}$$

Indeed, if for vertex $\mathbf{v}$, $i=i(\mathbf{v})$ is the number of bits that $u_r=v_r=1,\ r=1,\ldots,k$ and $j=j(\mathbf{v})$ is the number of bits where $u_r=0$ and $v_r=1$, then the probability of an edge between $\mathbf{u}$ and $\mathbf{v}$ equals

$$p_{\mathbf{uv}} = \alpha^i\beta^{j+l-i}\gamma^{k-l-j}.$$

Hence,

$$\begin{aligned}
\mathbb{E}X_{\mathbf{u}} = \sum_{\mathbf{v}\in V} p_{\mathbf{uv}} &= \sum_{i=0}^{l}\sum_{j=0}^{k-l}\binom{l}{i}\binom{k-l}{j}\alpha^i\beta^{j+l-i}\gamma^{k-l-j} \\
&= \sum_{i=0}^{l}\binom{l}{i}\alpha^i\beta^{l-i}\sum_{j=0}^{k-l}\beta^j\gamma^{k-l-l}
\end{aligned}$$

and (8.19) follows. So, if $l < k/2 + k^{2/3}$, then assuming that $\alpha \ge \beta \ge \gamma$,

$$\begin{aligned}
\mathbb{E}X_{\mathbf{u}} \le &(\alpha+\beta)^{k/2+k^{2/3}}(\beta+\gamma)^{k-(k/2+k^{2/3})} \\
= &((\alpha+\beta)(\beta+\gamma))^{k/2}\left(\frac{\alpha+\beta}{\beta+\gamma}\right)^{k^{2/3}} \\
= &(1-\varepsilon)^{k/2}\left(\frac{\alpha+\beta}{\beta+\gamma}\right)^{k^{2/3}} \\
= &o(1). \qquad (8.20)
\end{aligned}$$

Suppose now that $l \ge k/2 + k^{2/3}$ and let Y be the number of 1's in the label of a randomly chosen vertex of $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$. Since $\mathbb{E}Y = k/2$, the Chernoff bound (see (27.26)) implies that

$$\mathbb{P}\left(Y \ge \frac{k}{2} + k^{2/3}\right) \le e^{-k^{4/3}/(3k/2)} \le e^{-k^{1/3}/2} = o(1).$$

Therefore, there are $o(n)$ vertices with $l \ge k/2 + k^{2/3}$. It then follows from (8.20) that the expected number of non-isolated vertices in $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ is $o(n)$ and the Markov inequality then implies that this number is $o(n)$ w.h.p.

Next, when $\alpha+\beta = \beta+\gamma = 1$, which implies that $\alpha = \beta = \gamma = 1/2$, then random graph $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ is equivalent to $\mathbb{G}_{n,p}$ with $p = 1/n$ and so by Theorem 2.21 it does not have a component of order n, w.h.p.

To prove that the condition $(\alpha+\beta)(\beta+\gamma) > 1$ is sufficient we show that the subgraph of $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ induced by the vertices of H of weight $l \ge k/2$ is connected w.h.p. This will suffice as there are at least $n/2$ such vertices. Notice that for any vertex $\mathbf{u} \in H$ its expected degree, by (8.19), is at least

$$((\alpha+\beta)(\beta+\gamma))^{k/2} \gg \log n. \qquad (8.21)$$

We first show that for $\mathbf{u} \in V$,

$$\sum_{\mathbf{v}\in H} p_{\mathbf{uv}} \ge \frac{1}{4}\sum_{\mathbf{v}\in V} p_{\mathbf{uv}}. \qquad (8.22)$$

For the given vertex $\mathbf{u}$ let l be the weight of $\mathbf{u}$. For a vertex $\mathbf{v}$ let $i(\mathbf{v})$ be the number of bits where $u_r = v_r = 1,\ r = 1,\dots,k$, while $j(\mathbf{v})$ stands for the number of bits where $u_r = 0$ and $v_r = 1$. Consider the partition

$$V \setminus H = S_1 \cup S_2 \cup S_3,$$

where

$$S_1 = \{\mathbf{v} : i(\mathbf{v}) \geq l/2,\ j(\mathbf{v}) < (k-l)/2\},$$

$$S_2 = \{\mathbf{v} : i(\mathbf{v}) < l/2,\ j(\mathbf{v}) \geq (k-l)/2\},$$

$$S_3 = \{\mathbf{v} : i(\mathbf{v}) < l/2,\ j(\mathbf{v}) < (k-l)/2\}.$$

Next, take a vertex $\mathbf{v} \in S_1$ and turn it into $\mathbf{v}'$ by flipping the bits of $\mathbf{v}$ which correspond to 0's of $\mathbf{u}$. Surely, $i(\mathbf{v}') = i(\mathbf{v})$ and

$$j(\mathbf{v}') \geq (k-l)/2 > j(\mathbf{v}).$$

Notice that the weight of $\mathbf{v}'$ is at least $k/2$ and so $\mathbf{v}' \in H$. Notice also that $\alpha \geq \beta \geq \gamma$ implies that $p_{\mathbf{uv}'} \geq p_{\mathbf{uv}}$. Different vertices $\mathbf{v} \in S_1$ map to different $\mathbf{v}'$. Hence

$$\sum_{\mathbf{v}\in H} p_{\mathbf{uv}} \geq \sum_{\mathbf{v}\in S_1} p_{\mathbf{uv}}. \tag{8.23}$$

The same bound (8.23) holds for S_2 and S_3 in place of S_1. To prove the same relationship for S_2 one has to flip the bits of $\mathbf{v}$ corresponding to 1's in $\mathbf{u}$, while for S_3 one has to flip all the bits of $\mathbf{v}$. Adding up these bounds over the partition of $V \setminus H$ we get

$$\sum_{\mathbf{v}\in V\setminus H} p_{\mathbf{uv}} \leq 3 \sum_{\mathbf{v}\in H} p_{\mathbf{uv}}$$

and so the bound (8.22) follows.
Notice that combining (8.22) with the bound given in (8.21) we get that for $\mathbf{u} \in H$ we have

$$\sum_{\mathbf{v}\in H} p_{\mathbf{uv}} > 2c\log n, \tag{8.24}$$

where c can be a large as needed.

We finish the proof by showing that a subgraph of $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$ induced by vertex set H is connected w.h.p. For that we make use of Theorem 8.3. So, we will show that for any cut $(S, H \setminus S)$

$$\sum_{\mathbf{u}\in S}\sum_{\mathbf{v}\in H\setminus S} p_{\mathbf{uv}} \geq 10\log n.$$

Without loss of generality assume that vertex $\mathbf{1} \in S$. Equation (8.24) implies that for any vertex $\mathbf{u} \in H$ either

$$\sum_{\mathbf{v} \in S} p_{\mathbf{uv}} \geq c \log n, \tag{8.25}$$

or

$$\sum_{\mathbf{v} \in H \setminus S} p_{\mathbf{uv}} \geq c \log n. \tag{8.26}$$

If there is a vertex $\mathbf{u}$ such that (8.26) holds then since $\mathbf{u} \leq \mathbf{1}$ and $\alpha \geq \beta \geq \gamma$,

$$\sum_{\mathbf{u} \in S} \sum_{\mathbf{v} \in H \setminus S} p_{\mathbf{uv}} \geq \sum_{\mathbf{v} \in H \setminus S} p_{\mathbf{1v}} \geq \sum_{\mathbf{v} \in H \setminus S} p_{\mathbf{uv}} > c \log n.$$

Otherwise, (8.25) is true for every vertex $\mathbf{u} \in H$. Since at least one such vertex is in $H \setminus S$, we have

$$\sum_{\mathbf{u} \in S} \sum_{\mathbf{v} \in H \setminus S} p_{\mathbf{uv}} \geq c \log n,$$

and the Theorem follows. $\square$

8.4 Exercises

8.4.1 Prove Theorem 8.3 (with $c = 10$) using the result of Karger and Stein [546] that in any weighted graph on n vertices the number of r-minimal cuts is $O\left((2n)^{2r}\right)$. (A cut $(S, V \setminus S), S \subseteq V$, in a weighted graph G is called *r-minimal* if its weight, i.e., the sum of weights of the edges connecting S with $V \setminus S$, is at most r times the weight of minimal weighted cut of G).

8.4.2 Suppose that the entries of an $n \times n$ symmetric matrix A are all non-negative. Show that for any positive constants $c_1, c_2, \ldots, c_n$, the largest eigenvalue $\lambda(A)$ satisfies

$$\lambda(A) \leq \max_{1 \leq i \leq n} \left(\frac{1}{c_i} \sum_{j=1}^{n} c_j a_{i,j} \right).$$

8.4.3 Let A be the adjacency matrix of $\mathbb{G}_{n,\mathbf{P}^{\mathbf{w}}}$ and for a fixed value of x let

$$c_i = \begin{cases} w_i & w_i > x \\ x & w_i \leq x \end{cases}.$$

Let $m = \max\{w_i : i \in [n]\}$. Let $X_i = \frac{1}{c_i} \sum_{j=1}^{n} c_j a_{i,j}$. Show that

$$\mathbb{E} X_i \leq \overline{w^2} + x \text{ and } \operatorname{Var} X_i \leq \frac{m}{x} \overline{w^2} + x.$$

8.4.4 Apply Theorem 27.11 with a suitable value of x to show that w.h.p.

$$\lambda(A) \leq \overline{w^2} + (6(m\log n)^{1/2}(\overline{w^2} + \log n))^{1/2} + 3(m\log n)^{1/2}.$$

8.4.5 Show that if $\overline{w^2} > m^{1/2}\log n$ then w.h.p. $\lambda(A) = (1+o(1))\overline{w^2}$.

8.4.6 Suppose that $1 \leq w_i \ll W^{1/2}$ for $1 \leq i \leq n$ and that $w_i w_j \overline{w^2} \gg W \log n$. Show that w.h.p. $diameter(\mathbb{G}_{n,\mathbf{P^w}}) \leq 2$.

8.4.7 Prove, by the Second Moment Method, that if $\alpha+\beta = \beta+\gamma = 1$ then w.h.p. the number Z_d of the vertices of degree d in the random graph $\mathbb{G}_{n,\mathbf{P^{[k]}}}$, is concentrated around its mean, i.e., $Z_d = (1+o(1))\,\mathbb{E}Z_d$.

8.4.8 Fix $d \in \mathbb{N}$ and let Z_d denote the number of vertices of degree d in the Kronecker random graph $\mathbb{G}_{n,\mathbf{P^{[k]}}}$. Show that

$$\mathbb{E}Z_d = (1+o(1))\sum_{w=0}^{k}\binom{k}{w}\frac{(\alpha+\beta)^{dw}(\beta+\gamma)^{d(k-w)}}{d!}\times$$
$$\times \exp\left(-(\alpha+\beta)^{w}(\beta+\gamma)^{k-w}\right) + o(1).$$

8.4.9 Depending on the configuration of the parameters $0 < \alpha, \beta, \gamma < 1$, show that we have either

$$\mathbb{E}Z_d = \Theta\left(\left((\alpha+\beta)^d + (\beta+\gamma)^d\right)^k\right),$$

or

$$\mathbb{E}Z_d = o(2^k).$$

8.5 Notes

General model of inhomogeneous random graph

The most general model of inhomogeneous random graph was introduced by Bollobás, Janson and Riordan in their seminal paper [168]. They concentrate on the study of the phase transition phenomenon of their random graphs, which includes as special cases the models presented in this chapter as well as, among others, Dubins' model (see Kalikow and Weiss [540] and Durrett [315]), the mean-field scale-free model (see Riordan [762]), the CHKNS model (see Callaway, Hopcroft, Kleinberg, Newman and Strogatz [207]) and Turova's model (see [833], [834] and [835]).

The model of Bollobás, Janson and Riordan is an extension of one defined by Söderberg [809]. The formal description of their model goes as follows. Consider a *ground space* being a pair $(\mathscr{S},\mu)$, where $\mathscr{S}$ is a separable metric space and μ is a Borel probability measure on $\mathscr{S}$. Let $\mathscr{V}=(\mathscr{S},\mu,(\mathbf{x}_n)_{n\geq 1})$ be the *vertex space*, where $(\mathscr{S},\mu)$ is a ground space and $(\mathbf{x}_n)_{n\geq 1})$ is a random sequence $(x_1,x_2,\ldots,x_n)$ of n points of $\mathscr{S}$ satisfying the condition that for every μ-continuity set A, $A\subseteq\mathscr{S}$, $|\{i:x_i\in A\}|/n$ converges in probability to $\mu(A)$. Finally, let κ be a kernel on the vertex space $\mathscr{V}$ (understood here as a kernel on a ground space $(\mathscr{S},\mu)$), i.e., a symmetric non-negative (Borel) measurable function on $S\times S$. Given the (random) sequence $(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$ we let $G_{\mathscr{V}}(n,\kappa)$ be the random graph $G_{\mathscr{V}}(n,(p_{ij}))$ with $p_{ij}:=\min\{\kappa(\mathbf{x}_i,\mathbf{x}_j)/n,1\}$. In other words, $G_{\mathscr{V}}(n,\kappa)$ has n vertices and, given $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$, an edge $\{i,j\}$ (with $i\neq j$) exists with probability p_{ij}, independently for all other unordered pairs $\{i,j\}$.

Bollobás, Janson and Riordan present in [168] a wide range of results describing various properties of the random graph $G_{\mathscr{V}}(n,\kappa)$. They give a necessary and sufficient condition for the existence of a giant component, show its uniqueness and determine the asymptotic number of edges in the giant component. They also study the stability of the component, i.e., they show that its size does not change much if we add or delete a few edges. They also establish bounds on the size of small components, the asymptotic distribution of the number of vertices of given degree and study the distances between vertices (diameter). Finally they turn their attention to the phase transition of $G_{\mathscr{V}}(n,\kappa)$ where the giant component first emerges.

Janson and Riordan [511] study the susceptibility, i.e., the mean size of the component containing a random vertex, in a general model of inhomogeneous random graphs. They relate the susceptibility to a quantity associated to a corresponding branching process, and study both quantities in various examples.

Devroye and Fraiman [287] find conditions for the connectivity of inhomogeneous random graphs with intermediate density. They draw n independent points X_i from a general distribution on a separable metric space, and let their indices form the vertex set of a graph. An edge ij is added with probability $\min\{1,\kappa(X_i,X_j)\log n/n\}$, where $\kappa>0$ is a fixed kernel. They show that, under reasonably weak assumptions, the connectivity threshold of the model can be determined.

Lin and Reinert [625] show via a multivariate normal and a Poisson process approximation that, for graphs which have independent edges, with a possibly inhomogeneous distribution, only when the degrees are large can we reasonably approximate the joint counts for the number of vertices with given degrees as independent (note that in a random graph, such counts will typically be dependent). The proofs are based on Stein's method and the Stein–Chen method (see Chapter 26.3) with a new size-biased coupling for such inhomogeneous random graphs.

Rank one model

An important special case of the general model of Bollobás, Janson and Riordan is the so called *rank one model*, where the kernel κ has the form $\kappa(x,y) = \psi(x)\psi(y)$, for some function $\psi > 0$ on $\mathscr{S}$. In particular, this model includes the Chung-Lu model (*expected degree model*) discussed earlier in this Chapter. Recall that in their approach we attach edges (independently) with probabilities

$$p_{ij} = \min\left\{\frac{w_i w_j}{W}, 1\right\} \text{ where } W = \sum_{k=1}^{n} w_k.$$

Similarly, Britton, Deijfen and Martin-Löf [197] define edge probabilities as

$$p_{ij} = \frac{w_i w_j}{W + w_i w_j},$$

while Norros and Reittu [719] attach edges with probabilities

$$p_{ij} = \exp\left(-\frac{w_i w_j}{W}\right).$$

For those models several characteristics are studied, such as the size of the giant component ([221], [222] and [719]) and its volume ([221]) as well as spectral properties ([225] and [226]). It should be also mentioned here that Janson [504] established conditions under which all those models are asymptotically equivalent.

Recently, van der Hofstad [482], Bhamidi, van der Hofstad and Hooghiemstra[109], van der Hofstad, Kliem and van Leeuwaarden [484] and Bhamidi, Sen and Wang [110] undertake systematic and detailed studies of various aspects of the rank one model in its general setting.

Finally, consider *random dot product graphs* (see Young and Scheinerman [864]) where to each vertex a vector in $\mathbb{R}^d$ is assigned and we allow each edge to be present with probability proportional to the inner product of the vectors assigned to its endpoints. The paper [864] treats these as models of social networks.

Kronecker Random Graph

Radcliffe and Young [757] analysed the connectivity and the size of the giant component in a generalized version of the Kronecker random graph. Their results imply that the threshold for connectivity in $\mathbb{G}_{n,\mathbf{P}[\mathbf{k}]}$ is $\beta+\gamma = 1$. Tabor [827] proved that it is also the threshold for a k-factor. Kang, Karoński, Koch and Makai [543] studied the asymptotic distribution of small subgraphs (trees and cycles) in $\mathbb{G}_{n,\mathbf{P}[\mathbf{k}]}$.

Leskovec, Chakrabarti, Kleinberg and Faloutsos [623] and [624] have shown empirically that Kronecker random graphs resemble several real world networks.

Later, Leskovec, Chakrabarti, Kleinberg, Faloutsos and Ghahramani [624] fitted the model to several real world networks such as the Internet, citation graphs and online social networks.

The R-MAT model, introduced by Chakrabarti, Zhan and Faloutsos [211], is closely related to the Kronecker random graph. The vertex set of this model is also $\mathbb{Z}_2^n$ and one also has parameters α, β, γ. However, in this case one needs the additional condition that $\alpha + 2\beta + \gamma = 1$.

The process of generating a random graph in the R-MAT model creates a multigraph with m edges and then merges the multiple edges. The advantage of the R-MAT model over the random Kronecker graph is that it can be generated significantly faster when m is small. The degree sequence of this model has been studied by Groër, Sullivan and Poole [454] and by Seshadhri, Pinar and Kolda [798] when $m = \Theta(2^n)$, i.e. the number of edges is linear in the number of vertices. They have shown, as in Kang, Karoński, Koch and Makai [543] for $\mathbb{G}_{n,\mathbf{P}^{[\mathbf{k}]}}$, that the degree sequence of the model does not follow a power law distribution. However, no rigorous proof exists for the equivalence of the two models and in the Kronecker random graph there is no restriction on the sum of the values of α, β, γ.

Further extensions of Kronecker random graphs can be found [130] and [131].

Chapter 9

Fixed Degree Sequence

The graph $\mathbb{G}_{n,m}$ is chosen uniformly at random from the set of graphs with vertex set $[n]$ and m edges. It is of great interest to refine this model so that all the graphs chosen have a fixed degree sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$. Of particular interest is the case where $d_1 = d_2 = \cdots = d_n = r$, i.e., the graph chosen is a uniformly random r-regular graph. It is not obvious how to do this and this is the subject of the current chapter. We discuss the configuration model in the next section and show its usefulness in (i) estimating the number of graphs with a given degree sequence and (ii) showing that w.h.p. random d-regular graphs are connected w.h.p., for $3 \le d = O(1)$.

We finish by showing in Section 9.5 how for large r, $\mathbb{G}_{n,m}$ can be embedded in a random r-regular graph. This allows one to extend some results for $\mathbb{G}_{n,m}$ to the regular case.

9.1 Configuration Model

Let $\mathbf{d} = (d_1, d_2, \dots, d_n)$ where $d_1 + d_2 + \cdots + d_n = 2m$ is even. Let

$$\mathscr{G}_{n,\mathbf{d}} = \{\text{simple graphs with vertex set } [n] \text{ s.t. degree } d(i) = d_i,\ i \in [n]\}$$

and let $\mathbb{G}_{n,\mathbf{d}}$ be chosen randomly from $\mathscr{G}_{n,\mathbf{d}}$. We assume that

$$d_1, d_2, \dots, d_n \ge 1 \text{ and } \sum_{i=1}^{n} d_i(d_i - 1) = \Omega(n).$$

We describe a generative model of $\mathbb{G}_{n,\mathbf{d}}$ due to Bollobás [146]. It is referred to as the *configuration model*. Let $W_1, W_2, \dots, W_n$ be a partition of a set of *points* W, where $|W_i| = d_i$ for $1 \le i \le n$ and call the W_i's *cells*. We will assume some total order $<$ on W and that $x < y$ if $x \in W_i, y \in W_j$ where $i < j$. For $x \in W$ define $\varphi(x)$

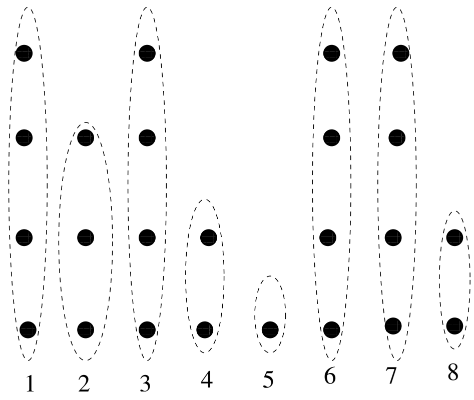


Figure 9.1: Partition of W into cells $W_1, \dots, W_8$.

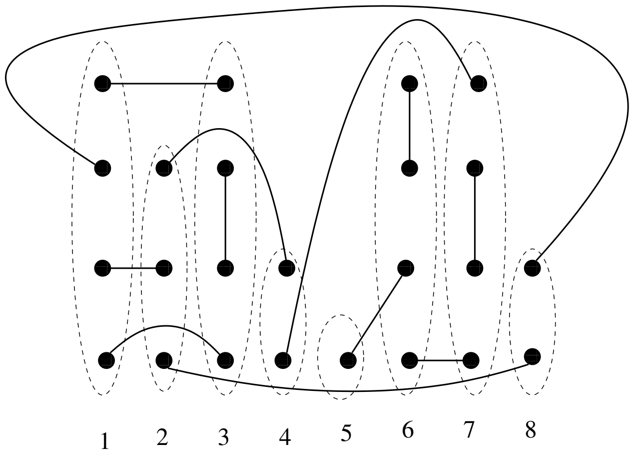


Figure 9.2: A partition F of W into $m = 12$ pairs

by $x \in W_{\varphi(x)}$. Let F be a partition of W into m pairs (*a configuration*). Given F we define the (multi)graph $\gamma(F)$ as

$$\gamma(F) = ([n], \{(\varphi(x), \varphi(y)) : (x,y) \in F\}).$$

Let us consider the following example of $\gamma(F)$. Let $n = 8$ and $d_1 = 4, d_2 = 3, d_3 = 4, d_4 = 2, d_5 = 1, d_6 = 4, d_7 = 4, d_8 = 2$. The accompanying diagrams, Figures 9.1, 9.2, 9.3 show a partition of W into $W_1, \dots, W_8$, a configuration and its corresponding multi-graph.

Denote by Ω the set of all configurations defined above for $d_1 + \cdots + d_n = 2m$ and notice that

$$|\Omega| = \frac{(2m)!}{m!2^m} = (2m-1)!!. \tag{9.1}$$

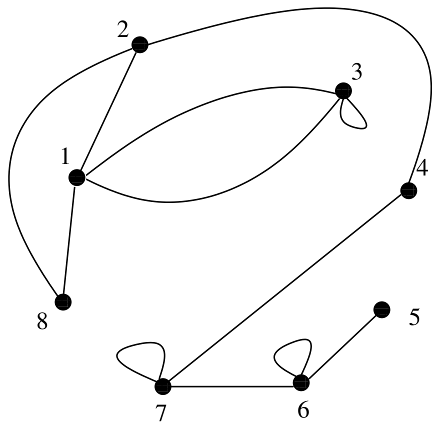


Figure 9.3: Graph $\gamma(F)$

To see this, take d_i "distinct" copies of i for $i = 1, 2, \ldots, n$ and take a permutation $\sigma_1, \sigma_2, \ldots, \sigma_{2m}$ of these $2m$ symbols. Read off F, pair by pair $\{\sigma_{2i-1}, \sigma_{2i}\}$ for $i = 1, 2, \ldots, m$. Each distinct F arises in $m!2^m$ ways.

We can also give an algorithmic, construction of a random element F of the family Ω.

Algorithm *F-GENERATOR*
    begin
    $U \longleftarrow W,\ F \longleftarrow \emptyset$
   for $t = 1, 2, \ldots, m$ **do**
      begin
          Choose x arbitrarily from U;
          Choose y randomly from $U \setminus \{x\}$;
          $F \longleftarrow F \cup \{(x, y)\}$;
          $U \longleftarrow U \setminus \{(x, y)\}$
      end
end

Note that F arises with probability $1/[(2m-1)(2m-3)\cdots 1] = |\Omega|^{-1}$.

Observe that the following relationship between a simple graph $G \in \mathcal{G}_{n,\mathbf{d}}$ and the number of configurations F for which $\gamma(F) = G$.

Lemma 9.1. *If $G \in \mathcal{G}_{n,\mathbf{d}}$, then*

$$|\gamma^{-1}(G)| = \prod_{i=1}^{n} d_i!\ .$$

Proof. Arrange the edges of G in lexicographic order. Now go through the sequence of $2m$ symbols, replacing each i by a new member of W_i. We obtain all F for which $\gamma(F) = G$. $\square$

The above lemma implies that we can use random configurations to "approximate" random graphs with a given degree sequence.

Corollary 9.2. *If F is chosen uniformly at random from the set of all configurations Ω and $G_1, G_2 \in \mathscr{G}_{n,\mathbf{d}}$ then*

$$\mathbb{P}(\gamma(F) = G_1) = \mathbb{P}(\gamma(F) = G_2).$$

So instead of sampling from the family $\mathscr{G}_{n,\mathbf{d}}$ and counting graphs with a given property, we can choose a random F and accept $\gamma(F)$ iff there are no loops or multiple edges, i.e. iff $\gamma(F)$ is a simple graph.

This is only a useful exercise if $\gamma(F)$ is simple with sufficiently high probability. We will assume for the remainder of this section that

$$\Delta = \max\{d_1, d_2, \ldots, d_n\} \leq n^\alpha, \alpha < 1/7.$$

We will prove later (see Lemma 9.7 and Corollary 9.8) that if F is chosen uniformly (at random) from Ω,

$$\mathbb{P}(\gamma(F) \text{ is simple}) = (1 + o(1))e^{-\lambda(\lambda+1)}, \tag{9.2}$$

where

$$\lambda = \frac{\sum d_i(d_i - 1)}{2\sum d_i}.$$

Hence, (9.1) and (9.2) will tell us not only how large is $\mathscr{G}_{n,\mathbf{d}}$, (Theorem 9.5) but also lead to the following conclusion.

Theorem 9.3. *Suppose that $\Delta \leq n^\alpha, \alpha < 1/7$. For any (multi)graph property $\mathscr{P}$*

$$\mathbb{P}(\mathbb{G}_{n,\mathbf{d}} \in \mathscr{P}) \leq (1 + o(1))e^{\lambda(\lambda+1)}\mathbb{P}(\gamma(F) \in \mathscr{P}),$$

The above statement is particularly useful if $\lambda = O(1)$, e.g., for random r-regular graphs, where r is a constant, since then $\lambda = \frac{r-1}{2}$. In the next section we will apply the above result to establish the connectedness of random regular graphs.

Before proving (9.2) for $\Delta \leq n^\alpha$, we feel it useful to give a simpler proof for the case of $\Delta = O(1)$.

Lemma 9.4. *If* $\Delta = O(1)$ *then* (9.2) *holds.*

Proof. Let L denote the number of loops and let D denote the number of *non-adjacent* double edges in $\gamma(F)$. Lemma 9.6 below shows that w.h.p. there are no adjacent double edges. We first estimate that for $k = O(1)$,

$$\begin{aligned}
\mathbb{E}\left(\binom{L}{k}\right) &= \sum_{\substack{S\subseteq[n]\\|S|=k}} \prod_{i\in S} \frac{d_i(d_i-1)}{4m-O(1)} \qquad (9.3)\\
&= \frac{1}{k!}\left(\sum_{i=1}^{n}\frac{d_i(d_i-1)}{4m}\right)^k + O\left(\frac{\Delta^4}{m}\right)\\
&\approx \frac{\lambda^k}{k!}.
\end{aligned}$$

Explanation for (9.3)**:** We assume that F-Generator begins with pairing up points in S. Therefore the random choice here is always from a set of size $2m - O(1)$.

It follows from Theorem 26.11 that L is asymptotically Poisson and hence that

$$\Pr(L=0) \approx e^{-\lambda}. \qquad (9.4)$$

We now show that D is also asymptotically Poisson and asymptotically independent of L. So, let $k = O(1)$. If $\mathscr{D}_k$ denotes the set of collections of $2k$ configuration points making up k double edges, then

$$\begin{aligned}
\mathbb{E}\left(\binom{D}{k}\,\middle|\, L=0\right) &= \sum_{\mathscr{D}_k} \Pr(\mathscr{D}_k \subseteq F \mid L=0)\\
&= \sum_{\mathscr{D}_k} \frac{\Pr(L=0 \mid \mathscr{D}_k \subseteq F)\Pr(\mathscr{D}_k \subseteq F)}{\Pr(L=0)}.
\end{aligned}$$

Now because $k = O(1)$, we see that the calculations that give us (9.4) will give us $\Pr(L=0 \mid \mathscr{D}_k \subseteq F) \approx \Pr(L=0)$. So,

$$\begin{aligned}
\mathbb{E}\left(\binom{D}{k}\,\middle|\, L=0\right) &\approx \sum_{\mathscr{D}_k} \Pr(\mathscr{D}_k \subseteq F)\\
&= \frac{1}{2}\sum_{\substack{S,T\subseteq[n]\\|S|=|T|=k\\S\cap T=\emptyset}} \sum_{\varphi:S\to T}\prod_{i\in S}\frac{2\binom{d_i}{2}\binom{d_{\varphi(i)}}{2}}{(2m-O(1))^2}\\
&= \frac{1}{k!}\left(\sum_{i=1}^{n}\frac{d_i(d_i-1)}{4m}\right)^{2k} + O\left(\frac{\Delta^8}{m}\right)
\end{aligned}$$

$$\approx \frac{\lambda^{2k}}{k!}.$$

It follows from Theorem 26.11 that

$$\Pr(D=0 \mid L=0) \approx e^{-\lambda^2} \tag{9.5}$$

and the lemma follows from (9.4) and (9.5). $\square$

Bender and Canfield [91] gave an asymptotic formula for $|\mathscr{G}_{n,\mathbf{d}}|$ when $\Delta = O(1)$. The paper [146] by Bollobás gives the same asymptotic formula when $\Delta < (2\log n)^{1/2}$. The following theorem allows for some more growth in Δ. Its proof uses the notion of *switching*. Switchings were introduced by McKay [671] and McKay and Wormald [672] and independently by Frieze [387], The bound $\alpha < 1/7$ is not optimal. For example, $\alpha < 1/2$ in [672].

Theorem 9.5. *Suppose that $\Delta \leq n^\alpha, \alpha < 1/7$.*

$$|\mathscr{G}_{n,\mathbf{d}}| \approx e^{-\lambda(\lambda+1)} \frac{(2m)!!}{\prod_{i=1}^n d_i!}.$$

In preparation we first prove

Lemma 9.6. *Suppose that $\Delta \leq n^\alpha$ where $\alpha < 1/7$. Let F be chosen uniformly (at random) from Ω. Then w.h.p. $\gamma(F)$ has*

(a) No double loops.

(b) At most $\Delta \log n$ loops.

(c) No adjacent loops.

(d) No adjacent double edges.

(e) No triple edges.

(f) At most $\Delta^2 \log n$ double edges.

(g) No vertex incident to a loop and a double edge.

(h) There are at most $\Delta^3 \log n$ triangles.

(i) No vertex is adjacent to two distinct vertices that have loops.

(j) No edge joining two distinct loops.

Proof. We will use the following inequality repeatedly.
Let $f_i = \{x_i, y_i\}, i = 1,2,\dots,k$ be k pairwise disjoint pairs of points. Then,

$$\mathbb{P}(f_i \in F, i = 1,2,\dots,k) \le \frac{1}{(2m-2k)^k}. \tag{9.6}$$

This follows immediately from

$$\mathbb{P}(f_i \in F \mid f_1, f_2, \dots, f_{i-1} \in F) = \frac{1}{2m-2i+1}.$$

This follows from considering **Algorithm** *F-GENERATOR* with $x = x_i$ and $y = y_i$ in the main loop.
(a) Using (9.6) we obtain

$$\mathbb{P}(F \text{ contains a pair of double loops}) \le \sum_{i=1}^{n} \binom{d_i}{2}^2 \left(\frac{1}{2m-8}\right)^2 \le \frac{\Delta^4 n}{(2m-8)^2} = o(1).$$

(b) Let $k_1 = \Delta \log n$.

$$\begin{aligned}
&\mathbb{P}(F \text{ has at least } k_1 \text{ loops}) \\
&\le o(1) + \sum_{\substack{x_1+\cdots+x_n=k_1,\\ x_i=0,1}} \prod_{i=1}^{n} \left(\binom{d_i}{2} \cdot \frac{1}{2m-2k_1}\right)^{x_i} \qquad (9.7)\\
&\le o(1) + \left(\frac{\Delta}{2m}\right)^{k_1} \sum_{\substack{x_1+\cdots+x_n=k_1,\\ x_i=0,1}} \prod_{i=1}^{n} d_i^{x_i} \\
&\le o(1) + \left(\frac{\Delta}{2m}\right)^{k_1} \frac{(d_1+\cdots+d_n)^{k_1}}{k_1!} \\
&\le o(1) + \left(\frac{\Delta e}{k_1}\right)^{k_1} \\
&= o(1).
\end{aligned}$$

The $o(1)$ term in (9.7) accounts for the probability of having a double loop.
(c)

$$\mathbb{P}(F \text{ contains a pair of adjacent loops})$$

(d)

$$\mathbb{P}(F \text{ contains a pair of adjacent double edges}) \le \sum_{i=1}^{n} \binom{d_i}{2}^2 \left(\frac{\Delta}{2m-8}\right)^2 \le$$

$$\frac{\Delta^5 m}{(2m-8)^2} = o(1).$$

(e)

$$\mathbb{P}(F \text{ contains a triple edge}) \leq \sum_{1\leq i<j\leq n} 6\binom{d_i}{3}\binom{d_j}{3}\left(\frac{1}{2m-6}\right)^3 \leq \frac{\Delta^5 m}{(2m-6)^3} = o(1).$$

(f) Let $k_2 = \Delta^2 \log n$.

$$\begin{aligned}
&\mathbb{P}(F \text{ has at least } k_2 \text{ double edges}) \\
&\leq o(1) + \sum_{\substack{x_1+\cdots+x_n=k_2,\\ x_i=0,1}} \prod_{i=1}^{n} \left(\binom{d_i}{2}\cdot\frac{\Delta}{2m-4k_2}\right)^{x_i} \qquad (9.8)\\
&\leq o(1) + \left(\frac{\Delta^2}{m}\right)^{k_2} \sum_{\substack{x_1+\cdots+x_n=k_2,\\ x_i=0,1}} \prod_{i=1}^{n} d_i^{x_i} \\
&\leq o(1) + \left(\frac{\Delta^2}{m}\right)^{k_2} \frac{(d_1+\cdots+d_n)^{k_2}}{k_2!} \\
&\leq o(1) + \left(\frac{2\Delta^2 e}{k_2}\right)^{k_2} \\
&= o(1).
\end{aligned}$$

The $o(1)$ term in (9.8) accounts for adjacent multiple edges and triple edges. The $\Delta/(2m-4k_2)$ term can be justified as follows: We have chosen two points x_1, x_2 in W_a in $\binom{d_i}{2}$ ways and this term bounds the probability that x_2 chooses a partner in the same cell as x_1.

(g)

$$\begin{aligned}
&\mathbb{P}(\exists \text{ vertex } v \text{ incident to a loop and a multiple edge} \\
&\leq \sum_{i=1}^{n} \binom{d_i}{2}^2 \frac{1}{2m-1}\frac{\Delta}{2m-5} \\
&\leq \frac{\Delta^4 m}{(2m-1)(2m-5)} \\
&= o(1).
\end{aligned}$$

(h) Let X denote the number of triangles in F. Then

$$\mathbb{E}(F) \le \sum_{i=1}^{n} \frac{\binom{d_i}{2}\Delta^2}{2m-4} \le \frac{\Delta^3 m}{2m-4} \le \Delta^3.$$

Now use the Markov inequality.

(i) The probability that there is a vertex adjacent to two loops is at most

$$\sum_{i=1}^{n} d_i \left(\frac{1}{2}\sum_{i=1}^{n} d_i(d_i-1)\right)^2 \left(\frac{\Delta}{M_1 - O(1)}\right)^4 \le \frac{M(\Delta M)^2\Delta^4}{(M-O(1))^4} = o(1).$$

(i) The probability that there is an edge joining two loops is at most

$$\sum_{i \ne j=1}^{n} \binom{d_i}{2}\binom{d_j}{2}\frac{(d_i-2)(d_j-2)}{(M_1-O(1))^3} \le \frac{(M_1\Delta)^2\Delta^4}{(M_1-O(1))^3} = o(1).$$

□

Let now $\Omega_{i,j}$ be the set of all $F \in \Omega$ such that F has i loops; j double edges, at most $\Delta^3 \log n$ triangles and no double loops or triple edges and no vertex incident with two double edges or with a loop and a multiple edge.

Lemma 9.7 (Switching Lemma)**.** *Suppose that $\Delta \le n^{\alpha}, \alpha < 1/7$. Let $M_1 = 2m$ and $M_2 = \sum_i d_i(d_i-1)$. For $i \le k_1 + 2k_2$ and $j \le k_2$, where $k_1 = \Delta \log n$ and $k_2 = \Delta^2 \log n$,*

$$\frac{|\Omega_{i+2,j-1}|}{|\Omega_{i,j}|} = \frac{j}{(i+2)(i+1)},$$

and

$$\frac{|\Omega_{i-1,0}|}{|\Omega_{i,0}|} = \frac{2iM_1}{M_2}\left(1 + O\left(\frac{\Delta^5 \log n}{M_1}\right)\right).$$

The corollary that follows is an immediate consequence of the Switching Lemma. It immediately implies Theorem 9.5.

Corollary 9.8. *Suppose that $\Delta \le n^{\alpha}$ where $\alpha < 1/7$. Then,*

$$\frac{|\Omega_{0,0}|}{|\Omega|} = (1+o(1))e^{-\lambda(\lambda+1)},$$

where

$$\lambda = \frac{M_2}{2M_1}.$$

It follows from the Switching Lemma that $i \leq k_1$ and $j \leq k_2$ implies

$$\frac{|\Omega_{i,j}|}{|\Omega_{0,0}|} = \left(1 + \tilde{O}\left(\frac{\Delta^3 k_2}{n}\right)\right)\frac{\lambda^{i+2j}}{i!\,j!} = (1+o(1))\frac{\lambda^{i+2j}}{i!\,j!}.$$

Lemma 9.6 implies that

$$\begin{aligned}(1-o(1))|\Omega| &= (1+o(1))|\Omega_{0,0}|\sum_{i=0}^{k_1}\sum_{j=0}^{k_2}\frac{\lambda^{i+2j}}{i!\,j!}\\ &= (1+o(1))|\Omega_{0,0}|e^{\lambda(\lambda+1)}.\end{aligned}$$

□

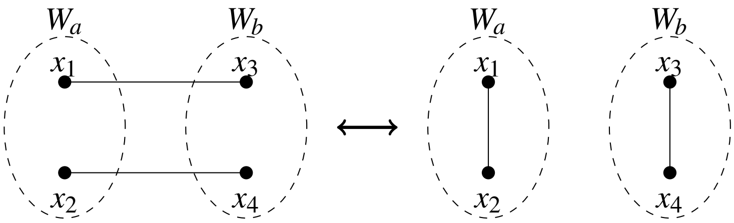


Figure 9.4: d-switch

To prove the Switching Lemma we need to introduce two specific operations on configurations, called a "d-switch" and an "ℓ-switch".
Figure 9.4 illustrates the "double edge removal switch" ("d-switch") operation. Here we have four points x_1, x_2, x_3, x_4 and a double edge associated with the pairs $\{x_1, x_3\}$, $\{x_2, x_4\} \in F$ where x_1, x_2 are in cell W_a and x_3, x_4 are in cell W_b. The d-switch operation replaces these pairs by a new set of pairs: $\{x_1, x_2\}$, $\{x_3, x_4\}$. This replaces a multiple edge by two loops and no other multiple edges are created.

In general, a *forward* d-switch operation takes F, a member of $\Omega_{i,j}$, to F', a member $\Omega_{i+2,j-1}$, see Figure 9.4. A *reverse* d-switch operation takes F', a member of $\Omega_{i+2,j-1}$, to F', a member $\Omega_{i,j}$. The number of choices η_f for a forward d-switch is j and the number of choices η_r for a reverse d-switch is $(i+2)(i+1)$. Lemma 9.6 implies that no triple edges are produced by the reverse d-switch.

Now for $F \in \Omega_{i,j}$ let $d_L(F) = j$ denote the number of $F' \in \Omega_{i+2,j-1}$ that can be obtained from F by an d-switch. Similarly, for $F' \in \Omega_{i+2,j-1}$ let $d_R(F') = (i+1)(i+2)$ denote the number of $F \in \Omega_{i,j}$ that can be transformed into F' by an d-switch. Then,

$$\sum_{F\in\Omega_{i,j}} d_L(F) = \sum_{F'\in\Omega_{i+2,j-1}} d_R(F').$$

So,

$$\frac{|\Omega_{i+2,j-1}|}{|\Omega_{i,j}|} = \frac{j}{(i+1)(i+2)},$$

which shows that the first statement of the Switching Lemma holds.

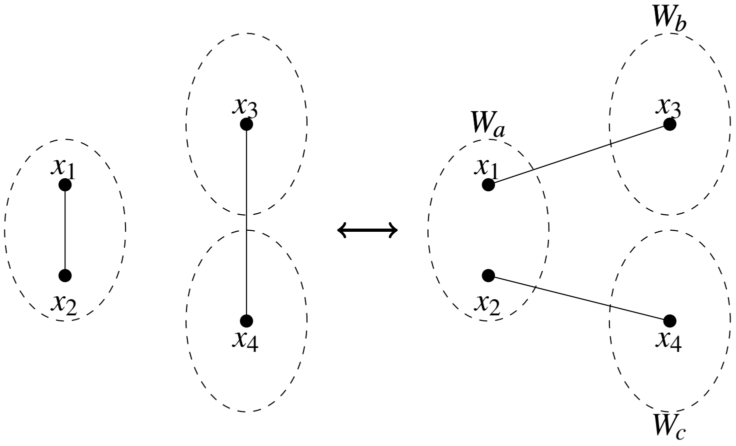


Figure 9.5: ℓ-switch

Now consider the second operation on configurations, described as a "loop removal switch"("ℓ-switch"), Figure 9.5. Here we have four points x_1,x_2,x_3,x_4 from three different cells, where x_1 and x_2 are in cell W_a, x_3 is in cell W_b and x_4 is in cell W_c. $\{x_1,x_2\} \in F$ forms a loop and $\{x_3,x_4\} \in F$. The ℓ-switch operation replaces these pairs by new pairs: $\{x_1,x_3\}$, $\{x_2,x_4\}$ or $\{x_1,x_4\}$, $\{x_2,x_3\}$ if in these operations no double edge is created.

We estimate the number of choices η during an ℓ-switch of $F \in \Omega_{i,0}$. For a forward switching operation

$$i\left(M_1 - 2\Delta^2 - 2i\right) \le \eta \le iM_1, \tag{9.9}$$

while, for the reverse procedure,

$$\frac{M_2}{2} - 3\Delta^3 \log n - i\Delta^2 - i\Delta^3 \le \eta \le \frac{M_2}{2}. \tag{9.10}$$

Proof of (9.9)**:**
To see why the above bounds hold, note that in the case of the forward loop removal switch, we have i choices for $\{x_1,x_2\}$ and at most M_1 choices for $\{x_3,x_4\}$ and there are two choices given these points. This explains the upper bound in

(9.9). To get the lower bound we subtract the number of "bad" choices. We can enumerate these bad choices as follows: We consider a fixed loop $\{x_1,x_2\}$ contained in cell W_a and we choose a pair $x_3 \in W_b$ and $x_4 \in W_c$. The transformation is bad only if there is $x \in W_a \setminus \{x_1,x_2\}$ ($\leq \Delta$ choices) that is paired in F with $y \in (W_b \setminus \{x_3\}) \cup (W_c \setminus \{x_4\})$ ($\leq 2\Delta$ choices). We also subtract $2i$ to account for avoiding the other $i-1$ loops in the choice of x_3,x_4.
Proof of (9.10)**:**
In the reverse procedure, we choose a pair $\{x_1,x_2\} \subseteq W_a$ in $M_2/2$ ways to arrive at the upper bound. The points $x_3 \in W_b, x_4 \in W_c$ are those paired with x_1,x_2 in F'. For the lower bound, a choice is bad only if (a,b,c) is a triangle. In this case, we create a double edge. There are at most $\Delta^3 \log n$ choices for the triangle and then three choices for a. We subtract a further $i\Delta^2$ to avoid creating another loop. Finally, we subtract $i\Delta^3$ in order to avoid increasing the number of triangles by choosing an edge that is within distance two of the loop. We also note here that forward d-switches do not increase the number of triangles.

Now for $F \in \Omega_{0,j}$ let $d_L(F)$ denote the number of $F' \in \Omega_{i-1,0}$ that can be obtained from F by an ℓ-switch. Similarly, for $F' \in \Omega_{i-1,0}$ let $d_R(F')$ denote the number of $F \in \Omega_{i,0}$ that can be transformed into F' by an ℓ-switch. Then,

$$\sum_{F \in \Omega_{i,0}} d_L(F) = \sum_{F' \in \Omega_{i-1,0}} d_R(F').$$

But, Lemma 9.6 implies that $i \leq 2\Delta^2 \log n$ and so

$$iM_1|\Omega_{i,0}|\left(1 - \frac{2\Delta^2 + 2\Delta^2 \log n}{M_1}\right) \leq \sum_{F \in \Omega_{i,0}} d_L(F) \leq iM_1|\Omega_{i,0}|,$$

while

$$\left(\frac{M_2}{2} - 3\Delta^3 \log n - 2\Delta^2 \log n(\Delta^2 + \Delta^3)\right) \leq |\Omega_{i-1,0}| \leq \sum_{F' \in \Omega_{i-1,0}} d_R(F') \leq \frac{M_2}{2}|\Omega_{i-1,0}|.$$

So

$$\frac{|\Omega_{i-1,0}|}{|\Omega_{i,0}|} = \frac{2iM_1}{M_2}\left(1 + O\left(\frac{\Delta^5 \log n}{M_1}\right)\right).$$

□

9.2 Connectivity of Regular Graphs

For an excellent survey of results on random regular graphs, see Wormald [859].

Bollobás [146] used the configuration model to prove the following: Let $\mathbb{G}_{n,r}$ denote a random r-regular graph with vertex set $[n]$ and $r \geq 3$ constant.

Theorem 9.9. *$\mathbb{G}_{n,r}$ is r-connected, w.h.p.*

Since an r-regular, r-connected graph, with n even, has a perfect matching, the above theorem immediately implies the following Corollary.

Corollary 9.10. *Let $\mathbb{G}_{n,r}$ be a random r-regular graph, $r \geq 3$ constant, with vertex set $[n]$ even. Then w.h.p. $\mathbb{G}_{n,r}$ has a perfect matching.*

Proof. (*of Theorem* 9.9)
Partition the vertex set $V = [n]$ of $\mathbb{G}_{n,r}$ into three parts, K, L and $V \setminus (K \cup L)$, such that $L = N(K)$, i.e., such that L separates K from $V \setminus (K \cup L)$ and $|L| = l \leq r-1$. We will show that w.h.p. there are no such K, L for k ranging from 2 to $n/2$. We will use the configuration model and the relationship stated in Theorem 9.3. We will divide the whole range of k into three parts.

(i) $2 \leq k \leq 3$.

Put $S := K \cup L,\ s = |S| = k + l \leq k + r - 1$. The set S contains at least $2r-1$ edges ($k = 2$) or at least $3r-3$ edges ($k = 3$). In both cases this is at least $s+1$ edges.

$$
\begin{aligned}
&\mathbb{P}(\exists S, s = |S| \leq r+2 :\ S \text{ contains } s+1 \text{ edges}) \\
&\leq \sum_{s=4}^{r+2} \binom{n}{s} \binom{rs}{s+1} \left(\frac{rs}{rn}\right)^{s+1} \qquad (9.11) \\
&\leq \sum_{s=4}^{r+2} n^s 2^{rs} s^{s+1} n^{-s-1} \\
&= o(1).
\end{aligned}
$$

Explanation for (9.11)**:** Having chosen a set of s vertices, spanning rs points R, we choose $s+1$ of these points T. $\frac{rs}{rn}$ bounds the probability that one of these points in T is paired with something in a cell associated with S. This bound holds conditional on other points of R being so paired.

(ii) $4 \leq k \leq ne^{-10}$.

The number of edges incident with the set K, $|K| = k$, is at least $(rk+l)/2$. Indeed let a be the number of edges contained in K and b be the number of $K : L$ edges. Then $2a + b = rk$ and $b \geq l$. This gives $a + b \geq (rk+l)/2$. So,

$$
\mathbb{P}(\exists K, L) \leq \sum_{k=4}^{ne^{-10}} \sum_{l=0}^{r-1} \binom{n}{k} \binom{n}{l} \binom{rk}{\frac{rk+l}{2}} \left(\frac{r(k+l)}{rn}\right)^{(rk+l)/2}
$$

$$\leq \sum_{k=4}^{ne^{-10}} \sum_{l=0}^{r-1} n^{-\left(\frac{r}{2}-1\right)k+\frac{l}{2}} \frac{e^{k+l}}{k^k l^l} 2^{rk} (k+l)^{(rk+l)/2}$$

Now

$$\left(\frac{k+l}{l}\right)^{l/2} \leq e^{k/2} \text{ and } \left(\frac{k+l}{k}\right)^{k/2} \leq e^{l/2},$$

and so

$$(k+l)^{(rk+l)/2} \leq l^{l/2} k^{rk/2} e^{(lr+k)/2}.$$

Therefore, with C_r a constant,

$$\begin{aligned}
\mathbb{P}(\exists K, L) &\leq C_r \sum_{k=4}^{ne^{-10}} \sum_{l=0}^{r-1} n^{-\left(\frac{r}{2}-1\right)k+\frac{l}{2}} e^{3k/2} 2^{rk} k^{(r-2)k/2} \\
&= C_r \sum_{k=4}^{ne^{-10}} \sum_{l=0}^{r-1} \left(n^{-\left(\frac{r}{2}-1\right)+\frac{l}{2k}} e^{3/2} 2^r k^{\frac{r}{2}-1}\right)^k \\
&= o(1).
\end{aligned}$$

(iii) $ne^{-10} < k \leq n/2$

Assume that there are a edges between sets L and $V \setminus (K \cup L)$. Denote also

$$\varphi(2m) = \frac{(2m)!}{m!\, 2^m} \approx 2^{1/2} \left(\frac{2m}{e}\right)^m.$$

Then, remembering that $r, l, a = O(1)$ we can estimate that

$$\begin{aligned}
&\mathbb{P}(\exists K, L) \\
&\leq \sum_{k,l,a} \binom{n}{k} \binom{n}{l} \binom{rl}{a} \frac{\varphi(rk+rl-a)\varphi(r(n-k-l)+a)}{\varphi(rn)} \qquad (9.12) \\
&\leq C_r \sum_{k,l,a} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{l}\right)^l \times \\
&\qquad\qquad \frac{(rk+rl-a)^{rk+rl-a}(r(n-k-l)+a)^{r(n-k-l)+a}}{(rn)^{rn}} \\
&\leq C'_r \sum_{k,l,a} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{l}\right)^l \frac{(rk)^{rk+rl-a}(r(n-k))^{r(n-k-l)+a}}{(rn)^{rn}} \\
&\leq C''_r \sum_{k,l,a} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{l}\right)^l \left(\frac{k}{n}\right)^{rk} \left(1-\frac{k}{n}\right)^{r(n-k)}
\end{aligned}$$

$$\leq C_r'' \sum_{k,l,a} \left(\left(\frac{k}{n}\right)^{r-1} e^{1-r/2}\, n^{r/k} \right)^k$$
$$= o(1).$$

Explanation of (9.12)**:** Having chosen K,L we choose a points in $W_{K\cup L} = \bigcup_{i\in K\cup L} W_i$ that will be paired outside $W_{K\cup L}$. This leaves $rk+rl-a$ points in $W_{K\cup L}$ to be paired up in $\varphi(rk+rl-a)$ ways and then the remaining points can be paired up in $\varphi(r(n-k-l)+a)$ ways. We then multiply by the probability $1/\varphi(rn)$ of the final pairing. □

9.3 Existence of a giant component

Molloy and Reed [685] provide an elegant and very useful criterion for when $\mathbb{G}_{n,\mathbf{d}}$ has a giant component. Suppose that there are $\lambda_i n + o(n^{3/4})$ vertices of degree $i = 1,2,\ldots,L$. We will assume that $L = O(1)$ and that the $\lambda_i, i \in [L]$ are constants independent of n. The paper [685] allows for $L = O(n^{1/4-\varepsilon})$. We will assume that $\lambda_1 + \lambda_2 + \cdots + \lambda_L = 1$.

Theorem 9.11. *Let* $\Lambda = \sum_{i=1}^{L} \lambda_i i(i-2)$. *Let* $\varepsilon > 0$ *be arbitrary.*

(a) *If* $\Lambda < -\varepsilon$ *then w.h.p. the size of the largest component in* $\mathbb{G}_{n,\mathbf{d}}$ *is* $O(\log n)$.

(b) *If* $\Lambda > \varepsilon$ *then w.h.p. there is a unique giant component of linear size* $\approx \Theta n$ *where* Θ *is defined as follows: let* $K = \sum_{i=1}^{L} i\lambda_i$ *and*

$$f(\alpha) = K - 2\alpha - \sum_{i=1}^{L} i\lambda_i \left(1 - \frac{2\alpha}{K}\right)^{i/2}. \tag{9.13}$$

Let ψ *be the smallest positive solution to* $f(\alpha) = 0$. *Then*

$$\Theta = 1 - \sum_{i=1}^{L} \lambda_i \left(1 - \frac{2\psi}{K}\right)^{i/2}.$$

If $\lambda_1 = 0$ *then* $\Theta = 1$, *otherwise* $0 < \Theta < 1$.

(c) *In Case (b), the degree sequence of the graph obtained by deleting the giant component satisfies the conditions of (a).*

Proof. We consider the execution of F-GENERATOR. We keep a sequence of partitions $U_t, A_t, E_t, t = 1,2,\ldots,m$ of W. Initially $U_0 = W$ and $A_0 = E_0 = \emptyset$. The $(t+1)$th iteration of F-GENERATOR is now executed as follows: it is designed so that we construct $\gamma(F)$ component by component. A_t is the set of points associated with the *partially exposed* vertices of the current component. These are vertices in the current component, not all of whose points have been paired. U_t is the set of unpaired points associated with the *entirely unexposed* vertices that have not been added to any component so far. E_t is the set of paired points. Whenever possible, we choose to make a pairing that involves the current component.

(i) If $A_t = \emptyset$ then choose x from U_t. Go to (iii).
We begin the exploration of a new component of $\gamma(F)$.

(ii) if $A_t \neq \emptyset$ choose x from A_t. Go to (iii).
Choose a point associated with a partially exposed vertex of the current component.

(iii) Choose y randomly from $(A_t \cup U_t) \setminus \{x\}$.

(iv) $F \leftarrow F \cup \{(x,y)\}$; $E_{t+1} \leftarrow E_t \cup \{x,y\}$; $A_{t+1} \leftarrow A_t \setminus \{x\}$.

(v) If $y \in A_t$ then $A_{t+1} \leftarrow A_{t+1} \setminus \{y\}$; $U_{t+1} \leftarrow U_t$.
y is associated with a vertex in the current component.

(vi) If $y \in U_t$ then $A_{t+1} \leftarrow A_t \cup (W_{\varphi(y)} \setminus y)$; $U_{t+1} \leftarrow U_t \setminus W_{\varphi(y)}$.
y is associated with a vertex $v = \varphi(y)$ not in the current component. Add all the points in $W_v \setminus \{y\}$ to the active set.

(vii) Goto (i).

(a) We fix a vertex v and estimate the size of the component containing v. We keep track of the size of A_t for $t = O(\log n)$ steps. Observe that

$$\mathbb{E}(|A_{t+1}| - |A_t| \mid |A_t| > 0) \lesssim \frac{\sum_{i=1}^{L} i\lambda_i n(i-2)}{M_1 - 2t - 1} = \frac{\Lambda n}{M_1 - 2t - 1} \leq -\frac{\varepsilon}{L}. \tag{9.14}$$

Here $M_1 = \sum_{i=1}^{L} i\lambda_i n$ as before. The explanation for (9.14) is that $|A|$ increases only in Step (vi) and there it increases by $i-2$ with probability $\lesssim \frac{i\lambda_i n}{M_1 - 2t}$. The two points x,y are missing from A_{t+1} and this explains the -2.

Let $\varepsilon_1 = \varepsilon/L$ and let

$$Y_t = \begin{cases} |A_t| + \varepsilon_1 t & |A_1|, |A_2|, \ldots, |A_t| > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

It follows from (9.14) that if $t = O(\log n)$ and $Y_1, Y_2, \ldots, Y_t > 0$ then

$$\mathbb{E}(Y_{t+1} \mid Y_1, Y_2, \ldots, Y_t) = \mathbb{E}(|A_{t+1}| + \varepsilon_1(t+1) \mid Y_1, Y_2, \ldots, Y_t) \leq |A_t| + \varepsilon_1 t = Y_t.$$

Otherwise, $\mathbb{E}(Y_{t+1} \mid \cdot) = 0 = Y_t$. It follows that the sequence (Y_t) is a super-martingale. Next let $Z_1 = 0$ and $Z_t = Y_t - Y_{t-1}$ for $t \geq 1$. Then we have (i) $-2 \leq Z_i \leq L$ and (ii) $\mathbb{E}(Z_i) \leq -\varepsilon_1$ for $i = 1, 2, \ldots, t$. Now,

$$\mathbb{P}(A_\tau \neq \emptyset, 1 \leq \tau \leq t) \leq \mathbb{P}(Y_t = Z_1 + Z_2 + \cdots + Z_t > 0),$$

It follows from Lemma 27.16 that if $Z = Z_1 + Z_2 + \cdots + Z_t$ then

$$\mathbb{P}(Z > 0) \leq P(Z - \mathbb{E}(Z) \geq t\varepsilon_1) \leq \exp\left\{-\frac{\varepsilon_1^2 t^2}{8t}\right\}.$$

It follows that with probability $1 - O(n^{-2})$ that A_t will become empty after at most $16\varepsilon_1^{-2}\log n$ rounds. Thus for any fixed vertex v, with probability $1 - O(n^{-2})$ the component contain v has size at most $4\varepsilon_1^{-2}\log n$. (We can expose the component containing v through our choice of x in Step (i).) Thus the probability there is a component of size greater than $16\varepsilon_1^{-2}\log n$ is $O(n^{-1})$. This completes the proof of (a).

(b)
If $t \leq \delta n$ for a small positive constant $\delta \ll \varepsilon/L^3$ then

$$\begin{aligned}\mathbb{E}(|A_{t+1}| - |A_t|) &\geq \frac{-2|A_t| + (1 + o(1))\sum_{i=1}^{L} i(\lambda_i n - 2t)(i-2)}{M_1 - 2\delta n} \\ &\geq \frac{-2L\delta n + (1 + o(1))(\Lambda n - 2\delta L^3 n)}{M_1 - 2\delta n} \geq \frac{\varepsilon}{2L}. \end{aligned} \tag{9.15}$$

Let $\varepsilon_2 = \varepsilon/2L$ and let

$$Y_t = \begin{cases} |A_t| - \varepsilon_2 t & |A_1|, |A_2|, \ldots, |A_t| > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

It follows from (9.14) that if $t \leq \delta n$ and $Y_1, Y_2, \ldots, Y_t > 0$ then

$$\mathbb{E}(Y_{t+1} \mid Y_1, Y_2, \ldots, Y_t) = \mathbb{E}(|A_{t+1}| - \varepsilon_2(t+1) \mid Y_1, Y_2, \ldots, Y_t) \geq |A_t| - \varepsilon_2 t = Y_t.$$

Otherwise, $\mathbb{E}(Y_{t+1} \mid \cdot) = 0 = Y_t$. It follows that the sequence (Y_t) is a sub-martingale. Next let $Z_1 = 0$ and $Z_t = Y_t - Y_{t-1}$ for $t \geq 1$. Then we have (i) $-2 \leq Z_i \leq L$ and (ii) $\mathbb{E}(Z_i) \geq \varepsilon_2$ for $i = 1, 2, \ldots, t$. Now,

$$\mathbb{P}(A_t \neq \emptyset) \geq \mathbb{P}(Y_t = Z_1 + Z_2 + \cdots + Z_t > 0),$$

It follows from Lemma 27.16 that if $Z = Z_1 + Z_2 + \cdots + Z_t$ then

$$\mathbb{P}(Z \leq 0) \leq P(Z - \mathbb{E}(Z) \geq t\varepsilon_2) \leq \exp\left\{-\frac{\varepsilon_2^2 t^2}{2t}\right\}.$$

It follows that if $L_0 = 100\varepsilon_2^{-2}$ then

$$\mathbb{P}\left(\exists L_0 \log n \leq t \leq \delta n : Z \leq \frac{\varepsilon_2 t}{2}\right) \leq$$
$$\mathbb{P}\left(\exists L_0 \log n \leq t \leq \delta n : Z - \mathbb{E}(Z) \geq \frac{\varepsilon_2 t}{2}\right)$$
$$\leq n \exp\left\{-\frac{\varepsilon_2^2 L_0 \log n}{8}\right\} = O(n^{-2}).$$

It follows that if $t_0 = \delta n$ then w.h.p. $|A_{t_0}| = \Omega(n)$ and there is a giant component and that the edges exposed between time $L_0 \log n$ and time t_0 are part of exactly one giant.

We now deal with the special case where $\lambda_1 = 0$. There are two cases. If in addition we have $\lambda_2 = 1$ then w.h.p. $\mathbb{G}_{\mathbf{d}}$ is the union of $O(\log n)$ vertex disjoint cycles, see Exercise 10.5.1. If $\lambda_1 = 0$ and $\lambda_2 < 1$ then the only solutions to $f(\alpha) = 0$ are $\alpha = 0, K/2$. For then $0 < \alpha < K/2$ implies

$$\sum_{i=2}^{L} i\lambda_i \left(1 - \frac{2\alpha}{K}\right)^{i/2} < \sum_{i=2}^{L} i\lambda_i \left(1 - \frac{2\alpha}{K}\right) = K - 2\alpha.$$

This gives $\Theta = 1$. Exercise 10.5.2 asks for a proof that w.h.p. in this case, $\mathbb{G}_{n,\mathbf{d}}$ consists a giant component plus a collection of small components that are cycles of size $O(\log n)$.

Assume now then that $\lambda_1 > 0$. We show that w.h.p. there are $\Omega(n)$ isolated edges. This together with the rest of the proof implies that $\Psi < K/2$ and hence that $\Theta < 1$. Indeed, if Z denotes the number components that are isolated edges, then

$$\mathbb{E}(Z) = \binom{\lambda_1 n}{2} \frac{1}{2M_1 - 1} \text{ and } \mathbb{E}(Z(Z-1)) = \binom{\lambda_1 n}{4} \frac{6}{(2M_1 - 1)(2M_1 - 3)}$$

and so the Chebyshev inequality (26.3) implies that $Z = \Omega(n)$ w.h.p.

Now for i such that $\lambda_i > 0$, we let $X_{i,t}$ denote the number of entirely unexposed vertices of degree i. We focus on the number of unexposed vertices of a give degree. Then,

$$\mathbb{E}(X_{i,t+1} - X_{i,t}) = -\frac{iX_{i,t}}{M_1 - 2t - 1}. \tag{9.16}$$

This suggests that we employ the differential equation approach of Section 28 in order to keep track of the $X_{i,t}$. We would expect the trajectory of $(t/n, X_{i,t}/n)$ to follow the solution to the differential equation

$$\frac{dx}{d\tau} = -\frac{ix}{K - 2\tau} \tag{9.17}$$

$x(0) = \lambda_i$. Note that $K = M_1/n$.

The solution to (9.17) is

$$x = \lambda_i \left(1 - \frac{2\tau}{K}\right)^{i/2}. \tag{9.18}$$

In what follows, we use the notation of Section 28, except that we replace λ_0 by $\xi_0 = n^{-1/4}$ to avoid confusion with λ_i.

(P0) $D = \left\{(\tau, x) : 0 < \tau < \frac{\Theta - \varepsilon}{2}, 2\xi_0 < x < 1\right\}$ where ε is small and positive.

(P1) $C_0 = 1$.

(P2) $\beta = L$.

(P3) $f(\tau, x) = -\frac{ix}{K-2\tau}$ and $\gamma = 0$.

(P4) The Lipschitz constant $L_1 = 2K/(K-2\Theta)^2$. This needs justification and follows from

$$\frac{x}{K-2\tau} - \frac{x'}{K-2\tau'} = \frac{K(x-x') + 2\tau(x-x') + 2x(\tau-\tau')}{(K-2\tau)(K-2\tau')}.$$

Theorem 28.1 then implies that with probability $1 - O(n^{1/4} e^{-\Omega(n^{1/4})})$,

$$\left| X_{i,t} - ni\lambda_i \left(1 - \frac{2t}{K}\right)^{i/2} \right| = O(n^{3/4}), \tag{9.19}$$

up to a point where $X_{i,t} = O(\xi_0 n)$. (The $o(n^{3/4})$ term for the number of vertices of degree i is absorbed into the RHS of (9.19).)

Now because

$$|A_t| = M_1 - 2t - \sum_{i=1}^{L} iX_{i,t} = Kn - 2t - \sum_{i=1}^{L} iX_{i,t},$$

we see that w.h.p.

$$\begin{aligned} |A_t| &= n\left(K - \frac{2t}{n} - \sum_{i=1}^{L} i\lambda_i \left(1 - \frac{2t}{Kn}\right)^{i/2}\right) + O(n^{3/4}) \\ &= nf\left(\frac{t}{n}\right) + O(n^{3/4}), \end{aligned} \tag{9.20}$$

so that w.h.p. the first time after time $t_0 = \delta n$ that $|A_t| = O(n^{3/4})$ is as at time $t_1 = \Psi n + O(n^{3/4})$. This shows that w.h.p. there is a component of size at least $\Theta n +$

$O(n^{3/4})$. Indeed, we simply subtract the number of entirely unexposed vertices from n to obtain this.

To finish, we must show that this component is unique and no larger than $\Theta n + O(n^{3/4})$. We can do this by proving (c), i.e. showing that the degree sequence of the graph $\mathbb{G}_U$ induced by the unexposed vertices satisfies the condition of Case (a). For then by Case (a), the giant component can only add $O(n^{3/4} \times \log n) = o(n)$ vertices from t_1 onwards.

We observe first that the above analysis shows that w.h.p. the degree sequence of $\mathbb{G}_U$ is asymptotically equal to $n\lambda_i'$, $i = 1, 2, \ldots, L$, where

$$\lambda_i' = \lambda_i \left(1 - \frac{2\Psi}{K}\right)^{i/2}.$$

(The important thing here is that the number of vertices of degree i is asymptotically proportional to λ_i'.) Next choose $\varepsilon_1 > 0$ sufficiently small and let $t_{\varepsilon_1} = \max\{t : |A_t| \geq \varepsilon_1 n\}$. There must exist $\varepsilon_2 < \varepsilon_1$ such that $t_{\varepsilon_1} \leq (\Psi - \varepsilon_2)n$ and $f'(\Psi - \varepsilon_2) \leq -\varepsilon_1$, else f cannot reach zero. Recall that $\Psi < K/2$ here and then,

$$\begin{aligned}
-\varepsilon_1 \geq f'(\Psi - \varepsilon_2) &= -2 + \frac{1}{K - 2(\Psi - \varepsilon_2)} \sum_{i \geq 1} i^2 \lambda_i \left(1 - \frac{2\Psi - 2\varepsilon_2}{K}\right)^{i/2} \\
&= -2 + \frac{1 + O(\varepsilon_2)}{K - 2\Psi} \sum_{i \geq 1} i^2 \lambda_i \left(1 - \frac{2\Psi}{K}\right)^{i/2} \\
&= \frac{1 + O(\varepsilon_2)}{K - 2\Psi} \left(-2 \sum_{i \geq 1} i \lambda_i \left(1 - \frac{2\Psi}{K}\right)^{i/2} + \sum_{i \geq 1} i^2 \lambda_i \left(1 - \frac{2\Psi}{K}\right)^{i/2}\right) \\
&= \frac{1 + O(\varepsilon_2)}{K - 2\Psi} \sum_{i \geq 1} i(i-2) \lambda_i \left(1 - \frac{2\Psi}{K}\right)^{i/2} \\
&= \frac{1 + O(\varepsilon_2)}{K - 2\Psi} \sum_{i \geq 1} i(i-2) \lambda_i'.
\end{aligned} \tag{9.21}$$

This completes the proofs of (b), (c). $\square$

9.4 $\mathbb{G}_{n,r}$ is asymmetric

In this section, we prove that w.h.p. $\mathbb{G}_{n,r}, r \geq 3$ only has one isomorphism, viz. the identity isomorphism. This was proved by Bollobás [152]. For a vertex v we let $d_k(v)$ denote the number of vertices at graph distance k from v in $\mathbb{G}_{n,r}$. We show that if $k_0 = \left\lceil \frac{3}{5} \log_{r-1} n \right\rceil$ then w.h.p. no two vertices have the same sequence $(d_k(v), k = 1, 2, \ldots, k_0)$. In the following $G = \mathbb{G}_{n,r}$.

Lemma 9.12.
Let $\ell_0 = \lceil 100 \log_{r-1} \log n \rceil$. *Then w.h.p.,* $e_G(S) \leq |S|$ *for all* $S \subseteq [n], |S| \leq 2\ell_0$.

Proof. Arguing as for (9.11), we have that

$$\begin{aligned}\mathbb{P}(\exists S : |S| \leq 2\ell_0, e_G(S) \geq |S|+1) &\leq \sum_{s=4}^{2\ell_0} \binom{n}{s} \binom{sr}{s+1} \left(\frac{sr}{rn-4\ell_0}\right)^{s+1} \\ &\leq \sum_{s=4}^{2\ell_0} \left(\frac{ne}{s}\right)^s (er)^{s+1} \left(\frac{s}{n-o(n)}\right)^{s+1} \\ &\leq \frac{1}{n} \sum_{s=4}^{2\ell_0} se^{2s+1+o(1)} = o(1).\end{aligned}$$

□

Let $\mathscr{E}$ denote the high probability event in Lemma 9.12. We will condition on the occurence of $\mathscr{E}$.

Now for $v \in [n]$ let $S_k(v)$ denote the set of vertices at distance k from v and let $S_{\leq k}(v) = \bigcup_{j \leq k} S_j(v)$. We note that

$$|S_k(v)| \leq r(r-1)^{k-1} \text{ for all } v \in [n], k \geq 1. \tag{9.22}$$

Furthermore, Lemma 9.12 implies that w.h.p. we have that for all $v, w \in [n], 1 \leq k \leq \ell_0$,

$$|S_k(v)| \geq (r-2)(r+1)(r-1)^{k-2}. \tag{9.23}$$

$$|S_k(v) \setminus S_k(w)| \geq (r-2)(r-1)^{k-1}. \tag{9.24}$$

This is because there can be at most one cycle in $S_{\leq \ell_0}(v)$ and the sizes of the relevant sets are reduced by having the cycle as close to v, w as possible.

Now consider $k > \ell_0$. Consider doing breadth first search from v or v, w exposing the configuration pairing as we go. Let an edge be *dispensable* if exposing it joins two vertices already known to be in $S_{\leq k}$. Lemma 9.12 implies that w.h.p. there is at most one dispensable edge in $S_{\leq \ell_0}$.

Lemma 9.13. *With probability* $1 - o(n^{-2})$, *(i) at most 20 of the first* $n^{2/5}$ *exposed edges are dispensable and (ii) at most* $n^{1/4}$ *of the first* $n^{3/5}$ *exposed edges are dispensable.*

Proof. The probability that the kth edge is dispensable is at most $\frac{(k-1)r}{rn-2k}$, independent of the history of the process. Hence,

$$\mathbb{P}(\exists \text{ 20 dispensable edges in first } n^{2/5}) \leq \binom{n^{2/5}}{20} \left(\frac{rn^{2/5}}{rn-o(n)}\right)^{20} = o(n^{-2}).$$

$$\mathbb{P}(\exists\, n^{1/4} \text{ dispensable edges in first } n^{3/5}) \leq \binom{n^{3/5}}{n^{1/4}} \left(\frac{rn^{3/5}}{rn-o(n)}\right)^{n^{1/4}} = o(n^{-2}).$$

$\square$

Now let $\ell_1 = \left\lceil \log_{r-1} n^{2/5} \right\rceil$ and $\ell_2 = \left\lceil \log_{r-1} n^{3/5} \right\rceil$. Then we have that, conditional on $\mathscr{E}$, with probability $1-o(n^{-2})$,

$$|S_k(v)| \geq ((r-2)(r+1)(r-1)^{\ell_0-2}-40))(r-1)^{k-\ell_0} : \ \ell_0 < k \leq \ell_1.$$
$$|S_k(v)| \geq ((r-2)(r+1)(r-1)^{\ell_1-1}-40(r-1)^{\ell_1-\ell_0}-2n^{1/4})(r-1)^{k-\ell_1} : \ \ell_1 < k \leq \ell_2.$$
$$|S_k(w)\setminus S_k(v)| \geq ((r-2)(r-1)^{\ell_0-1}-40)(r-1)^{k-\ell_0} : \ \ell_0 < k \leq \ell_1.$$
$$|S_k(w)\setminus S_k(v)| \geq ((r-2)(r-1)^{\ell_1-1}-40(r-1)^{\ell_1-\ell_0}-2n^{1/4})(r-1)^{k-\ell_1} : \ \ell_1 < k \leq \ell_2.$$

We deduce from this that if $\ell_3 = \left\lceil \log_{r-1} n^{4/7} \right\rceil$ and $k = \ell_3 + a, a = O(1)$ then with probability $1-o(n^{-2})$,

$$|S_k(w)| \geq ((r-2)(r+1)-o(1))(r-1)^{k-2} \approx (r-2)(r+1)(r-1)^{a-2}n^{4/7}.$$
$$|S_k(w)\setminus S_k(v)| \geq (r-2-o(1))(r-1)^{k-1} \approx (r-2)(r-1)^{a-1}n^{4/7}.$$

Suppose now that we consider the execution of breadth first search up until we have determined $S_k(v)$, but we have only determined $S_{k-1}(w)$. When we expose the unused edges of $S_{k-1}(w)$, some of these pairings will fall in $S_{\leq k}(v) \cup S_{k-1}(w)$. Expose any such pairings and condition on the outcome. There are at most $n^{1/4}$ such pairings and the size of $|S_k(v) \cap S_k(w)|$ is now determined. Then in order to have $d_k(v) = d_k(w)$ there has to be an exact outcome t for $|S_k(w) \setminus S_k(v)|$. There must now be $s = \Theta(n^{4/7})$ pairings between $W_x, x \in S_{k-1}(w)$ and $W_y, y \notin S_{\leq k}(v) \cup S_{k-1}(w)$. Furthermore, to have $d_k(v) = d_k(w)$ these s pairings must involve exactly t of the sets $W_y, y \notin S_k(v) \cup S_k(w)$, where t is determined *before* the choice of these s pairings. The following lemma will then be used to show that G is asymmetric w.h.p.

Lemma 9.14. *Let $R = \bigcup_{i=1}^m R_i$ be a partitioning of an rm set R into m subsets of size r. Suppose that S is a random s-subset of R, where $m^{5/9} < s < m^{3/5}$. Let X_S denote the number of sets R_i intersected by S. Then*

$$\max_j \mathbb{P}(X_S = j) \leq \frac{c_0 m^{1/2}}{s},$$

for some constant c_0.

Proof. We may assume that $s \geq m^{1/2}$. The probability that S has at least 3 elements in some set R_i is at most

$$\frac{m\binom{r}{3}\binom{rm-3}{s-3}}{\binom{rm}{s}} \leq \frac{s^3}{m^2} \leq \frac{m^{1/2}}{s}.$$

But

$$\mathbb{P}(X_S = j) \leq \mathbb{P}\left(\max_i |S \cap R_i| \geq 3\right) + \mathbb{P}\left(X_S = j \text{ and } \max_i |S \cap R_i| \leq 2\right).$$

So the lemma will follow if we prove that for every j,

$$P_j = \mathbb{P}\left(X_S = j \text{ and } \max_i |S \cap R_i| \leq 2\right) \leq \frac{c_1 m^{1/2}}{s}, \tag{9.25}$$

for some constant c_1.

Clearly, $P_j = 0$ if $j < s/2$ and otherwise

$$P_j = \frac{\binom{m}{j}\binom{j}{s-j} r^{2j-s} \binom{r}{2}^{s-j}}{\binom{rm}{s}}. \tag{9.26}$$

Now for $s/2 \leq j < s$ we have

$$\frac{P_{j+1}}{P_j} = \frac{(m-j)(s-j)}{(2j+2-s)(2j+1-s)} \frac{2r}{r-1}. \tag{9.27}$$

We note that if $s - j \geq \frac{10s^2}{m}$ then $\frac{P_{j+1}}{P_j} \geq \frac{10(r-1)}{3r} \geq 2$ and so the j maximising P_j is of the form $s - \frac{\alpha s^2}{m}$ where $\alpha \leq 10$. If we substitute $j = s - \frac{\alpha s^2}{m}$ into (9.27) then we see that

$$\frac{P_{j+1}}{P_j} \in \frac{2\alpha r}{r-1}\left[1 \pm c_2 \frac{s}{m}\right]$$

for some absolute constant $c_2 > 0$.

It follows that if j_0 is the index maximising P_j then

$$\left| j_0 - \left(s - \frac{(r-1)s^2}{2rm}\right)\right| \leq 1.$$

Furthermore, if $j_1 = j_0 - \frac{s}{m^{1/2}}$ then

$$\frac{P_{j+1}}{P_j} \leq 1 + c_3 \frac{m^{1/2}}{s} \text{ for } j_1 \leq j \leq j_0,$$

for some absolute constant $c_3 > 0$.

This implies that

$$P_j \geq P_{j_0}\left(1+c_3\frac{m^{1/2}}{s}\right)^{-(j_0-j_1)} = P_{j_0}\exp\left\{-(j_0-j_1)\left(c_3\frac{m^{1/2}}{s}+O\left(\frac{m}{s^2}\right)\right)\right\} \geq P_{j_0}e^{-2c_3}.$$

It follows from this that

$$P_{j_0} \leq \frac{e^{2c_3}m^{1/2}}{s}.$$

□

We apply Lemma 9.14 with $m=n, s=\Theta(n^{4/7}), j=t$ to show that

$$\mathbb{P}(d_k(v)=d_k(w), k\in[\ell_3,\ell_3+14]) \leq \left(\frac{c_0n^{1/2}}{n^{4/7}}\right)^{15} = o(n^{-2}).$$

This proves

Theorem 9.15. *W.h.p. $\mathbb{G}_{n,r}$ has a unique trivial automorphism.*

9.5 $\mathbb{G}_{n,r}$ versus $\mathbb{G}_{n,p}$

The configuration model is most useful when the maximum degree is bounded. When r is large, one can learn a lot about random r-regular graphs from the following theorem of Kim and Vu [566]. They proved that if $\log n \ll r \ll n^{1/3}/(\log n)^2$ then there is a joint distribution $G_0, G=\mathbb{G}_{n,r}, G_1$ such that w.h.p. (i) $G_0 \subseteq G$, (ii) the maximum degree $\Delta(G_1 \setminus G) \leq \frac{(1+o(1))\log n}{\log(\varphi(r)/\log n)}$ where $\varphi(r)$ is any function satisfying $(r\log n)^{1/2} \leq \varphi(r) \ll r$. Here $G_i = \mathbb{G}_{n,p_i}, i=0,1$ where $p_0=(1-o(1))\frac{r}{n}$ and $p_1=(1+o(1))\frac{r}{n}$. In this way we can deduce properties of $\mathbb{G}_{n,r}$ from $\mathbb{G}_{n,r/n}$. For example, G_0 is Hamiltonian w.h.p. implies that $\mathbb{G}_{n,r}$ is Hamiltonian w.h.p. Recently, Dudek, Frieze, Ruciński and Šileikis [309] have increased the range of r for which (i) holds. The cited paper deals with random hypergraphs and here we describe the simpler case of random graphs.

Theorem 9.16. *There is a positive constant C such that if*

$$C\left(\frac{r}{n}+\frac{\log n}{r}\right)^{1/3} \leq \gamma=\gamma(n) < 1,$$

and $m=\lfloor(1-\gamma)nr/2\rfloor$, then there is a joint distribution of $\mathbb{G}(n,m)$ and $\mathbb{G}_{n,r}$ such that

$$\mathbb{P}(\mathbb{G}_{n,m} \subset \mathbb{G}_{n,r}) \to 1.$$

Corollary 9.17. *Let $\mathscr{Q}$ be an increasing property of graphs such that $\mathbb{G}_{n,m}$ satisfies $\mathscr{Q}$ w.h.p. for some $m = m(n)$, $n\log n \ll m \ll n^2$. Then $\mathbb{G}_{n,r}$ satisfies $\mathscr{Q}$ w.h.p. for $r = r(n) \approx \frac{2m}{n}$.*

Our approach to proving Theorem 9.16 is to represent $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,r}$ as the outcomes of two graph processes which behave similarly enough to permit a good coupling. For this let $M = nr/2$ and define

$$\mathbb{G}_M = (\varepsilon_1, \dots, \varepsilon_M)$$

to be an ordered random uniform graph on the vertex set $[n]$, that is, $\mathbb{G}_{n,M}$ with a random uniform ordering of edges. Similarly, let

$$\mathbb{G}_r = (\eta_1, \dots, \eta_M)$$

be an ordered random r-regular graph on $[n]$, that is, $\mathbb{G}_{n,r}$ with a random uniform ordering of edges. Further, write $\mathbb{G}_M(t) = (\varepsilon_1, \dots, \varepsilon_t)$ and $\mathbb{G}_r(t) = (\eta_1, \dots, \eta_t)$, $t = 0, \dots, M$.

For every ordered graph G of size t and every edge $e \in K_n \setminus G$ we have

$$\Pr(\varepsilon_{t+1} = e \,|\, \mathbb{G}_M(t) = G) = \frac{1}{\binom{n}{2} - t}.$$

This is not true if we replace $\mathbb{G}_M$ by $\mathbb{G}_r$, except for the very first step $t = 0$. However, it turns out that for most of time the conditional distribution of the next edge in the process $\mathbb{G}_r(t)$ is approximately uniform, which is made precise in the lemma below. For $0 < \varepsilon < 1$, and $t = 0, \dots, M$ consider the inequalities

$$\Pr(\eta_{t+1} = e \,|\, \mathbb{G}_r(t)) \geq \frac{1-\varepsilon}{\binom{n}{2} - t} \text{ for every } e \in K_n \setminus \mathbb{G}_r(t), \tag{9.28}$$

and define a stopping time

$$T_\varepsilon = \max\{u : \forall t \leq u \text{ condition (9.28) holds}\}.$$

Lemma 9.18. *There is a positive constant C' such that if*

$$C' \left(\frac{r}{n} + \frac{\log n}{r} \right)^{1/3} \leq \varepsilon = \varepsilon(n) < 1, \tag{9.29}$$

then

$$T_\varepsilon \geq (1-\varepsilon)M \quad \textit{w.h.p.}$$

From Lemma 9.18, which is proved in Section 9.5, we deduce Theorem 9.16 using a coupling.

Proof of Theorem 9.16. Let $C = 3C'$, where C' is the constant from Lemma 9.18. Let $\varepsilon = \gamma/3$. The distribution of $\mathbb{G}_r$ is uniquely determined by the conditional probabilities

$$p_{t+1}(e|G) := \Pr\left(\eta_{t+1} = e \,|\, \mathbb{G}_r(t) = G\right), \quad t = 0, \ldots, M-1. \tag{9.30}$$

Our aim is to couple $\mathbb{G}_M$ and $\mathbb{G}_r$ up to the time T_ε. For this we will define a graph process $\mathbb{G}'_r := (\eta'_t), t = 1, \ldots, M$ such that the conditional distribution of (η'_t) coincides with that of (η_t) and w.h.p. (η'_t) shares many edges with $\mathbb{G}_M$.

Suppose that $G_r = G'_r(t)$ and $G_M = G_M(t)$ have been exposed and for every $e \notin G_r$ the inequality

$$p_{t+1}(e|G_r) \geq \frac{1-\varepsilon}{\binom{n}{2} - t} \tag{9.31}$$

holds (we have such a situation, in particular, if $t \leq T_\varepsilon$). Generate a Bernoulli $(1-\varepsilon)$ random variable ξ_{t+1} independently of everything that has been revealed so far; expose the edge ε_{t+1}. Moreover, generate a random edge $\zeta_{t+1} \in K_n \setminus G_r$ according to the distribution

$$\mathbb{P}(\zeta_{t+1} = e|\mathbb{G}'_r(t) = G_r, \mathbb{G}_M(t) = G_M) := \frac{p_{t+1}(e|G_r) - \frac{1-\varepsilon}{\binom{n}{2}-t}}{\varepsilon} \geq 0,$$

where the inequality holds because of the assumption (9.31). Observe also that

$$\sum_{e \notin G_r} \mathbb{P}(\zeta_{t+1} = e|\mathbb{G}'_r(t) = G_r, \mathbb{G}_M(t) = G_M) = 1,$$

so ζ_{t+1} has a well-defined distribution. Finally, fix a bijection $f_{G_r,G_M} : G_r \setminus G_M \to G_M \setminus G_r$ between the sets of edges and define

$$\eta'_{t+1} = \begin{cases} \varepsilon_{t+1}, & \text{if } \xi_{t+1} = 1, \varepsilon_{t+1} \notin G_r, \\ f_{G_r,G_M}(\varepsilon_{t+1}), & \text{if } \xi_{t+1} = 1, \varepsilon_{t+1} \in G_r, \\ \zeta_{t+1}, & \text{if } \xi_{t+1} = 0. \end{cases}$$

Note that

$$\xi_{t+1} = 1 \qquad \Rightarrow \qquad \varepsilon_{t+1} \in \mathbb{G}'_r(t+1). \tag{9.32}$$

We keep generating ξ_t's even after the stopping time has passed, that is, for $t > T_\varepsilon$, whereas η'_{t+1} is then sampled according to probabilities (9.30), without coupling. Note that ξ_t's are i.i.d. and independent of $\mathbb{G}_M$. We check that

$$\begin{aligned} &\mathbb{P}(\eta'_{t+1} = e \mid \mathbb{G}'_r(t) = G_r, \mathbb{G}_M(t) = G_M) \\ &= \mathbb{P}(\varepsilon_{t+1} = e)\,\mathbb{P}(\xi_{t+1} = 1) + \mathbb{P}(\zeta_{t+1} = e)\,\mathbb{P}(\xi_{t+1} = 0) \end{aligned}$$

$$= \frac{1-\varepsilon}{\binom{n}{2}-t} + \left(\frac{p_{t+1}(e|G_r) - \frac{1-\varepsilon}{\binom{n}{2}-t}}{\varepsilon} \right) \varepsilon$$
$$= p_{t+1}(e|G_r)$$

for all admissible $\mathbb{G}_r, \mathbb{G}_M$, i.e., such that $\mathbb{P}(\mathbb{G}_r(t) = G_r, \mathbb{G}_M(t) = G_M) > 0$, and for all $e \notin G_r$.

Further, define a set of edges which are potentially shared by $\mathbb{G}_M$ and $\mathbb{G}_r$:

$$S := \{\varepsilon_i : \xi_i = 1\ , 1 \le i \le (1-\varepsilon)M\}.$$

Note that

$$|S| = \sum_{i=1}^{\lfloor (1-\varepsilon)M \rfloor} \xi_i$$

is distributed as $\mathrm{Bin}(\lfloor (1-\varepsilon)M \rfloor, 1-\varepsilon)$.

Since (ξ_i) and (ε_i) are independent, conditioning on $|S| \ge m$, the first m edges in the set S comprise a graph which is distributed as $\mathbb{G}_{n,m}$. Moreover, if $T_\varepsilon \ge (1-\varepsilon)M$, then by (9.32) we have $S \subset \mathbb{G}_r$, therefore

$$\mathbb{P}(\mathbb{G}_{n,m} \subset \mathbb{G}_{n,r}) \ge \mathbb{P}(|S| \ge m, T_\varepsilon \ge (1-\varepsilon)M).$$

We have $\mathbb{E}|S| \ge (1-2\varepsilon)M$. Recall that $\varepsilon = \gamma/3$ and therefore $m = \lfloor (1-\gamma)M \rfloor = \lfloor (1-3\varepsilon)M \rfloor$. Applying the Chernoff bounds and our assumption on ε, we get

$$\mathbb{P}(|S| < m) \le e^{-\Omega(\gamma^2 m)} = o(1).$$

Finally, by Lemma 9.18 we have $T_\varepsilon \ge (1-\varepsilon)M$ w.h.p., which completes the proof of the theorem. □

Proof of Lemma 9.18

In all proofs of this section we will assume the condition (9.29). To prove Lemma 9.18 we will start with a fact which allows one to control the degrees of the evolving graph $\mathbb{G}_r(t)$.

For a vertex $v \in [n]$ and $t = 0, \ldots, M$, let

$$\deg_t(v) = |\{i \le t : v \in \eta_i\}|.$$

Lemma 9.19. *Let $\tau = 1 - t/M$. We have that w.h.p.*

$$\forall t \le (1-\varepsilon)M, \quad \forall v \in [n], \quad |\deg_t(v) - tr/M| \le 6\sqrt{\tau r \log n}. \tag{9.33}$$

In particular w.h.p.

$$\forall t \le (1-\varepsilon)M, \quad \forall v \in [n], \quad \deg_t(v) \le (1-\varepsilon/2)r. \tag{9.34}$$

Proof. Observe that if we fix an r-regular graph H and condition $\mathbb{G}_r$ to be a permutation of the edges of H, then $X := \deg_t(v)$ is a hypergeometric random variable with expected value $tr/M = (1-\tau)r$. Using the result of Section 27.5 and Theorem 27.11, and checking that the variance of X is at most τr, we get

$$\mathbb{P}(|X - tr/M| \geq x) \leq 2\exp\left\{-\frac{x^2}{2(\tau r + x/3)}\right\}.$$

Let $x = 6\sqrt{\tau r \log n}$. From (9.29), assuming $C' \geq 1$, we get

$$\frac{x}{\tau r} = 6\sqrt{\frac{\log n}{\tau r}} \leq 6\sqrt{\frac{\log n}{\varepsilon r}} \leq 6\varepsilon,$$

and so $x \leq 6\tau r$. Using this, we obtain

$$\frac{1}{2}\mathbb{P}(|X - tr/M| \geq x) \leq \exp\left\{-\frac{36\tau r \log n}{2(\tau r + 2\tau r)}\right\} = n^{-6}.$$

Inequality (9.33) now follows by taking a union bound over $nM \leq n^3$ choices of t and v.

To get (9.34), it is enough to prove the inequality for $t = (1-\varepsilon)M$. Inequality (9.33) implies

$$\deg_{(1-\varepsilon)M}(v) \leq (1-\varepsilon)r + 6\sqrt{\varepsilon r \log n}.$$

Thus it suffices to show that

$$6\sqrt{\varepsilon r \log n} \leq \varepsilon r/2,$$

or, equivalently, $\varepsilon \geq 144 \log n/r$, which is implied by (9.29) with $C' \geq 144$. □

Given an ordered graph $G = (e_1, \ldots, e_t)$, we say that an ordered r-regular graph H is an *extension* of G if the first t edges of H are equal to G. Let $\mathscr{G}_G(n,r)$ be the family of extensions of G and $\mathbb{G}_G = \mathbb{G}_G(n,r)$ be a graph chosen uniformly at random from $\mathscr{G}_G(n,r)$.

Further, for a graph $H \in \mathscr{G}_G(n,r)$ and $u, v \in [n]$ let

$$\deg_{H|G}(u,v) = |\{w \in [n] : \{u,w\} \in H \setminus G, \{v,w\} \in H\}|.$$

Note that $\deg_{H|G}(u,v)$ is not in general symmetric in u and v, but for $G = \emptyset$ coincides with the usual co-degree in a graph H.

The next fact is used in the proof of Lemma 9.21 only.

Lemma 9.20. *Let graph G with $t \leq (1-\varepsilon)M$ edges be such that $\mathscr{G}_G(n,r)$ is nonempty. For each $e \notin G$ we have*

$$\mathbb{P}(e \in \mathbb{G}_G) \leq \frac{4r}{\varepsilon n}. \tag{9.35}$$

Moreover, if $l \geq l_0 := 4r^2/(\varepsilon n)$, then for every $u,v \in [n]$ we have

$$\mathbb{P}\left(\deg_{\mathbb{G}_G|G}(u,v) > l\right) \leq 2^{-(l-l_0)}. \tag{9.36}$$

Proof. To prove (9.35) define the families

$$\mathscr{G}_{e\in} = \{H \in \mathscr{G}_G(n,r) : e \in H\} \quad \text{and} \quad \mathscr{G}_{e\notin} = \left\{H' \in \mathscr{G}_G(n,r) : e \notin H'\right\}.$$

Let us define an auxiliary bipartite graph B between $\mathscr{G}_{e\in}$ and $\mathscr{G}_{e\notin}$ in which $H \in \mathscr{G}_{e\in}$ is connected to $H' \in \mathscr{G}_{e\notin}$ whenever H' can be obtained from H by the following switching operation. Fix an ordered edge $\{w,x\}$ in $H \setminus G$ which is disjoint from $e = \{u,v\}$ and such that there are no edges between $\{u,v\}$ and $\{w,x\}$ and replace the edges $\{u,v\}$ and $\{w,x\}$ by $\{u,w\}$ and $\{v,x\}$ to obtain H'. Writing $f(H)$ for the number of graphs $H' \in \mathscr{G}_{e\notin}$ which can be obtained from H by a switching, and $b(H')$ for the number of graphs $H \in \mathscr{G}_{e\in}$ such that H' can be obtained H by a switching, we get that

$$|\mathscr{G}_{e\in}| \min_H f(H) \leq |E(B)| \leq |\mathscr{G}_{e\notin}| \max_{H'} b(H'). \tag{9.37}$$

We have $b(H') \leq \deg_{H'}(u)\deg_{H'}(v) \leq r^2$. On the other hand, recalling that $t \leq (1-\varepsilon)M$, for every $H \in \mathscr{G}_{e\in}$ we get

$$f(H) \geq M - t - 2r^2 \geq \varepsilon M \left(1 - \frac{2r^2}{\varepsilon M}\right) \geq \frac{\varepsilon M}{2},$$

because, assuming $C' \geq 8$, we have

$$\frac{2r^2}{\varepsilon M} \leq \frac{4r}{C'n}\left(\frac{n}{r}\right)^{1/3} \leq \frac{4}{C'} \leq \frac{1}{2}.$$

Therefore (9.37) implies that

$$\mathbb{P}(e \in \mathbb{G}_G) \leq \frac{|\mathscr{G}_{e\in}|}{|\mathscr{G}_{e\notin}|} \leq \frac{2r^2}{\varepsilon M} = \frac{4r}{\varepsilon n},$$

which concludes the proof of (9.35).

To prove (9.36), fix $u,v \in [n]$ and define the families

$$\mathscr{G}(l) = \left\{H \in \mathscr{G}_G(n,r) : \deg_{H|G}(u,v) = l\right\}, \qquad l = 0,1,\ldots.$$

We compare sizes of $\mathscr{G}(l)$ and $\mathscr{G}(l-1)$ in a similar way as above. For this we define the following switching which maps a graph $H \in \mathscr{G}(l)$ to a graph $H' \in$

Figure 9.6: Switching between $\mathscr{G}(l)$ and $\mathscr{G}(l-1)$: Before and after.

$\mathscr{G}(l-1)$. Select a vertex w contributing to $\deg_{H|G}$, that is, such that $\{u,w\} \in H \setminus G$ and $\{v,w\} \in H$; pick an ordered pair $u',w' \in [n] \setminus \{u,v,w\}$ such that $\{u',w'\} \in H \setminus G$ and there are no edges of H between $\{u,v,w\}$ and $\{u',w'\}$; replace edges $\{u,w\}$ and $\{u',w'\}$ by $\{u,u'\}$ and $\{w,w'\}$ (see Figure 9.6).

The number of ways to apply a forward switching to H is

$$f(H) \geq 2l(M-t-3r^2) \geq 2l\varepsilon M \left(1 - \frac{3r^2}{\varepsilon M}\right) \geq l\varepsilon M,$$

since, assuming $C' \geq 12$ we have

$$\frac{3r^2}{\varepsilon M} = \frac{6r}{\varepsilon n} \leq \frac{6}{C'}\left(\frac{r}{n}\right)^{2/3} \leq \frac{1}{2},$$

and the number of ways to apply a backward switching is $b(H) \leq r^3$. So,

$$\frac{|\mathscr{G}(l)|}{|\mathscr{G}(l-1)|} \leq \frac{\max_{H \in \mathscr{G}(l-1)} b(H)}{\min_{H \in \mathscr{G}(l)} f(H)} \leq \frac{2r^2}{\varepsilon l n} \leq \frac{1}{2},$$

by the assumption $l \geq l_0 := 4r^2/(\varepsilon n)$. Then

$$\begin{aligned} \mathbb{P}\left(\deg_{\mathbb{G}_G|G}(u,v) > l\right) &\leq \sum_{i>l} \frac{|\mathscr{G}(i)|}{|\mathscr{G}_G(n,r)|} \leq \sum_{i>l} \frac{|\mathscr{G}(i)|}{|\mathscr{G}(l_0)|} \\ &= \sum_{i>l} \prod_{j=l_0+1}^{i} \frac{|\mathscr{G}(j)|}{|\mathscr{G}(j-1)|} \leq \sum_{i>l} 2^{-(i-l_0)} = 2^{-(l-l_0)}, \end{aligned}$$

which completes the proof of (9.36). □

For the last lemma, which will be directly used in Lemma 9.18, we need to provide a few more definitions regarding random r-regular multigraphs.

Let G be an ordered graph with t edges. Let $\mathbb{M}_G(n,r)$ be a random *multigraph extension* of G to an ordered r-regular multigraph. Namely, $\mathbb{M}_G(n,r)$ is a sequence of M edges (some of which may be loops), the first t of which comprise G, while the remaining ones are generated by taking a uniform random permutation Π of the multiset $\{1,\dots,1,\dots,n,\dots,n\}$ with multiplicities $r - \deg_G(v)$, $v \in [n]$, and splitting it into consecutive pairs.

Recall that the number of such permutations is

$$N_G := \frac{(2(M-t))!}{\prod_{v\in[n]}(r-\deg_G(v))!},$$

and note that if a multigraph extension H of G has l loops, then

$$\mathbb{P}(\mathbb{M}_G(n,r)=H) = 2^{M-t-l}/N_G. \tag{9.38}$$

Thus, $\mathbb{M}_G(n,r)$ is not uniformly distributed over all multigraph extensions of G, but it is uniform over $\mathscr{G}_G(n,r)$. Thus, $\mathbb{M}_G(n,r)$, conditioned on being simple, has the same distribution as $\mathbb{G}_G(n,r)$. Further, for every edge $e \notin G$, let us write

$$\mathbb{M}_e = \mathbb{M}_{G\cup e}(n,r) \quad \text{and} \quad \mathscr{G}_e = \mathscr{G}_{G\cup e}(n,r). \tag{9.39}$$

The next claim shows that the probabilities of simplicity $\mathbb{P}(\mathbb{M}_e \in \mathscr{G}_e)$ are asymptotically the same for all $e \notin G$.

Lemma 9.21. *Let G be graph with $t \le (1-\varepsilon)M$ edges such that $\mathscr{G}_G(n,r)$ is nonempty. If $\Delta_G \le (1-\varepsilon/2)r$, then for every $e', e'' \notin G$ we have*

$$\frac{\mathbb{P}(\mathbb{M}_{e''} \in \mathscr{G}_{e''})}{\mathbb{P}(\mathbb{M}_{e'} \in \mathscr{G}_{e'})} \ge 1 - \frac{\varepsilon}{2}.$$

Proof. Set

$$\mathbb{M}' = \mathbb{M}_{e'} \quad \mathbb{M}'' = \mathbb{M}_{e''}, \quad \mathscr{G}' = \mathscr{G}_{e'}, \quad \text{and} \quad \mathscr{G}'' = \mathscr{G}_{e''}, \tag{9.40}$$

for convenience. We start by constructing a coupling of $\mathbb{M}'$ and $\mathbb{M}''$ in which they differ in at most three positions (counting in the replacement of e' by e'' at the $(t+1)$st position).

Let $e' = \{u', v'\}$ and $e'' = \{u'', v''\}$. Suppose first that e' and e'' are disjoint. Let Π' be the permutation underlying the multigraph $\mathbb{M}'$. Let Π^* be obtained from Π' by replacing a uniform random copy of u'' by u' and a uniform random copy of v'' by v'. If e' and e'' share a vertex, then assume, without loss of generality, that $v' = v''$, and define Π^* by replacing only a random u'' in Π' by u'. Then define $\mathbb{M}^*$ by splitting Π^* into consecutive pairs and appending them to $G \cup e''$.

It is easy to see that Π^* is uniform over permutations of the multiset $\{1,\ldots, 1,\ldots,n,\ldots,n\}$ with multiplicities $d - \deg_{G\cup e''}(v), v \in [n]$, and therefore $\mathbb{M}^*$ has the same distribution as $\mathbb{M}''$. Thus, we will further identify $\mathbb{M}^*$ and $\mathbb{M}''$.

Observe that if we condition $\mathbb{M}'$ on being a simple graph H, then $\mathbb{M}^* = \mathbb{M}''$ can be equivalently obtained by choosing an edge incident to u'' in $H \setminus (G \cup e')$ uniformly at random, say, $\{u'', w\}$, and replacing it by $\{u', w\}$, and then repeating this operation for v'' and v'. The crucial idea is that such a switching of edges is unlikely to create loops or multiple edges.

It is, however, possible, that for certain H this is not true. For example, if $e'' \in H \setminus (G \cup e')$, then the random choice of two edges described above is unlikely to destroy this e'', but e' in the non-random part will be replaced by e'', thus creating a double edge e''. Moreover, if almost every neighbor of u'' in $H \setminus (G \cup e')$ is also a neighbor of u', then most likely the replacement of u'' by u' will create a double edge. To avoid such instances, we want to assume that

(i) $e'' \notin H$

(ii) $\max\left(\deg_{H|G\cup e'}(u', u''), \deg_{H|G\cup e'}(v', v'')\right) \leq l_0 + \log_2 n$,

where $l_0 = 4r^2/\varepsilon n$ is as in Lemma 9.20. Define the following subfamily of simple extensions of $G \cup e'$:

$$\mathscr{G}'_{\text{nice}} = \left\{H \in \mathscr{G}' : H \text{ satisfies (i) and (ii)}\right\}.$$

Since $\mathbb{M}'$, conditioned on $\mathbb{M}' \in \mathscr{G}'$, is distributed as $\mathbb{G}_{G\cup e'}(n,r)$, by Lemma 9.20 and the assumption (9.29) with $C' \geq 20$,

$$\begin{aligned}\Pr\left(\mathbb{M}' \notin \mathscr{G}'_{\text{nice}} \,|\, \mathbb{M}' \in \mathscr{G}'\right) &= \mathbb{P}\left(\mathbb{G}_{G\cup e'}(n,r) \notin \mathscr{G}'_{\text{nice}}\right) \\ &\leq \frac{4r}{\varepsilon n} + 2 \times 2^{-\log_2 n} \leq \frac{\varepsilon}{4}. \end{aligned} \tag{9.41}$$

We have

$$\begin{aligned}&\Pr\left(\mathbb{M}'' \in \mathscr{G}'' \,|\, \mathbb{M}' \in \mathscr{G}'_{\text{nice}}\right) \Pr\left(\mathbb{M}' \in \mathscr{G}'_{\text{nice}} \,|\, \mathbb{M}' \in \mathscr{G}'\right) = \\ &\qquad \frac{\mathbb{P}(\mathbb{M}'' \in \mathscr{G}'', \mathbb{M}' \in \mathscr{G}'_{\text{nice}})}{\mathbb{P}(\mathbb{M}' \in \mathscr{G}'_{\text{nice}})} \cdot \frac{\mathbb{P}(\mathbb{M}' \in \mathscr{G}'_{\text{nice}}, \mathbb{M}' \in \mathscr{G}')}{\mathbb{P}(\mathbb{M}' \in \mathscr{G}')} \leq \\ &\qquad\qquad \frac{\mathbb{P}\left(\mathbb{M}'' \in \mathscr{G}''\right)}{\mathbb{P}\left(\mathbb{M}' \in \mathscr{G}'\right)}. \end{aligned} \tag{9.42}$$

To complete the proof of the claim, it suffices to show that

$$\Pr\left(\mathbb{M}'' \in \mathscr{G}'' \,|\, \mathbb{M}' \in \mathscr{G}'_{\text{nice}}\right) \geq 1 - \frac{\varepsilon}{4}, \tag{9.43}$$

since plugging (9.41) and (9.43) into (9.42) will complete the proof of the statement.

To prove (9.43), fix $H \in \mathscr{G}'_{\text{nice}}$ and condition on $\mathbb{M}' = H$. A loop can only be created in $\mathbb{M}''$ when u'' is incident to u' in $H \setminus (G \cup e')$ and the randomly chosen edge is $\{u', u''\}$, or, provided $v' \neq v''$, when v'' is incident to v' in $H \setminus (G \cup e')$ and we randomly choose $\{v', v''\}$. Therefore, recalling that $\Delta_G \leq (1-\varepsilon/2)r$, we get

$$\Pr\left(\mathbb{M}'' \text{ has a loop} \,|\, \mathbb{M}' = H\right) \leq \frac{1}{\deg_{H\setminus(G\cup e')}(u'')} + \frac{1}{\deg_{H\setminus(G\cup e')}(v'')} \leq \frac{4}{\varepsilon r} \leq \frac{\varepsilon}{8}, \quad (9.44)$$

where the second term is present only if $e' \cap e'' = \emptyset$, and the last inequality is implied by (9.29).

A multiple edge can be created in three ways: (i) by choosing, among the edges incident to u'', an edge $\{u'', w\} \in H \setminus (G \cup e')$ such that $\{u', w\} \in H$; (ii) similarly for v'' (if $v' \neq v''$); (iii) choosing both edges $\{u'', v'\}$ and $\{v'', u'\}$ (provided they exist in $H \setminus (G \cup e')$). Therefore, by (ii) and assumption $\Delta_G \leq (1-\varepsilon/2)r$,

$$\begin{aligned}
&\Pr\left(\mathbb{M}'' \text{ has a multiple edge } \,|\, \mathbb{M}' = H\right) \\
&\qquad \leq \frac{\deg_{H|G\cup e'}(u'', u')}{\deg_{H\setminus(G\cup e')}(u'')} + \frac{\deg_{H|G\cup e'}(v'', v')}{\deg_{H\setminus(G\cup e')}(v'')} \\
&\qquad\qquad + \frac{1}{\deg_{H\setminus(G\cup e')}(u'')\deg_{H\setminus(G\cup e')}(v'')} \\
&\qquad\qquad \leq 2\left(\frac{8r}{\varepsilon^2 n} + \frac{2\log_2 n}{\varepsilon r}\right) + \frac{4}{\varepsilon^2 r^2} \leq \frac{\varepsilon}{8}, \quad (9.45)
\end{aligned}$$

because (9.29) implies $\varepsilon > C'(r/n)^{1/3}$ and

$$\varepsilon > C'(\log n/r)^{1/3} > C'(\log n/r)^{1/2}$$

and we can choose arbitrarily large C'. (Again, in case when $|e' \cap e''| = 1$, the R-H-S of (9.45) reduces to only the first summand.)

Combining (9.44) and (9.45), we have shown (9.43). □

Proof of Lemma 9.18. In view of Lemma 9.19 it suffices to show that

$$\Pr(\eta_{t+1} = e \,|\, \mathbb{G}_r(t) = G) \geq \frac{1-\varepsilon}{\binom{n}{2} - t}, \qquad e \notin G.$$

for every $t \leq (1-\varepsilon)M$ and G such that

$$r(\tau+\delta) \geq r - \deg_G(v) \geq r(\tau-\delta) \geq \frac{\varepsilon r}{2}, \qquad v \in [n], \quad (9.46)$$

where

$$\tau = 1 - t/M, \qquad \delta = 6\sqrt{\tau \log n / r}.$$

For every $e', e'' \notin G$ we have (recall the definitions (9.39) and (9.40))

$$\frac{\Pr\left(\eta_{t+1} = e'' \,|\, \mathbb{G}_r(t) = G\right)}{\Pr\left(\eta_{t+1} = e' \,|\, \mathbb{G}_r(t) = G\right)} = \frac{|\mathscr{G}_{G\cup e''}(n,r)|}{|\mathscr{G}_{G\cup e'}(n,r)|} = \frac{|\mathscr{G}''|}{|\mathscr{G}'|}. \tag{9.47}$$

By (9.38) we have

$$\mathbb{P}\left(\mathbb{M}' \in \mathscr{G}'\right) = \frac{|\mathscr{G}'|2^{M-t}}{N_G} = \frac{|\mathscr{G}'|2^{M-t}\prod_{v\in[n]}(r-\deg_{G\cup e'}(v))!}{(2(M-t))!},$$

and similarly for the family $\mathscr{G}''$. This yields, after a few cancellations, that

$$\frac{|\mathscr{G}''|}{|\mathscr{G}'|} = \frac{\prod_{v\in e''\setminus e'}(r-\deg_G(v))}{\prod_{v\in e'\setminus e''}(r-\deg_G(v))} \cdot \frac{\mathbb{P}\left(\mathbb{M}'' \in \mathscr{G}''\right)}{\mathbb{P}\left(\mathbb{M}' \in \mathscr{G}'\right)} \tag{9.48}$$

By (9.46), the ratio of products in (9.48) is at least

$$\left(\frac{\tau-\delta}{\tau+\delta}\right)^2 \geq \left(1-\frac{2\delta}{\tau}\right)^2 \geq 1 - 24\sqrt{\frac{\log n}{\tau r}} \geq 1 - 24\sqrt{\frac{\log n}{\varepsilon r}} \geq 1 - \frac{\varepsilon}{2},$$

where the last inequality holds by the assumption (9.29). Since by Lemma 9.21 the ratio of probabilities in (9.48) is

$$\frac{\mathbb{P}\left(\mathbb{M}'' \in \mathscr{G}''\right)}{\mathbb{P}\left(\mathbb{M}' \in \mathscr{G}'\right)} \geq 1 - \frac{\varepsilon}{2},$$

we have obtained that

$$\frac{\Pr\left(\eta_{t+1} = e'' \,|\, \mathbb{G}_r(t) = G\right)}{\Pr\left(\eta_{t+1} = e' \,|\, \mathbb{G}_r(t) = G\right)} \geq 1 - \varepsilon.$$

Finally, noting that

$$\max_{e'\notin G} \Pr\left(\eta_{t+1} = e' \,|\, \mathbb{G}_r(t) = G\right)$$

is at least as large as the average over all $e' \notin G$, which is $\frac{1}{\binom{n}{2}-t}$, we conclude that for every $e \notin G$

$$\begin{aligned}\Pr\left(\eta_{t+1} = e \,|\, \mathbb{G}_r(t) = G\right) &\geq (1-\varepsilon)\max_{e'\notin G} \Pr\left(\eta_{t+1} = e' \,|\, \mathbb{G}_r(t) = G\right) \\ &\geq \frac{1-\varepsilon}{\binom{n}{2}-t},\end{aligned}$$

which finishes the proof. □

9.6 Exercises

9.6.1 Show that w.h.p. a random 2-regular graph on n vertices consists of $O(\log n)$ vertex disjoint cycles.

9.6.2 Suppose that in the notation of Theorem 9.11, $\lambda_1 = 0, \lambda_2 < 1$. Show that w.h.p. $\mathbb{G}_{n,\mathbf{d}}$ consists of a giant component plus a collection of small components of size $O(\log n)$.

9.6.3 Let H be a subgraph of $\mathbb{G}_{n,r}, r \geq 3$ obtained by independently including each vertex with probability $\frac{1+\varepsilon}{r-1}$, where $\varepsilon > 0$ is small and positive. Show that w.h.p. H contains a component of size $\Omega(n)$.

9.6.4 Let $\mathbf{x} = (x_1, x_2, \dots, x_{2m})$ be chosen uniformly at random from $[n]^{2m}$. Let $G_{\mathbf{x}}$ be the multigraph with vertex set $[n]$ and edges $(x_{2i-1}, x_{2i}), i = 1, 2, \dots, m$. Let $d_{\mathbf{x}}(i)$ be the number of times that i appears in $\mathbf{x}$.

Show that conditional on $d_{\mathbf{x}}(i) = d_i, i \in [n]$, $G_{\mathbf{x}}$ has the same distribution as the multigraph $\gamma(F)$ of Section 9.1.

9.6.5 Suppose that we condition on $d_{\mathbf{x}}(i) \geq k$ for some non-negative integer k. For $r \geq 0$, let

$$f_r(x) = e^x - 1 - x - \cdots - \frac{x^{k-1}}{(k-1)!}.$$

Let Z be a random variable taking values in $\{k, k+1, \dots, \}$ such that

$$\mathbb{P}(Z = i) = \frac{\lambda^i e^{-\lambda}}{i! f_k(\lambda)} \qquad \text{for } i \geq k,$$

where λ is arbitrary and positive.

Show that the degree sequence of $G_{\mathbf{x}}$ is distributed as independent copies $Z_1, Z_2, \dots, Z_n$ of Z, subject to $Z_1 + Z_2 + \cdots + Z_n = 2m$.

9.6.6 Show that

$$\mathbb{E}(Z) = \frac{\lambda f_{k-1}(\lambda)}{f_k(\lambda)}.$$

Show using the Local Central Limit Theorem (see e.g. Durrett [318]) that if $\mathbb{E}(Z) = \frac{2m}{n}$ then

$$\mathbb{P}\left(\sum_{j=1}^{\nu} Z_j = 2m - k\right) = \frac{1}{\sigma\sqrt{2\pi n}}\left(1 + O((k^2+1)\nu^{-1}\sigma^{-2})\right)$$

where $\sigma^2 = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2$ is the variance of Z.

9.6.7 Use the model of (i)–(iii) to show that if $c = 1+\varepsilon$ and ε is sufficiently small and $\omega \to \infty$ then w.h.p. the 2-core of $G_{n,p}, p = c/n$ does not contain a cycle C, $|C| = \omega$ in which more than 10% of the vertices are of degree three or more.

9.6.8 Let $\mathbb{G} = \mathbb{G}_{n,r}, r \geq 3$ be the random r-regular configuration multigraph of Section 9.2. Let X denote the number of Hamilton cycles in $\mathbb{G}$. Show that

$$\mathbb{E}(X) \approx \sqrt{\frac{\pi}{2n}} \left((r-1) \left(\frac{r-2}{r} \right)^{(r-2)/2} \right)^n .$$

9.6.9 Show that if graph $G = G_1 \cup G_2$ then its rainbow connection satisfies $rc(G) \leq rc(G_1) + rc(G_2) + |E(G_1) \cap E(G_2)|$. Using the contiguity of $\mathbb{G}_{n,r}$ to the union of r independent matchings, (see Chapter 25), show that $rc(\mathbb{G}_{n,r}) = O(\log_r n)$ for $r \geq 6$.

9.6.10 Show that w.h.p. $\mathbb{G}_{n,3}$ is not planar.

9.7 Notes

Giant Components and Cores

Hatami and Molloy [471] discuss the size of the largest component in the scaling window for a random graph with a fixed degree sequence.

Cooper [245] and Janson and Luczak [506] discuss the sizes of the cores of random graphs with a given degree sequence.

Hamilton cycles

Robinson and Wormald [767], [768] showed that random r-regular graphs are Hamiltonian for $3 \leq r = O(1)$. In doing this, they introduced the important new method of small subgraph conditioning. It is a refinement on the Chebyshev inequality. Somewhat later Cooper, Frieze and Reed [272] and Krivelevich, Sudakov, Vu Wormald [606] removed the restriction $r = O(1)$. Frieze, Jerrum, Molloy, Robinson and Wormald [396] gave a polynomial time algorithm that w.h.p. finds a Hamilton cycle in a random regular graph. Cooper, Frieze and Krivelevich [266] considered the existence of Hamilton cycles in $G_{n,\mathbf{d}}$ for certain classes of degree sequence.

Chromatic number

Frieze and Łuczak [405] proved that w.h.p. $\chi(\mathbb{G}_{n,r}) = (1+o_r(1))\frac{r}{2\log r}$ for $r = O(1)$. Here $o_r(1) \to 0$ as $r \to \infty$. Achlioptas and Moore [5] determined the chromatic number of a random r-regular graph to within three values, w.h.p. Kemkes, Pérez-Giménez and Wormald [563] reduced the range to two values. Shi and Wormald [804], [805] consider the chromatic number of $\mathbb{G}_{n,r}$ for small r. In particular they show that w.h.p. $\chi(\mathbb{G}_{n,4}) = 3$. Frieze, Krivelevich and Smyth [401] gave estimates for the chromatic number of $\mathbb{G}_{n,\mathbf{d}}$ for certain classes of degree sequence.

Eigenvalues

The largest eigenvalue of the adjacency matrix of $\mathbb{G}_{n,r}$ is always r. Kahn and Szemerédi [538] showed that w.h.p. the second eigenvalue is of order $O(r^{1/2})$. Friedman [380] proved that w.h.p. the second eigenvalue is at most $2(r-1)^{1/2} + o(1)$. Broder, Frieze, Suen and Upfal [200] considered $\mathbb{G}_{n,\mathbf{d}}$ where $C^{-1}d \le d_i \le Cd$ for some constant $C > 0$ and $d \le n^{1/10}$. They show that w.h.p. the second eigenvalue of the adjacency matrix is $O(d^{1/2})$.

First Order Logic

Haber and Krivelevich [459] studied the first order language on random d-regular graphs. They show that if $r = \Omega(n)$ or $r = n^{\alpha}$ where α is irrational, then $\mathbb{G}_{n,r}$ obeys a 0-1 law.

Rainbow Connection

Dudek, Frieze and Tsourakakis [310] studied the rainbow connection of random regular graphs. They showed that if $4 \le r = O(1)$ then $rc(\mathbb{G}_{n,r}) = O(\log n)$. This is best possible up to constants, since $rc(\mathbb{G}_{n,r}) \ge diam(\mathbb{G}_{n,r}) = \Omega(\log n)$. Kamčev, Krivelevich and Sudakov [541] gave a simpler proof when $r \ge 5$, with a better hidden constant.

Chapter 10

Intersection Graphs

Let G be a (finite, simple) graph. We say that G is an *intersection graph* if we can assign to each vertex $v \in V(G)$ a set S_v, so that $\{v, w\} \in E(G)$ exactly when $S_v \cap S_w \neq \emptyset$. In this case, we say G is the intersection graph of the family of sets $\mathscr{S} = \{S_v : v \in V(G)\}$.

Although all graphs are intersection graphs (see Marczewski [653]) some classes of intersection graphs are of special interest. Depending on the choice of family $\mathscr{S}$, often reflecting some geometric configuration, one can consider, for example, *interval graphs* defined as the intersection graphs of intervals on the real line, *unit disc graphs* defined as the intersection graphs of unit discs on the plane etc. In this chapter we will discuss properties of *random intersection graphs*, where the family $\mathscr{S}$ is generated in a random manner.

10.1 Binomial Random Intersection Graphs

Binomial random intersection graphs were introduced by Karoński, Scheinerman and Singer-Cohen in [554] as a generalisation of the classical model of the binomial random graph $\mathbb{G}_{n,p}$.

Let n, m be positive integers and let $0 \leq p \leq 1$. Let $V = \{1, 2, \ldots, n\}$ be the set of vertices and for every $1 \leq k \leq n$, let S_k be a random subset of the set $M = \{1, 2, \ldots, m\}$ formed by selecting each element of M independently with probability p. We define a *binomial random intersection graph* $G(n, m, p)$ as the intersection graph of sets S_k, $k = 1, 2, \ldots n$. Here $S_1, S_2, \ldots, S_n$ are generated independently. Hence two vertices i and j are adjacent in $G(n, m, p)$ if and only if $S_i \cap S_j \neq \emptyset$.

There are other ways to generate binomial random intersection graphs. For example, we may start with a classical bipartite random graph $\mathbb{G}_{n,m,p}$, with vertex set bipartition

$$(V,M), V = \{1,2,\dots,n\}, M = \{1,2,\dots,m\},$$

where each edge between V and M is drawn independently with probability p. Next, one can generate a graph $G(n,m,p)$ with vertex set V and vertices i and j of $G(n,m,p)$ connected if and only if they share a common neighbor (in M) in the random graph $\mathbb{G}_{n,m,p}$. Here the graph $\mathbb{G}_{n,m,p}$ is treated as a *generator* of $G(n,m,p)$.

One observes that the probability that there is an edge $\{i,j\}$ in $G(n,m,p)$ equals $1-(1-p^2)^m$, since the probability that sets S_i and S_j are disjoint is $(1-p^2)^m$, however, in contrast with $\mathbb{G}_{n,p}$, the edges do not occur independently of each other.

Another simple observation leads to some natural restrictions on the choice of probability p. Note that the expected number of edges of $G(n,m,p)$ is,

$$\binom{n}{2}(1-(1-p^2)^m) \approx n^2mp^2,$$

provided $mp^2 \to 0$ as $n \to \infty$. Therefore, if we take $p = o((n\sqrt{m})^{-1})$ then the expected number of edges of $G(n,m,p)$ tends to 0 as $n \to \infty$ and therefore w.h.p. $G(n,m,p)$ is empty.

On the other hand the expected number of non-edges in $G(n,m,p)$ is

$$\binom{n}{2}(1-p^2)^m \le n^2e^{-mp^2}.$$

Thus if we take $p = (2\log n + \omega(n))/m)^{1/2}$, where $\omega(n) \to \infty$ as $n \to \infty$, then the random graph $G(n,m,p)$ is complete w.h.p. One can also easily show that when $\omega(n) \to -\infty$ then $G(n,m,p)$ is w.h.p. not complete. So, when studying the evolution of $G(n,m,p)$ we may restrict ourselves to values of p in the range between $\omega(n)/(n\sqrt{m})$ and $((2\log n - \omega(n))/m)^{1/2}$, where $\omega(n) \to \infty$.

Equivalence

One of the first interesting problems to be considered is the question as to when the random graphs $G(n,m,p)$ and $\mathbb{G}_{n,p}$ have asymptotically the same properties. Intuitively, it should be the case when the edges of $G(n,m,p)$ occur "almost independently", i.e., when there are no vertices of degree greater than two in M in the generator $\mathbb{G}_{n,m,p}$ of $G(n,m,p)$. Then each of its edges is induced by a vertex of degree two in M, "almost" independently of other edges. One can show that this happens w.h.p. when $p = o\left(1/(nm^{1/3})\right)$, which in turn implies that both random graphs are asymptotically equivalent for all graph properties $\mathscr{P}$. Recall that a graph property $\mathscr{P}$ is defined as a subset of the family of all labeled graphs on vertex set $[n]$, i.e., $\mathscr{P} \subseteq 2^{\binom{n}{2}}$. The following equivalence result is due to Rybarczyk [783] and Fill, Scheinerman and Singer-Cohen [360].

Theorem 10.1. *Let* $0 \leq a \leq 1$, $\mathscr{P}$ *be any graph property,* $p = o\left(1/(nm^{1/3})\right)$ *and*

$$\hat{p} = 1 - \exp\left(-mp^2(1-p)^{n-2}\right). \tag{10.1}$$

Then

$$\mathbb{P}(\mathbb{G}_{n,\hat{p}} \in \mathscr{P}) \to a$$

if and only if

$$\mathbb{P}(G(n,m,p) \in \mathscr{P}) \to a$$

as $n \to \infty$.

Proof. Let X and Y be random variables taking values in a common finite (or countable) set S. Consider the probability measures $\mathscr{L}(X)$ and $\mathscr{L}(Y)$ on S whose values at $A \subseteq S$ are $\mathbb{P}(X \in A)$ and $\mathbb{P}(Y \in A)$. Define the total variation distance between $\mathscr{L}(X)$ and $\mathscr{L}(Y)$ as

$$d_{TV}(\mathscr{L}(X), \mathscr{L}(Y)) = \sup_{A \subseteq S} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|,$$

which is equivalent to

$$d_{TV}(\mathscr{L}(X), \mathscr{L}(Y)) = \frac{1}{2} \sum_{s \in S} |\mathbb{P}(X = s) - \mathbb{P}(Y = s)|.$$

Notice (see Fact 4 of [360]) that if there exists a probability space on which random variables X' and Y' are both defined, with $\mathscr{L}(X) = \mathscr{L}(X')$ and $\mathscr{L}(Y) = \mathscr{L}(Y')$, then

$$d_{TV}(\mathscr{L}(X), \mathscr{L}(Y)) \leq \mathbb{P}(X' \neq Y'). \tag{10.2}$$

Furthermore (see Fact 3 in [360]) if there exist random variables Z and Z' such that $\mathscr{L}(X|Z = z) = \mathscr{L}(Y|Z' = z)$, for all z, then

$$d_{TV}(\mathscr{L}(X), \mathscr{L}(Y)) \leq 2 d_{TV}(\mathscr{L}(Z), \mathscr{L}(Z')). \tag{10.3}$$

We will need one more observation. Suppose that a random variable X has distribution the $\text{Bin}(n, p)$, while a random variable Y has the Poisson distribution, and $\mathbb{E} X = \mathbb{E} Y$. Then

$$d_{TV}(X, Y) = O(p). \tag{10.4}$$

We leave the proofs of (10.2), (10.3) and (10.4) as exercises.
To prove Theorem 10.1 we also need some auxiliary results on a special coupon collector scheme.

Let Z be a non-negative integer valued random variable, r a non-negative integer and γ a real, such that $r\gamma \leq 1$. Assume we have r coupons $Q_1, Q_2, \dots, Q_r$ and one blank coupon B. We make Z independent draws (with replacement), such that in each draw,

$$\mathbb{P}(Q_i \text{ is chosen}) = \gamma, \text{ for } i = 1, 2, \dots, r,$$

and

$$\mathbb{P}(B \text{ is chosen}) = 1 - r\gamma.$$

Let $N_i(Z),\ i = 1, 2, \dots, r$ be a random variable counting the number of times that coupon Q_i was chosen. Furthermore, let

$$X_i(Z) = \begin{cases} 1 & \text{if } N_i(Z) \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The number of different coupons selected is given by

$$X(Z) = \sum_{i=1}^{r} X_i(Z). \tag{10.5}$$

With the above definitions we observe that the following holds.

Lemma 10.2. *If a random variable Z has the Poisson distribution with expectation λ then $N_i(Z),\ i = 1, 2, \dots, r$, are independent and identically Poisson distributed random variables, with expectation $\lambda\gamma$. Moreover the random variable $X(Z)$ has the distribution $Bin(r, 1 - e^{-\lambda\gamma})$.*

Let us consider the following special case of the scheme defined above, assuming that $r = \binom{n}{2}$ and $\gamma = 1/\binom{n}{2}$. Here each coupon represents a distinct edge of K_n.

Lemma 10.3. *Suppose $p = o(1/n)$ and let a random variable Z be the $Bin\left(m, \binom{n}{2} p^2 (1-p)^{n-2}\right)$ distributed, while a random variable Y be the $Bin\left(\binom{n}{2}, 1 - e^{-mp^2(1-p)^{n-2}}\right)$ distributed. Then*

$$d_{TV}\left(\mathscr{L}(X(Z)), \mathscr{L}(Y)\right) = o(1).$$

Proof. Let Z' be a Poisson random variable with the same expectation as Z, i.e.,

$$\mathbb{E} Z' = m \binom{n}{2} p^2 (1-p)^{n-2}.$$

By Lemma 10.2, $X(Z')$ has the binomial distribution

$$\mathrm{Bin}\left(\binom{n}{2}, 1-e^{-mp^2(1-p)^{n-2}}\right),$$

and so, by (10.3) and (10.4), we have

$$\begin{aligned} d_{TV}(\mathscr{L}(Y), \mathscr{L}(X(Z))) \\ = d_{TV}(\mathscr{L}(X(Z')), \mathscr{L}(X(Z))) \le 2d_{TV}(\mathscr{L}(Z'), \mathscr{L}(Z)) \\ \le O\left(\binom{n}{2} p^2(1-p)^{n-2}\right) = O\left(n^2p^2\right) = o(1). \end{aligned}$$

□

Now define a random intersection graph $G_2(n,m,p)$ as follows. Its vertex set is $V = \{1,2,\ldots,n\}$, while $e = \{i,j\}$ is an edge in $G_2(n,m,p)$ iff in a (generator) bipartite random graph $\mathbb{G}_{n,m,p}$, there is a vertex $w \in M$ of degree two such that both i and j are connected by an edge with w.

To complete the proof of our theorem, notice that,

$$\begin{aligned} d_{TV}(\mathscr{L}(G(n,m,p)), \mathscr{L}(\mathbb{G}_{n,\hat{p}})) \le \\ d_{TV}(\mathscr{L}(G(n,m,p)), \mathscr{L}(G_2(n,m,p))) + d_{TV}(\mathscr{L}(G_2(n,m,p)), \mathscr{L}(\mathbb{G}_{n,\hat{p}})) \end{aligned}$$

where $\hat{p}$ is defined in (10.1). Now, by (10.2)

$$\begin{aligned} d_{TV}(\mathscr{L}(G(n,m,p)), \mathscr{L}(G_2(n,m,p))) \\ \le \mathbb{P}(\mathscr{L}(G(n,m,p)) \ne \mathscr{L}(G_2(n,m,p))) \\ \le \mathbb{P}(\exists w \in M \text{ of } \mathbb{G}_{n,m,p} \text{ s.t. } deg(w) > 2) \le m\binom{n}{3} p^3 = o(1), \end{aligned}$$

for $p = o(1/(nm^{1/3})$.

Hence it remains to show that

$$d_{TV}(\mathscr{L}(G_2(n,m,p)), \mathscr{L}(\mathbb{G}_{n,\hat{p}})) = o(1). \tag{10.6}$$

Let Z be distributed as $\mathrm{Bin}\left(m, \binom{n}{2}p^2(1-p)^{n-2}\right)$, $X(Z)$ is defined as in (10.5) and let Y be distributed as $\mathrm{Bin}\left(\binom{n}{2}, 1-e^{-mp^2(1-p)^{n-2}}\right)$. Then the number of edges $|E(G_2(n,m,p))| = X(Z)$ and $|E(\mathbb{G}_{n,\hat{p}}))| = Y$. Moreover for any two graphs G and G' with the same number of edges

$$\mathbb{P}(G_2(n,m,p) = G) = \mathbb{P}(G_2(n,m,p) = G')$$

and

$$\mathbb{P}(\mathbb{G}_{n,\hat{p}} = G) = \mathbb{P}(\mathbb{G}_{n,\hat{p}} = G').$$

Equation (10.6) now follows from Lemma 10.3. The theorem follows immediately. □

For monotone properties (see Chapter 1) the relationship between the classical binomial random graph and the respective intersection graph is more precise and was established by Rybarczyk [783].

Theorem 10.4. *Let* $0 \le a \le 1$*,* $m = n^\alpha, \alpha \ge 3$*, Let* $\mathscr{P}$ *be any monotone graph property. For* $\alpha > 3$*, assume*

$$\Omega(1/(nm^{1/3})) = p = O(\sqrt{\log n/m})$$

while for $\alpha = 3$ *assume* $\left(1/(nm^{1/3})\right) = o(p)$*. Let*

$$\hat{p} = 1 - \exp\left(-mp^2(1-p)^{n-2}\right).$$

If for all $\varepsilon = \varepsilon(n) \to 0$

$$\mathbb{P}(\mathbb{G}_{n,(1+\varepsilon)\hat{p}} \in \mathscr{P}) \to a,$$

then

$$\mathbb{P}(G(n,m,p) \in \mathscr{P}) \to a$$

as $n \to \infty$*.*

Small subgraphs

Let H be any fixed graph. A *clique cover* $\mathscr{C}$ is a collection of subsets of vertex set $V(H)$ such that, each induces a complete subgraph (*clique*) of H, and for every edge $\{u,v\} \in E(H)$, there exists $C \in \mathscr{C}$, such that $u,v \in C$. Hence, the cliques induced by sets from $\mathscr{C}$ exactly cover the edges of H. A clique cover is allowed to have more than one copy of a given set. We say that $\mathscr{C}$ is *reducible* if for some $C \in \mathscr{C}$, the edges of H induced by C are contained in the union of the edges induced by $\mathscr{C} \setminus C$, otherwise $\mathscr{C}$ is *irreducible*. Note that if $C \in \mathscr{C}$ and $\mathscr{C}$ is irreducible, then $|C| \ge 2$.

In this section, $|\mathscr{C}|$ stands for the number of cliques in $\mathscr{C}$, while $\sum \mathscr{C}$ denotes the sum of clique sizes in $\mathscr{C}$, and we put $\sum \mathscr{C} = 0$ if $\mathscr{C} = \emptyset$.

Let $\mathscr{C} = \{C_1, C_2, \ldots, C_k\}$ be a clique cover of H. For $S \subseteq V(H)$ define the following two *restricted clique covers*

$$\mathscr{C}_t[S] := \{C_i \cap S : |C_i \cap S| \ge t, \;\; i = 1,2,\ldots,k\},$$

where $t = 1,2$. For a given S and $t = 1,2$, let

$$\tau_t = \tau_t(H,\mathscr{C},S) = \left(n^{|S|/\Sigma\mathscr{C}_t[S]} m^{|\mathscr{C}_t[S]|/\Sigma\mathscr{C}_t[S]}\right)^{-1}.$$

Finally, let

$$\tau(H) = \min_{\mathscr{C}} \max_{S \subseteq V(H)} \{\tau_1, \tau_2\},$$

where the minimum is taken over all clique covers $\mathscr{C}$ of H. We can in this calculation restrict our attention to irreducible covers.
Karoński, Scheinerman and Singer-Cohen [554] proved the following theorem.

Theorem 10.5. *Let H be a fixed graph and $mp^2 \to 0$. Then*

$$\lim_{n\to\infty} \mathbb{P}(H \subseteq G(n,m,p)) = \begin{cases} 0 & \text{if } p/\tau(H) \to 0 \\ 1 & \text{if } p/\tau(H) \to \infty. \end{cases}$$

As an illustration, we will use this theorem to show the threshold for complete graphs in $G(n,m,p)$, when $m = n^\alpha$, for different ranges of $\alpha > 0$.

Corollary 10.6. *For a complete graph K_h with $h \geq 3$ vertices and $m = n^\alpha$, we have*

$$\tau(K_h) = \begin{cases} n^{-1}m^{-1/h} & \text{for } \alpha \leq 2h/(h-1) \\ n^{-1/(h-1)}m^{-1/2} & \text{for } \alpha \geq 2h/(h-1). \end{cases}$$

Proof. There are many possibilities for clique covers to generate a copy of a complete graph K_h in $G(n,m,p)$. However in the case of K_h only two play a dominating role. Indeed, we will show that for $\alpha \leq \alpha_0$, $\alpha_0 = 2h/(h-1)$ the clique cover $\mathscr{C} = \{V(K_h)\}$ composed of one set containing all h vertices of K_h only matters, while for $\alpha \geq \alpha_0$ the clique cover $\mathscr{C} = \binom{K_h}{2}$, consisting of $\binom{h}{2}$ pairs of endpoints of the edges of K_h, takes the leading role.

Let $V = V(K_h)$ and denote those two clique covers by $\{V\}$ and $\{E\}$, respectively. Observe that for the cover $\{V\}$ the following equality holds.

$$\max_{S\subseteq V}\{\tau_1(K_h,\{V\},S), \tau_2(K_h,\{V\},S)\} = \tau_1(K_h,\{V\},V). \tag{10.7}$$

To see this, check first that for $|S| = h$,

$$\tau_1(K_h,\{V\},V) = \tau_2(K_h,\{V\},V) = n^{-1}m^{-1/h}.$$

For S of size $|S| = s,\ 2 \leq s \leq h-1$ restricting the clique cover $\{V\}$ to S, gives a single s-clique, so for $t = 1,2$

$$\tau_t(K_h, \{V\}, S) = n^{-1}m^{-1/s} < n^{-1}m^{-1/h}.$$

Finally, when $|S| = 1$, then $\tau_1 = (nm)^{-1}$, while $\tau_2 = 0$, both smaller than $n^{-1}m^{-1/h}$, and so equation (10.7) follows.
For the edge-clique cover $\{E\}$ we have a similar expression, viz.

$$\max_{S \subseteq V}\{\tau_1(K_h, \{E\}, S), \tau_2(K_h, \{E\}, S)\} = \tau_1(K_h, \{E\}, V). \tag{10.8}$$

To see this, check first that for $|S| = h$,

$$\tau_1(K_h, \{E\}, V) = n^{-1/(h-1)}m^{-1/2}.$$

Let $S \subset V$, with $s = |S| \leq h-1$, and consider restricted clique covers with cliques of size at most two, and exactly two.

For τ_1, the clique cover restricted to S is the edge-clique cover of K_s, plus a 1-clique for each of the $h-s$ external edges for each vertex of K_s, so

$$\begin{aligned}
&\tau_1(K_h, \{E\}, S) \\
&= \left(n^{s/[s(s-1)+s(h-s)]}m^{[s(s-1)/2+s(h-s)]/[s(s-1)+s(h-s)]}\right)^{-1} \\
&= \left(n^{1/(h-1)}m^{[h-(s+1)/2]/(h-1)}\right)^{-1} \\
&\leq \left(n^{1/(h-1)}m^{h/(2(h-1))}\right)^{-1} \\
&< \left(n^{1/(h-1)}m^{1/2}\right)^{-1},
\end{aligned}$$

while for τ_2 we have

$$\tau_2(K_h, \{E\}, S) = \left(n^{1/(s-1)}m^{1/2}\right)^{-1} < \left(n^{1/(h-1)}m^{1/2}\right)^{-1},$$

thus verifying equation (10.8).

Let $\mathscr{C}$ be any irreducible clique cover of K_h (hence each clique has size at least two). We will show that for any fixed α

$$\tau_1(K_h, \mathscr{C}, V) \geq \begin{cases} \tau_1(K_h, \{V\}, V) & \text{for } \alpha \leq 2h/(h-1) \\ \tau_1(K_h, \{E\}, V) & \text{for } \alpha \geq 2h/(h-1). \end{cases}$$

Thus,

$$\tau_1(K_h, \mathscr{C}, V) \geq \min\{\tau_1(K_h, \{V\}, V), \tau_1(K_h, \{E\}, V)\}. \tag{10.9}$$

Because $m = n^\alpha$ we see that

$$\tau_1(K_h, \mathscr{C}, V) = n^{-x_{\mathscr{C}}(\alpha)},$$

where

$$x_{\mathscr{C}}(\alpha) = \frac{h}{\sum \mathscr{C}} + \frac{|\mathscr{C}|}{\sum \mathscr{C}}\alpha, \quad x_{\{V\}}(\alpha) = 1 + \frac{\alpha}{h}, \quad x_{\{E\}}(\alpha) = \frac{1}{h-1} + \frac{\alpha}{2}.$$

(To simplify notation, below we have replaced $x_{\{V\}}, x_{\{E\}}$ by x_V, x_E, respectively). Notice, that for $\alpha_0 = 2h/(h-1)$ exponents

$$x_V(\alpha_0) = x_E(\alpha_0) = 1 + \frac{2}{h-1}.$$

Moreover, for all values of $\alpha < \alpha_0$ the function $x_V(\alpha) > x_E(\alpha)$, while for $\alpha > \alpha_0$ the function $x_V(\alpha) < x_E(\alpha)$.

Now, observe that $x_{\mathscr{C}}(0) = \frac{h}{\sum \mathscr{C}} \leq 1$ since each vertex is in at least one clique of $\mathscr{C}$. Hence $x_{\mathscr{C}}(0) \leq x_V(0) = 1$. We will show also that $x_{\mathscr{C}}(\alpha) \leq x_V(\alpha)$ for $\alpha > 0$. To see this we need to bound $|\mathscr{C}|/\sum \mathscr{C}$.

Suppose that $u \in V(K_h)$ appears in the fewest number of cliques of $\mathscr{C}$, and let r be the number of cliques $C_i \in \mathscr{C}$ to which u belongs. Then

$$\sum \mathscr{C} = \sum_{i: C_i \ni u} |C_i| + \sum_{i: C_i \not\ni u} |C_i| \geq ((h-1) + r) + 2(|\mathscr{C}| - r),$$

where $h-1$ counts all other vertices aside from u since they must appear in some clique with u.
For any $v \in V(K_h)$ we have

$$\begin{aligned}
\sum \mathscr{C} + |\{i : C_i \ni v\}| - (h-1) &\geq \sum \mathscr{C} + r - (h-1) \\
&\geq (h-1) + r + 2(|\mathscr{C}| - r) + r - (h-1) \\
&= 2|\mathscr{C}|.
\end{aligned}$$

Summing the above inequality over all $v \in V(K_h)$,

$$h \sum \mathscr{C} + \sum \mathscr{C} - h(h-1) \geq 2h|\mathscr{C}|,$$

and dividing both sides by $2h\sum \mathscr{C}$, we finally get

$$\frac{|\mathscr{C}|}{\sum \mathscr{C}} \leq \frac{h+1}{2h} - \frac{h-1}{2\sum \mathscr{C}}.$$

Now, using the above bound,

$$x_{\mathscr{C}}(\alpha_0) = \frac{h}{\sum \mathscr{C}} + \frac{|\mathscr{C}|}{\sum \mathscr{C}} \left(\frac{2h}{h-1} \right)$$

$$\le \frac{h}{\sum\mathscr{C}} + \left(\frac{h+1}{2h} - \frac{h-1}{2\sum\mathscr{C}}\right)\left(\frac{2h}{h-1}\right)$$
$$= 1 + \frac{2}{h-1}$$
$$= x_V(\alpha_0).$$

Now, since $x_{\mathscr{C}}(\alpha) \le x_V(\alpha)$ at both $\alpha = 0$ and $\alpha = \alpha_0$, and both functions are linear, $x_{\mathscr{C}}(\alpha) \le x_V(\alpha)$ throughout the interval $(0, \alpha_0)$.

Since $x_E(\alpha_0) = x_V(\alpha_0)$ we also have $x_{\mathscr{C}}(\alpha_0) \le x_E(\alpha_0)$. The slope of $x_{\mathscr{C}}(\alpha)$ is $\frac{|\mathscr{C}|}{\sum\mathscr{C}}$, and by the assumption that $\mathscr{C}$ consists of cliques of size at least 2, this is at most 1/2. But the slope of $x_E(\alpha)$ is exactly 1/2. Thus for all $\alpha \ge \alpha_0$, $x_{\mathscr{C}}(\alpha) \le x_E(\alpha)$. Hence the bounds given by formula (10.9) hold.

One can show (see [781]) that for any irreducible clique-cover $\mathscr{C}$ that is not $\{V\}$ nor $\{E\}$,

$$\max_S \{\tau_1(K_h, \mathscr{C}, S), \tau_2(K_h, \mathscr{C}, S)\} \ge \tau_1(K_h, \mathscr{C}, V).$$

Hence, by (10.9),

$$\max_S \{\tau_1(K_h, \mathscr{C}, S), \tau_2(K_h, \mathscr{C}, S)\} \ge \min\{\tau_1(K_h, \{V\}, V), \tau_1(K_h, \{E\}, V)\}.$$

This implies that

$$\tau(K_h) = \begin{cases} n^{-1}m^{-1/h} & \text{for } \alpha \le \alpha_0 \\ n^{-1/(h-1)}m^{-1/2} & \text{for } \alpha \ge \alpha_0, \end{cases}$$

which completes the proof of Corollary 10.6. □

To add to the picture of asymptotic behavior of small cliques in $G(n,m,p)$ we will quote the result of Rybarczyk and Stark [781], who with use of Stein's method (see Chapter 26.3) obtained an upper bound on the total variation distance between the distribution of the number of h-cliques and a respective Poisson distribution for any fixed h.

Theorem 10.7. *Let $G(n,m,p)$ be a random intersection graph, where $m = n^{\alpha}$. Let $c > 0$ be a constant and $h \ge 3$ a fixed integer, and X_n be the random variable counting the number of copies of a complete graph K_h in $G(n,m,p)$.*

(i) If $\alpha < \frac{2h}{h-1}$, $p \approx cn^{-1}m^{-1/h}$ then

$$\lambda_n = \mathbb{E}X_n \approx c^h/h!$$

and

$$d_{TV}(\mathscr{L}(X_n), \mathrm{Po}(\lambda_n)) = O\left(n^{-\alpha/h}\right);$$

(ii) If $\alpha = \frac{2h}{h-1}, p \approx cn^{-(h+1)/(h-1)}$ *then*

$$\lambda_n = \mathbb{E}X_n \approx \left(c^h + c^{h(h-1)}\right)/h!$$

and

$$d_{TV}(\mathscr{L}(X_n), \mathrm{Po}(\lambda_n)) = O\left(n^{-2/(h-1)}\right);$$

(iii) If $\alpha > \frac{2h}{h-1}, p \approx cn^{-1/(h-1)}m^{-1/2}$ *then*

$$\lambda_n = \mathbb{E}X_n \approx c^{h(h-1)}/h!$$

and

$$d_{TV}(\mathscr{L}(X_n), \mathrm{Po}(\lambda_n)) = O\left(n^{\left(h-\frac{\alpha(h-1)}{2}-\frac{2}{h-1}\right)} + n^{-1}\right).$$

10.2 Random Geometric Graphs

The graphs we consider in this section are the intersection graphs that we obtain from the intersections of balls in the d-dimensional unit cube, $D = [0,1]^d$ where $d \geq 2$. For simplicity we will only consider $d = 2$ in the text.

We let $\mathscr{X} = \{X_1, X_2, \ldots, X_n\}$ be independently and uniformly chosen from $D = [0,1]^2$. For $r = r(n)$ let $G_{\mathscr{X},r}$ be the graph with vertex set $\mathscr{X}$. We join X_i, X_j by an edge iff X_j lies in the disk

$$B(X_i, r) = \left\{X \in [0,1]^2 : |X - X_i| \leq r\right\}.$$

Here $|\ |$ denotes Euclidean distance.

For a given set $\mathscr{X}$ we see that increasing r can only add edges and so thresholds are usually expressed in terms of upper/lower bounds on the size of r.

The book by Penrose [733] gives a detailed exposition of this model. Our aim here is to prove some simple results that are not intended to be best possible.

Connectivity

The threshold (in terms of r) for connectivity was shown to be identical with that for minimum degree one, by Gupta and Kumar [456]. This was extended to k-connectivity by Penrose [732]. We do not aim for tremendous accuracy. The simple proof of connectivity was provided to us by Tobias Müller [699].

Theorem 10.8. *Let $\varepsilon > 0$ be arbitrarily small and let $r_0 = r_0(n) = \sqrt{\frac{\log n}{\pi n}}$. Then w.h.p.*

$$G_{\mathscr{X},r} \text{ contains isolated vertices if } r \leq (1-\varepsilon)r_0 \tag{10.10}$$
$$G_{\mathscr{X},r} \text{ is connected if } r \geq (1+\varepsilon)r_0 \tag{10.11}$$

Proof. First consider (10.10) and the degree of X_1. Then

$$\mathbb{P}(X_1 \text{ is isolated}) \geq (1-\pi r^2)^{n-1}.$$

The factor $(1-\pi r^2)^{n-1}$ bounds the probability that none of $X_2, X_3, \ldots, X_n$ lie in $B(X_1, r)$, given that $B(X_1, r) \subseteq D$. It is exact for points far enough from the boundary of D.

Now

$$(1-\pi r^2)^{n-1} \geq \left(1 - \frac{(1-\varepsilon)\log n}{n}\right)^n = n^{\varepsilon-1+o(1)}.$$

So if I is the set of isolated vertices then $\mathbb{E}(|I|) \geq n^{\varepsilon-1+o(1)} \to \infty$. Now

$$\mathbb{P}(X_1 \in I \mid X_2 \in I) \leq \left(1 - \frac{\pi r^2}{1-\pi r^2}\right)^{n-2} \leq (1+o(1))\,\mathbb{P}(X_1 \in I).$$

The expression $\left(1 - \frac{\pi r^2}{1-\pi r^2}\right)$ is the probability that a random point does not lie in $B(X_1, r)$, given that it does not lie in $B(X_2, r)$, and that $|X_2 - X_1| \geq 2r$. Equation (10.10) now follows from the Chebyshev inequality (26.3).

Now consider (10.11). Let $\eta \ll \varepsilon$ be a sufficiently small constant and divide D into ℓ_0^2 sub-squares of side length ηr, where $\ell_0 = 1/\eta r$. We refer to these sub-squares as cells. We can assume that η is chosen so that ℓ_0 is an integer. We say that a cell is *good* if contains at least $i_0 = \eta^3 \log n$ members of $\mathscr{X}$ and *bad* otherwise. We next let $K = 100/\eta^2$ and consider the number of bad cells in a $K \times K$ square block of cells.

Lemma 10.9. *Let B be a $K \times K$ square block of cells. The following hold w.h.p.:*

(a) If B is further than $100r$ from the closest boundary edge of D then B contains at most $k_0 = (1-\varepsilon/10)\pi/\eta^2$ bad cells.

(b) If B is within distance $100r$ of exactly one boundary edge of D then B contains at most $k_0/2$ bad cells.

(c) If B is within distance $100r$ *of two boundary edges of D then B contains no bad cells.*

Proof. (a) There are less than $\ell_0^2 < n$ such blocks. Furthermore, the probability that a fixed block contains k_0 or more bad cells is at most

$$\binom{K^2}{k_0}\left(\sum_{i=0}^{i_0}\binom{n}{i}(\eta^2r^2)^i(1-\eta^2r^2)^{n-i}\right)^{k_0}$$
$$\leq \left(\frac{K^2e}{k_0}\right)^{k_0}\left(2\left(\frac{ne}{i_0}\right)^{i_0}(\eta^2r^2)^{i_0}e^{-\eta^2r^2(n-i_0)}\right)^{k_0}. \quad (10.12)$$

Here we have used Corollary 27.4 to obtain the LHS of (10.12).

Now

$$\left(\frac{ne}{i_0}\right)^{i_0}(\eta^2r^2)^{i_0}e^{-\eta^2r^2(n-i_0)}$$
$$\leq n^{O(\eta^3\log(1/\eta)-\eta^2(1+\varepsilon-o(1))/\pi} \leq n^{-\eta^2(1+\varepsilon/2)/\pi}, \quad (10.13)$$

for η sufficiently small. So we can bound the RHS of (10.12) by

$$\left(\frac{2K^2en^{-\eta^2(1+\varepsilon/2)/\pi}}{(1-\varepsilon/10)\pi/\eta^2}\right)^{(1-\varepsilon/10)\pi/\eta^2} \leq n^{-1-\varepsilon/3}. \quad (10.14)$$

Part (a) follows after inflating the RHS of (10.14) by n to account for the number of choices of block.

(b) Replacing k_0 by $k_0/2$ replaces the LHS of (10.14) by

$$\left(\frac{4K^2en^{-\eta^2(1+\varepsilon/2)/\pi}}{(1-\varepsilon/10)\pi/2\eta^2}\right)^{(1-\varepsilon/10)\pi/2\eta^2} \leq n^{-1/2-\varepsilon/6}. \quad (10.15)$$

Observe now that the number of choices of block is $O(\ell_0) = o(n^{1/2})$ and then Part (b) follows after inflating the RHS of (10.15) by $o(n^{1/2})$ to account for the number of choices of block.

(c) Equation (10.13) bounds the probability that a single cell is bad. The number of cells in question in this case is $O(1)$ and (c) follows. $\square$

We now do a simple geometric computation in order to place a lower bound on the number of cells within a ball $B(X,r)$.

Lemma 10.10. *A half-disk of radius $r_1 = r(1-\eta\sqrt{2})$ with diameter part of the grid of cells contains at least $(1-2\eta^{1/2})\pi/2\eta^2$ cells.*

Proof. We place the half-disk in a $2r_1 \times r_1$ rectangle. Then we partition the rectangle into $\zeta_1 = r_1/r\eta$ rows of $2\zeta_1$ cells. The circumference of the circle will cut the ith row at a point which is $r_1(1-i^2\eta^2)^{1/2}$ from the centre of the row. Thus the ith row will contain at least $2\left\lfloor r_1(1-i^2\eta^2)^{1/2}/r\eta \right\rfloor$ complete cells. So the half-disk contains at least

$$\begin{aligned}\frac{2r_1}{r\eta}\sum_{i=1}^{1/\eta}((1-i^2\eta^2)^{1/2}-\eta) &\geq \frac{2r_1}{r\eta}\int_{x=1}^{1/\eta-1}((1-x^2\eta^2)^{1/2}-\eta)dx \\ &= \frac{2r_1}{r\eta^2}\int_{\theta=\arcsin(\eta)}^{\arcsin(1-\eta)}(\cos^2(\theta)-\eta\cos(\theta))d\theta \\ &\geq \frac{2r_1}{r\eta^2}\left[\frac{\theta}{2}-\frac{\sin(2\theta)}{4}-\eta\right]_{\theta=\arcsin(\eta)}^{\arcsin(1-\eta)}.\end{aligned}$$

Now

$$\arcsin(1-\eta) \geq \frac{\pi}{2} - 2\eta^{1/2} \text{ and } \arcsin(\eta) \leq 2\eta.$$

So the number of cells is at least

$$\frac{2r_1}{r\eta^2}\left(\frac{\pi}{4} - \eta^{1/2} - \eta\right).$$

This completes the proof of Lemma 10.10. □

We deduce from Lemmas 10.9 and 10.10 that

$$X \in \mathscr{X} \text{ implies that } B(X, r_1) \cap D \text{ contains at least one good cell.} \tag{10.16}$$

Now let Γ be the graph whose vertex set consists of the good cells and where cells c_1, c_2 are adjacent iff their centres are within distance r_1. Note that if c_1, c_2 are adjacent in Γ then any point in $\mathscr{X} \cap c_1$ is adjacent in $G_{\mathscr{X},r}$ to any point in $\mathscr{X} \cap c_2$. It follows from (10.16) that all we need to do now is show that Γ is connected.

It follows from Lemma 10.9 that at most π/η^2 rows of a $K \times K$ block contain a bad cell. Thus more than 95% of the rows and of the columns of such a block are free of bad cells. Call such a row or column good. The cells in a good row or column of some $K \times K$ block form part of the same component of Γ. Two neighboring blocks must have two touching good rows or columns so the cells in a good row or column of some block form part of a single component of Γ.

Any other component C must be in a block bounded by good rows and columns. But the existence of such a component means that it is surrounded by bad cells. Now consider the half disk H of radius r_1 that is centered at c. We can assume (i)

H is contained entirely in B and (ii) at least $(1-2\eta^{1/2})\pi/2\eta^2 - (1-\eta\sqrt{2})/\eta \geq (1-3\eta^{1/2})\pi/2\eta^2$ cells in H are bad. Property (i) arises because cells above c whose centers are at distance at most r_1 are all bad and for (ii) we have discounted any bad cells on the diameter through c that might be in C. This provides half the claimed bad cells. We obtain the rest by considering a lowest cell of C. Near the boundary, we only need to consider one half disk with diameter parallel to the closest boundary. Finally observe that there are no bad cells close to a corner. □

Hamiltonicity

The first inroads on the Hamilton cycle problem were made by Diaz, Mitsche and Pérez-Giménez [291]. Best possible results were later given by Balogh, Bollobás, Krivelevich, Müller and Walters [64] and by Müller, Pérez-Giménez and Wormald [700]. As one might expect Hamiltonicity has a threshold at r close to r_0. We now have enough to prove the result from [291].

We start with a simple lemma, taken from [64].

Lemma 10.11. *The subgraph Γ contains a spanning tree of maximum degree at most six.*

Proof. Consider a spanning tree T of γ that minimises the sum of the lengths of the edges joining the centres of the cells. Then T does not have any vertex of degree greater than 6. This is because, if centre v were to have degree at least 7, then there are two neighboring centres u,w of v such that the angle between the line segments $[v,u]$ and $[v,w]$ is strictly less than 60 degrees. We can assume without loss of generality that $[v,u]$ is shorter than $[v,w]$. Note that if we remove the edge $\{v,w\}$ and add the edge $\{u,w\}$ then we obtain another spanning tree but with strictly smaller total edge-length, a contradiction. Hence T has maximum degree at most 6. □

Theorem 10.12. *Suppose that $r \geq (1+\varepsilon)r_0$. Then w.h.p. $G_{\mathscr{X},r}$ is Hamiltonian.*

Proof. We begin with the tree T promised by Lemma 10.11. Let c be a good cell. We partition the points of $\mathscr{X} \cap c$ into $2d$ roughly equal size sets $P_1, P_2, \ldots, P_{2d}$ where $d \leq 6$ is the degree of c in T. Since, the points of $\mathscr{X} \cap c$ form a clique in $G = G_{\mathscr{X},r}$ we can form $2d$ paths in G from this partition.

We next do a walk W through T e.g. by Breadth First Search that goes through each edge of T twice and passes through each node of Γ a number of times equal to twice its degree in Γ. Each time we pass through a node we traverse the vertices of a new path described in the previous paragraph. In this way we create a cycle H that goes through all the points in $\mathscr{X}$ that lie in good cells.

Now consider the points P in a bad cell c with centre x. We create a path in G through P with endpoints x,y, say. Now choose a good cell c' contained in the ball $B(x,r_1)$ and then choose an edge $\{u,v\}$ of H in the cell c'. We merge the points in P into H by deleting $\{u,v\}$ and adding $\{x,u\},\{y,v\}$. To make this work, we must be careful to ensure that we only use an edge of H at most once. But there are $\Omega(\log n)$ edges of H in each good cell and there are $O(1)$ bad cells within distance $2r$ say of any good cell and so this is easily done. $\square$

Chromatic number

We look at the chromatic number of $G_{\mathscr{X},r}$ in a limited range. Suppose that $n\pi r^2 = \frac{\log n}{\omega_r}$ where $\omega_r \to \infty, \omega_r = O(\log n)$. We are below the threshold for connectivity here. We will show that w.h.p.

$$\chi(G_{\mathscr{X},r}) \approx \Delta(G_{\mathscr{X},r}) \approx cl(G_{\mathscr{X},r})$$

where will use cl to denote the size of the largest clique. This is a special case of a result of McDiarmid [665].

We first bound the maximum degree.

Lemma 10.13.

$$\Delta(G_{\mathscr{X},r}) \approx \frac{\log n}{\log \omega_r} \text{ w.h.p.}$$

Proof. Let Z_k denote the number of vertices of degree k and let $Z_{\geq k}$ denote the number of vertices of degree at least k. Let $k_0 = \frac{\log n}{\omega_d}$ where $\omega_d \to \infty$ and $\omega_d = o(\omega_r)$. Then

$$\mathbb{E}(Z_{\geq k_0}) \leq n\binom{n}{k_0}(\pi r^2)^{k_0} \leq n\left(\frac{ne\omega_d \log n}{n\omega_r \log n}\right)^{\frac{\log n}{\omega_d}} = n\left(\frac{e\omega_d}{\omega_r}\right)^{\frac{\log n}{\omega_d}}.$$

So,

$$\log(\mathbb{E}(Z_{\geq k_0})) \leq \frac{\log n}{\omega_d}(\omega_d + 1 + \log \omega_d - \log \omega_r). \tag{10.17}$$

Now let $\varepsilon_0 = \omega_r^{-1/2}$. Then if

$$\omega_d + \log \omega_d + 1 \leq (1-\varepsilon_0)\log \omega_r$$

then (10.17) implies that $\mathbb{E}(Z_k) \to 0$. This verifies the upper bound on Δ claimed in the lemma.

Now let $k_1 = \frac{\log n}{\widehat{\omega}_d}$ where $\widehat{\omega}_d$ is the solution to

$$\widehat{\omega}_d + \log \widehat{\omega}_d + 1 = (1+\varepsilon_0)\log \omega_r.$$

Next let M denote the set of vertices that are at distance greater than r from any edge of D. Let M_k be the set of vertices of degree k in M. If $\widehat{Z}_k = |M_k|$ then

$$\mathbb{E}(\widehat{Z}_{k_1}) \geq n\,\mathbb{P}(X_1 \in M) \times \binom{n-1}{k_1}(\pi r^2)^{k_1}(1-\pi r^2)^{n-1-k_1}.$$

$\mathbb{P}(X_1 \in M) \geq 1-4r$ and so

$$\begin{aligned}\mathbb{E}(\widehat{Z}_{k_1}) &\geq (1-4r)\frac{n}{3k_1^{1/2}}\left(\frac{(n-1)e}{k_1}\right)^{k_1}(\pi r^2)^{k_1} e^{-n\pi r^2/(1-\pi r^2)} \\ &\geq (1-o(1))\frac{n^{1-1/\omega_r}}{3k_1^{1/2}}\left(\frac{e\widehat{\omega}_d}{\omega_r}\right)^{\frac{\log n}{\widehat{\omega}_d}}.\end{aligned}$$

So,

$$\begin{aligned}&\log(\mathbb{E}(\widehat{Z}_{k_1})) \geq \\ &\quad -o(1) - O(\log\log n) + \frac{\log n}{\widehat{\omega}_d}\left(\widehat{\omega}_d + 1 + \log \widehat{\omega}_d - \log \omega_r - \frac{\widehat{\omega}_d}{\omega_r}\right) \\ &\qquad = \Omega\left(\frac{\varepsilon_0 \log n \log \omega_r}{\widehat{\omega}_d}\right) = \Omega\left(\frac{\log n}{\omega_r^{1/2}}\right) \to \infty.\end{aligned}$$

An application of the Chebyshev inequality finishes the proof of the lemma. Indeed,

$$\begin{aligned}&\mathbb{P}(X_1, X_2 \in M_k) \leq \mathbb{P}(X_1 \in M)\,\mathbb{P}(X_2 \in M) \times \\ &\quad \left(\mathbb{P}(X_2 \in B(X_1, r)) + \left(\binom{n-1}{k_1}(\pi r^2)^{k_1}(1-\pi r^2)^{n-2k_1-2}\right)^2\right) \\ &\qquad \leq (1+o(1))\,\mathbb{P}(X_1 \in M_k)\,\mathbb{P}(X_2 \in M_k).\end{aligned}$$

□

Now $\mathrm{cl}(G_{\mathscr{X},r}) \leq \Delta(G_{\mathscr{X},r}) + 1$ and so we now lower bound $\mathrm{cl}(G_{\mathscr{X},r})$ w.h.p. But this is easy. It follows from Lemma 10.13 that w.h.p. there is a vertex X_j with at least $(1-o(1))\frac{\log n}{\log(4\omega_r)}$ vertices in its $r/2$ ball $B(X_j, r/2)$. But such a ball provides a clique of size $(1-o(1))\frac{\log n}{\log(4\omega_r)}$. We have therefore proved

Theorem 10.14. *Suppose that $n\pi r^2 = \frac{\log n}{\omega_r}$ where $\omega_r \to \infty, \omega_r = O(\log n)$. Then w.h.p.*

$$\chi(G_{\mathscr{X},r}) \approx \Delta(G_{\mathscr{X},r}) \approx cl(G_{\mathscr{X},r}) \approx \frac{\log n}{\log \omega_r}.$$

We now consider larger r.

Theorem 10.15. *Suppose that $n\pi r^2 = \omega_r \log n$ where $\omega_r \to \infty, \omega_r = o(n/\log n)$. Then w.h.p.*

$$\chi(G_{\mathscr{X},r}) \approx \frac{\omega_r \sqrt{3} \log n}{2\pi}.$$

Proof. First consider the triangular lattice in the plane. This is the set of points $T = \{m_1 a + m_2 b : m_1, m_2 \in \mathbb{Z}\}$ where $a = (0,1), b = (1/2, \sqrt{3}/2)$, see Figure 10.1.

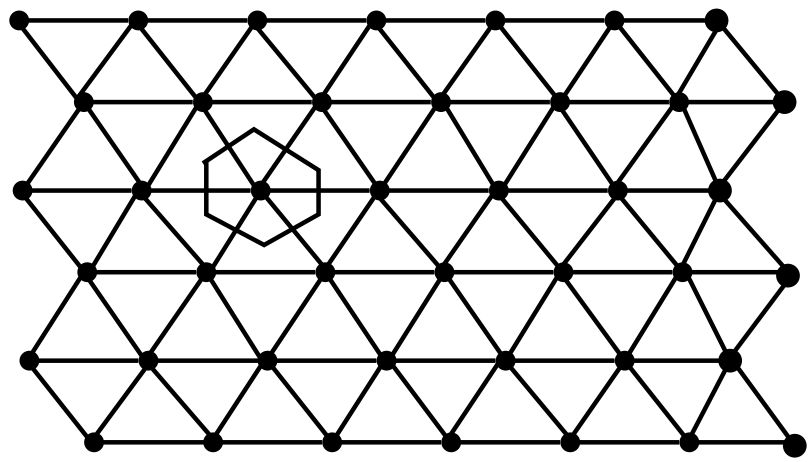


Figure 10.1: The small hexagon is an example of a C_v.

As in the diagram, each $v \in T$ can be placed at the centre of a hexagon C_v. The C_v's intersect on a set of measure zero and each C_v has area $\sqrt{3}/2$ and is contained in $B(v, 1/\sqrt{3})$. Let $\Gamma(T,d)$ be the graph with vertex set T where two vertices $x, y \in T$ are joined by an edge if their Euclidean distance $|x-y| < d$.

Lemma 10.16. *[McDiarmid and Reed [667]]*

$$\chi(\Gamma(T,d)) \leq (d+1)^2.$$

Proof. Let $\delta = \lceil d \rceil$. Let R denote a $\delta \times \delta$ rhombus made up of triangles of T with one vertex at the origin. This rhombus has δ^2 vertices, if we exclude those at the top and right hand end. We give each of these vertices a distinct color and then tile the plane with copies of R. This is a proper coloring, by construction. $\square$

Armed with this lemma we can easily get an upper bound on $\chi(G_{\mathscr{X},r})$. Let $\delta = 1/\omega_r^{1/3}$ (with this value $ns^2 \gg \log n$) and let $s = \delta r$. Let sT be the contraction of the lattice T by a factor s i.e. $sT = \{sx : x \in T\}$. Then if $v \in sT$ let sC_v be the hexagon with centre v, sides parallel to the sides of C_v but reduced by a factor s. $|\mathscr{X} \cap sC_v|$ is distributed as $\text{Bin}(n, s^2\sqrt{3}/2)$. So the Chernoff bounds imply that with probability $1 - o(n^{-1})$,

$$sC_v \text{ contains } \leq \theta = \left\lceil (1 + \omega_r^{-1/8}) ns^2\sqrt{3}/2 \right\rceil \text{ members of } \mathscr{X}. \tag{10.18}$$

Let $\rho = r + 2s/\sqrt{3}$. We note that if $x \in C_v$ and $y \in C_w$ and $|x - y| \leq r$ then $|v - w| \leq \rho$. Thus, given a proper coloring φ of $\Gamma(sT, \rho)$ with colors $[q]$ we can w.h.p. extend it to a coloring ψ of $G_{\mathscr{X},r}$ with color's $[q] \times [\theta]$. If $x \in sC_v$ and $\varphi(x) = a$ then we let $\psi(x) = (a, b)$ where b ranges over $[\theta]$ as x ranges over $sC_v \cap \mathscr{X}$. So, w.h.p.

$$\chi(G_{\mathscr{X},r}) \leq \theta\chi(\Gamma(sT,\rho)) = \theta\chi(\Gamma(T,\rho/s)) \leq \theta\left(\frac{\rho}{s} + 1\right)^2 \approx \frac{ns^2\sqrt{3}}{2} \times \frac{r^2}{s^2} = \frac{\omega_r\sqrt{3}\log n}{2\pi}. \tag{10.19}$$

For the lower bound we use a classic result on packing disks in the plane.

Lemma 10.17. *Let $A_n = [0,n]^2$ and $\mathscr{C}$ be a collection of disjoint disks of unit area that touch A_n. Then $|\mathscr{C}| \leq (1 + o(1))\pi n^2/\sqrt{12}$.*

Proof. Thue's theorem states that the densest packing of disjoint same size disks in the plane is the hexagonal packing which has density $\lambda = \pi/\sqrt{12}$. Let $\mathscr{C}'$ denote the disks that are contained entirely in A_n. Then we have

$$|\mathscr{C}'| \geq |\mathscr{C}| - O(n) \text{ and } |\mathscr{C}'| \leq \frac{\pi n^2}{\sqrt{12}}.$$

The first inequality comes from the fact that if $C \in \mathscr{C} \setminus \mathscr{C}'$ then it is contained in a perimeter of width $O(1)$ surrounding A_n. $\square$

Now consider the subgraph H of $G_{\mathscr{X},r}$ induced by the points of $\mathscr{X}$ that belong to the square with centre $(1/2, 1/2)$ and side $1 - 2r$. It follows from Lemma 10.17 that if $\alpha(H)$ is the size of the largest independent set in H then $\alpha(H) \leq (1 + o(1))2/r^2\sqrt{3}$. This is because if S is an independent set of H then the disks

$B(x,r/2)$ for $x \in S$ are necessarily disjoint. Now using the Chernoff bounds, we see that w.h.p. H contains at least $(1-o(1))n$ vertices. Thus

$$\chi(G_{\mathscr{X},r}) \geq \chi(H) \geq \frac{|V(H)|}{\alpha(H)} \geq (1-o(1))\frac{r^2\sqrt{3}n}{2} = (1-o(1))\frac{\omega_r\sqrt{3}\log n}{2\pi}.$$

This completes the proof of Theorem 10.15. □

10.3 Exercises

10.3.1 Show that if $p = \omega(n)/\left(n\sqrt{m}\right)$, and $\omega(n) \to \infty$, then $G(n,m,p)$ has w.h.p. at least one edge.

10.3.2 Show that if $p = (2\log n + \omega(n))/m)^{1/2}$ and $\omega(n) \to -\infty$ then w.h.p. $G(n,m,p)$ is not complete.

10.3.3 Prove that the bound (10.2) holds.

10.3.4 Prove that the bound (10.3) holds.

10.3.5 Prove that the bound (10.4) holds.

10.3.6 Prove the claims in Lemma 10.2.

10.3.7 Let X denotes the number of isolated vertices in the binomial random intersection graph $G(n,m,p)$, where $m = n^\alpha,\ \alpha > 0$. Show that if

$$p = \begin{cases} (\log n + \varphi(n))/m & \text{when } \alpha \leq 1 \\ \sqrt{(\log n + \varphi(n))/(nm)} & \text{when } \alpha > 1, \end{cases}$$

then $\mathbb{E}X \to e^{-c}$ if $\lim_{n\to\infty} \varphi(n) \to c$, for any real c.

10.3.8 Find the variance of the random variable X counting isolated vertices in $G(n,m,p)$.

10.3.9 Let Y be a random variable which counts vertices of degree greater than one in $G(n,m,p)$, with $m = n^\alpha$ and $\alpha > 1$. Show that for $p^2m^2n \gg \log n$

$$\lim_{n\to\infty} \mathbb{P}\left(Y > 2p^2m^2n\right) = 0.$$

10.3.10 Suppose that $r \geq (1+\varepsilon)r_0$, as in Theorem 10.8. Show that if $1 \leq k = O(1)$ then $G_{\mathscr{X},r}$ is k-connected w.h.p.

10.3.11 Show that if $2 \le k = O(1)$ and $r \gg n^{-\frac{k}{2(k-1)}}$ then w.h.p. $G_{\mathscr{X},r}$ contains a k-clique. On the other hand, show that if $r = o(n^{-\frac{k}{2(k-1)}})$ then $G_{\mathscr{X},r}$ contains no k-clique.

10.3.12 Suppose that $r \gg \sqrt{\frac{\log n}{n}}$. Show that w.h.p. the diameter of $G_{\mathscr{X},r} = \Theta\left(\frac{1}{r}\right)$.

10.3.13 Suppose that $r \ge (1+\varepsilon)r_0$, as in Theorem 10.8. Show that if $2 \le k = O(1)$ then $G_{\mathscr{X},r}$ has k edge disjoint Hamilton cycles w.h.p.

10.3.14 Given $\mathscr{X}$ and an integer k we define the k-nearest neighbor graph $G_{k-NN,\mathscr{X}}$ as follows: We add an edge between x and y of $\mathscr{X}$ iff y is one of x's k nearest neighbors, in Euclidean distance or vice-versa. Show that if $k \ge C\log n$ for a sufficiently large C then $G_{k-NN,\mathscr{X}}$ is connected w.h.p.

10.3.15 Suppose that we independently deposit n random black points $\mathscr{X}_b$ and n random white points $\mathscr{X}_w$ into D. Let $B_{\mathscr{X}_b,\mathscr{X}_w,r}$ be the bipartite graph where we connect $x \in \mathscr{X}_b$ with $\mathscr{X}_w$ iff $|x-y| \le r$. Show that if $r \gg \sqrt{\frac{\log n}{n}}$ then w.h.p. $B_{\mathscr{X}_b,\mathscr{X}_w,r}$ contains a perfect matching.

10.4 Notes

Binomial Random Intersection Graphs

For $G(n,m,p)$ with $m = n^{\alpha}$, α constant, Rybarczyk and Stark [782] provided a condition, called *strictly α-balanced* for the Poisson convergence for the number of induced copies of a fixed subgraph, thus complementing the results of Theorem 10.5 and generalising Theorem 10.7. (Thresholds for small subgraphs in a related model of random intersection digraph are studied by Kurauskas [614]).

Rybarczyk [784] introduced a coupling method to find thresholds for many properties of the binomial random intersection graph. The method is used to establish *sharp threshold functions* for k-connectivity, the existence of a perfect matching and the existence of a Hamilton cycle.

Stark [815] determined the distribution of the degree of a typical vertex of $G(n,m,p)$, $m = n^{\alpha}$ and showed that it changes sharply between $\alpha < 1, \alpha = 1$ and $\alpha > 1$.

Behrisch [86] studied the evolution of the order of the largest component in $G(n,m,p)$, $m = n^{\alpha}$ when $\alpha \ne 1$. He showed that when $\alpha > 1$ the random graph $G(n,m,p)$ behaves like $\mathbb{G}_{n,p}$ in that a giant component of size order n appears w.h.p. when the expected vertex degree exceeds one. This is not the case when $\alpha < 1$. There is a jump in the order of size of the largest component, but not to one of linear size. Further study of the component structure of $G(n,m,p)$ for $\alpha = 1$ is due to Lageras and Lindholm in [616].

Behrisch, Taraz and Ueckerdt [87] study the evolution of the chromatic number of a random intersection graph and showed that, in a certain range of parameters, these random graphs can be colored optimally with high probability using various greedy algorithms.

Uniform Random Intersection Graphs

Uniform random intersection graphs differ from the binomial random intersection graph in the way a subset of the set M is defined for each vertex of V. Now for every $k = 1,2,\ldots,n$, each S_k has fixed size r and is randomly chosen from the set M. We use the notation $G(n,m,r)$ for an r-uniform random intersection graph. This version of a random intersection graph was introduced by Eschenauer and Gligor [340] and, independently, by Godehardt and Jaworski [439].

Bloznelis, Jaworski and Rybarczyk [123] determined the emergence of the giant component in $G(n,m,r)$ when $n(\log n)^2 = o(m)$. A precise study of the phase transition of $G(n,m,r)$ is due to Rybarczyk [785]. She proved that if $c > 0$ is a constant, $r = r(n) \geq 2$ and $r(r-1)n/m \approx c$, then if $c < 1$ then w.h.p. the largest component of $G(n,m,r)$ is of size $O(\log n)$, while if $c > 1$ w.h.p. there is a single giant component containing a constant fraction of all vertices, while the second largest component is of size $O(\log n)$.

The connectivity of $G(n,m,r)$ was studied by various authors, among them by Eschenauer and Gligor [340] followed by DiPietro, Mancini, Mei, Panconesi and Radhakrishnan [298],
Blackbourn and Gerke [112] and Yagan and Makowski [861]. Finally, Rybarczyk [785] determined the sharp threshold for this property. She proved that if $c > 0$ is a constant, $\omega(n) \to \infty$ as $n \to \infty$ and $r^2n/m = \log n + \omega(n)$, then similarly as in $\mathbb{G}_{n,p}$, the uniform random intersection graph $G(n,m,r)$ is disconnected w.h.p. if $\omega(n) \to \infty$, is connected w.h.p. if $\omega(n) \to \infty$, while the probability that $G(n,m,r)$ is connected tends to $e^{-e^{-c}}$ if $\omega(n) \to c$. The Hamiltonicity of $G(n,m,r)$ was studied in [126] and by Nicoletseas, Raptopoulos and Spirakis [718].

If in the uniform model we require $|S_i \cap S_j| \geq s$ to connect vertices i and j by an edge, then we denote this random intersection graph by $G_s(n,m,r)$. Bloznelis, Jaworski and Rybarczyk [123] studied phase transition in $G_s(n,m,r)$. Bloznelis and Łuczak [125] proved that w.h.p. for even n the threshold for the property that $G_s(n,m,r)$ contains a perfect matching is the same as that for $G_s(n,m,r)$ being connected. Bloznelis and Rybarczyk [127] show that w.h.p. the edge density threshold for the property that each vertex of $G_s(n,m,r)$ has degree at least k is the same as that for $G_s(n,m,r)$ being k-connected (for related results see [866]).

Generalized Random Intersection Graphs

Godehardt and Jaworski [439] introduced a model which generalizes both the binomial and uniform models of random intersection graphs. Let P be a probability measure on the set $\{0,1,2,\ldots,m\}$. Let $V=\{1,2,\ldots,n\}$ be the vertex set. Let $M=\{1,2,\ldots,m\}$ be the set of *attributes*. Let $S_1,S_2,\ldots,S_n$ be independent random subsets of M such that for any $v\in V$ and $S\subseteq M$ we have $\mathbb{P}(S_v=S)=P(|S|)/\binom{m}{|S|}$. If we put an edge between any pair of vertices i and j when $S_i\cap S_j\neq\emptyset$, then we denote such a random intersection graph as $G(n,m,P)$, while if the edge is inserted if $|S_i\cap S_i|\geq s,\ s\geq 1$, the respective graph is denoted as $G_s(n,m,P)$. Bloznelis [116] extends these definitions to random intersection digraphs.

The study of the degree distribution of a typical vertex of $G(n,m,P)$ is given in [524], [280] and [114], see also [525]. Bloznelis (see [115] and [117]) shows that the order of the largest component L_1 of $G(n,m,P)$ is asymptotically equal to $n\rho$, where ρ denotes the non-extinction probability of a related multi-type Poisson branching process. Kurauskas and Bloznelis [615] study the asymptotic order of the clique number of the sparse random intersection graph $G_s(n,m,P)$.

Finally, a dynamic approach to random intersection graphs is studied by Barbour and Reinert [77], Bloznelis and Karoński [124], Bloznelis and Goetze [121] and Britton, Deijfen, Lageras and Lindholm [196].

One should also notice that some of the results on the connectivity of random intersection graphs can be derived from the corresponding results for random hyperghraphs, see for example [585], [795] and [440].

Inhomogeneous Random Intersection Graphs

Nicoletseas, Raptopoulos and Spirakis [717] have introduced a generalisation of the binomial random intersection graph $G(n,m,p)$ in the following way. As before let n,m be positive integers and let $0\leq p_i\leq 1, i=1,2,\ldots,m$. Let $V=\{1,2,\ldots,n\}$ be the set of vertices of our graph and for every $1\leq k\leq n$, let S_k be a random subset of the set $M=\{1,2,\ldots,m\}$ formed by selecting ith element of M independently with probability p_i. Let $\mathbf{p}=(p_i)_{i=1}^m$. We define the *inhomogeneous random intersection graph* $G(n,m,\mathbf{p})$ as the intersection graph of sets $S_k,\ k=1,2,\ldots n$. Here two vertices i and j are adjacent in $G(n,m,\mathbf{p})$ if and only if $S_i\cap S_j\neq\emptyset$. Several asymptotic properties of the random graph $G(n,m,\mathbf{p})$ were studied, such as: large independent sets (in [718]), vertex degree distribution (by Bloznelis and Damarackas in [118]), sharp threshold functions for connectivity, matchings and Hamiltonian cycles (by Rybarczyk in [784]) as well as the size of the largest component (by Bradonjić, Elsässer, Friedrich, Sauerwald and Stauffer in [193]).

To learn more about different models of random intersection graphs and about other results we refer the reader to recent review papers [119] and [120].

Random Geometric Graphs

McDiarmid and Müller [666] gives the leading constant for the chromatic number when the average degree is $\Theta(\log n)$. The paper also shows a "surprising" phase change for the relation between χ and ω. Also the paper extends the setting to arbitrary dimensions. Müller [698] proves a two-point concentration for the clique number and chromatic number when $nr^2 = o(\log n)$.

Blackwell, Edmonson-Jones and Jordan [113] studied the spectral properties of the adjacency matrix of a random geometric graph (RGG). Rai [758] studied the spectral measure of the transition matrix of a simple random walk. Preciado and Jadbabaie [751] studied the spectrum of RGG's in the context of the spreading of viruses.

Sharp thresholds for monotone properties of RGG's were shown by McColm [659] in the case $d = 1$ viz. a graph defined by the intersection of random subintervals. And for all $d \geq 1$ by Goel, Rai and Krishnamachari [441].

First order expressible properties of random points
$\mathscr{X} = \{X_1, X_2, \ldots, X_n\}$ on a unit circle were studied by McColm [658]. The graph has vertex set $\mathscr{X}$ and vertices are joined by an edge if and only if their angular distance is less than some parameter d. He showed among other things that for each fixed d, the set of a.s. FO sentences in this model is a complete non-categorical theory. McColm's results were anticipated in a more precise paper [438] by Godehardt and Jaworski, where the case $d = 1$, i.e., the evolution a random interval graph, was studied.

Diaz, Penrose, Petit and Serna [294] study the approximability of several layout problems on a family of RGG's. The layout problems that they consider are bandwidth, minimum linear arrangement, minimum cut width, minimum sum cut, vertex separation, and edge bisection. Diaz, Grandoni and Marchetti-Spaccemela [293] derive a constant expected approximation algorithm for the β-balanced cut problem on random geometric graphs: find an edge cut of minimum size whose two sides contain at least βn vertices each.

Bradonjić, Elsässer, Friedrich, Sauerwald and Stauffer [192] studied the broadcast time of RGG's. They study a regime where there is likely to be a single giant component and show that w.h.p. their broadcast algorithm only requires $O(n^{1/2}/r + \log n)$ rounds to pass information from a single vertex, to every vertex of the giant. They show on the way that the diameter of the giant is $\Theta(n^{1/2}/r)$ w.h.p. Friedrich, Sauerwald and Stauffer [382] extended this to higher dimensions.

A recent interesting development can be described as *Random Hyperbolic Graphs*. These are related to the graphs of Section 10.2 and are posed as models of real world networks. Here points are randomly embedded into hyperbolic, as opposed to Euclidean space. See for example Bode, Fountoulakis and Müller [128], [129]; Candellero and Fountoulakis [208]; Chen, Fang, Hu and Mahoney [218]; Friedrich and Krohmer [381]; Krioukov, Papadopolous, Kitsak, Vahdat and Boguñá [588]; Fountoulakis [371]; Gugelmann, Panagiotou and Peter [455]; Papadopolous, Krioukov, Boguñá and Vahdat [727]. One version of this model is described in [371]. The models are a little complicated to describe and we refer the reader to the above references.

Chapter 11

Digraphs

In graph theory, we sometimes orient edges to create a directed graph or digraph. It is natural to consider randomly generated digraphs and this chapter discusses the component size and connectivity of the simplest model $\mathbb{D}_{n,p}$. Hamiltonicity is discussed in the final section.

11.1 Strong Connectivity

In this chapter we study the random digraph $\mathbb{D}_{n,p}$. This has vertex set $[n]$ and each of the $n(n-1)$ possible edges occurs independently with probability p. We will first study the size of the strong components of $\mathbb{D}_{n,p}$.

Recall the definition of strong components: Given a digraph $D = (V,A)$ we define the relation ρ on V by $x\rho y$ if there is a path from x to y in D and there is a path from y to x in D. It is easy to show that ρ is an equivalence relation and the equivalence classes are called the strong components of D.

Strong component sizes: sub-critical region.

Theorem 11.1. *Let $p = c/n$, where c is a constant, $c < 1$. Then w.h.p.*

(i) all strong components of $\mathbb{D}_{n,p}$ are either cycles or single vertices,

(ii) the number of vertices on cycles is at most ω, for any $\omega = \omega(n) \to \infty$.

Proof. The expected number of cycles is

$$\sum_{k=2}^{n} \binom{n}{k} (k-1)! \left(\frac{c}{n}\right)^k \leq \sum_{k=2}^{n} \frac{c^k}{k} = O(1).$$

Part (ii) now follows from the Markov inequality.
To tackle (i) we observe that if there is a component that is not a cycle or a single vertex then there is a cycle C and vertices $a,b \in C$ and a path P from a to b that is internally disjoint from C.
However, the expected number of such subgraphs is bounded by

$$\sum_{k=2}^{n}\sum_{l=0}^{n-k}\binom{n}{k}(k-1)!\left(\frac{c}{n}\right)^{k}k^2\binom{n}{l}l!\left(\frac{c}{n}\right)^{l+1} \leq \sum_{k=2}^{\infty}\sum_{l=0}^{\infty}\frac{k^2c^{k+l+1}}{kn} = O(1/n).$$

Here l is the number of vertices on the path P, excluding a and b. $\square$

Strong component sizes: super-critical region.

We will prove the following beautiful theorem that is a directed analogue of the existence of a giant component in $\mathbb{G}_{n,p}$. It is due to Karp [556].

Theorem 11.2. *Let $p = c/n$, where c is a constant, $c > 1$, and let x be defined by $x < 1$ and $xe^{-x} = ce^{-c}$. Then w.h.p. $\mathbb{D}_{n,p}$ contains a unique strong component of size $\approx \left(1 - \frac{x}{c}\right)^2 n$. All other strong components are of logarithmic size.*

We will prove the above theorem through a sequence of lemmas.
For a vertex v vet

$$\begin{aligned} D^+(v) &= \{w : \exists \text{ path } v \text{ to } w \text{ in } \mathbb{D}_{n,p}\} \\ D^-(v) &= \{w : \exists \text{ path } w \text{ to } v \text{ in } \mathbb{D}_{n,p}\}. \end{aligned}$$

We will first prove

Lemma 11.3. *There exist constants α,β, dependent only on c, such that w.h.p. $\not\exists$ v such that $|D^{\pm}(v)| \in [\alpha \log n, \beta n]$.*

Proof. If there is a v such that $|D^+(v)| = s$ then $\mathbb{D}_{n,p}$ contains a tree T of size s, rooted at v such that

(i) all arcs are oriented away from v, and

(ii) there are no arcs oriented from $V(T)$ to $[n] \setminus V(T)$.

The expected number of such trees is bounded above by

$$s\binom{n}{s}s^{s-2}\left(\frac{c}{n}\right)^{s-1}\left(1-\frac{c}{n}\right)^{s(n-s)} \leq \frac{n}{cs}\left(ce^{1-c+s/n}\right)^{s}.$$

Now $ce^{1-c} < 1$ for $c \neq 1$ and so there exists β such that when $s \leq \beta n$ we can bound $ce^{1-c+s/n}$ by some constant $\gamma < 1$ (γ depends only on c). In which case

$$\frac{n}{cs}\gamma^s \leq n^{-3} \text{ for } \frac{4}{\log 1/\gamma}\log n \leq s \leq \beta n.$$

□

Fix a vertex $v \in [n]$ and consider a directed breadth first search from v. Let $S_0^+ = S_0^+(v) = \{v\}$ and given $S_0^+, S_1^+ = S_1^+(v), \ldots, S_k^+ = s_k^+(v) \subseteq [n]$ let $T_k^+ = T_k^+(v) = \bigcup_{i=1}^k S_i^+$ and let

$$S_{k+1}^+ = \left\{w \notin T_k^+ : \exists x \in T_k^+ \text{ such that } (x,w) \in E(\mathbb{D}_{n,p})\right\}.$$

We similarly define $S_0^- = S_0^-(v), S_1^- = S_1^-(v), \ldots, S_k^- = S_k^-, T_k^-(v) \subseteq [n]$ with respect to a directed breadth first search into v.

Not surprisingly, we can show that the subgraph Γ_k induced by T_k^+ is close in distribution to the tree defined by the first $k+1$ levels of a Galton-Watson branching process with $\mathrm{Po}(c)$ as the distribution of the number of offspring from a single parent. See Chapter 29 for some salient facts about such a process. Here $\mathrm{Po}(c)$ is the Poisson random variable with mean c i.e.

$$\mathbb{P}(\mathrm{Po}(c) = k) = \frac{c^k e^{-c}}{k!} \qquad \text{for } k = 0, 1, 2, \ldots, .$$

Lemma 11.4. *If $\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_k$ and $\hat{T}_k$ are defined with respect to the Galton-Watson branching process and if $k \leq k_0 = (\log n)^3$ and $s_0, s_1, \ldots, s_k \leq (\log n)^4$ then*

$$\mathbb{P}\left(|S_i^+| = s_i, 0 \leq i \leq k\right) = \left(1 + O\left(\frac{1}{n^{1-o(1)}}\right)\right)\mathbb{P}\left(|\hat{S}_i| = s_i, 0 \leq i \leq k\right).$$

Proof. We use the fact that if $\mathrm{Po}(a), \mathrm{Po}(b)$ are independent then $\mathrm{Po}(a) + \mathrm{Po}(b)$ has the same distribution as $\mathrm{Po}(a+b)$. It follows that

$$\mathbb{P}\left(|\hat{S}_i| = s_i, 0 \leq i \leq k\right) = \prod_{i=1}^k \frac{(cs_{i-1})^{s_i} e^{-cs_{i-1}}}{s_i!}.$$

Furthermore, putting $t_{i-1} = s_0 + s_1 + \ldots + s_{i-1}$ we have for $v \notin T_{i-1}^{+}$,

$$\mathbb{P}(v \in S_i^{+}) = 1 - (1-p)^{s_{i-1}} = s_{i-1} p \left(1 + O\left(\frac{(\log n)^7}{n}\right)\right). \tag{11.1}$$

$$\begin{aligned} &\mathbb{P}\left(|S_i^{+}| = s_i, 0 \le i \le k\right) = \\ &= \prod_{i=1}^{k} \binom{n - t_{i-1}}{s_i} \left(\frac{s_{i-1}c}{n}\left(1 + O\left(\frac{(\log n)^7}{n}\right)\right)\right)^{s_i} \\ &\qquad \times \left(1 - \frac{s_{i-1}c}{n}\left(1 + O\left(\frac{(\log n)^7}{n}\right)\right)\right)^{n - t_{i-1} - s_i} \end{aligned} \tag{11.2}$$

Here we use the fact that given s_{i-1}, t_{i-1}, the distribution of $|S_i^{+}|$ is the binomial with $n - t_{i-1}$ trials and probability of success given in (11.1). The lemma follows by simple estimations. □

Lemma 11.5. *For* $1 \le i \le (\log n)^3$

$$\begin{aligned} &(a) \;\; \mathbb{P}\left(|S_i^{+}| \ge s \log n \,\big|\, |S_{i-1}^{+}| = s\right) \le n^{-10} \\ &(b) \;\; \mathbb{P}\left(|\hat{S}_i| \ge s \log n \,\big|\, |\hat{S}_{i-1}| = s\right) \le n^{-10}. \end{aligned}$$

Proof.

$$\begin{aligned} (a) \;\; \mathbb{P}\left(|S_i^{+}| \ge s \log n \,\big|\, |S_{i-1}^{+}| = s\right) &\le \mathbb{P}\left(\mathrm{Bin}(sn, c/n) \ge s \log n\right) \\ &\le \binom{sn}{s \log n}\left(\frac{c}{n}\right)^{s \log n} \\ &\le \left(\frac{snec}{sn \log n}\right)^{s \log n} \\ &\le \left(\frac{ec}{\log n}\right)^{\log n} \\ &\le n^{-10}. \end{aligned}$$

The proof of (b) is similar. □

Keeping v fixed we next let

$$\begin{aligned} \mathscr{F} &= \left\{\exists\, i : |T_i^{+}| > (\log n)^2\right\} \\ &= \left\{\exists\, i \le (\log n)^2 : |T_0^{+}|, |T_1^{+}|, \ldots, |T_{i-1}^{+}| < (\log n)^2 < |T_i^{+}|\right\}. \end{aligned}$$

Lemma 11.6.

$$\mathbb{P}(\mathscr{F}) = 1 - \frac{x}{c} + o(1).$$

Proof. Applying Lemma 11.4 we see that

$$\mathbb{P}(\mathscr{F}) = \mathbb{P}(\hat{\mathscr{F}}) + o(1), \tag{11.3}$$

where $\hat{\mathscr{F}}$ is defined with respect to the branching process.
Now let $\hat{\mathscr{E}}$ be the event that the branching process eventually becomes extinct. We write

$$\mathbb{P}(\hat{\mathscr{F}}) = \mathbb{P}(\hat{\mathscr{F}}|\neg\hat{\mathscr{E}})\mathbb{P}(\neg\hat{\mathscr{E}}) + \mathbb{P}(\hat{\mathscr{F}} \cap \hat{\mathscr{E}}). \tag{11.4}$$

To estimate (11.4) we use Theorem 29.1. Let

$$G(z) = \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} z^k = e^{cz-c}$$

be the probability generating function of $Po(c)$. Then Theorem 29.1 implies that $\rho = \mathbb{P}(\hat{\mathscr{E}})$ is the smallest non-negative solution to $G(\rho) = \rho$. Thus

$$\rho = e^{c\rho - c}.$$

Substituting $\rho = \frac{\xi}{c}$ we see that

$$\mathbb{P}(\hat{\mathscr{E}}) = \frac{\xi}{c} \text{ where } \frac{\xi}{c} = e^{\xi - c}, \tag{11.5}$$

and so $\xi = x$.
The lemma will follow from (11.4) and (11.5) and $\mathbb{P}(\hat{\mathscr{F}}|\neg\hat{\mathscr{E}}) = 1$ and

$$\mathbb{P}(\hat{\mathscr{F}} \cap \hat{\mathscr{E}}) = o(1).$$

This in turn follows from

$$\mathbb{P}(\hat{\mathscr{E}} \mid \hat{\mathscr{F}}) = o(1), \tag{11.6}$$

which will be established using the following lemma.

Lemma 11.7. *Each member of the branching process has probability at least $\varepsilon > 0$ of producing $(\log n)^2$ descendants at depth $\log n$. Here $\varepsilon > 0$ depends only on c.*

Proof. If the current population size of the process is s then the probability that it reaches size at least $\frac{c+1}{2}s$ in the next round is

$$\sum_{k \geq \frac{c+1}{2}s} \frac{(cs)^k e^{-cs}}{k!} \geq 1 - e^{-\alpha s}$$

for some constant $\alpha > 0$ provided $s \geq 100$, say.
Now there is a positive probability ε_1 say that a single member spawns at least 100 descendants and so there is a probability of at least

$$\varepsilon_1 \left(1 - \sum_{s=100}^{\infty} e^{-\alpha s}\right)$$

that a single object spawns

$$\left(\frac{c+1}{2}\right)^{\log n} \gg (\log n)^2$$

descendants at depth $\log n$.

$\square$

Given a population size between $(\log n)^2$ and $(\log n)^3$ at level i_0, let s_i denote the population size at level $i_0 + i \log n$. Then Lemma 11.7 and the Chernoff bounds imply that

$$\mathbb{P}\left(s_{i+1} \leq \frac{1}{2}\varepsilon s_i (\log n)^2\right) \leq \exp\left\{-\frac{1}{8}\varepsilon^2 s_i (\log n)^2\right\}.$$

It follows that

$$\begin{aligned} \mathbb{P}(\hat{\mathscr{E}} \mid \hat{\mathscr{F}}) &\leq \mathbb{P}\left(\exists i:\ s_i \leq \left(\frac{1}{2}\varepsilon(\log n)^2\right)^i s_0 \middle| s_0 \geq (\log n)^2\right) \\ &\leq \sum_{i=1}^{\infty} \exp\left\{-\frac{1}{8}\varepsilon^2 \left(\frac{1}{2}\varepsilon(\log n)^2\right)^i (\log n)^2\right\} = o(1). \end{aligned}$$

This completes the proof (11.6) and of Lemma 11.6. $\square$

We must now consider the probability that both $D^+(v)$ and $D^-(v)$ are large.

Lemma 11.8.

$$\mathbb{P}\left(|D^-(v)| \geq (\log n)^2 \mid |D^+(v)| \geq (\log n)^2\right) = 1 - \frac{x}{c} + o(1).$$

Proof. Expose $S_0^+, S_1^+, \ldots, S_k^+$ until either $S_k^+ = \emptyset$ or we see that $|T_k^+| \in [(\log n)^2, (\log n)^3]$. Now let S denote the set of edges/vertices defined by
$S_0^+, S_1^+, \ldots, S_k^+$.

Let $\mathscr{C}$ be the event that there are no edges from T_l^- to S_k^+ where T_l^- is the set of vertices we reach through our BFS into v, up to the point where we first realise that $D^-(v) < (\log n)^2$ (because $S_i^- = \emptyset$ and $|T_i^-| \le (\log n)^2$) or we realise that $D^-(v) \ge (\log n)^2$. Then

$$\mathbb{P}(\neg\mathscr{C}) = O\left(\frac{(\log n)^4}{n}\right) = \frac{1}{n^{1-o(1)}}$$

and, as in (11.2),

$$\begin{aligned}
&\mathbb{P}\left(|S_i^-| = s_i,\ 0 \le i \le k \mid \mathscr{C}\right) = \\
&= \prod_{i=1}^{k} \binom{n'-t_{i-1}}{s_i} \left(\frac{s_{i-1}c}{n}\left(1+O\left(\frac{(\log n)^7}{n}\right)\right)\right)^{s_i} \\
&\qquad\qquad \times \left(1-\frac{s_{i-1}c}{n}\left(1+O\left(\frac{(\log n)^7}{n}\right)\right)\right)^{n'-t_{i-1}-s_i}
\end{aligned}$$

where $n' = n - |T_k^+|$.
Given this we can prove a conditional version of Lemma 11.4 and continue as before. □

We have now shown that if α is as in Lemma 11.3 and if

$$S = \left\{v : |D^+(v)|, |D^-(v)| > \alpha \log n\right\}$$

then the expectation

$$\mathbb{E}(|S|) = (1+o(1))\left(1-\frac{x}{c}\right)^2 n.$$

We also claim that for any two vertices v, w

$$\mathbb{P}(v, w \in S) = (1+o(1))\,\mathbb{P}(v \in S)\,\mathbb{P}(w \in S) \tag{11.7}$$

and therefore the Chebyshev inequality implies that w.h.p.

$$|S| = (1+o(1))\left(1-\frac{x}{c}\right)^2 n.$$

But (11.7) follows in a similar manner to the proof of Lemma 11.8.
All that remains of the proof of Theorem 11.2 is to show that

$$S \text{ is a strong component w.h.p.} \tag{11.8}$$

Recall that any $v \notin S$ is in a strong component of size $\le \alpha \log n$ and so the second part of the theorem will also be done.

We prove (11.8) by arguing that

$$\mathbb{P}\left(\exists\, v,w \in S : w \notin D^+(v)\right) = o(1). \tag{11.9}$$

In which case, we know that w.h.p. there is a path from each $v \in S$ to every other vertex $w \neq v$ in S.

To prove (11.9) we expose $S_0^+, S_1^+, \ldots, S_k^+$ until we find that $|T_k^+(v)| \geq n^{1/2}\log n$. At the same time we expose $S_0^-, S_1^-, \ldots, S_l^-$ until we find that $|T_l^-(w)| \geq n^{1/2}\log n$. If $w \notin D^+(v)$ then this experiment will have tried at least $\left(n^{1/2}\log n\right)^2$ times to find an edge from $D^+(v)$ to $D^-(w)$ and failed every time. The probability of this is at most

$$\left(1 - \frac{c}{n}\right)^{n(\log n)^2} = o(n^{-2}).$$

This completes the proof of Theorem 11.2. □

Threshold for strong connectivity

Here we prove

Theorem 11.9. *Let $\omega = \omega(n)$, $c > 0$ be a constant, and let $p = \frac{\log n + \omega}{n}$. Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{D}_{n,p} \textit{ is strongly connected}) = \begin{cases} 0 & \textit{if } \omega \to -\infty \\ e^{-2e^{-c}} & \textit{if } \omega \to c \\ 1 & \textit{if } \omega \to \infty. \end{cases}$$

$$= \lim_{n\to\infty} \mathbb{P}(\nexists\, v \textit{ s.t. } d^+(v) = 0 \textit{ or } d^-(v) = 0)$$

Proof. We leave as an exercise to prove that

$$\lim_{n\to\infty} \mathbb{P}(\exists\, v \text{ s.t. } d^+(v) = 0 \text{ or } d^-(v) = 0) = \begin{cases} 1 & \text{if } \omega \to -\infty \\ 1 - e^{-2e^{-c}} & \text{if } \omega \to c \\ 0 & \text{if } \omega \to \infty. \end{cases}$$

Given this, one only has to show that if $\omega \not\to -\infty$ then w.h.p. there does not exist a set S such that (i) $2 \leq |S| \leq n/2$ and (ii) $E(S : \bar{S}) = \emptyset$ or $E(\bar{S} : S) = \emptyset$ and (iii) S induces a connected component in the graph obtained by ignoring orientation. But, here with $s = |S|$,

$$\mathbb{P}(\exists\, S) \leq 2\sum_{s=2}^{n/2} \binom{n}{s} s^{s-2} (2p)^{s-1} (1-p)^{s(n-s)}$$

$$
\begin{aligned}
&\leq \frac{2n}{\log n}\sum_{s=2}^{n/2}\left(\frac{ne}{s}\right)^s s^{s-2}\left(\frac{2\log n}{n}\right)^s n^{-s(1-s/n)}e^{\omega s/n}\\
&\leq \frac{2n}{\log n}\sum_{s=2}^{n/2}(2n^{-(1-s/n)}e^{\omega/n}\log n)^s\\
&= o(1).
\end{aligned}
$$

□

11.2 Hamilton Cycles

Existence of a Hamilton Cycle

Here we prove the following remarkable inequality: It is due to McDiarmid [661]

Theorem 11.10.

$$\mathbb{P}(\mathbb{D}_{n,p} \textit{ is Hamiltonian}) \geq \mathbb{P}(\mathbb{G}_{n,p} \textit{ is Hamiltonian})$$

Proof. We consider an ordered sequence of random digraphs $\Gamma_0, \Gamma_1, \Gamma_2, \dots, \Gamma_N$, $N = \binom{n}{2}$ defined as follows: Let $e_1, e_2, \dots, e_N$ be an enumeration of the edges of the complete graph K_n. Each $e_i = \{v_i, w_i\}$ gives rise to two directed edges $\overrightarrow{e_i} = (v_i, w_i)$ and $\overleftarrow{e_i} = (w_i, v_i)$. In Γ_i we include $\overrightarrow{e_j}$ and $\overleftarrow{e_j}$ independently of each other, with probability p, for $j \leq i$. While for $j > i$ we include both or neither with probability p. Thus Γ_0 is just $\mathbb{G}_{n,p}$ with each edge $\{v,w\}$ replaced by a pair of directed edges $(v,w),(w,v)$ and $\Gamma_N = \mathbb{D}_{n,p}$. Theorem 11.10 follows from

$$\mathbb{P}(\Gamma_i \text{ is Hamiltonian}) \geq \mathbb{P}(\Gamma_{i-1} \text{ is Hamiltonian}).$$

To prove this we condition on the existence or otherwise of directed edges associated with $e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_N$. Let $\mathscr{C}$ denote this conditioning.
Either

(a) $\mathscr{C}$ gives us a Hamilton cycle without arcs associated with e_i, or

(b) not (a) and there exists a Hamilton cycle if at least one of $\overrightarrow{e_i}, \overleftarrow{e_i}$ is present, or

(c) $\nexists$ a Hamilton cycle even if both of $\overrightarrow{e_i}, \overleftarrow{e_i}$ are present.

(a) and (c) give the same conditional probability of Hamiltonicity in Γ_i, Γ_{i-1}. In Γ_{i-1} (b) happens with probability p. In Γ_i we consider two cases (i) exactly one of $\overrightarrow{e_i}, \overleftarrow{e_i}$ yields Hamiltonicity and in this case the conditional probability is p and (ii) either of $\overrightarrow{e_i}, \overleftarrow{e_i}$ yields Hamiltonicity and in this case the conditional probability is $1-(1-p)^2 > p$.
Note that we will never require that **both** $\overrightarrow{e_i}, \overleftarrow{e_i}$ occur. □

Theorem 11.10 was subsequently improved by Frieze [386], who proved the equivalent of Theorem 6.5.

Theorem 11.11. *Let $p = \frac{\log n + c_n}{n}$. Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{D}_{n,p} \text{ has a Hamilton cycle}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-2e^{-c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

Number of Distinct Hamilton Cycles

Here we give an elegant result of Ferber, Kronenberg and Long [354].

Theorem 11.12. *Let $p = \omega\left(\frac{\log^2 n}{n}\right)$. Then w.h.p. $\mathbb{D}_{n,p}$ contains $e^{o(n)} n! p^n$ directed Hamilton cycles.*

Proof. The upper bound follows from the fisrt moment method. Let X_H denote the number of Hamilton cycles in $D = \mathbb{D}_{n,p}$. Now $\mathbb{E} X_H = (n-1)! p^n$, and therefore the Markov inequality implies that w.h.p. we have $X_H \leq e^{o(n)} n! p^n$.

For the lower bound let $\alpha := \alpha(n)$ be a function tending slowly to infinity with n. Let $S \subseteq V(G)$ be a fixed set of size s, where $s \approx \frac{n}{\alpha \log n}$ and let $V' = V \setminus S$. Moreover, assume that s is chosen so that $|V'|$ is divisible by integer $\ell = 2\alpha \log n$. From now on the set S will be fixed and we will use it for closing Hamilton cycles. Our strategy is as follows: we first expose all the edges within V', and show that one can find the "correct" number of distinct families $\mathscr{P}$ consisting of $m := |V'|/\ell$ vertex-disjoint paths which span V'. Then, we expose all the edges with at least one endpoint in S, and show that w.h.p. one can turn "most" of these families into Hamilton cycles and that all of these cycles are distinct.

We take a random partitioning $V' = V_1 \cup \ldots \cup V_\ell$ such that all the V_i's are of size m. Let us denote by D_j the bipartite graph with parts V_j and V_{j+1}. Observe that D_j is distributed as $\mathbb{G}_{m,m,p}$, and therefore, since $p = \omega\left(\frac{\log n}{m}\right)$, by Exercise 11.3.2, with probability $1 - n^{-\omega(1)}$ we conclude that D_j contains $(1-o(1))m$ edge-disjoint perfect matchings (in particular, a $(1-o(1))m$ regular subgraph). The Van der Waerden conjecture proved by Egorychev [328] and by Falikman

[347] implies the following: Let $G = (A \cup B, E)$ be an r-regular bipartite graph with part sizes $|A| = |B| = n$. Then, the number of perfect matchings in G is at least $\left(\frac{r}{n}\right)^n n!$.

Applying this and the union bound, it follows that w.h.p. each D_j contains at least $(1-o(1))^m m! p^m$ perfect matchings for each j. Taking the union of one perfect matching from each of the D_j's we obtain a family $\mathscr{P}$ of m vertex disjoint paths which spans V'. Therefore, there are

$$((1-o(1))^m m! p^m)^\ell = (1-o(1))^{n-s} (m!)^\ell p^{n-s}$$

distinct families $\mathscr{P}$ obtained from this partitioning in this manner. Since this occurs w.h.p. we conclude (applying the Markov inequality to the number of partitions for which the bound fails) that this bound holds for $(1-o(1))$-fraction of such partitions. Since there are $\frac{(n-s)!}{(m!)^\ell}$ such partitions, one can find at least

$$\begin{aligned}(1-o(1))\frac{(n-s)!}{(m!)^\ell}(1-o(1))^{n-s} (m!)^\ell p^{n-s}\\ = (1-o(1))^{n-s}(n-s)! p^{n-s} = (1-o(1))^n n! p^n\end{aligned}$$

distinct families, each of which consists of exactly m vertex-disjoint paths of size ℓ (for the last equality, we used the fact that $s = o(n/\log n)$).

We show next how to close a given family of paths into a Hamilton cycle. For each such family $\mathscr{P}$, let $A := A(\mathscr{P})$ denote the collection of all pairs (s_P, t_P) where s_P is a starting point and t_P is the endpoint of a path $P \in \mathscr{P}$, and define an auxiliary directed graph $D(A)$ as follows. The vertex set of $D(A)$ is $V(A) = S \cup \{z_P = (s_P, t_P) : z_P \in A\}$. Edges of $D(A)$ are determined as follows: if $u, v \in S$ and $(u, v) \in E(D)$ then (u, v) is an edge of $D(A)$. The in-neighbors (out-neighbors) of vertices z_P in S are the in-neighbors of s_P in D (out-neighbors of t_P). Lastly, (z_P, z_Q) is an edge of $D(A)$ if (t_P, s_Q) is an edge D.

Clearly $D(A)$ is distributed as $\mathbb{D}_{s+m,p}$, and that a Hamilton cycle in $D(A)$ corresponds to a Hamilton cycle in D after adding the corresponding paths between each s_P and t_P. Now distinct families $\mathscr{P} \neq \mathscr{P}'$ yield distinct Hamilton cycles (to see this, just delete the vertices of S from the Hamilton cycle, to recover the paths). Using Theorem 11.11 we see that for $p = \omega(\log n/(s+m)) = \omega(\log(s+m)/(s+m))$, the probability that $D(A)$ does not have a Hamilton cycle is $o(1)$. Therefore, using the Markov inequality we see that for almost all of the families $\mathscr{P}$, the corresponding auxiliary graph $D(A)$ is indeed Hamiltonian and we have at least $(1-o(1))^n n! p^n$ distinct Hamilton cycles, as desired. $\square$

11.3 Exercises

11.3.1 Let $p = \frac{\log n + (k-1)\log\log n + \omega}{n}$ for a constant $k = 1, 2, \ldots$. Show that w.h.p. $\mathbb{D}_{np}$ is k-strongly connected.

11.3.2 The Gale-Ryser theorem states: Let $G = (A \cup B, E)$ be a bipartite graph with parts of sizes $|A| = |B| = n$. Then, G contains an r-factor if and only if for every two sets $X \subseteq A$ and $Y \subseteq B$, we have

$$e_G(X,Y) \geq r(|X| + |Y| - n).$$

Show that if $p = \omega(\log n/n)$ then with probability $1 - n^{-\omega(1)}$, $\mathbb{G}_{n,n,p}$ contains $(1 - o(1))np$ edge disjoint perfect matchings.

11.3.3 Show that if $p = \omega((\log n)^2/n)$ then w.h.p. $\mathbb{G}_{n,p}$ contains $e^{o(n)} n! p^n$ distinct Hamilton cycles.

11.3.4 A *tournament* T is an orientation of the complete graph K_n. In a random tournament, edge $\{u,v\}$ is oriented from u to v with probability 1/2 and from v to u with probability 1/2. Show that w.h.p. a random tournament is strongly connected.

11.3.5 Let T be a random tournament. Show that w.h.p. the size of the largest acyclic sub-tournament is asymptotic to $2\log_2 n$. (A tournament is acyclic if it contains no directed cycles).

11.3.6 Suppose that $0 < p < 1$ is constant. Show that w.h.p. the size of the largest acyclic tournament contained in $\mathbb{D}_{n,p}$ is asymptotic to $2\log_b n$ where $b = 1/p$.

11.3.7 Let $mas(D)$ denote the number of vertices in the largest acyclic subgraph of a digraph D. Suppose that $0 < p < 1$ is constant. Show that w.h.p. $mas(\mathbb{D}_{n,p}) \leq \frac{4\log n}{\log q}$ where $q = \frac{1}{1-p}$.

11.3.8 Consider the random digraph D_n obtained from $G_{n,1/2}$ by orienting edge (i,j) from i to j when $i < j$. This can be viewed as a partial order on $[n]$ and is called a *Random Graph Order*. Show that w.h.p. D_n contains a path of length at least $0.51n$. (In terms of partial orders, this bounds the height of the order).

11.3.9 Show that if $np \geq \log^{10} n$ then w.h.p. $D_{n,p}$ is $\frac{1}{2} \pm o(1)$ resilient, i.e. $\left(\frac{1}{2} - \varepsilon\right) np \leq \Delta_{\mathscr{H}} \leq \left(\frac{1}{2} + \varepsilon\right) np$. (Hint: just tweak the proof of Theorem 20.4 so that the lemmas refer to digraphs.)

11.3.10 Let O represent an orientation of the edges of a Hamilton cycle. Show that

$$\mathbb{P}(D_{n,p} \text{ has a Hamilton cycle with orientation } O) \geq \mathbb{P}(G_{n,p} \text{ has a Hamilton cycle}).$$

11.4 Notes

Packing

The paper of Frieze [386] was in terms of the hitting time for a digraph process $\mathbb{D}_t$. It proves that the first time that the $\delta^+(G_t), \delta^-(G_t) \geq k$ is w.h.p. the time when G_t has k edge disjoint Hamilton cycles. The paper of Ferber, Kronenberg and Long [354] shows that if $p = \omega((\log n)^4/n)$ then w.h.p. $\mathbb{D}_{n,p}$ contains $(1-o(1))np$ edge disjoint Hamilton cycles.

Long Cycles

The papers by Hefetz, Steger and Sudakov [479] and by Ferber, Nenadov, Noever, Peter and Škorić [357] study the local resilience of having a Hamilton cycle. In particular, [357] proves that if $p \gg \frac{(\log n)^8}{n}$ then w.h.p. one can delete any subgraph H of $\mathbb{D}_{n,p}$ with maximum degree at most $(\frac{1}{2} - \varepsilon)np$ and still leave a Hamiltonian subgraph.

Krivelevich, Lubetzky and Sudakov [597] proved that w.h.p. the random digraph $\mathbb{D}_{n,p}, p = c/n$ contains a directed cycle of length $(1-(1+\varepsilon_c)e^{-c})n$ where $\varepsilon_c \to 0$ as $c \to \infty$.

Cooper, Frieze and Molloy [268] showed that a random regular digraph with indegree = outdegree = r is Hamiltonian w.h.p. iff $r \geq 3$.

Connectivity

Cooper and Frieze [255] studied the size of the largest strong component in a random digraph with a given degree sequence. The strong connectivity of an inhomogeneous random digraph was studied by Bloznelis, Götze and Jaworski in [122].

Chapter 12

Hypergraphs

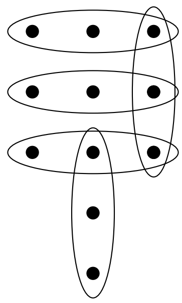
In this chapter we discuss random k-uniform hypergraphs, where $k \geq 3$. We are concerned with the models $\mathbb{H}_{n,p;k}$ and $\mathbb{H}_{n,m;k}$. For $\mathbb{H}_{n,p;k}$ we consider the hypergraph with vertex set $[n]$ in which each possible k-set in $\binom{[n]}{k}$ is included as an edge with probability p. In $\mathbb{H}_{n,m;k}$ the edge set is a random m-subset of $\binom{[n]}{k}$. The parameter k is fixed and independent of n throughout this chapter.

Many of the properties of $\mathbb{G}_{n,p}$ and $\mathbb{G}_{n,m}$ have been generalized without too much difficulty to these models of hypergraphs. Hamilton cycles have only recently been tackled with any success. Surprisingly enough, in some cases it is enough to use the Chebyshev inequality and we will describe these cases. We begin however with a more basic question. That is as to when is there a giant component and when are $\mathbb{H}_{n,m;k}, \mathbb{H}_{n,p;k}$ connected?

12.1 Component Size

We remind the reader that $k \geq 3$ here. Suppose that $p = \frac{c}{\binom{n-1}{k-1}}$ and c is constant. We will prove that if $c < \frac{1}{k-1}$ then w.h.p. the maximum component size is $O(\log n)$ and if $c > \frac{1}{k-1}$ then w.h.p. there is a unique giant component of size $\Omega(n)$. This generalises the main result of Chapter 2. We will assume then that $p = \frac{c}{\binom{n-1}{k-1}}$ for the remainder of this section.

Many of the components of a sparse random graph are small trees. The corresponding objects here are called *hypertrees*. The *size* of a hypertree is the number of edges that it contains. An edge is a hypertree of size one. We obtain a hypertree of size $k+1$ by choosing a hypertree C of size k and a vertex $v \in V(C)$ and then adding a new edge $\{v, v_2, v_3, \ldots, v_k\}$ where $v_2, v_3, \ldots, v_k \notin V(C)$.



A hypertree of size 5.

Lemma 12.1. *A hypertree of size ℓ contains $\ell(k-1)+1$ vertices.*

Proof. By induction on ℓ. It is clearly true if $\ell = 1$ and if we increase the size by one, then we add exactly $k-1$ new vertices. □

A proof of the following generalisation of Cayley's formula for the number of spanning hypertrees of the complete graph can be found in Selivanov [801] and Sivasubramanian [807] .

Lemma 12.2. *The number $N_k(\ell)$ of distinct hypertrees with ℓ edges contained in the complete k-uniform hypergraph on $[(k-1)\ell+1]$ is given by*

$$N_k(\ell) = ((k-1)\ell+1)^{\ell-1}\frac{((k-1)\ell)!}{\ell!\,((k-1)!)^{\ell}}.$$

We now look into the structure of small components of $\mathbb{H}_{n,p;k}$.

Lemma 12.3. *Let $p = \frac{c}{\binom{n-1}{k-1}}$ where $c \neq \frac{1}{k-1}$. Suppose that $S \subseteq [n]$ and $|S| = s \leq \log^4 n$. Suppose also that S contains a hypertree C with $t = (k-1)\ell+1$ vertices and ℓ edges. Suppose that in addition, S contains a vertices and b edges that are not part of C. Suppose also that there are no edges joining C to $V \setminus S$. Then, for some $A = A(c) > 0$, we have that w.h.p.*
(a) $b \leq 1$ and $b(k-2)+1 \geq a+1 \geq b(k-1)$.
(b) $\ell < A\log n$ and either (i) $a = b = 0$ or (ii) $b = 1$.
(c) $|\{S : \ell \leq A\log n, b = 1\}| = O(\log^{k+1} n)$.

Proof. We bound the expected number of sets S that violate this by

$$\binom{n}{s}\binom{s}{a}\binom{s}{k}^{b} t^{\ell-1}\frac{((k-1)\ell)!}{\ell!((k-1)!)^{\ell}}\left(\frac{c}{\binom{n-1}{k-1}}\right)^{\ell+b}\left(1-\frac{c}{\binom{n-1}{k-1}}\right)^{t\binom{n-s}{k-1}} \tag{12.1}$$

$$\leq \left(\frac{ne}{s}\right)^{s}\frac{s^{a+kb}}{k!^{b}}t^{\ell-1}\left(\frac{\ell^{k-2}(k-1)^{k-1}}{e^{k-2+o(1)}(k-1)!}\right)^{\ell}\left(\frac{e^{o(1)}c(k-1)!}{n^{k-1}}\right)^{\ell+b} e^{-(c+o(1))t}$$

$$
\begin{aligned}
&\le \frac{s^{kb-1}c^b n^{a+1-b(k-1)}e^{a+1-c}}{k^b t}\left(\frac{e^{1-(k-1)c+o(1)}c(k-1)^{k-1}\ell^{k-2}t}{s^{k-1}}\right)^{\ell} \\
&\le \frac{s^{kb-1}c^b n^{a+1-b(k-1)}e^{a+1-c}}{k^b t}\left(\frac{e^{o(1)}e^{1-(k-1)c}c(k-1)^{k}\ell^{k-1}}{((k-1)\ell)^{k-1}}\right)^{\ell} \\
&= \frac{s^{kb-1}c^b n^{a+1-b(k-1)}e^{a+1-c}}{k^b t}\left(e^{o(1)}ce^{1-(k-1)c}(k-1)\right)^{\ell} \qquad (12.2)
\end{aligned}
$$

Now $ce^{-(k-1)c}$ is maximised at $c = 1/(k-1)$ and for $c \neq 1/(k-1)$ we have

$$
e^{o(1)}ce^{1-(k-1)c}(k-1) \le 1-\varepsilon_c.
$$

We may therefore upper bound the expression in (12.2) by

$$
n^{a+1-b(k-1)+O(b\log\log n/\log n)}e^{-\varepsilon_c\ell} \le n^{1-b+O(b\log\log n/\log n)}e^{-\varepsilon_c\ell}. \qquad (12.3)
$$

For the second expression, we used the fact that $a \le b(k-2)$. The second expression in (12.3) tends to zero if $b > 1$ and so we can assume that $b \le 1$. The first expression in (12.3) then tends to zero if $a+1 < b(k-1)$ and this verifies Part (a). Because $a+1-b(k-1) \le 1$, we see that Part (b) of the lemma follows with $A = 2/\varepsilon_c$. Part (c) follows from the Markov inequality. □

Lemma 12.4. *W.h.p. there are no components in the range* $[A\log n, \log^4 n]$ *and all but* $\log^{k+1} n$ *of the small components (of size at most* $A\log n$*) are hypertrees.*

Proof. Now suppose that S is a small component of size s which is not a hypertree and let C be the set of vertices of a maximal hypertree with $t = (k-1)\ell+1$ vertices and ℓ edges that is a subgraph of S. (Maximal in the sense that it is not contained in any other hypertree of S.) Lemma 12.3 implies that w.h.p. there is at most one edge e in S that is not part of C but is incident with at least two vertices of C. Furthermore, Lemma 12.3(c) implies that the number of sets with $b = 1$ is $O(\log^{k+1} n)$ w.h.p. □

Lemma 12.5. *If* $c < \frac{1}{k-1}$ *then w.h.p. the largest component has size* $O(\log n)$.

Proof. Fix an edge e and do a Breadth First Search (BFS) on the edges starting with e. We start with $L_1 = e$ and let L_t denote the number of vertices at depth t in the search i.e the neighbors of L_{t-1} that are not in L_{t-1}. Then $|L_{t+1}|$ is dominated by $(k-1)\mathrm{Bin}\left(|L_t|\binom{n}{k-1}, p\right)$. So,

$$
\mathbb{E}(|L_{t+1}| \mid L_t) \le (k-1)|L_t|\binom{n}{k-1}p \le \theta|L_t|
$$

where $\theta = (k-1)c < 1$. It follows that if $t_0 = \frac{2\log n}{\log 1/\theta}$ then

$$\Pr(\exists e : L_{t_0} \neq \emptyset) \leq nk\theta^{t_0} = o(1).$$

So, w.h.p. there are at most t_0 levels. Furthermore, if $|L_t|$ ever reaches $\log^2 n$ then the Chernoff bounds imply that w.h.p. $|L_{t+1}| \leq |L_t|$. This implies that the maximum size of a component is $O(\log^3 n)$ and hence, by Lemma 12.4, the maximum size is $O(\log n)$. □

Lemma 12.6. *If $c > \frac{1}{k-1}$ then w.h.p. there is a unique giant component of size $\Omega(n)$ and all other components are of size $O(\log n)$.*

Proof. We consider BFS as in Lemma 12.5. We choose a vertex v and begin to grow a component from it. Sometimes the component we grow has size $O(\log n)$, in which case, we choose a new vertex, not yet seen in the search and grow another component. We argue that w.h.p. (i) after $O(\log n)$ attempts, we grow a component of size at least $\log^2 n$ and (ii) this component is of size $\Omega(n)$.

We say that we *explore* $L_t \setminus L_{t-1}$ when we determine L_{t+1}. With L_t as in Lemma 12.5, we see that if we have explored at most $\log^4 n$ vertices then given L_t, $|L_{t+1}|$ dominates $Y_t = (k-1)\text{Bin}\left(|L_t|\binom{n-o(n)}{k-1}, p\right) - X_t$ where X_t is an overcount due to vertices outside L_t that are in two edges containing vertices of L_t. We have

$$\mathbb{E}(X_t) \leq n|L_t|^2\binom{n}{k-3}p = O\left(\frac{\log^8 n}{n}\right).$$

Now

$$\mathbb{E}(Y_t) = (k-1)|L_t|\binom{n-o(n)}{k-1}p = (1-o(1))c|L_t| = \theta|L_t| \text{ where } \theta > 1.$$

The Chernoff bounds $\left(\text{applied to Bin}\left(|L_t|\binom{n-o(n)}{k-1}, p\right)\right)$ imply that

$$\mathbb{P}\left(Y_t \leq \frac{1+\theta}{2}|L_t|\right) \leq \exp\left\{-\frac{(\theta-1)^2|L_t|}{3(k-1)}\right\}.$$

So, for $t = O(\log\log n)$ we have

$$\mathbb{P}\left(|L_t| \geq \left(\frac{1+\theta}{2}\right)^t\right) \\ \geq \prod_{\ell=1}^{t}\left(1-\exp\left\{-\frac{(\theta-1)^2\left(\frac{1+\theta}{2}\right)^\ell}{3(k-1)}\right\}\right) - O\left(\frac{t\log^8 n}{n}\right) \tag{12.4}$$

$$\geq \gamma,$$

where $\gamma > 0$ is a positive constant. This lower bound of γ follows from the fact that $\sum_{\ell=1}^{\infty} \exp\left\{-\frac{(\theta-1)^2\left(\frac{1+\theta}{2}\right)^{\ell}}{3(k-1)}\right\}$ converges, see Apostol [51], Theorem 8.55. Note that if $t = \frac{2\log\log n}{\log\frac{1+\theta}{2}}$ then $\left(\frac{1+\theta}{2}\right)^t = \log^2 n$. It follows that the probability we fail to grow a component of size $\log^2 n$ after s attempts is at most $(1-\gamma)^s$. Choosing $s = \frac{2\log n}{\log 1/(1-\gamma)}$ we see that after exploring $O(\log^3 n)$ vertices, we find a component of size at least $\log^2 n$, with probability $1-n^{-(1-o(1))}$.

We show next that with (conditional) probability $1-O(n^{-(1-o(1))})$ the component of size at least $\log^2 n$ will in fact have at least $n_0 = n^{(k-1)/k}\log n$ vertices. We handle X_t, Y_t exactly as above. Going back to (12.4), if we run BFS for another $O(\log n)$ steps then, starting with $|L_{t_0}| \approx \log^2 n$, we have

$$\begin{aligned}
&\mathbb{P}\left(|L_{t_0+t}| \geq \left(\frac{1+\theta}{2}\right)^t \log^2 n\right) \\
&\geq \prod_{\ell=1}^{t}\left(1-\exp\left\{-\frac{(\theta-1)^2\log^2 n}{3(k-1)}\right\}\right) - O\left(\frac{\log^9 n}{n}\right) \\
&= 1-n^{-(1-o(1))}. \qquad (12.5)
\end{aligned}$$

It follows that w.h.p. $\mathbb{H}_{n,p;k}$ only contains components of size $O(\log n)$ and $\Omega(n_o)$. For this we use the fact that we only need to apply (12.5) less than n/n_0 times.

We now prove that there is a unique giant component. This is a simple sprinkling argument. Suppose that we let $(1-p) = (1-p_1)(1-p_2)$ where $p_2 = \frac{1}{\omega n^{k-1}}$ for some $\omega \to \infty$ slowly. Then we know from Lemma 12.4 that there is a gap in component sizes for $\mathbb{H}(n, p_1, k)$. Now add in the second round of edges with probability p_2. If C_1, C_2 are distinct components of size at least n_0 then the probability there is no $C_1 : C_2$ edge added is at most

$$(1-p_2)^{\sum_{i=1}^{k-1}\binom{n_0}{i}\binom{n_0}{k-i}} \leq \exp\left\{-\frac{\sum_{i=1}^{k-1}\binom{n_0}{i}\binom{n_0}{k-i}}{\omega n^{k-1}}\right\} \leq$$
$$\exp\left\{-\frac{2^{k-1}n_0^k}{\omega n^{k-1}}\right\} = \exp\left\{-\frac{2^{k-1}\log^k n}{\omega}\right\}.$$

So, w.h.p. all components of size at least n_0 are merged into one component. $\square$

We now look at the size of this giant component. The fact that almost all small components are hypertrees when $c < \frac{1}{k-1}$ yields the following lemma. The proof follows that of Lemma 2.13 and is left as Exercise 13.4.9. For $c>0$ we define

$$x = x(c) = \begin{cases} c & c \le \frac{1}{k-1} \\ \text{The solution in } \left(0, \frac{1}{k-1}\right) \text{ to } xe^{-(k-1)x} = ce^{-(k-1)c} & c > \frac{1}{k-1} \end{cases}.$$

Lemma 12.7. *Suppose that $c>0$ is constant and $c(k-1) \neq 1$ and x is as defined above. Then*

$$\sum_{\ell=0}^{\infty} \frac{c^{\ell}((k-1)\ell+1)^{\ell} e^{-c((k-1)\ell+1)}}{\ell!((k-1)\ell+1)} = \left(\frac{x}{c}\right)^{1/(k-1)}.$$

It follows from Lemmas 12.6 and 12.7 that we have

Lemma 12.8. *If $c > \frac{1}{k-1}$ then w.h.p. there is a unique giant component of size $\approx \left(1 - \left(\frac{x}{c}\right)^{1/(k-1)}\right) n$.*

The connectivity threshold for $\mathbb{H}_{n,p;k}$ coincides with minimum degree at least one. The proof is straightforward and is left as Exercise 13.4.10.

Lemma 12.9. *Let $p = \frac{\log n + c_n}{\binom{n-1}{k-1}}$. Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{H}_{n,p;k} \text{ is connected}) = \begin{cases} 0 & c_n \to -\infty. \\ e^{-e^{-c}} & c_n \to c. \\ 1 & c_n \to \infty. \end{cases}$$

12.2 Hamilton Cycles

Suppose that $1 \le r < k$. An *r-overlapping Hamilton cycle C* in a k-uniform hypergraph $H = (V, \mathscr{E})$ on n vertices is a collection of $m_r = n/(k-r)$ edges of H such that for some cyclic order of $[n]$ every edge consists of k consecutive vertices and for every pair of consecutive edges E_{i-1}, E_i in C (in the natural ordering of the edges) we have $|E_{i-1} \cap E_i| = r$. Thus, in every r-overlapping Hamilton cycle the sets $C_i = E_i \setminus E_{i-1}$, $i = 1, 2, \dots, m_r$, are a partition of V into sets of size $k-r$. Hence, $m_r = n/(k-r)$. We thus always assume, when discussing r-overlapping Hamilton cycles, that this necessary condition, $k-r$ divides n, is fulfilled. In the literature, when $r = k-1$ we have a *tight* Hamilton cycle and when $r = 1$ we have a *loose* Hamilton cycle.

In this section we will restrict our attention to the case $r = k-1$ i.e. tight Hamilton cycles. So when we say that a hypergraph is Hamiltonian, we mean that it contains a tight Hamilton cycle. The proof extends easily to $r \ge 2$. The case $r = 1$ poses some more problems and is discussed in Frieze [390], Dudek, Frieze

[305] and Dudek, Frieze, Loh and Speiss [307] and in Ferber [352]. Also, see Section 22. The following theorem is from Dudek and Frieze [306]. Furthermore, we assume that $k \geq 3$.

Theorem 12.10.

(i) If $p \leq (1-\varepsilon)e/n$, then w.h.p. $\mathbb{H}_{n,p;k}$ is not Hamiltonian.

(ii) If $k = 3$ and $np \to \infty$ then $\mathbb{H}_{n,p;k}$ is Hamiltonian w.h.p.

(iii) For all fixed $\varepsilon > 0$, if $k \geq 4$ and $p \geq (1+\varepsilon)e/n$, then w.h.p. $\mathbb{H}_{n,p;k}$ is Hamiltonian.

Proof. We will prove parts (i) and (ii) and leave the proof of (iii) as an exercise, with a hint.

Let $([n], \mathscr{E})$ be a k-uniform hypergraph. A permutation π of $[n]$ is *Hamilton cycle inducing* if

$$E_\pi(i) = \{\pi(i-1+j) : j \in [k]\} \in \mathscr{E} \text{ for all } i \in [n].$$

(We use the convention $\pi(n+r) = \pi(r)$ for $r > 0$.) Let the term *hamperm* refer to such a permutation.

Let X be the random variable that counts the number of hamperms π for $\mathbb{H}_{n,p;k}$. Every Hamilton cycle induces at least one hamperm and so we can concentrate on estimating $\mathbb{P}(X > 0)$.

Now

$$\mathbb{E}(X) = n!p^n.$$

This is because π induces a Hamilton cycle if and only if a certain n edges are all in $\mathbb{H}_{n,p;k}$.

For part (i) we use Stirling's formula to argue that

$$\mathbb{E}(X) \leq 3\sqrt{n}\left(\frac{np}{e}\right)^n \leq 3\sqrt{n}(1-\varepsilon)^n = o(1).$$

This verifies part (i).

We see that

$$\mathbb{E}(X) \geq \left(\frac{np}{e}\right)^n \to \infty \tag{12.6}$$

in parts (ii) and (iii).

Fix a hamperm π. Let $H(\pi) = (E_\pi(1), E_\pi(2), \ldots, E_\pi(n))$ be the Hamilton cycle induced by π. Then let $N(b,a)$ be the number of permutations π' such that $|E(H(\pi)) \cap E(H(\pi'))| = b$ and $E(H(\pi)) \cap E(H(\pi'))$ consists of a edge disjoint paths. Here a path is a maximal sub-sequence $F_1, F_2, \ldots, F_q$ of the edges of $H(\pi)$

such that $F_i \cap F_{i+1} \neq \emptyset$ for $1 \leq i < q$. The set $\bigcup_{j=1}^{q} F_j$ may contain other edges of $H(\pi)$. Observe that $N(b,a)$ does not depend on π.

Note that

$$\frac{\mathbb{E}(X^2)}{\mathbb{E}(X)^2} = \frac{n!N(0,0)p^{2n}}{\mathbb{E}(X)^2} + \sum_{b=1}^{n}\sum_{a=1}^{b} \frac{n!N(b,a)p^{2n-b}}{\mathbb{E}(X)^2}.$$

Since trivially, $N(0,0) \leq n!$, we obtain,

$$\frac{\mathbb{E}(X^2)}{\mathbb{E}(X)^2} \leq 1 + \sum_{b=1}^{n}\sum_{a=1}^{b} \frac{n!N(b,a)p^{2n-b}}{\mathbb{E}(X)^2}. \tag{12.7}$$

We show that

$$\sum_{b=1}^{n}\sum_{a=1}^{b} \frac{n!N(b,a)p^{2n-b}}{\mathbb{E}(X)^2} = \sum_{b=1}^{n}\sum_{a=1}^{b} \frac{N(b,a)p^{n-b}}{\mathbb{E}(X)} = o(1). \tag{12.8}$$

The Chebyshev inequality implies that

$$\mathbb{P}(X=0) \leq \frac{\mathbb{E}(X^2)}{\mathbb{E}(X)^2} - 1 = o(1),$$

as required.

It remains to show (12.8). First we find an upper bound on $N(b,a)$. Choose a vertices v_i, $1 \leq i \leq a$, on π. We have at most

$$n^a \tag{12.9}$$

choices. Let

$$b_1 + b_2 + \cdots + b_a = b,$$

where $b_i \geq 1$ is an integer for every $1 \leq i \leq a$. Note that this equation has exactly

$$\binom{b-1}{a-1} < 2^b \tag{12.10}$$

solutions. For every i, we choose a path of length b_i in $H(\pi)$ which starts at v_i. Suppose a path consists of edges $F_1, F_2, \dots, F_q$, $q = b_i$. Assuming that $F_1, \dots, F_j$ are chosen, we have at most k possibilities for F_{j+1}. Hence, every such a path can be selected in most k^{b_i} ways. Consequently, we have at most

$$\prod_{i=1}^{a} k^{b_i} = k^b$$

choices for all a paths.

Thus, by the above considerations we can find a edge disjoint paths in $H(\pi)$ with the total of b edges in at most

$$n^a(2k)^b \tag{12.11}$$

many ways.

Let $P_1, P_2, \dots, P_a$ be any collection of the above a paths. Now we count the number of permutations π' containing these paths.

First we choose for every P_i a sequence of vertices inducing this path in π'. We see the vertices in each edge of P_i in at most $k!$ orderings. Crudely, every such sequence can be chosen in at most $(k!)^{b_i}$ ways. Thus, we have

$$\prod_{i=1}^{a}(k!)^{b_i} = (k!)^b \tag{12.12}$$

choices for all a sequences.

Now we bound the number of permutations containing these sequences. First note that

$$|V(P_i)| \geq b_i + k - 1.$$

Thus we have at most

$$n - \sum_{i=1}^{a}(b_i + k - 1) = n - b - a(k-1) \tag{12.13}$$

vertices not in $V(P_1) \cup \dots \cup V(P_a)$. We choose a permutation σ of $V \setminus (V(P_1) \cup \dots \cup V(P_a))$. Here we have at most

$$(n - b - a(k-1))!$$

choices. Now we extend σ to a permutation of $[n]$. We mark a positions on σ and then insert the sequences. We can do it in

$$\binom{n}{a} a! < n^a$$

ways. Therefore, the number of permutations containing $P_1, P_2, \dots, P_a$ is smaller than

$$(k!)^b (n - b - a(k-1))! n^a. \tag{12.14}$$

Thus, by (12.11) and (12.14) and the Stirling formula we obtain

$$N(b,a) < n^{2a}(2k!k)^b (n - b - a(k-1))! < n^{2a}(2k!k)^b \sqrt{2\pi n} \left(\frac{n}{e}\right)^{n-b-a(k-1)}.$$

Since
$$\mathbb{E}(X) = n!p^n = \sqrt{(2+o(1))\pi n}\left(\frac{n}{e}\right)^n p^n,$$
we get
$$\frac{N(b,a)p^{n-b}}{\mathbb{E}(X)} < (1+o(1))n^{2a}(2k!k)^b\left(\frac{e}{n}\right)^{b+a(k-1)} p^{-b}.$$
Finally, since $a \leq b$ we estimate $e^{b+a(k-1)} \leq e^{kb}$, and consequently,
$$\frac{N(b,a)p^{n-b}}{\mathbb{E}(X)} < \left(\frac{2k!ke^k}{np}\right)^b \frac{1+o(1)}{n^{a(k-3)}}. \tag{12.15}$$

Proof of (ii):
Here $k = 3$ and $np \geq \omega$. Hence, we obtain in (12.15)
$$\frac{N(b,a)p^{n-b}}{\mathbb{E}(X)} \leq (1+o(1))\left(\frac{2k!ke^k}{\omega}\right)^b.$$
Thus,
$$\sum_{b=1}^{n}\sum_{a=1}^{b}\frac{N(b,a)p^{n-b}}{\mathbb{E}(X)} < (1+o(1))\sum_{b=1}^{n} b\left(\frac{2k!ke^k}{\omega}\right)^b = o(1). \tag{12.16}$$

This completes the proof of part (ii).

We prove Part (iii) by estimating $N(b,a)$ more carefully, see Exercise 12.4.2 at the end of the chapter. □

Before leaving Hamilton cycles, we note that Allen, Böttcher, Kohayakawa, and Person [25] describe a polynomial time algorithm for finding a tight Hamilton cycle in $\mathbb{H}_{n,p;k}$ w.h.p. when $p = n^{-1+\varepsilon}$ for a constant $\varepsilon > 0$.

There is a weaker notion of Hamilton cycle due to Berge [101] viz. an alternating sequence $v_1, e_1, v_2, \ldots, v_n, e_n$ of vertices and edges such that (i) $v_1, v_2, \ldots, v_n$ are distinct and (ii) $v_i \in e_{i-1} \cap e_i$ for $i = 1, 2, \ldots, n$. The cycle is weak if we do not insist that the edges are distinct. Poole [745] proves that the threshold for the existence of a weak Hamilton cycle in $\mathbb{H}_{n,m;k}$ is equal to the threshold for minimum degree one.

12.3 Perfect matchings in r-regular s-uniform hypergraphs

In this section we discuss the use of a powerful method called *small subgraph conditioning*. This was first employed by Robinson and Wormald [768] to show that random regular graphs are Hamiltonian w.h.p.

Given a hypergraph H with vertex set V, for $v \in V$, let $d_H(v) = |\{i : v \in X_i\}|$ be the *degree* of v. We call H $r-$*regular* if $d_H(v) = r$ for all $v \in V$. Let now $V = [sn]$, where $[k] = \{1, 2, \ldots, k\}$ for all positive integers k, and let $\mathscr{G} = \mathscr{G}(n, r, s) = \{G = (V, E) : G \text{ is } r-\text{regular and } s-\text{uniform }\}$. Let $H = H_{n,r,s}$ be chosen uniformly at random from $\mathscr{G}$. The main aim of this section is to prove the following result of Cooper, Frieze, Molloy and Reed [269].

Theorem 12.11. *Suppose r, s are fixed positive integers, $r \geq 3$, then*

$$\lim_{n \to \infty} \Pr(H_{n,r,s} \text{ has a perfect matching }) = \begin{cases} 0 & s > \sigma_r \\ 1 & s < \sigma_r \end{cases}$$

where

$$\sigma_r = \frac{\log r}{(r-1)\log\left(\frac{r}{r-1}\right)} + 1.$$

We note that if r is at least 3, then σ_r is never an integer, and so this result is best possible.

Next let $f(s) = \min\{r : s < \sigma_r\}$. Thus $f(s)$ gives the threshold in terms of degree for a s-uniform hypergraph to almost surely have a perfect matching. The first few values of $f(s)$ are shown in Table 12.1. For s large, note that $f(s)$ is approximately e^{s-1}, for example $e^8 = 2980.1$ and $e^9 = 8103.1$.

s	2	3	4	5	6	7	8	9	10
$f(s)$	3	7	19	53	146	401	1094	2977	8098

Table 12.1:

Configurations

Let $W_v = \{v\} \times [r]$ for $v \in V = [sn]$ and $W = \bigcup_{v \in V} W_v$. Each W_v should be regarded as a *block* of r *fractional edges* for each $v \in V$, thus generalising the concept of half-edges arising from the use of configurations in the context of graphs. In this paper, a *configuration* is a partition of W into $m = rn$ subsets S of size s. Equivalently, a configuration is a set of m disjoint subsets of W, each of size s. Let $\Omega = \Omega(n, r, s)$ be the set of all such configurations, and let $F = F(n, r, s)$ be chosen randomly from Ω.

For $x = (v, i) \in W$ we let $V(x) = v$. If $F \in \Omega$ and $S \in F$ we let $V(S) = \{V(x) : x \in S\}$. We define the multigraph $\gamma(F) = (V, \{V(S) : S \in F\})$.

A configuration F is said to be *simple*, if $S \in F$ implies $|V(S)| = s$ and any two distinct sets $S_1, S_2 \in F$ satisfy $V(S_1) \neq V(S_2)$. Thus $\gamma(F)$ is s-uniform if and only if F is simple.

For us the main properties of the connection between configurations and graphs are the following.

(A) Each $G \in \mathscr{G}$ arises from precisely $(r!)^{sn}$ simple configurations F.

(B) $\lim_{n\to\infty} \mathbf{Pr}(F \text{ is simple }) = e^{-(s-1)(r-1)/2}$

The assertion **(B)** follows from (12.19) with $k = 1$, applied to Lemma 12.14 (as $|V(S)| < s$ is a 1-cycle in the context of this paper), and the observation that

$$\lim_{n\to\infty} \mathbf{Pr}(\exists S_1, S_2 \in F \text{ with } V(S_1) = V(S_2)) = 0.$$

We will say that a perfect matching of F is a set $\{S_i : i \in I\} \subseteq F$ such that
(i) $|V(S_i)| = s$, for all $i \in I$,
(ii) $i, j \in I, i \neq j$ implies $V(S_i) \cap V(S_j) = \emptyset$, and
(iii) $\bigcup_{i\in I} V(S_i) = V$.

Thus if F is simple, it has a perfect matching if and only if $\gamma(F)$ has a perfect matching. Hence Theorem 12.11 will follow immediately from (**A**) and (**B**) above and the theorem below.

Theorem 12.12.

$$\lim_{n\to\infty} \mathbf{Pr}(F \textit{ has a perfect matching }) = \begin{cases} 0 & s > \sigma_r \\ 1 & s < \sigma_r \end{cases}$$

Outline of a Proof of Theorem 12.12

We use the notation $\alpha \approx \beta$ to mean $\alpha = (1 + o(1))\beta$ where the $o(1)$ term tends to zero as n tends to infinity. All subsequent inequalities are only claimed to hold for sufficiently large n.

Suppose that F is chosen randomly from Ω. Let $Z(F)$ denote the number of perfect matchings in F. The proof of the following lemma is left as Exercise 12.4.3.

Lemma 12.13.

$$\mathbb{E}(Z) \approx \sqrt{s}\left(r\left(\frac{r-1}{r}\right)^{(s-1)(r-1)}\right)^n, \tag{12.17}$$

$$\frac{\mathbb{E}(Z^2)}{\mathbb{E}(Z)^2} \approx \sqrt{\frac{r-1}{r-s}}, \qquad \textit{if } s < \sigma_r. \tag{12.18}$$

Notice that the (easy) first part of Theorem 1 now follows immediately since the right-hand side of (1) tends to zero exponentially fast when $s > \sigma_r$.

To apply the Analysis of Variance technique, we have to decide on a partition of Ω. We proceed analogously to Robinson and Wormald. For the moment let b,x be arbitrarily large *fixed* positive integers.

We now define a $k-$cycle of F for integer $k \geq 1$.

$k = 1$: $S \in F$ is a $1-$cycle if $|V(S)| < s$.

$k = 2$: $S_1, S_2 \in F$ form a $2-$cycle if $|V(S_1) \cap V(S_2)| \geq 2$.

$k \geq 3$: $S_1, S_2, \dots, S_k \in F$ form a $k-$cycle if there exist distinct $v_1, v_2, \dots, v_k \in V$ such that $v_i \in V(S_i) \cap V(S_{i+1})$ for $1 \leq i \leq k$, $(S_{k+1} \equiv S_1)$.

Observe that F is simple if and only if it has no 1-cycles and yields no repeated edges.

Next let C_k denote the number of $k-$cycles of F for $k \geq 1$. For $\mathbf{c} = (c_1, c_2, \dots, c_b) \in N^b$, where $N = \{0, 1, 2, \dots\}$, let $\Omega_{\mathbf{c}} = \{F \in \Omega : C_k = c_k, 1 \leq k \leq b\}$. Let

$$\lambda_k = \frac{((s-1)(r-1))^k}{2k}. \tag{12.19}$$

The proof of the following lemma is left as Exercise 12.4.4.

Lemma 12.14. *Let* $\mathbf{c}$ *be fixed, then*

$$\pi_c = \mathbf{Pr}(F \in \Omega_{\mathbf{c}}) \approx \prod_{k=1}^{b} \frac{\lambda_k^{c_k} e^{-\lambda_k}}{k!}.$$

Now, for $x > 0$, define

$$S(x) = \{\mathbf{c} \in N^b : |c_k - \lambda_k| \leq x\lambda_k^{2/3}, 1 \leq k \leq b\},$$

and

$$\overline{\Omega} = \bigcup_{\mathbf{c} \notin S(x)} \Omega_{\mathbf{c}}.$$

Let

$$\overline{\pi} = \mathbf{Pr}(F \in \overline{\Omega}).$$

For $\mathbf{c} \in N^b$ let

$$E_{\mathbf{c}} = \mathbb{E}(Z \mid F \in \Omega_{\mathbf{c}})$$

and

$$V_{\mathbf{c}} = \mathbf{Var}(Z \mid F \in \Omega_{\mathbf{c}}).$$

Then we have

$$\mathbf{E}(Z^2) = \sum_{\mathbf{c} \in N^b} \pi_{\mathbf{c}} V_{\mathbf{c}} + \sum_{\mathbf{c} \in N^b} \pi_{\mathbf{c}} E_{\mathbf{c}}^2. \tag{12.20}$$

The following two lemmas contain the most important observations. Lemma 12.15 shows that for most groups, the group mean is large and Lemma 19.6 shows that most of the variance can be explained by the *variance between groups*. The proof of the following lemma is left as Exercise 12.4.5.

Lemma 12.15. *For all sufficiently large x the following assertions hold.*
(a) $\overline{\pi} \leq e^{-\alpha x}$ for some absolute constant $\alpha > 0$.
(b) $\mathbf{c} \in S(x)$ implies

$$E_{\mathbf{c}} \geq e^{-(\beta+\gamma x)}\,\mathbb{E}(Z),$$

for some absolute constants $\beta, \gamma > 0$. □

The proof of the following lemma is left as Exercise 12.4.6.

Lemma 12.16. *If x is sufficiently large then*

$$\sum_{\mathbf{c}\in S(x)} \pi_{\mathbf{c}} E_{\mathbf{c}}^2 \geq \left(1 - be^{-3\gamma x}\right)\left(1 - \left(\frac{s-1}{r-1}\right)^b\right)\left(\sqrt{\frac{r-1}{r-s}}\right)\mathbb{E}(Z)^2.$$

where γ is as in Lemma 12.15. □

Hence we have from (12.18) and (12.20),

$$\sum_{\mathbf{c}\in N^b} \pi_{\mathbf{c}} V_{\mathbf{c}} \leq \delta\,\mathbb{E}(Z)^2, \tag{12.21}$$

where $\delta = \left(be^{-3\gamma x} + (\frac{s-1}{r-1})^b\right)$ The rest is an application of the Chebyshev inequality. Define the random variable $\hat{Z}(F)$ by

$$\hat{Z}(F) = E_{\mathbf{c}},\ \text{if } F \in \Omega_{\mathbf{c}}.$$

Then for any $t > 0$

$$\begin{aligned}
\mathbf{Pr}(|Z-\hat{Z}| \geq t) &\leq \mathbb{E}((Z-\hat{Z})^2/t^2) \\
&= \sum_{\mathbf{c}\in N^b} \pi_{\mathbf{c}} V_{\mathbf{c}}/t^2 \\
&\leq \delta\,\mathbb{E}(Z)^2/t^2, \qquad (12.22)
\end{aligned}$$

where the last inequality follows from (12.21).

Now put $t = e^{-(\beta+\gamma x)}\,\mathbb{E}(Z)/2$ where β, γ are from Lemma 12.15. Applying Lemma 12.15 we obtain

$$\begin{aligned}
\mathbf{Pr}(Z \neq 0) &\geq \mathbf{Pr}(Z \geq e^{-(\beta+\gamma x)}\,\mathbb{E}(Z)/2) \\
&\geq \mathbf{Pr}(\left|Z-\hat{Z}\right| \leq t \wedge (F \notin \overline{\Omega}))
\end{aligned}$$

$$\geq \quad 1-4\delta e^{2(\beta+\gamma x)}-\overline{\pi}$$

Hence, using Lemma 12.15,

$$\lim_{n\to\infty} \mathbf{Pr}(Z=0) \leq 4e^{2\beta}\left(be^{-\gamma x}+\left(\frac{s-1}{r-1}\right)^{b} e^{2\gamma x}\right)+e^{-\alpha x} \tag{12.23}$$

Note that $(s-1)/(r-1)<1/2$ so putting $b=3\gamma x/\log 2$ and choosing x large enough the right-hand side of (12.23) becomes as small as we like. Hence (12.23) implies that $\lim_{n\to\infty}\mathbf{Pr}(Z=0)=0$, proving Theorem 12.12.

At this point we have achieved our objective of introducing the notion of small subgraph conditioning.

12.4 Exercises

12.4.1 Show that if $m=cn\log n$ then (i) $ck<1$ implies that $\mathbb{H}_{n,m;k}$ has isolated vertices w.h.p. and (ii) if $ck>1$ then $\mathbb{H}_{n,m;k}$ is connected w.h.p.

12.4.2 Use the configuration model to k-uniform, $k\geq 3$, hypergraphs. Use it to show that if $r=O(1)$ then the number of r-regular, k-uniform hypergraphs with vertex set $[n]$, $k|n$ is asymptotically equal to

$$\frac{(rn)!}{(k!)^{rn/k}r!^{n}(rn/k)!}e^{-(k-1)(r-1)/2}.$$

12.4.3 Prove Lemma 12.13.

12.4.4 Prove Lemma 12.14.

12.4.5 Prove Lemma 12.15.

12.4.6 Prove lemma 12.16.

12.4.7 Generalise the notion of switchings to k-uniform hypergraphs. Use them to extend the result of Exercise 12.4.2 to $r=n^{\varepsilon}$, for sufficiently small $\varepsilon>0$.

12.4.8 Extend the result of Exercise 12.4.2 to k-uniform, $k\geq 3$, hypergraphs with a fixed degree sequence $\mathbf{d}=(d_1,d_2,\ldots,d_n)$, and maximum degree $\Delta=O(1)$. Let $M_1=\sum_{i=1}^{n}d_i$ where $k|M_1$ and $M_2=\sum_{i=1}^{n}d_i(d_i-1)$. Show that the number of k-uniform hypergraphs with degree sequence $\mathbf{d}$ is asymptotically equal to

$$\frac{M_1!}{k!^{M_1/k}\left(\prod_{i=1}^{n}d_i!\right)(M_1/k)!}\exp\left\{-\frac{(k-1)M_2}{2M_1}\right\}.$$

12.4.9 Prove Lemma 12.7.

12.4.10 Prove Lemma 12.9.

12.4.11 Let $\mathbb{H}_{n,\mathbf{d}}$ be a random k-uniform hypergraph with a fixed degree sequence. Suppose that there are $\lambda_i n$ vertices of of degree $i = 1,2,\ldots,L = O(1)$. Let $\Lambda = \sum_{i=1}^{L} \lambda_i i((i-1)(k-1)-1)$. Prove that for $\varepsilon > 0$, we have

(a) If $\Lambda < -\varepsilon$ then w.h.p. the size of the largest component in $\mathbb{H}_{n,\mathbf{d}}$ is $O(\log n)$.

(b) If $\Lambda > \varepsilon$ then w.h.p. there is a unique giant component of linear size $\approx \Theta n$ where Θ is defined as follows: let $K = \sum_{i=1}^{L} i\lambda_i$ and

$$f(\alpha) = K - k\alpha - \sum_{i=1}^{L} i\lambda_i \left(1 - \frac{k\alpha}{K}\right)^{i/k}. \tag{12.24}$$

Let ψ be the smallest positive solution to $f(\alpha) = 0$. Then

$$\Theta = 1 - \sum_{i=1}^{L} \lambda_i \left(1 - \frac{k\psi}{K}\right)^{i/k}.$$

If $\lambda_1 = 0$ then $\Theta = 1$, otherwise $0 < \Theta < 1$.

(c) In Case (b), the degree sequence of the graph obtained by deleting the giant component satisfies the conditions of (a).

12.4.12 Prove Part (iii) of Theorem 12.10 by showing that

$$\begin{aligned} N(b,a) &\leq n^{2a} \binom{b-1}{a-1} \sum_{t \geq 0} 2^{t+a} (n-b-a(k-1)-t)!(k!)^{a+t} \\ &\leq c_k (2k!)^a (n-b-a(k-1))!, \end{aligned}$$

where c_k depends only on k.
Then use (12.7).

12.4.13 In a *directed* k-uniform hypergraph, the vertices of each edge are totally order. Thus each k-set has $k!$ possible orientations. Given a permutation $i_1, i_2, \ldots, i_n$ of $[n]$ we construct a directed ℓ-overlapping Hamilton cycle $\vec{E}_1 = (i_1, \ldots, i_k), \vec{E}_2 = (i_{k-\ell+1} \ldots, i_{2k-\ell}), \ldots, \vec{E}_{m_\ell} = (i_{n-(k-\ell)+1}, \ldots, i_\ell)$. Let $\vec{\mathbb{H}}_{n,p:k}$ be the directed hypergraph in which each possible directed edge is included with probability p. Use the idea of McDiarmid in Section 11.2 to show (see Ferber [352]) that

$$\begin{aligned} &\mathbb{P}(\vec{\mathbb{H}}_{n,p:k} \text{ contains a directed } r\text{-overlapping Hamilton cycle}) \\ &\qquad \geq \mathbb{P}(\mathbb{H}_{n,p;k} \text{ contains an } r\text{-overlapping Hamilton cycle}). \end{aligned}$$

12.4.14 A hypergraph $H = (V, E)$ is 2-colorable if there exists a partition of V into two non-empty sets A, B such that $e \cap A \neq \emptyset$, $e \cap B \neq \emptyset$ for all $e \in E$. Let $m = \binom{n}{k} p$. Show that if c is sufficiently large and $m = c2^k n$ then w.h.p. $H_{n,p;k}$ is not 2-colorable.

12.4.15 Verify (22.3) and (22.4)

12.4.16 Given a hypergraph H, let a vertex coloring be *strongly proper* if no edge contains two vertices of the same color. The strong chromatic number $\chi_1(H)$ is the minimum number of color's in a strongly proper coloring. Suppose that $k \geq 3$ and $0 < p < 1$ is constant. Show that w.h.p.

$$\chi_1(H_{n,p;k}) \approx \frac{d}{2\log d} \qquad \text{where } d = \frac{n^{k-1}p}{(k-2)!}.$$

12.5 Notes

Components and cores

If $H = (V, E)$ is a k-uniform hypergraph and $1 \leq j \leq k-1$ then two sets $J_1, J_2 \in \binom{V}{j}$ are said to be j-connected if there is a sequence of serts $E_1, E_2, \ldots, E_\ell$ such that $J_1 \subseteq E_1, J_2 \subseteq E_\ell$ and $|E_i \cap E_{i+1}| \geq j$ for $1 \leq i < \ell$. This defines an equivalence relation on $\binom{V}{j}$ and the equivalance classes are called *j-components*. Karoński and Łuczak [549] studied the sizes of the 1-components of the random hypergraph $H_{n,m;k}$ and proved the existence of a phase transition at $m \approx \frac{n}{k(k-1)}$. Cooley, Kang and Koch [240] generalised this to j-components and proved the existence of a phase transition at $m \approx \frac{\binom{n}{k}}{\left(\binom{k}{j}-1\right)\binom{n}{k-j}}$. As usual, a phase transition corresponds to the emergence of a unique giant, i.e. one of order $\binom{n}{j}$.

The notion of a core extends simply to hypergraphs and the sizes of cores in random hypergraphs has been considered by Molloy [684]. The r-core is the largest sub-hypergraph with minimum degree r. Molloy proved the existence of a constant $c_{k,r}$ such that if $c < c_{r,k}$ then w.h.p. $H_{n,cn;k}$ has no r-core and that if $c > c_{r,k}$ then w.h.p. $H_{n,cn;k}$ has a r-core. The efficiency of the peeling algorithm for finding a core has been considered by Jiang, Mitzenmacher and Thaler [527]. They show that w.h.p. the number of rounds in the peeling algorithm is asymptotically $\frac{\log\log n}{\log(k-1)(r-1)}$ if $c < c_{r,k}$ and $\Omega(\log n)$ if $c > c_{r,k}$. Gao and Molloy [423] show that for $|c - c_{r,k}| \leq n^{-\delta}$, $0 < \delta < 1/2$, the number of rounds grows like $\tilde{\Theta}(n^{\delta/2})$. In this discussion, $(r,k) \neq (2,2)$.

Chromatic number

Krivelevich and Sudakov [602] studied the chromatic number of the random k-uniform hypergraph $H_{n,p;k}$. For $1 \leq \gamma \leq k-1$ we say that a set of vertices S is γ-independent in a hypergraph H if $|S \cap e| \leq \gamma$. The γ-chromatic number of a hypergraph $H = (V,E)$ is the minimum number of sets in a partition of V into γ-independent sets. They show that if $d^{(\gamma)} = \gamma\binom{k-1}{\gamma}\binom{n-1}{k-1}p$ is sufficiently large then w.h.p. $\frac{d^{(\gamma)}}{(\gamma+1)\log d^{(\gamma)}}$ is a good estimate of the γ-chromatic number of $H_{n,p;k}$.

Dyer, Frieze and Greenhill [322] extended the results of [6] to hypergraphs. Let $u_{k,\ell} = \ell^{k-1}\log \ell$. They show that if $u_{k,\ell-1} < c < u_{k,\ell}$ then w.h.p. the (weak $(\gamma = k-1)$) chromatic number of $H_{n,cn;k}$ is either k or $k+1$.

Achlioptas, Kim, Krivelevich and Tetali [3] studied the 2-colorability of $H = H_{n,p;k}$. Let $m = \binom{n}{k}p$ be the expected number of edges in H. They show that if $m = c2^k n$ and $c > \frac{\log 2}{2}$ then w.h.p. H is not 2-colorable. They also show that if c is a small enough constant then w.h.p. H is 2-colorable.

Orientability

Gao and Wormald [425], Fountoulakis, Khosla and Panagiotou [373] and Lelarge [619] discuss the orientability of random hypergraphs. Suppose that $0 < \ell < k$. To ℓ-orient an edge e of a k-uniform hypergraph $H = (V,E)$, we assign positive signs to ℓ of its vertices and $k-\ell$ negative signs to the rest. An (ℓ,r)-orientation of H consists of an ℓ-orientation of each of its edges so that each vertex receives at most r positive signs due to incident edges. This notion has uses in load balancing. The papers establish a threshold for the existence of an (ℓ,r)-orientation. Describing it it is somewhat complex and we refer the reader to the papers themselves.

VC-dimension

Ycart and Ratsaby [862] discuss the VC-dimension of $H = H_{n,p;k}$. Let $p = cn{-}\alpha$ for constants c,α. They give the likely VC-dimension of H for various values of α. For example if $h \in [k]$ and $\alpha = k - \frac{h(h-1)}{h+1}$ then the VC-dimension is h or $h-1$ w.h.p.

Erdős-Ko-Rado

The famous Erdős-Ko-Rado theorem states that if $n > 2k$ then the maximum size of a family of mutually intersecting k-subsets of $[n]$ is $\binom{n-1}{k-1}$ and this is achieved by all the subsets that contain the element 1. Such collections will be called *stars*. Balogh, Bohman and Mubayi [62] considered this problem in relation to the random hypergraph $H_{n,p;k}$. They consider for what values of k, p is it true that maximum size intersecting family of edges is w.h.p. a star. More recently Hamm and Kahn [462], [463] have answered some of these questions. For many ranges of k, p the answer is as yet unknown.

Bohman, Cooper, Frieze, Martin and Ruszinko [133] and Bohman, Frieze, Martin, Ruszinko and Smyth [136] studied the k-uniform hypergraph H obtained by adding random k-sets one by one, only adding a set if it intersects all previous sets. They prove that w.h.p. H is a star for $k = o(n^{1/3})$ and were able to analyse the structure of H for $k = o(n^{5/12})$.

Hamilton cycles in regular hypergraphs

Dudek, Frieze, Ruciński and Šileikis [308] made some progress on loose Hamilton cycles in random regular hypergraphs. Their approach was to find an embeddding of $\mathbb{H}_{n,m;k}$ in a random regular k-uniform hypergraph.

Chapter 13

Random Subgraphs of the Hypercube

While most work on random graphs has been on random subgraphs of K_n, a complete graph on n vertices, it is true to say that there has also been a good deal of work on random subgraphs of the n-cube, $\mathbb{Q}^n$. This has vertex set $V_n = \{0,1\}^n$ and an edge between $u, v \in V_n$ iff they differ in exactly one coordinate. Notice that $\mathbb{Q}^n$ has $N = 2^n$ vertices and $m = n2^{n-1}$ edges. To obtain a subgraph, we independently delete edges of $\mathbb{Q}^n$, each with probability $1-p$ or, alternatively, we retain each of its edges independently and with probability p. We denote the resulting subgraph by $\mathbb{Q}_p^n$.

13.1 The Evolution

As we have already learned from Chapter 2 the evolution of Erdős - Rényi-type random graphs has clearly distinguishable phases. In fact, this is also a characteristic feature of the evolution of a random subgraph $\mathbb{Q}_p^n$. Indeed, in the first phase (called subcritical) all connected components are small, of order not exceeding $O(n)$, while after passing a critical edge probability p_c and so entering its next phase (called supercritical), a giant connected component emerges which dominates the structure of $\mathbb{Q}_p^n$. The next theorem, due to Ajtai, Komlós and Szemerédi [10] and Bollobás, Kohayakaya and Łuczak [170], describes this process in details. A recent, elegant, compact and self-contained proof presented here is due to Krivelevich [590]

Theorem 13.1. *Let $c > 0$ be constant, and let $p = \frac{c}{n}$. Then*

(i) if $c < 1$ then w.h.p. the order of the largest component of $\mathbb{Q}_p^n$ is at most $O(n)$,

(ii) *if $c > 1$ then w.h.p. $\mathbb{Q}_p^n$ has a unique giant connected component, whose order is asymptotic to $y2^n$, where $y = y(c)$ is the (unique) solution in $(0,1)$ of the equation $y = 1 - \exp\{-cy\}$, while all its remaining connected components are of order at most $O(n)$.*

Proof. We start with a short proof of statement (i).
Let $p = \frac{c}{n}$ where $c < 1$. Put $c = 1 - \varepsilon$, where ε is sufficiently small. Let us run the Breadth First Search (BFS) algorithm on $\mathbb{Q}_p^n$ guided by a sequence of random bits $\bar{X} = (X_i)_{i=1}^m$, where $m = n2^{n-1}$, and the X_i's are independent $Bernoulli(p)$ random variables. We move step by step exposing edges via BFS, each time deciding whether a given edge queried at step i is an edge of $\mathbb{Q}_p^n$ if and only if the respective random bit $X_i = 1$. Suppose that, after finishing the search, $\mathbb{Q}_p^n$ has a connected component K with more than k vertices, where k is a positive integer. Suppose now that we are in that moment of the process that we has just discovered the $(k+1)$-th vertex of K. In the phase of discovering K up until this moment, only the edges of $\mathbb{Q}^n$ incident to the set K_0 of the first k discovered vertices of K have been queried, and $k-1$ edges of $\mathbb{Q}_p^n$ induced by this set of vertices have been revealed. So there are at least $k-1$ edges of $\mathbb{Q}^n$ spanned by K_0. Hence, the total number of edges of $\mathbb{Q}^n$ queried so far in this phase is at most $kn - (k-1) = k(n-1) + 1$.

It follows that the sequence $\bar{X}$ contains $k(n-1)+1$ consecutive bits X_i out of which at least k are equal to 1. Probability of such event can be bounded by standard Chernoff- type bound (see Section 27.4) as follows

$$\mathbb{P}\left(Binom\left(k(n-1)+1,(1-\varepsilon)/n\right) \geq k\right) \leq e^{-\varepsilon^2 k/4}.$$

Therefore the probability of having at least one interval of length $k(n-1)+1$ in the sequence of $m = n2^{n-1} < 2^{2n}$ random bits with at least k ones is at most $2^{2n}e^{-\varepsilon^2 k/4}$. Taking $k = 9n/\varepsilon^2$, we see that w.h.p. there is no such interval in $\bar{X}$, and so w.h.p. $\mathbb{Q}_p^n$ has no component larger than k when $c < 1$.

Now we proceed to the proof of statement (ii) of Theorem 13.1. Let $p = \frac{c}{n}$ where $c > 1$ and assume that $t > 1$ is a fixed integer. Let $v \in V(\mathbb{Q}^n)$, and denote by $C(v)$ the connected component $\mathbb{Q}_p^n$ containing v. Finally, let

$$W = \{v : |C(v)| \geq n^t\}.$$

The main objective of the first part of the proof of statement (ii) is to show that w.h.p.

(**a**) all connected components outside W are of order $O(n)$,

(**b**) $|W| = (1+o(1))y2^n$, where $y = y(c)$ is the (unique) solution in $(0,1)$ of the equation $y = 1 - \exp\{-cy\}$, (**c**) every vertex $v \in V(\mathbb{Q}^n)$ is at distance at most 2 (in $\mathbb{Q}^n$) from W.

To prove (**a**), let $|C(v)| = k$ where $k \in \left[\frac{n}{c-1-\ln c}, n^t\right]$. Obviously the connected component $C(v)$ spans a tree and no vertex of such a tree connects with any vertex lying outside the component. So, for such an event to occur, there should be a tree T of order k in $\mathbb{Q}^n$ with no edges of $\mathbb{Q}^n$ between $V(T)$ and $V(\mathbb{Q}^n) \setminus V(T)$ present in $\mathbb{Q}^n_p$. There are at least $k(n - 2\log_2 k)$ such edges in $\mathbb{Q}^n$ (Harper's isoperimetric inequality, see Exercise 13.4.2). Notice also that the number of trees T rooted at vertex v can be bounded by $(en)^{k-1}$ (see Exercise 13.4.1). Combining both observations, we get

$$\begin{aligned}
\mathbb{P}\left(\frac{2n}{c-1-\log c} \le |C(v)| \le n^t\right) &\le \sum_{k=\frac{2n}{c-1-\log c}}^{n^t} (en)^{k-1} p^{k-1} (1-p)^{k(n-2\log_2 k)} \\
&= \sum_{k=\frac{2n}{c-1-\log c}}^{n^t} (en)^{k-1} \left(\frac{c}{n}\right)^{k-1} e^{-\frac{c}{n}(kn - 2k\log_2 k)} \\
&\le \sum_{k=\frac{2n}{c-1-\log c}}^{n^t} \left(ece^{-c+\frac{2c\log_2 k}{n}}\right)^k \\
&= \sum_{k=\frac{2n}{c-1-\log c}}^{n^t} \left(e^{-c+1+\log c + o(1)}\right)^k = o(2^{-n}).
\end{aligned}$$

Applying the union bound over all 2^n vertices we see that $\mathbb{Q}^n_p$ has no components of order between $\frac{2n}{c-1-\ln c}$ and n^t, and so the statement (**a**) that w.h.p. all components outside W are of the order $O(n)$ holds.

To see that statement (**b**) also holds we first estimate the probability that $|C(v)|$ is between $n^{1/2}$ and n^t. One can easily see, following the computations above, that

$$\mathbb{P}\left(n^{1/2} \le |C(v)| \le n^t\right) \le \sum_{k=n^{1/2}}^{n^t} \left(ece^{-c+\frac{2c\log_2 k}{n}}\right)^k = o(1), \tag{13.1}$$

since $ece^{-c} < 1$ for $c \neq 1$.

Return now to the BFS algorithm on $\mathbb{Q}^n$ introduced in the proof of statement (i), starting from v and feeding it with independent Bernoulli(p) bits, one for each queried edge of $\mathbb{Q}^n$. For as long as $|C(v)| \le n^{1/2}$, every vertex $u \in V(\mathbb{Q}^n)$, queried for neighbors outside of the current component $C(v)$, has at least $n - n^{1/2}$ potential neighbors to query. Hence the exploration process can be coupled with a Galton-Watson tree rooted at v with offspring distribution Binom$(n - n^{1/2}, p)$. Since $(n - n^{1/2})p = c - o(1)$ the component $C(v)$ grows to $n^{1/2}$ with the same

probability as the probability of survival of the respective Galton-Watson tree, i.e., with probability asymptotic to y, as specified in (**a**) (see, for example, Durrett's book [317]).
This, combined with (13.1) leads to the conclusion, that

$$\mathbb{P}\left(|C(v)| \geq n^t\right) = (1 - o(1))y,$$

and hence that

$$\mathbb{E}(|W|) = (1 + o(1))y2^n.$$

One can easily check that $|W|$ is highly concentrated around its expectation. Indeed, applying edge-exposure martingale (see Section 27.7), we see that adding or deleting an edge can change the value of $|W|$ by at most $2n^t$. Hence by the standard martingale inequality (see Theorem 27.16),

$$\mathbb{P}\left(\mid |W| - \mathbb{E}(|W|) \mid \geq 2^{2n/3}\right) \leq 2\exp\left(-\frac{2^{4n/3}}{2n2^{n-1}(2n^t)^2}\right) = o(1).$$

Therefore w.h.p.

$$|W| = (1 + o(1))y2^n.$$

which completes the proof of (**b**).

Finally, to see that (**c**) holds we choose a vertex v_0 with all its coordinates equal to 0 and show that it is at distance at most two from W with probability $1 - o(2^{-n})$. Set $\varepsilon = (c-1)/c$ and let I be the set of first $\lfloor \varepsilon n/2 \rfloor$ vertex coordinates. For $i \neq j \in I$, consider the subcube of $\mathbb{Q}^n$

$$H_{ij}^n = \{v \in \mathbb{Q}^n : v_i = v_j = 1, v_k = 0 \text{ for all } k \in I \setminus \{i, j\}\}.$$

Let u_{ij} be the vertex with 1 in coordinates i and j and 0 in all other coordinates. Clearly, u_{ij} is at distance 2 from v_0 and u_{ij} belongs to H_{ij}^n. Moreover, the subcubes H_{ij}^n are vertex disjoint and have dimension $n' = n - \lfloor \varepsilon n/2 \rfloor \geq (1 - \varepsilon/2)n$. Generating a random subgraph of H_{ij}^n with edge probability $p = c/n$, $c > 1$ and observing that $n'p \geq (1 - \varepsilon/2)np = (1 - \varepsilon/2)c = (c+1)/2$, we have that the components of H_{ij}^n containing u_{ij} behave the same as the components $C(v)$ of $\mathbb{Q}_p^n$ (see (**a**)). Therefore

$$\mathbb{P}\left(|C(u_{ij})| \geq n^t\right) = (1 + o(1))y(n'p) = (1 + o(1))\left(\frac{c+1}{2}\right) \geq ;$$

for some $\delta = \delta(c) > 0$.
These events are independent for distinct pairs $\{i, j\}$. Hence the probability that v_0 is not at distance at most two (in $\mathbb{Q}^n$) from W is at most

$$\mathbb{P}\left(Binom(\varepsilon^2 n^2/5,) = 0\right) \leq e^{-\Omega(n^2)} = o(2^{-n}).$$

Taking the union bound over vertices v leads to the conclusion that every vertex $v \in \mathbb{Q}^n$ is w.h.p. at distance at most two from W and (**c**) follows.

In the final step of the proof we will show that in fact the vertices composing the set W belong not to separate large components but are members of the same giant component of the order asymptotic to $y2^n$. To this goal we apply the well known coupling technique, introduced in the first section of Chapter 1, to generate a random subgraph $\mathbb{Q}_p^n$ in two steps.

Let $p = c/n,\ c > 1$ and probabilities p_1 and p_2 be defined as follows. Let $p_2 = 1/n^5$ and let p_1 be such that $(1-p_1)(1-p_2) = 1-p$. Let us generate two independent random subgraphs $\mathbb{Q}_{p_1}^n$ and $\mathbb{Q}_{p_2}^n$ on the same set of 2^n vertices, and then take their union. Note that $\mathbb{Q}_{p_1}^n \cup \mathbb{Q}_{p_2}^n$ has the same distribution as $\mathbb{Q}_p^n$ and that $p_1 = \frac{c-o(1)}{n}$.

Define sets $C_1(v)$ and W_1 in $\mathbb{Q}_{p_1}^n$ in an analogous way to $C(v)$ and W in $\mathbb{Q}_p^n$, i.e., $C_1(v)$ denotes the connected component of a random subgraph $\mathbb{Q}_{p_1}^n$ containing vertex v, while

$$W_1 = \{v : |C_1(v)| \geq n^t\}.$$

Note that $\mathbb{Q}_{p_1}^n$ differs only slightly from $\mathbb{Q}_p^n$, and therefore the set W_1 shares the properties (**a**), (**b**) and (**c**) with set W, i.e, all connected components outside W_1 are of order $O(n)$, $|W_1| = (1+o(1))y2^n$, where $y = y(c)$ is the (unique) solution in $(0,1)$ of the equation $y = 1-\exp\{-cy\}$, and every vertex $v \in V(\mathbb{Q}^n)$ is at distance at most two (in $\mathbb{Q}^n$) from W_1.

Now, fix $t = 41$ and generate and expose the edges of $\mathbb{Q}_{p_2}^n$. Under these circumstances we will show that all components $C_1(v)$ with at least n^{41} vertices merge into one component in the union of $\mathbb{Q}_{p_1}^n \cup \mathbb{Q}_{p_1}^n \sim \mathbb{Q}_p^n$.

Assume that it is not true. Then we can partition the components $\mathbb{Q}_{p_1}^n[W_1]$ of $\mathbb{Q}_{p_1}^n$ induced by vertex set W_1 into two families $\mathscr{A}$ and $\mathscr{B}$ such that there are no paths in $\mathbb{Q}_{p_2}^n$ between sets

$$A := \bigcup_{C_1 \in \mathscr{A}} V(C_1) \quad \text{and} \quad B := \bigcup_{C_1 \in \mathscr{B}} V(C_1).$$

Let $s \leq 2^n$ be the total number of components in $\mathbb{Q}_{p_1}^n[W_1]$, and let $r \leq s/2$ be the number of components in the family with fewer components. Then $|A|,|B| \geq rn^{41}$. Now, since every vertex of $\mathbb{Q}^n$ is at distance at most two from $A \cup B$, we can partition $V(\mathbb{Q}^n)$ into sets A', B', where $A \subseteq A'$ and A' contains all vertices of $V(\mathbb{Q}^n) \setminus B$ at distance at most two from A, and every vertex in $B' := V(\mathbb{Q}^n) \setminus A'$ is at distance at most two from B. By Harper's edge isoperimetric inequality (see Exercise 13.4.3), $\mathbb{Q}^n$ has at least $\min\{|A'|,|B'|\} \geq rn^{41}$ edges between A' and B'. Since every vertex in A' is at distance at most two from A and every vertex in B' is at distance at most two from B, we can extend each edge crossing between A'

and B' to a path of length at most five between A and B. As every edge of $\mathbb{Q}^n$ is contained in less than $5n^4$ paths of length at most five, by applying a simple greedy argument we obtain a family of

$$\frac{rn^{41}}{5\cdot 5n^4}=\frac{rn^{37}}{25}$$

edge disjoint paths of length at most five between A and B. The probability that none of these paths is in $\mathbb{Q}^n_{p_2}$ is at most

$$\left(1-p_2^5\right)^{rn^{37}/25}\le e^{rn^{12}/25}.$$

Also $s<2^n$, and hence the number of ways to partition the components $\mathbb{Q}^n_{p_1}[W_1]$ into $\mathscr{A}$ and $\mathscr{B}$ with r components in one of the families is at most $\binom{s}{r}$. Therefore the probability that the components of $\mathbb{Q}^n_{p_1}[W_1]$ do not merge into a single component in $\mathbb{Q}^n_p$ is at most

$$\sum_{r=1}^{s/2}\binom{s}{r}e^{-rn^{12}/25}\le\sum_{r=1}^{s/2}\left(\frac{se^{1-n^{12}/25}}{r}\right)^r=o(1).$$

Hence w.h.p. all components of W_1 merge into a single giant component L_1 with at least $(1+o(1))y2^n$ vertices.
It remains to check the orders of components outside of the set W_1. Define an auxilliary graph Γ whose vertices are the components of $\mathbb{Q}^n_{p_1}$ outside of W_1, where two vertices (components) are connected by an edge if $\mathbb{Q}^n_{p_2}$ contains at least one edge between them. By the property (**a**) all components of $\mathbb{Q}^n_{p_1}$ outside of W_1 are of order $O(n)$, and thus, using a very crude estimate, there are w.h.p. $O(n^2)$ edges of $\mathbb{Q}^n$ between every pair of connected components. Hence every pair of vertices of Γ are connected by an edge independently and with probability at most

$$q=1-(1-p_2)^{n^2}=O(n^{-3}).$$

It follows that the maximum component size in Γ is at most n. Indeed, the maximum degree of Γ is $\Delta=O(n^3)$ and so

$$\Pr(\Gamma\text{ contains a component of order }k)\le 2^n(e\Delta)^kq^k=2^nO(kn^{-2})^k. \tag{13.2}$$

(At most $2^n(e\Delta)^k$ choices for a tree of size k in Γ (see Exercise 13.4.1).) Thus putting $k=n$ in (13.2) we see that we have the claimed bound on component size in Γ.

Recalling that w.h.p. every component of $\mathbb{Q}^n_{p_1}$ outside of W_1 has $O(n)$ vertices, it follows that w.h.p. all components of $\mathbb{Q}^p_n$ outside of W_1 are of the order $O(n^2)$.

Hence following the proof of property (**a**), we conclude that in fact all such components $L_i, i \geq 2$ satisfy
$$|L_i| \leq \frac{2n}{c-1-\ln c}.$$
Finally, by property (**b**), we can give a high probability upper bound on L_1 in $\mathbb{Q}_n^p$, which is
$$|L_1| \leq (1+o(1))y2^n,$$
and so the proof of statement (ii) is completed. □

13.2 Connectivity

The natural question as to whether connectivity of $\mathbb{Q}_p^n$ enjoys a threshold property was answered by Burtin in [205], (and, in the language of boolean functions, by Saposhenko in [789]) who proved that $p = 1/2$ is the threshold for connectedness.

Theorem 13.2.

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{Q}_p^n \textit{ is connected}) = \begin{cases} 0 & \textit{if } p < 1/2, \\ 1 & \textit{if } p > 1/2. \end{cases}$$

□

Next Erdős and Spencer [338] proved that the limiting probability that $\mathbb{Q}_p^n$ is connected tends to $1/e$ for $p = 1/2$ and later Bollobás in [158] established the *hitting time* for connectedness. To understand the nature of Bollobás' result and why it strengthen Burtin's threshold theorem, we have to recall the notion of *a random graph process* (see Section 2), now in the context of the subgraph $\mathbb{Q}^n$. Let $\mathbb{Q}_0$ be the empty (no edges) subgraph of $\mathbb{Q}^n$. At each step $1 \leq k \leq m$, where $m = n2^{n-1}$, $\mathbb{Q}_t^n$ is obtained from $\mathbb{Q}_{t-1}^n$ by adding uniformly at random a new edge from $\mathbb{Q}^n$. We call $\mathbb{Q}_t^n$ the *state* of a random hypercube process $\widetilde{\mathbb{Q}_t^n} = \{\mathbb{Q}_t^n\}_0^{n2^{n-1}}$ at *time* t. The *hitting time* of a monotone increasing non-empty graph property $\mathscr{P}$ is the random variable equal to the time τ for which $\mathbb{Q}_\tau^n \in \mathscr{P}$, but $\mathbb{Q}_{\tau-1}^n \notin \mathscr{P}$.

Theorem 13.3. *Consider the random hypercube process $\widetilde{\mathbb{Q}_t^n}$. Let τ_D denote the hitting time of minimum degree (at least) one, and let τ_C be the hitting time for connectivity of $\mathbb{Q}_p^n$. Then, w.h.p., $\tau_C = \tau_D$.*

To prove Bollobás' result we follow a recent , simplified and self contained version of the proof by Diskin and Krivelevich, given in [300]

Proof. (of Theorem 13.3)

Note that $\tau_C \geq \tau_D$ deterministically. To show that $\tau_C \leq \tau_D$ we will show that the final moments of the evolution of random hypercube before it becomes connected, resemble the analogous moments of the evolution of $\mathbb{G}_{n,p}$. Thus, we show that there exists t such that w.h.p. $\mathbb{Q}^n_t$ has a unique connected component of order $2^n - o(2^n)$ and all other vertices of $\mathbb{Q}^n_t$ are isolated vertices whose distance between them (in $\mathbb{Q}^n$) is at least two. Indeed, then any new edge added in the process either falls within unique giant component or connects an isolated vertex to this component. It implies that a random hypercube becomes connected as soon as its last isolated vertex disappears, so $\tau_C \leq \tau_D$, which implies that w.h.p. $\tau_C = \tau_D$.

Let $\varepsilon > 0$ be a sufficiently small constant, and let $p \geq \frac{1}{2} - \varepsilon$. Then it is easily seen that w.h.p. every pair of isolated vertices in $\mathbb{Q}^n_p$ are at distance at least two in $\mathbb{Q}^n$. To see this, fix an edge $\{u, v\} \in \mathbb{Q}^n$. The probability that both u and v are isolated vertices in in $\mathbb{Q}^n_p$ is

$$(1-p)^{2n-1} = \frac{1}{2^{2n-1}}(1-2\varepsilon)^{2n-1} \leq 2^{-3n/2}.$$

There are $n2^{n-1}$ edges to consider, thus , by the union bound, the probability that there are two isolated vertices in $\mathbb{Q}^n_p$ at distance one in $\mathbb{Q}^n$ is at most $n2^{n-1}2^{-3n/2} = o(1)$.

To show that for $p \geq \frac{1}{2} - \varepsilon$ w.h.p. there is a unique giant connected component in $\mathbb{Q}^n_p$ on at least $(1-o(1))2^n$ vertices while w.h.p. all other components (if there are any) are isolated vertices, is a bit more complicated.

We will first show that if $p \geq \frac{1}{2} - \varepsilon$ then w.h.p. there are no components in $\mathbb{Q}^n_p$ whose order is between 2 and $2^{n/3}$.

Let $k \in [2, 2^{n/3}]$. Let $\mathscr{A}_k$ be the event that there exists a connected component on k vertices in $\mathbb{Q}^n_p$. Obviously such a component spans a tree and no vertex of such a tree connects with any vertex lying outside the component. So, for $\mathscr{A}_k$ to occur, there should be a tree T of order k in $\mathbb{Q}^n$ with no edges of $\mathbb{Q}^n$ between $V(T)$ and $V \setminus V(T)$ present in $\mathbb{Q}^n_p$. There are at least $k(n - 2\log_2 k)$ such edges in $\mathbb{Q}^n$ (see Exercise 13.4.2). Notice also that the number of trees T can be bounded by $2^n(e\Delta)^{k-1} = 2^n(en)^{k-1}$ (see Exercise 13.4.1). Combining both observations, by the union bound, we get

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{2\leq k\leq 2^{n/3}} \mathscr{A}_k\right) &\leq \sum_{k=2}^{2^{n/3}} 2^n (en)^{k-1}(1-p)^{k(n-2\log_2 k)} \\
&\leq \sum_{k=2}^{n} 2^n \left[en\left(\frac{1}{2}+\varepsilon\right)^{n-2\log_2 k}\right]^k + \sum_{n+1}^{2^{n/3}} 2^n \left[en\left(\frac{1}{2}+\varepsilon\right)^{n/3}\right]^k
\end{aligned}$$

$$\leq n2^{-n/3} + 2^{n/3}2^{-n^2/4}$$
$$= o(1).$$

To complete the proof we will apply again a useful coupling technique, which generates the random graph $\mathbb{Q}_p^n$ in two steps.

Let $p_2 = \varepsilon$, and let p_1 be such that

$$(1-p_1)(1-p_2) = 1-p = \frac{1}{2} + \varepsilon.$$

Note that $\mathbb{Q}_{p_1}^n \cup \mathbb{Q}_{p_2}^n$ has the same distribution as $\mathbb{Q}_p^n$ and that $p_1 \geq \frac{1}{2} - 2\varepsilon$. The expected number of isolated vertices in $\mathbb{Q}_{p_1}^n$ is $2^n(1-p_1)^n \leq 2^{9\varepsilon n}$. Thus, by Markov's inequality, there are w.h.p. at most $2^{10\varepsilon n}$ isolated vertices in $\mathbb{Q}_{p_1}^n$. Moreover, w.h.p. in $\mathbb{Q}_{p_1}^n$ there are no components of order between 2 and $2^{n/3}$. Let W be the set of vertices in components of $\mathbb{Q}_{p_1}^n$ of order at least $2^{n/3}$. Note that the number s of such components can not exceed $2^{2n/3}$. Let us move to the next step of the coupling procedure and "sprinkle" edges of $\mathbb{Q}^n$ with probability p_2. We will show that as a result w.h.p. all the components of W merge in $\mathbb{Q}_{p_1}^n \cup \mathbb{Q}_{p_2}^n$. Indeed, otherwise there would have been a component respecting partition $A \sqcup B = W$, with no edges between A and B in $\mathbb{Q}_{p_2}^n$. Suppose, without loss of generality, that $a = |A| \leq |B|$. Then, one can show (see Exercise 13.4.3) that the number of edges between A and its complement, $e(A, A^C) \geq |A|$. As there are at most $n2^{10\varepsilon n}$ edges in $\mathbb{Q}^n$ touching $V \setminus W$, we derive $e(A, A^C \cap W) \geq |A|/2$. Moreover, there are at most

$$\sum_{i=1}^{a2^{-n/3}} \binom{s}{i} \leq s^{a2^{-n/3}} \leq 2^{na2^{-n/3}}$$

ways to partition W into two parts one of which is of order a. Thus, by the union bound, the probability of this event is at most

$$\sum_{a=2^{n/3}}^{2^{n-1}} 2^{na2^{-n/3}}(1-\varepsilon)^{a/2} \leq \sum_{a=2^{n/3}}^{2^{n-1}} 2^{na2^{-n/3}} e^{-\varepsilon a/2} = o(1).$$

In summary, when $t = (2-\varepsilon)^n$ for small ε, the random graph $\mathbb{Q}_t^n$ consists w.h.p. of (i) a giant component C, (ii) a collection of isolated vertices I such that (iii) all the $\mathbb{Q}^n$ potential neighbors of I are in C, and vertices from I can not be joined by an edge from $\mathbb{Q}^n$, since they are at distance at least two from each other in $\mathbb{Q}^n$. Therefore, deterministically, adding more edges to $\mathbb{Q}_t^n$ either puts them into C or connects a vertex in I to C. Thus, $\tau_C = \tau_D$ equals the time when the last vertex of I is connected to C. $\square$

13.3 Perfect Matching

Notice that connectivity and the property of having a perfect matching share in a random subgraph a simple necessary condition of having no isolated vertices, which suggests that both properties have the same threshold edge probability. It was first observed in the case of a random graph $\mathbb{G}_{n,p}$ (see Chapter 4) and it holds, as it is shown in the next theorem due to Bollobás [158], in a random subgraph as well. Note that a hypercube $\mathbb{Q}_n$ is bipartite therefore the main tool to prove this result is Hall's Theorem giving the necessary and sufficient condition for a random subgraph of the hypercube to contain a perfect matching.

Theorem 13.4. *Consider the random hypercube process $\widetilde{\mathbb{Q}_t^n}$. Let τ_D denote the hitting time of minimum degree (at least) one, while let τ_M be the hitting time for having a perfect matching in $\mathbb{Q}_t^n$. Then, w.h.p., $\tau_M = \tau_D$.*

Proof. As in the case of connectivity $\tau_M \geq \tau_D$ deterministically. To complete the proof we have to show that $\tau_M \leq \tau_D$. We will apply a similar approach as in the proof of Theorem 13.3, i.e., we will consider properties of a random subgraph $\mathbb{Q}_p^n$ of $\mathbb{Q}^n$, and then use them to pass from $\mathbb{Q}_p^n$ to a random hypergraph process $\mathbb{Q}_t$.
Let us first introduce some necessary notation. We call a vertex $v \in \{0,1\}^n$ of the cube $\mathbb{Q}^n$ *even* if an v has an even number of digits equal to 1, and call it *odd* otherwise. Given a set $W \subset V(\mathbb{Q}^n)$, we write W_0 for the set of even vertices of W and $W_1 = W \setminus W_0$ for the set of odd vertices.
Because $\mathbb{Q}^n$ is a bipartite graph, to decide whether a spanning subgraph of $\mathbb{Q}^n$ has a perfect matching a necessary and sufficient condition given by Hall's Theorem has to be examined, i.e., one has to show that certain *obstructions* to the existence of a perfect matching (PM) are, in the case of $\mathbb{Q}_p^n$, w.h.p. not present.
For $k \geq 1$, a *k-obstruction* for a PM in $\mathbb{Q}_p^n$ is a set $W = W_0 \cup W_1$ of k vertices spanning a connected subgraph of $\mathbb{Q}^n$, together with an index $i \in \{0,1\}$ such that if $i=0$ then $|W_0| \geq |W_1|$ and all of the neighbors $N(W_0)$ of the vertices W_0 are in W_1, i.e., $N(W_0) \subset W_1$. Similarly, if $i=1$ then $|W_1| \geq |W_0|$ and $N(W_1) \subset W_0$. By Hall's Theorem, if $\mathbb{Q}_p^n$ fails to contain a perfect matching then $\mathbb{Q}_p^n$ has to have an k-obstruction for some odd k, in the range $1 \leq k \leq 2^n/2 = 2^{n-1}$. We are interested in minimal k-obstructions only. We first look for k-obstructions with $k \geq 3$ and leave the case of $k=1$ (isolated vertices) for further consideration.

Denote by $X_k = X_k(\mathbb{Q}_p^n)$ the random variable counting k-obstructions in $\mathbb{Q}_p^n$. So $\sum_{k=3}^{2^{n-1}} X_k$ counts all possible obstructions on at least three vertices. By the First Moment Method (see Lemma 26.2), to show that $\mathbb{Q}_p^n$ w.h.p. does not contain any such obstruction it is enough to prove that,

$$\sum_{k=3}^{2^{n-1}} \mathbb{E} X_k = o(1). \tag{13.3}$$

In fact, proving that it holds for

$$p = p(n) = \frac{1}{2}\left(1 - \frac{c(n)}{n}\right), \text{ where } c(n) = \log\log n, \tag{13.4}$$

constitutes the main part the proof of Theorem 13.4.

Estimation of the sum (13.3) is quite complicated and it will be divided into two parts:

$$\sum_{k=3}^{2^{n-1}} \mathbb{E}X_k = \sum_{k=3}^{\lfloor 2^n/n^9 \rfloor} \mathbb{E}X_k + \sum_{\lfloor 2^n/n^9 \rfloor + 1}^{2^{n-1}} \mathbb{E}X_k \tag{13.5}$$

First choose $3 \le k \le \lfloor 2^n/n^9 \rfloor$. Denote by C_k the number of sets $W \subset V(\mathbb{Q}^n)$, $|W| = k$, such that the graph $\mathbb{Q}^n[W]$ induced by vertex set W is connected. Observe that (see Exercise 13.4.5) there must be a set $A \subset W$ such that $|A| \le \lfloor |W|/3 \rfloor$ and the distance of any vertex of W from the set A is at most 2. Moreover there are at most $|A|(n + \binom{n}{2}) = |A|n(n+1)/2$ vertices not in A but within distance 2 of A and since $\lfloor k/3 \rfloor + \lceil 2k/3 \rceil = k$,

$$\begin{aligned} C_k &\le \binom{2^n}{\lfloor k/3 \rfloor}\binom{\lfloor k/3 \rfloor n(n+1)/2}{\lceil 2k/3 \rceil} \le \left(\frac{3e2^n}{k}\right)^{k/3}\left(\frac{en(n+1)}{4}\right)^{2k/3} \\ &\le \left(\frac{n^4 2^{n+1}}{k}\right)^{k/3}. \end{aligned} \tag{13.6}$$

Let $W = W_0 \cup W_1$ with $|W_i| \ge |W \setminus W_i|$ for either $i = 0$ or $i = 1$. Note that the subgraph $\mathbb{Q}^n[W]$ of $\mathbb{Q}^n$ has at most $\frac{k}{2}\log_2 k$ edges (see Exercise 13.4.4). $\mathbb{Q}^n$ is an n-regular graph and therefore $\mathbb{Q}^n$ has at least $\frac{k}{2}(n - \log_2 k)$ edges joining W_i to $V(\mathbb{Q}^n) \setminus W_i$. Hence $N(W_i) \subset W \setminus W_i$, if and only if none of these edges are in $\mathbb{Q}^n_p$ and so

$$\begin{aligned} \mathbb{P}(N(W_i) \subset W \setminus W_i) &\le (1-p)^{\frac{k}{2}(n-\log_2 k)} = 2^{-\frac{k}{2}(n-\log_2 k)}\left(1 + \frac{c(n)}{n}\right)^{\frac{k}{2}(n-\log_2 k)} \\ &\le \left(\frac{k}{2^n}\right)^{k/2}(\log n)^{k\left(1-\frac{\log_2 k}{n}\right)} \le \left(\frac{k\log^{2(1-\log_2 k/n)} n}{2^n}\right)^{k/2}. \end{aligned} \tag{13.7}$$

Combining (13.6) and (13.7) and recalling the definition of a k-obstruction we get

$$\mathbb{E}X_k \le 2\left(\frac{n^4 2^{n+1}}{k}\right)^{k/3}\left(\frac{k\log^2 n}{2^n}\right)^{k/2}$$

$$= 2\left(\frac{kn^{8/3}\log^6 n}{2^{n-2}}\right)^{k/6}.$$

Hence it implies that

$$\sum_{k=3}^{\lfloor 2^n/n^9\rfloor} \mathbb{E}X_k \leq 2\sum_{k=3}^{\lfloor 2^n/n^9\rfloor}\left(\frac{kn^{8/3}\log^6 n}{2^{n-2}}\right)^{k/6} = o(1), \tag{13.8}$$

and so the first part of the estimation in (13.5) is completed.
Estimation of the second sum, i.e., $\sum_{\lfloor 2^n/n^9\rfloor+1}^{2^{n-1}} \mathbb{E}X_k$, needs much longer computations. Let $\lfloor 2^n/n^9\rfloor + 1 \leq k \leq 2^{n-1}$ and set $\alpha = k/2^n$ so that $k = \alpha 2^n$. Denote by O_k the set of all candidates (W,i), $i = 0,1$, for a k-obstruction, with $W = W_0 \cup W_1$ already distinguished so that for the pair (W,i) to be a k-obstruction we require that no edge of $\mathbb{Q}_p^n$ should join W_i to a vertex not in W. Because $W \subset V(\mathbb{Q}^n)$, ignoring the condition that that $\mathbb{Q}^n[W]$ should be connected, we see that

$$|O_k| < 2\binom{2^n}{k} = 2\binom{2^n}{\alpha 2^n} \leq \left(\frac{e}{\alpha}\right)^{\alpha 2^n}. \tag{13.9}$$

Given a k-set $W \subset V(\mathbb{Q}_n)$ denote by $b = b(W) = \beta k$ the number of edges joining vertices in W to vertices not in W (i.e., the *edge-boundary* of W). Let us estimate the probability that (W,i) is an obstruction in $\mathbb{Q}_p^n$. For either $i = 0$ or $i = 1$ we have

$$\begin{aligned}\mathbb{P}(N(W_i) \subset W \setminus W_i) &\leq (1-p)^{b(W)/2} = (1-p)^{\beta k/2} = 2^{-\beta k/2}\left(1+\frac{c(n)}{n}\right)^{\beta k/2}\\ &\leq 2^{-(\beta - 2c(n)/n)k/2} \leq 2^{-(\beta-1)k/2}.\end{aligned} \tag{13.10}$$

So by (13.9) and (13.10) we get that if $\beta_0 = 2\log_2(2e/\alpha) + 2$ then

$$\begin{aligned} S_1 = &\sum_{\substack{|W|=k\\ b(W)=\beta k, \beta\geq\beta_0}} \mathbb{P}(N(W_i) \subset W \setminus W_i) \leq \left(\frac{e}{\alpha}\right)^k 2^{-(\beta-1)k/2}\\ &\leq 2^{-(\beta-1-2\log_2(e/\alpha))k/2} \leq 2^{-k/2}.\end{aligned} \tag{13.11}$$

So,

$$\Pr(\exists W : |W| = k, b(W) = \beta k,\, \beta \geq \beta_0) \leq \sum_{k\geq 2^n/n^9} 2^{-k/2} = o(1).$$

So, it remains to find bounds on

$$S_2 = \sum_{\substack{|W|=k\\ b(W)=\beta k, \beta\leq\beta_0}} \mathbb{P}(N(W_i) \subset W \setminus W_i), \tag{13.12}$$

To estimate S_2 we will again use the bound (13.7), which, for $k = \alpha 2^n$ takes the following form

$$\mathbb{P}(N(W_i) \subseteq W \setminus W_i) \leq \alpha^{k/2} (\log n)^{k(\log_2(1/\alpha))/n}. \tag{13.13}$$

However, when $b(W) \leq \beta_0 k$, we need a much better bound on the number of k-obstructions satisfying this condition.

Denote the set of such obstructions by O_{k,β_0}. Given a k-set $W \subset V(\mathbb{Q}^n)$, denote by b_j the number of vertices $v = (v_1, v_2, \dots, v_n)$ for which the vertex $v' = (v_1, \dots, v_{j-1}, v'_j, v_{j+1}, \dots, v_n)$ obtained from v by switching the jth digit, does not belong to W, so

$$b(W) = \sum_{j=1}^{n} b_j \leq \beta_0 k. \tag{13.14}$$

Denote by F a 3-element subset of the set $[n] = \{1, 2, \dots, n\}$ and call $W \subset V(\mathbb{Q}^n)$ an *F-cuboid* if

$$\sum_{j \in F} b_j \leq 3\beta_0 k/n,$$

and let $\mathbb{Q}_F$ be the set of F-cuboids. Clearly, if $F, F' \in \binom{[n]}{3}$ then $|\mathbb{Q}_F| = |\mathbb{Q}_{F'}|$, and if $(W, i) \in O_{k,\beta_0}$ then $W \in \mathbb{Q}_F$ for some $F \in \binom{[n]}{3}$. Therefore

$$|O_{k,\beta_0}| \leq 2\binom{n}{3}|\mathbb{Q}_F| \leq n^3 |\mathbb{Q}_F|. \tag{13.15}$$

So, to finish the computation of the bound on S_2 we just need to estimate $|\mathbb{Q}_F|$. For $F \in \binom{[n]}{3}$ an *F-cube* is a set of the form

$$F_b = \{a \in \{0,1\}^n : a_j = b_j \text{ for } j \in [n] \setminus F\},$$

where

$$b = (b_j) \in \{0,1\}^{[n] \setminus F}.$$

Note that an F-cube contains 8 vertices of the cube $\mathbb{Q}^n$, any two F-cubes are disjoint and and there are $2^{n-3} = 2^n/8$ F-cubes and $V(\mathbb{Q}^n)$ is the union of the F-cubes. We say that W *cuts* an F-cube F_b if $F_b \cap W \neq \emptyset$ and $F_b \setminus W \neq \emptyset$. If W cuts an F-cube F_b then F_b spans at least 3 boundary edges of W. So if $W \in \mathbb{Q}_F$ then, by the definition of F-cuboid, W cuts at most $\beta_0 k/n\ F - cubes$. Therefore, a set $W \in \mathbb{Q}_F$ contains at least $(k - 8\beta_0 k/n)/8$ F-cubes and at most $8\beta_0 k/n$ other vertices. So,

$$|\mathbb{Q}_F| \leq \binom{2^n/8}{k/8}\binom{2^n}{8\beta_0 k/n} \leq \left(\frac{e}{\alpha}\right)^{k/8} \left(\frac{en}{8\beta_0 \alpha}\right)^{8\beta_0 k/n}. \tag{13.16}$$

Combining (13.13), (13.15) and (13.16) we get

$$S_2 \le n^3 \left(\frac{e}{\alpha}\right)^{k/8} \left(\frac{en}{8\beta_0\alpha}\right)^{8\beta_0 k/n} \alpha^{k/2} (\log n)^{k(\log_2(1/\alpha))/n},$$

and so

$$\log_2 S_2 \le (k/8)[24(\log_2 n)/k + \log_2(e/\alpha) + \log_2(en/\alpha)64\beta_0/n \\ - 4\log_2(1/\alpha) + (\log\log n)\log_2(1/\alpha)/n],$$

which implies that

$$\log_2 S_2 \le -k/8 - 1, \tag{13.17}$$

since $1 < \log_2(1/\alpha) \le 9\log_2 n$.
Putting together inequalities (13.11) and (13.17) we have that for $\lfloor 2^n/n^9 \rfloor + 1 \le k \le 2^{n-1}$ we have

$$\mathbb{E} X_k < S_1 + S_2 \le 2^{-k/8}$$

and so

$$\sum_{\lfloor 2^n/n^9 \rfloor + 1}^{2^{n-1}} \mathbb{E} X_k \le 2^{-(\lfloor 2^n/n^9 \rfloor + 1)/8 + 4} \le 2^{-2^n/n^{10}}. \tag{13.18}$$

Combining (13.5), (13.8) and (13.18) we finally get that

$$\sum_{k=3}^{2^{n-1}} \mathbb{E} X_k = o(1),$$

so $\mathbb{Q}_p^n$ w.h.p. does not contain an k-obstruction for $3 \le k \le 2^{n-1}$.

To complete the picture we have to consider the remaining obstructions when $k = 1$, i.e., isolated vertices. Denote by $Y_0 = Y_0(\mathbb{Q}_p^n)$ the number of isolated vertices in $\mathbb{Q}_p^n$ and let

$$p = \frac{1}{2}\left(1 - \frac{c(n)}{n}\right), \text{ where } |c(n)| \le \log\log n.$$

Firstly, one can check, applying the Method of Moments (see Chapter 4 for an analogous result in $\mathbb{G}_{n,p}$), that Y_0 has an asymptotically Poisson distribution with expected value $\mu \sim e^{c(n)}$.

Secondly, the following observation allows us to transfer results on k-obstructions for a perfect matching in $\mathbb{Q}_p^n$ to analogous results in the random hypergraph process $\widetilde{\mathbb{Q}_t^n}$. Note that the number of edges in $\mathbb{Q}_p^n$ has the binomial distribution with mean $n2^{n-1}p$ and, by the Central Limit Theorem, it is concentrated around the mean as $n \to \infty$. Thus, in our case, we can assume that $\mathbb{Q}^n_t$ behaves in a similar fashion to $\mathbb{Q}^n_p$ when

$$t = \frac{1}{2}\left(1 - \frac{c(n)}{n}\right) n2^{n-1}.$$

Now we are ready to show that in a random hypercube process $\widetilde{\mathbb{Q}^n_t}$ the hitting time τ_M for a perfect matching does not exceed the hitting time τ_D of minimum degree at least one.

Let

$$c_0(n) = (\log\log n)/2 \text{ and } t_0 = \lfloor (1 - c_0(n)/n) n2^{n-1} \rfloor.$$

Then, since the number of isolated vertices in $\mathbb{Q}^n_{t_0}$ has an asymptotically Poisson distribution with expectation $e^{c_0(n)}$ (Exercise 12.4.6), w.h.p. $\mathbb{Q}^n_{t_0}$ contains an isolated vertex, and thus w.h.p.

$$\tau_D(\mathbb{Q}^n_t) \geq t_0.$$

Moreover, as we have proved above, $\mathbb{Q}^n_{t_0}$ w.h.p. has no k-obstructions for $3 \leq k \leq n2^{n-1}$, which implies that for process $\widetilde{\mathbb{Q}^n_t}$ the hitting time $\tau_D \geq \tau_M$. Indeed, let $\mathbb{Q}^n_t \supset \mathbb{Q}^n_{t_0}$ be such that the minimum degree $\delta(\mathbb{Q}^n_t) \geq 1$. Then $\mathbb{Q}^n_t$ has no 1-obstruction, and if $3 \leq k \leq 2^{n-1}$ then it also has no k-obstruction since $\mathbb{Q}^n_{t_0}$ has no k-obstruction for $3 \leq k \leq 2^{n-1}$. Thus there is no odd integer $k \leq 2^{n-1}$ for which $\mathbb{Q}^n_t$ contains a k obstruction, Therefore, by Hall's Theorem, $\mathbb{Q}^n_t$ has a perfect matching.

Hence, recalling that $\tau_D \leq \tau_M$ deterministically, we have that w.h.p.

$$\tau_D = \tau_M,$$

which completes the proof. □

Note that from Theorems 13.4 and 13.3we have that w.h.p.

$$\tau_C = \tau_M = \tau_D,$$

where τ_c is the hitting time for connectedness of $\mathbb{Q}^n_t$, i.e, it means that a random subgraph is w.h.p. connected and has a perfect matching as soon as the last isolated vertex disappears. Thus a simple necessary condition is once again sufficient w.h.p.

13.4 Exercises

13.4.1 (see [107] or Lemma 3 of [300]) Let G be a graph of maximum degree Δ, and let $k > 0$ be an integer. For every vertex $v \in V(G)$, let $t(v,k)$ be the number of trees in G on k vertices, rooted at v. Prove that $t(v,k) \leq (e\Delta)^{k-1}$.

13.4.2 (see [468] or Lemma 4 of [300]) Show that for every $S \subseteq V(\mathbb{Q}^n)$ the number $e(S,S^C)$ of edges between set S and its complement S^C, we have that

$$e(S,S^C) \geq |S|(n - 2\log_2 |S|).$$

13.4.3 (see [468] or Lemma 5 of [300]) Show that for every $S \subseteq V(\mathbb{Q}^n)$ with $|S| \leq 2^{n-1}$ we have that

$$e(S,S^C) \geq |S|.$$

13.4.4 (see [174]) Show that for $1 \leq k \leq 2^n$, a subgraph of $\mathbb{Q}^n$ spanned by k vertices has at most $\frac{k}{2}\log_2 k$ edges.

13.4.5 (see [158], Lemma 3) Denote the distance between two vertices u and v of a graph G by $d_G(u,v)$, and for a vertex $u \in V(G)$ and a set $S \subset V(G)$, write $d_G(u,S) = \min_{v \in S} d_G(u,v)$. Show that if G is a connected graph on n vertices and $1 \leq r < n$, then there is a set $A \subset V(G)$ such that $|A| \leq |V|/(r+1)$ and $d_G(u,A) \leq r$ for every vertex u of G.

13.4.6 $Y_0 = Y_0(\mathbb{Q}_t^n)$ the number of isolated vertices in $\mathbb{Q}_t^n$ and let

$$t = \frac{1}{2}\left(1 - \frac{c(n)}{n}\right) n2^{n-1}, \text{ where } |c(n)| \leq \log\log n.$$

Prove that Y_0 has asymptotically (as $n \to \infty$) Poisson distribution with mean

$$\mu = \left(1 + \frac{c(n)}{n}\right)^n \sim e^{c(n)}.$$

13.5 Notes

Recall, that in their breakthrough paper Ajtai, Komlós and Szemerédi [10] showed that if $p_e = (1+\varepsilon)/n$ then w.h.p. there will be a unique giant component of order 2^n. Next, their results were tightened in Bollobás, Kohayakawa and Łuczak [170] where the case $\varepsilon = o(1)$ was considered. A self-contained proof of both results, due to Krivelevich [590], is given in Section 13.1.
Further detailed study on the phase transition in $\mathbb{Q}_p^n$ extending results from [170] one can find in papers by van der Hofstad and Slade (see [485] and [486]), which in turn build on a sequence of papers by Borgs, Chayes, van der Hofstad, Slade and Spencer (see [181], [182] and [183]). A short overview of these results can be found in [808]. A final touch on studies of the asymptotic behavior of $\mathbb{Q}_p^n$ around the critical edge probability p_c was provided by van der Hofstad and Nachmias in [487]. From previous papers, quoted above, it was known that if

$p = p_c(1+O(2^{-n/3}))$ then w.h.p. the largest connected component L_1 of $\mathbb{Q}_p^n$ is of order $\Theta(2^{2n/3})$ and that this quantity is not concentrated. They show that for any sequence ε_n such that $\varepsilon_n = o(1)$, but $\varepsilon_n \gg 2^{-2n/3}$ and $p = p_c(1+\varepsilon_n)$, then w.h.p. $|L_1| = (2+o(1))\varepsilon_n 2^n$ and $|L_2| = o(\varepsilon_n 2^n)$, where L_2 is the second largest component. This resolves a conjecture from Borgs et al. paper [183].
Recently, Erde, Kang and Krivelevich in [329] show that w.h.p. the vertex-expansion of the giant component of $\mathbb{Q}_p^n$ is inverse polynomial in n. As a consequence they obtain a polynomial in n bound on the diameter of the giant component and the mixing time of the lazy random walk on the giant component, answering questions stated by Bollobás, Kohayakawa and Łuczak in [171].
McDiarmid, Scott and Whiters in [669] consider the supercrital period of the evolution of a random subgraph $\mathbb{Q}_p^n$ when the edge probability p is fixed and $p \in (0,1/2)$. Then, w.h.p. $\mathbb{Q}_p^n$ consists of one giant component together with many smaller components which form the "fragment". They investigate the fragment and, for example, they give asymptotic estimates for the mean numbers of components in the fragment of each size, and describe their asymptotic distributions, extending earlier work of Weber on components of $\mathbb{Q}_p^n$ (see [849],[850] and [851]).
Anastos, Diskin, Elboim and Krivelevich in [42] show that in $\mathbb{Q}_p^n$ there is a phase transition with respect to the length of a longest increasing path around $p = e/n$, where $P = \{v_1,\dots,v_k\}$ is an increasing path of length $k-1$ in $\mathbb{Q}^n$, if for every $i \in [k-1]$ the edge $v_i v_{i+1}$ is obtained by switching some zero coordinate in v_i to a one coordinate in v_{i+1}. In particular they proved that if α is a constant and $p = \alpha/n$ then, when $\alpha < e$, there exists $\delta \in [0,1)$ such that w.h.p. a longest increasing path in $\mathbb{Q}_p^n$ is of length at most δn. On the other hand, when $\alpha > e$, w.h.p. there is a path of length $n-2$.
Kronenberg and Spinka in [609] study the asymptotic number of independent sets in $\mathbb{Q}_p^n$ as $n \to \infty$ for a wide range of parameters p, including values of p tending to zero as fast as $C \log n/n^{1/3}$, constant values of p, and values of p tending to one.

In Section 13.2 we presented a new proof, due to Diskin and Krivelevich [300], of Burtin's result [205], referring also to contributions of Saposhenko [789] and Erdős and Spencer [338]. In further analysis, Bollobás, Kohayakawa and Łuczak [172] established that w.h.p. the hitting time for k-connectivity coincides with that of minimum degree at least k. If only vertices are deleted from $\mathbb{Q}_n$ with probability $1-p_v$, then the connectivity threshold is around $p_v = 1/2$, see Saposhenko [790] or Weber [848]. If both edges and vertices and vertices are deleted then the connectivity threshold is around $p_e p_v = 1/2$, see Dyer, Frieze and Foulds [320].

As we demonstrated in Section 13.3 the threshold for the existence of a prefect matching at around $p_e = 1/2$ was established by Bollobás [158], while Condon, Espuny Díaz, Girão, Kühn and Osthus in their breakthrough paper [236] solved

a long standing open conjecture of Bollobás, proving that the threshold probability for Hamiltonicity in $\mathbb{Q}_p^n$ equals 1/2. They also show that, w.h.p. for all fixed $p \in (0,1]$ $\mathbb{Q}_p^n$ contains an almost spanning cycle.
The G-random process on Cartesian products of graphs, Diskin and Geisler in [299] extends the result of Bollobás from [158] to the G-random process on Cartesian products of graphs, showing that w.h.p. the hitting times for minimum degree (at least) one, connectivity, and the existence of a (nearly-)perfect matching are the same.

Chapter 14

Randomly Perturbed Dense Graphs

In this chapter we are concerned with the following natural extension of the classical Erdős-Rényi-Gilbert models $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$, introduced by Bohman, Frieze and Martin in [134]. They start with a fixed graph $H = (V, E)$ on n vertices and a set R of m edges, chosen uniformly at random from the set $\overline{E}$ of all $\binom{n}{2} - |E|$ non-edges of H to form a random graph

$$G_{H,m} = (V, E \cup R).$$

Alternatively, one can take each non-edge of H and add it independently with probability p, creating the random graph $G_{H,p}$. This new class of random graphs would, for example, model graphs or networks which were basically deterministically produced, but for which there is some uncertainty about the complete structure. Finally, notice that if H is an empty graph then the model reduces to $\mathbb{G}_{n,m}$ or $\mathbb{G}_{n,p}$, respectively.

Since this new model starts with *any* graph H, we may not be able to get general results for a class produced by this random mechanism. Therefore it is quite natural to restrict the attention to a specific narrower and meaningful class of graphs. This is exactly what Bohman, Frieze and Martin did in their paper [134], which started this line of research. They consider the following scenario: Let $0 < d < 1$ be a fixed positive constant and let $\mathscr{G}(n,d)$ denote the set of graphs with vertex set $[n]$ which have minimum degree $\delta \geq dn$. Next choose H arbitrarily from $\mathscr{G}(n,d)$ and add a random set R of m edges to create the random graph $G_{H,m}$.

Obviously, for this new model we can ask similar questions as we do for classical random graphs. In particular, we may consider a general class $\mathscr{P}$ of *monotone* graph properties and ask how many random edges m one has to add to H so that the resulting graph $G_{H,m}$ will have property $\mathscr{P}$ w.h.p. We will address such questions in the next sections dealing with subgraphs, Hamiltonian cycles and vertex connectivity. We will also point out to another important line of research dealing with the Ramsey properties of such randomly perturbed dense graphs.

14.1 Subgraphs

Consider a monotone property $\mathscr{P}$ that a random graph $G_{H,m}$ (or $G_{H,p}$) contains an isomorphic copy of a fixed graph G. Finding a threshold for such a property in classic random graph was one of the fundamental questions raised by Erdős and Rényi in their seminal paper [332] and answered completely by Bollobás in [147] (see Chapter 5).

Recall first the notion of *maximum subgraph density* $m(G)$ of a fixed graph G defined as

$$m(G) = \max\{E(F)/V(F) : F \subseteq G\}.$$

For every positive integer r define

$$m_r(G) = \min_{V(G)=\bigcup_i V_i} \max_i \; m(G[V_i]),$$

where the minimum is over all partitions of $V(G)$ into at most r parts and $G_i = G[V_i]$ is a subgraph of G induced by V_i. Note also, that $m_r(G) < 1/2$ if and only if G is r-colorable.

Denote by $\overline{\mathscr{G}}(n,d)$ the set of all graphs on vertex set $[n]$ and *average* vertex degree at least dn, i.e., with at least $dn^2/2$ edges. Obviously, since $\mathscr{G}(n,d) \subset \overline{\mathscr{G}}(n,d)$, all results for a general class $\overline{\mathscr{G}}(n,d)$ are also valid for $\mathscr{G}(n,d)$.

The following result was proved by Krivelevich, Sudakov and Tetali in [605].

Theorem 14.1. *Let $0 < d < 1$ be a fixed constant and let $r \geq 2$ be the unique integer satisfying $d \in \left(\frac{r-2}{r-1}, \frac{r-1}{r}\right]$. Let G be a fixed graph.*

(i) *If $H \in \overline{\mathscr{G}}(n,d)$ and $m = \omega\left(n^{2-1/m_r(G)}\right)$ then $G_{H,m}$ contains a a copy of G w.h.p.*

(ii) *There exists a graph $H_0 \in \overline{\mathscr{G}}(n,d)$ with $d \sim \frac{r-1}{r}$ such that $m = o\left(n^{2-1/m_r(G)}\right)$, and w.h.p. $G_{H_0,m}$ fails to contain a copy of G.*

Before we reproduce the proof from [605] we need to discuss its two important elements. Firstly, the asymptotic equivalence of random graphs $G_{H,m}$ and $G_{H,p}$ stemming from their construction as a union of a dense graph with random edges. It is well known that for monotone graph properties, random graph $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$ are asymptotically equivalent when the expected number of edges of $\mathbb{G}_{n,p}$ is close to the number of edges of $\mathbb{G}_{n,m}$ (see Theorem 1.4). Similarly, it is also the case with $G_{H,m}$ and $G_{H,p}$ when $m = (\binom{n}{2} - |E(H)|)p$ (see Remark 1.1 of [605] for details). Since the binomial model is computationally much more tractable, we switch here and in other parts of this chapter from $G_{H,m}$ to $G_{H,p}$.

Another characteristic feature of the proofs of many results for randomly perturbed dense graphs is an application of the celebrated Szemerédi Regularity Lemma (see [823]). The Lemma was originally applied in the context of randomly perturbed random graphs by Bohman, Frieze, Krivelevich and Martin [135].

For two disjoint vertex sets A and B in graph G, let $e(A,B)$ denote the number of edges with one vertex in A and the other in B, while

$$d(A,B) = \frac{e(A,B)}{|A||B|}$$

is the *density* of the pair (A,B)

Definition 14.2. Let $\varepsilon > 0$. Given a graph G and two disjoint vertex sets $A \subset V(G)$ and $B \subset V(G)$, we say that the pair is ε-regular if for every $X \subset A$ and $Y \subset B$ satisfying $|X| > \varepsilon|A|$ and $|Y| > \varepsilon|B|$ we have $|d(X,Y) - d(A,B)| < \varepsilon$.

We state next two Lemmas without proofs (see [605] for details).

Lemma 14.3. *Let $r \geq 2$ be an integer and let $d > \frac{r-2}{r-1}$ be a fixed constant. Then there exist real constants ε, γ and an integer constant K, n_0 such that for all $n \geq n_0$, every graph G on n vertices with average degree at least dn contains r disjoint vertex sets $A_1, \ldots, A_r$ of cardinality*

$$|A_1| = \ldots = |A_r| \geq n/K$$

such that for each $1 \leq i \neq j \leq r$, the pair (A_i, A_j) is ε-regular of density at least γ.

Lemma 14.4. *For all real $\varepsilon, \gamma > 0$ and integers $r, T > 0$ there exists an integer n_0 and a real δ such that the following holds for all $n \geq n_0$. Let G be an r-partite graph with parts $A_1, \ldots, A_r$ of cardinality n. Assume that for each $1 \leq i \neq j \leq r$, the pair (A_i, A_j) is ε-regular of density at least γ. For each $1 \leq i \leq r$ choose a random subset U_i of cardinality T in A_i. Then with probability at least δ the r-partite subgraph of G with parts $U_1, \ldots, U_r$ is complete*

Proof. of Theorem 14.1:

In the light of our previous discussion about equivalence of models $G_{H,m}$ and $G_{H,p}$ it is enough to prove both statements of Theorem 14.1 for $G_{H,p}$ with

$$p \sim \frac{m}{\binom{n}{2} - \frac{dn}{2}} = o\left(n^{-1/m_r(G)}\right), \qquad \text{by assumption on } m.$$

We prove statement (ii) first. Let us randomly perturb a complete r-partite graph H_0 on n vertices with nearly equal parts $V_1, \ldots, V_r$, with probability p. Obviously, $H_0 \in \overline{\mathscr{G}}(n,d)$ with $d \sim \frac{r}{r-1}$ since,

$$|E(H_0)| \sim \frac{r}{r-1}\frac{n^2}{2}.$$

Notice that the random graph $G_{H_0,p}$, due to the choice of H_0, contains r independent copies of the random graph $\mathbb{G}_{n/r,p}$ on sets of vertices $V_1,\dots,V_r$.

Now, let $G_1,G_2,\dots$ be a family of subgraphs of G such that $m(G_i)\geq m_r(G)$ for all i. By Theorem 5.3 w.h.p. none of the parts $V_j,\ j=1,\dots,r$ contains any of the finitely many subgraphs G_i of G with maximum subgraph density $m(G_i)\geq m_r(G)$. The definition of $m_r(G)$, implies that for any partition of the vertex set of G into r parts one part should induce such a graph G_i, so we conclude that w.h.p. there is no copy G in $G_{H_0,p}$, which proves statement (ii).

To prove statement (i), fix a partition of the vertex set of graph G into r parts $W_1,\dots,W_r$ so that $m(G_j)\leq m_r(G)$, where G_j denotes the subgraph of G induced by a vertex set W_j, for $j=1,\dots,r$, and let $T=\max_{1\leq j\leq r}|W_j|$. Now, let $H\in\overline{\mathscr{G}}(n,d)$. Applying Lemma 14.3 to H, we get r disjoint sets $A_1,\dots,A_r\subset V(H)$ of linear size such that all pairs $(A_i,A_j),\ 1\leq i\neq j\leq r$, are ε-regular of positive density $\gamma>0$, for some ε,γ depending only on d.

Recall that the set R of random edges added to H to get $G_{H,p}$ contains every non-edge of H independently with probability $\omega\left(n^{-\frac{1}{m_r(G)}}\right)$. Thus, for an appropriately chosen function $\varphi=\varphi(n)\to\infty$, we can represent the set R as a union of φ independent sets of random edges R_i, where every non-edge of H is in R_i randomly and independently with probability which is still $\omega\left(n^{-\frac{1}{m_r(G)}}\right)$

Consider now what happens if we add the set R_i to H. By definition, for all i and j, the edges R_i inside A_j form a copy of the random graph $\mathbb{G}_{|A_j|,p}$ with $p=\omega\left(n^{-\frac{1}{m_r(G)}}\right)$ and all these copies are independent. Thus by Theorem 5.3, w.h.p. the random set R_i puts a copy of G_j inside A_j for each $j=1,\dots,r$. Assume that this is indeed the case, and let $U_j,\ j=1.\dots,r$ be the vertex set of such copy. The sets U_j are mutually independent and thus can be considered as random sets of size $|U_j|$ inside A_j. Recall that $|U_j|\leq T$ for all $j=1.\dots,r$. Therefore, By Lemma 14.4 , with probability at least δ, the r-partite graph $H[U_1\cup\dots\cup U_r]$ is complete, in which case the graph $H\cup R_i$ spans a copy of G on $U_1\cup\dots\cup U_r$. Thus with probability at least $(1-o(1))\delta>\delta/2$ there is a copy of G in $H\cup R_i$. Observe that $\delta>0$ is a constant depending only on d and G. As the random sets R_i are independent, it follows that the probability that $G_{H,p}=H\cup R$ does not have a copy of G is at most $(1-\delta/2)^{\varphi}=o(1)$ and statement (ii) holds. $\square$

14.2 Hamiltonicity

We will reproduce here proofs of two theorems, due to Bohman, Frieze and Martin [134] about the number of edges needed to have $G_{H,m}$ to be Hamiltonian w.h.p. Since $d\geq 1/2$ implies that H itself is Hamiltonian (Dirac's theorem), this could be considered to be a probabilistic generalization of Dirac's theorem to the case where $d < 1/2$. We henceforth assume that $d < 1/2$. Also, let

$$\theta = \ln d^{-1} \geq .69.$$

Theorem 14.5. *Suppose* $0 < d \leq 1/2$ *is constant,* $H \in \mathscr{G}(n,d)$*.Then*

(i) *If* $m \geq (30\theta + 13)n$ *then* $G_{H,m}$ *is Hamiltonian w.h.p..*

(ii) *For* $d \leq 1/10$ *there exist graphs* $H \in \mathscr{G}(n,d)$ *such that if* $m < \theta n/3$ *then w.h.p.* $G_{H,m}$ *is not Hamiltonian.*

Proof. We will assume from now on that m is exactly $\lceil 30\theta n \rceil + 13n$. We let the m random edges added to H be split into two sets R_1 and R_2, so that $R = R_1 \cup R_2$ where $|R_1| = m_1 = \lceil 30\theta n \rceil$. Then let $G_{H,m_1} = H \cup R_1$. We first show that

Lemma 14.6. G_{H,m_1} *is connected w.h.p..*

Proof. Let $N = \binom{n}{2}$. If $u, v \in [n]$ then either they are at distance one or two in H or

$$\mathbb{P}(dist_{G_{H,m_1}}(u,v) \geq 3) \leq \left(1 - \frac{|R_1|}{N}\right)^{d^2n^2} \leq e^{-60\theta d^2 n}.$$

Hence,

$$\mathbb{P}(diam(G_{H,m_1}) \geq 3) \leq n^2 e^{-60\theta d^2 n} = o(1).$$

□

Given a longest path P in a graph G with end-vertices x_0, y and an edge $\{y, v\}$ where v is an internal vertex of P, we obtain a new longest path $P' = x_0..vy..w$ where w is the neighbor of v on P between v and y. We say that P' is obtained from P by a *rotation* with x_0 *fixed*.

Let $END_G(x_0, P)$ be the set of end-vertices of longest paths of G which can be obtained from P by a sequence of rotations keeping x_0 as a fixed end-vertex. Let $END_G(P) = \{x_0\} \cup END_G(x_0, P)$. Note that if G is connected and non-Hamiltonian then there is no edge $\{x_0, y\}$ where $y \in END_G(x_0, P)$.

It follows from Pósa [746] that

$$|N_G(END_{\mathbb{G}}(P))| < 2|END_G(P)|, \tag{14.1}$$

where for a graph G and a set $S \subseteq V(\mathbb{G})$

$$N_G(S) = \{w \notin S : \exists v \in S \text{ such that } vw \in E(G)\}.$$

Lemma 14.7. Whp

$$|N_{G_{H,m_1}}(S)| \geq 3|S| \tag{14.2}$$

for all $S \subseteq [n]$, $|S| \leq n/5$.

Proof. Now $|N_H(S)| \geq 3|S|$ for all $S \subseteq [n]$, $|S| \leq dn/3$. So,

$$\begin{aligned}
\mathbb{P}(\exists |S| \leq n/5 : \ |N_{G_{H,m_1}}(S)| < 3|S|) &\leq \sum_{k=dn/3}^{n/5} \binom{n}{k}\binom{n}{3k}\left(1-\frac{m}{N}\right)^{k(n-4k)} \\
&\leq \sum_{k=dn/3}^{n/5} \left(\frac{n^4 e^4}{27k^4} e^{-12\theta}\right)^k \\
&= o(1).
\end{aligned}$$

□

It follows from Lemma 14.7 that for any longest path P in a graph G that contains G_{H,m_1} as a subgraph we have $n/5 \leq |END_G(P)| \leq |P|$.

Let P_0 be a longest path in G_{H,m_1} of length $l_0 \geq dn$. Proceeding as in Section 6.2 , we see that

$$\mathbb{P}(G_{H,m} \text{ is not Hamiltonian}) \leq o(1) + \Pr(Bin(13n, 2/25) \leq n) = o(1). \quad (14.3)$$

Proof of (ii) . Let $m = cn$ for some constant c and let H be the complete bipartite graph $K_{A,B}$ where $|A| = dn$ and $|B| = (1-d)n$. Let I be the set of vertices of B which are not incident with an edge in R. If $|I| > |A|$ then $G_{H,m}$ is not Hamiltonian. Instead of choosing m random edges for R, we choose each possible edge independently with probability $p = \frac{2m}{(d^2+(1-d)^2)n^2}$.
Then

$$\mathbb{E}(|I|) = (1-d)n(1-p)^{(1-d)n-1} \sim (1-d)\exp\left\{-\frac{2(1-d)m}{(d^2+(1-d)^2)n}\right\} n.$$

It follows from the Chebychev inequality that $|I|$ is concentrated around its mean and so $G_{H,m}$ will be non-Hamiltonian w.h.p. if c satisfies

$$c < \frac{1}{2(1-d)}(d^2+(1-d)^2)\ln(d^{-1}-1).$$

This verifies (ii).

□

So it seems that we have to add $\Theta(n)$ random edges in order to make $G_{H,m}$ Hamiltonian w.h.p.. Since a random member of $\mathscr{G}(n,d)$ is already likely to be Hamiltonian, this is a little disappointing. Why should we need so many edges in the worst-case? It turns out that this is due to the existence of a large independent set. Let $\alpha = \alpha(H)$ be the independence number of H.

Theorem 14.8. *Suppose $H \in \mathscr{G}(n,d)$ and $1 \leq \alpha(H) < d^2n/2$ and so $d > n^{-1/2}$ (d need not be constant in this theorem). If*

$$\frac{md^3}{\ln d^{-1}} \to \infty \tag{14.4}$$

then $G_{H,m}$ is Hamiltonian w.h.p..

Note that if d is constant then Theorem 14.8 implies that $m \to \infty$ is sufficient.

Proof. (of Theorem 12.12)
We will first show that we can decompose H into a few large cycles.

Lemma 14.9. *Suppose that $G_{H,m}$ has minimum degree dn where $d \leq 1/2$ and that $\alpha(G_{H,m}) < d^2n/2$. Let $k_0 = \left\lfloor \frac{2}{d} \right\rfloor$. Then the vertices of $G_{H,m}$ can be partitioned into at most k_0 vertex disjoint cycles.*

Proof. Let C_1 be the largest cycle in H. $|C_1| \geq dn+1$ and we now show that the graph $H \setminus C_1$ has minimum degree $\geq dn - \alpha$.

To see this, let $C_1 = v_1, \ldots, v_c, v_{c+1} = v_1$. Let $w \in V(H \setminus C_1)$. Because C_1 is maximum sized, no such w is adjacent to both v_i and v_{i+1}. Also, if $w \sim v_i$ and $w \sim v_j$ with $i < j$ and $v_{i-1} \sim v_{j-1}$, then

$$w, v_j, \ldots, v_c, v_1, \ldots, v_{i-1}, v_{j-1}, \ldots, v_i, w$$

is a larger cycle. So the predecessors of $N(w)$ in C_1 must form an independent set and $|N(w) \cap C_1| \leq \alpha$. Similar arguments are to be found in [227].

We can repeat the above argument to create disjoint cycles $C_1, \ldots, C_r$ where $|C_1| \geq |C_2| \geq \cdots \geq |C_r|$ and C_j is a maximum sized cycle in the graph $H_{j-1} = H \setminus \left(C_1 \cup \cdots \cup C_{j-1}\right)$ for $j = 1,2,\ldots,r$. Now H_k has minimum degree at least $dn - k\alpha$ and at most $n - dn - 1 - (dn - \alpha + 1) - \cdots - (dn - (k-1)\alpha + 1) = n - k(dn + 1 - (k-1)\alpha/2)$ vertices. Since $d^2n > 2\alpha$, H_{k_0}, if it existed, would have minimum degree at least 2 and a negative number of vertices. □

Let $C_1, \ldots, C_r$ be the cycles given by Lemma 14.9. In order to simplify the analysis, we assume the edges of R are chosen from $\overline{E}$ by including each $e \in \overline{E}$ independently with probability $p = \frac{m}{|\overline{E}|}$. Because Hamiltonicity is a monotone property, showing that $G_{H,p}$ is Hamiltonian w.h.p. implies the theorem. We get a further simplification in the analysis if we choose these random edges in rounds: set $R = R_1 \cup R_2 \cup \cdots \cup R_r$ where each edge set R_i is independently chosen by including $e \in \overline{E}$ with probability p, where $1 - (1-p)^r = \frac{m}{|\overline{E}|}$. Each R_i will be used to either extend a path or close a cycle and will only be used for one such attempt. In this way each such attempt is independent of the previous. To this end let $G_{H,m_t} = H \cup \bigcup_{s=1}^{t} R_t$ for $t = 0,1,\ldots,r$. Thus $G_{H,m_0} = H$ and $G_{H,m_r} = G_{H,m}$.

Let $e = \{x,y\}$ be an edge of C_r and let Q be the path $C_r - e$. In each phase of our procedure, we have a current path P with endpoints x,y together with a collection of vertex disjoint cycles $A_1, A_2, \ldots, A_s$ which cover V. Initially $P = Q$, $s = r-1$ and $A_i = C_i$, $i = 1,2,\ldots,r-1$.

Suppose a path P and collection of edge disjoint cycles have been constructed in $G_{H,m_{t-1}}$ (initially $t = 1$). Consider the set $Z = END_{G_{H,m_{t-1}}}(x,P)$ created from rotations with x as a fixed endpoint, as in the proof of Theorem 14.5. We identify the following possibilities:

Case 1: There exists $z_1 \in Z$, $z_2 \notin P$ such that $f = (z_1, z_2)$ is an edge of H. Let Q be the corresponding path with endpoints x, z_1 which goes through $V(P)$. Now suppose that $z_2 \in C_i$ and let $f' = (z_2, z_3)$ be an edge of C_i incident with z_2. Now replace P by the path Q, f, Q' where $Q' = C_i - f$. This construction reduces the number of cycles by one.

Case 2: $|V(P)| \leq n/2$ and $z \in Z$ implies that $N_{G_{H,m_{t-1}}}(z) \subseteq V(P)$. It follows from (14.1) that $|Z| \geq dn/3$. Now add the next set R_t of random edges. Since $|V(P)| \leq n/2$, the probability that no edge in R_t joins $z_1 \in Z$ to $z_2 \in V \setminus V(P)$ is at most $(1-p)^{(dn/3)(n/2)}$. If there is no such edge, we fail, otherwise we can use (z_1, z_2) to proceed as in Case 1. We also replace t by $t+1$.

Case 3: $|V(P)| > n/2$ and $z \in Z$ implies that $N_{G_{H,m_{t-1}}}(z) \subseteq V(P)$. Now we close P to a cycle. For each $z \in Z$ let $A_z = END_{G_{H,m_{t-1}}}(z, Q_z)$ where Q_z is as defined in Case 1. Each A_z is of size at least $dn/3$. Add in the next set R_t of random edges. The probability that R_t contains no edge of the form (z,z') where $z \in Z$ and $z' \in A_z$ is at most $(1-p)^{d^2n^2/10}$. If there is no such edge, we fail. Otherwise, we have constructed a cycle C through the set $V(P)$ in the graph G_{H,m_t}. If C is Hamiltonian we stop. Otherwise, we choose a remaining cycle C', distinct from C and replace P by $C' - e$ where e is any edge of C'. Now $|V(P)| < n/2$ and we can proceed to Case 1 or Case 2.

After at most r executions of each of the above three cases, we either fail or produce a Hamilton cycle. The probability of failure is bounded by

$$
\begin{aligned}
&k_0((1-p)^{(dn/3)(n/2)} + (1-p)^{d^2n^2/10}) \\
&\leq 2d^{-1}\left(\left(1 - \frac{m}{|\overline{E}|}\right)^{\frac{dn^2}{6r}} + \left(1 - \frac{m}{|\overline{E}|}\right)^{\frac{d^2n^2}{10r}}\right) \leq 4d^{-1}e^{-md^3/10} = o(1)
\end{aligned}
$$

provided (14.4) holds. □

An observation: We do not actually need the condition that $\alpha(H) \leq d^2n/2$ to complete this proof. The weaker condition that $d^2n/2$ bounds the independence number of the neighborhood of each vertex is enough.

Note that Aigner-Horev, Hefetz and Krivelevich in [11] prove several improvements and extensions of the Theorem 12.12. In particular, keeping the bound on the independence number $\alpha(H) = O(d^2 n)$ as in Theorem 12.12 and allowing for $d = \Omega(n^{-1/3})$, they determine the correct order of magnitude of the number of random edges whose addition to H w.h.p. yields a pancyclic graph.

14.3 Vertex Connectivity

The next property we consider is vertex connectivity. First note that if $d \geq 1/2$ then $H \in \mathscr{G}(n,d)$ is at least $(2\lceil dn \rceil - n + 2)$-connected. This can be seen by the fact that, by removing any set of size $2\lceil dn \rceil - n$, the resulting graph has minimum degree at least half the number of vertices. This means that the resulting graph has a Hamilton cycle and is, thus, at least 2-connected. Since there are graphs H which are disconnected but have minimum degree dn ($d < 1/2$), we focus on the number of random edges required to make $G_{H,m}$ k-connected for $k \leq cn$, for some constant $c = c(d) > 0$. The following result was proved by Bohman, Frieze, Krivelevich and Martin in [135].

Theorem 14.10.

(i) *Let $H \in \mathscr{G}(n,d)$. If $k = O(1)$ and $m = \omega(1)$ then $G_{H,m}$ is k-connected, w.h.p. If $\omega(1) \leq k \leq d^2 n/32$ and $m = 640k/d^2$ then $G_{H,m}$ is k-connected, w.h.p.*

(ii) *If $d < 1/2$ then there exists an $H_0 \in \mathscr{G}(n,d)$ such that w.h.p. $G_{H_0,m}$ fails to be k-connected for all k such that $m \leq \frac{k}{2}\lfloor \frac{n}{dn+1} \rfloor$.*

Let $\kappa(G)$ denote the vertex connectivity of graph G. We first prove the following lemma that may be of independent interest.

Lemma 14.11. *Let $H = (V,E)$ be a graph on n vertices with minimum degree $k > 0$. Then there exists a partition $V = V_1 \cup \ldots \cup V_t$ such that for every $1 \leq i \leq t$ the set V_i has at least $k/8$ vertices and the induced subgraph $H[V_i]$ is $k^2/(16n)$-connected.*

Proof. of Lemma 14.11.

Recall the following classical result on vertex connectivity.

Theorem 14.12 (Mader Theorem, see [295]). *Every graph of average degree at least k has a $k/4$-connected subgraph.*

Let $(C_1, \ldots, C_t)$ be a family of disjoint subsets of V with the property that each induced subgraph $H[C_i]$ is $k/8$-connected and that, among all such families of subsets, the set of vertices

$$C \stackrel{\text{def}}{=} \bigcup_{i=1}^{t} C_i$$

is maximal. According to Theorem 14.12, $t > 0$. Also, $|C_i| \geq k/8$ for all i and thus $t \leq 8n/k$.

Let now $(V_1, \dots, V_t)$ be a family of disjoint subsets of V such that $C_i \subseteq V_i$, the induced subgraph $H[V_i]$ is $k^2/(16n)$-connected for all $1 \leq i \leq t$ and that among all such families the set of vertices

$$U \stackrel{\text{def}}{=} \bigcup_{i=1}^{t} V_i$$

is maximal. We claim that $U = V$. Assume to the contrary that there exists a vertex $v \in V \setminus U$. If $|N(v) \cap V_i| \geq k^2/(16n)$ for some i, then adding v to V_i can be easily seen to keep $H[V_i]$ $k^2/(16n)$-connected, contradicting the maximality of U. Thus v has less than $k^2/(16n)$ neighbors in each of the $t \leq 8n/k$ sets V_i, and therefore $deg_{V \setminus U}(v) > k - (8n/k)(k^2/(16n)) = k/2$. We conclude that the minimum degree of the induced subgraph $H[V \setminus U]$ is at least $k/2$. Applying Theorem 14.12, this time to $H[V \setminus U]$, unveils a $k/8$-connected subgraph disjoint from C – a contradiction to the choice of $(C_1, \dots, C_t)$. Hence the family $(V_1, \dots, V_t)$ covers indeed all the vertices of H and thus forms a required partition. □

We remark that the above result is optimal up to constant multiplicative factors. To see this take $\lceil (n - k^2/n)/(k+1) \rceil$ disjoint cliques C_i of size $k+1$ each, add an independent set I on the (at most k^2/n) remaining vertices, and connect each vertex of I with roughly k^2/n arbitrarily chosen vertices of C_i, $1 \leq i \leq \lceil (n - k^2/n)/(k+1) \rceil$. Denote the obtained graph by H. Let $K \subseteq H$ be a subgraph of H containing some vertices from I. If K intersects two distinct cliques C_i, C_j, then deleting $V(K) \cap I$ disconnects $V(K) \cap C_i$ from $V(K) \cap C_j$, and thus the connectivity of K does not exceed $|V(K) \cap I| \leq |I| \leq k^2/n$. If K intersects a unique clique C_i, then deleting all neighbors of $v \in V(K) \cap I$ from C_i disconnects v from the rest of K, implying $\kappa(H[K]) \leq deg_{C_i}(v) \leq k^2/n$.

Proof of Theorem 14.10. Let us begin with part (i). Let H be a graph with minimum degree at least dn. Let $(V_1, \dots, V_t)$ be a partition of $V(H)$ such that $|V_i| \geq dn/8$ and $H[V_i]$ is $(d^2n/16)$-connected, $1 \leq i \leq t$. The existence of such a partition is guaranteed by Lemma 14.11. It is enough to show that the graph $G_{H,m}$ w.h.p. contains a matching of size k in each bipartite graph induced by (V_i, V_j). Let F_{ij} be a maximum matching between V_i and V_j in H. If $|F_{ij}| \geq k$ we are done. Assume therefore that $|F_{ij}| < k$. Choose a subset $A \subset V_i \setminus \bigcup_{e \in F_{ij}} e$ and a subset $B \subset V_j \setminus \bigcup_{e \in F_{ij}} e$ of cardinalities $|A| = |B| = 3dn/32$. Obviously H has no edges connecting A and B due to the maximality of F_{ij}.

Consider first the case $k = O(1)$. Then the set R contains w.h.p. $\omega(1)$ random edges between A and B. If $\omega(1) = o(\sqrt{n})$, then those random edges form w.h.p. a matching as required.

Let now $k = \omega(1)$. Instead of $G_{H,m}$ we may consider $G_{H,p}$ with $p = \frac{1280k}{d^2n^2}$. Then the probability that the set R of random edges does not have a matching of size k between A and B can be estimated from above by:

$$\sum_{i=0}^{k-1} \binom{\frac{3dn}{32}}{i}^2 i! p^i (1-p)^{\left(\frac{3dn}{32}-i\right)^2}$$

(This expression arises from first choosing the size $i < k$ of a maximum matching M between A and B in R, then choosing the vertices of M in A and B, then forming a pairing between them, then requiring all matching edges to be present in R, and finally requiring all edges lying outside the vertices of the matching to be absent). We can estimate the above expression from above by:

$$\begin{aligned} e^{-p\left(\frac{3dn}{32}-k\right)^2}\left(1+\sum_{i=1}^{k}\left(\frac{3edn}{32i}\right)^{2i} i^i p^i\right) &< e^{-p\left(\frac{3dn}{32}-\frac{d^2n}{32}\right)^2}\left(1+\sum_{i=1}^{k}\left[\left(\frac{3edn}{32i}\right)^2 ip\right]^i\right) \\ &< e^{-\frac{d^2n^2p}{256}}\left(1+\sum_{i=1}^{k}\left(\frac{9e^2d^2n^2p}{1024i}\right)^i\right) \\ &= e^{-5k}\left(1+\sum_{i=1}^{k}\left(\frac{45e^2k}{4i}\right)^i\right) \end{aligned}$$

In the last sum each summand is easily seen to be at least twice as much as the previous summand, and hence the above estimate is at most

$$2e^{-5k}\left(\frac{45e^2k}{4k}\right)^k = o(1).$$

As to part (ii), let H_0 consist of $\lfloor \frac{n}{dn+1} \rfloor$ disjoint cliques $C_1, \ldots, C_t$ each of size at least $dn+1$. If $H_0 \cup R$ is k-connected, then each C_i is incident to at least k edges, implying $|R| \geq \frac{kt}{2} = \frac{k}{2}\lfloor \frac{n}{dn+1} \rfloor$. □

14.4 Ramsey Properties

Let G_1 and G_2 be two graphs. Following the Erdős and Rado arrow notation we write that a graph $F \to (G_1, G_2)$ if every Red-Blue coloring of the edges of F contains a Red copy of G_1 or a Blue copy of G_2.

In this section, following Krivelevich, Sudakov and Tetali [605], we determine the threshold function for the number of random edges m one needs to add to a dense graph $H \in \overline{\mathscr{G}}(n,d)$, $0 < d < 1$, i.e., a graph with average degree at least dn, to guarantee that w.h.p. $G_{H,m} \to (G_1, G_2)$. In particular, we present the result from [605] which solves the problem for $G_1 = K_3$ and $G_2 = K_t$, for $t \geq 3$.

Theorem 14.13. *Let $0 < d < 1$ and $t \geq 3$ be an integer.*

(i) *Let $H \in \overline{\mathscr{G}}(n,d)$ and $m = \omega\left(n^{2-2/(t-1)}\right)$. Then $G_{H,m} \to (K_3, K_t)$ w.h.p.*

(ii) *There exists a graph H_0 on n vertices and with $n^2/4$ edges such that if $m = o\left(n^{2-2/(t-1)}\right)$, then $G_{H_0,m} \not\to (K_3, K_t)$ w.h.p.*

We need to first introduce an important definition used throughout the proof of the above theorem.

Definition 14.14. Let $G = (V,E)$ be a graph and let U be a subset of V. Denote by

$$N^*(U) = \{v \in V | (v,u) \in E, \text{ for every } u \in U\},$$

the *common neighborhood* of U in G. For a constant $0 < c < 1$. a subset $V_0 \subset V$ of verices is called $c-typical$ if its common neighborhood has size $|N^*(V_0)| \geq cn$.

Although the main result of this section is for $H \in \overline{\mathscr{G}}(n,d)$, in what follows we assume, without loss of generality, that $H \in \mathscr{G}(n,d/2)$ for fixed $0 < d < 1$, i.e., is a graph with minimum degree at least $dn/2$. Indeed, there is no loss of generality since only estimates on the number of edges in linear-sized (in n) will be used, whereas every graph H on n vertices with average degree dn contains a subgraph H' on $n' \geq dn/2$ vertices with minimum degree at least $dn/2$.

Lemma 14.15. *Let $0 < d, \alpha, \beta < 1$ be fixed constants. Let $H \in \mathscr{G}(n,d/2)$, and let $m = \omega\left(n^{2-2/(t-1)}\right)$. Then there exist constants c_1 and c_2 depending on d, such that such $G_{H,m}$ w.h.p.*

(i) *contains a c_1-typical copy of K_t , for $t \geq 4$,*

(ii) *every set U of size $|U| \geq \alpha n$ contains a c_2-typical copy of K_{t-1}, for $t \geq 4$,*

(iii) *every set U of size $|U| \geq \beta n$ satisfies: $G_{H,m}[U] \to (K_3, K_{t-2})$, for $t \geq 5$.*

Proof. We start with proving statement (ii) first, indicating the changes necessary for the proof of simpler statement (i).
For $v \in V(H)$, let $deg_H(v)$ denotes the degree of v in H. For a subset $U \subseteq V(H)$ we denote by $deg_H(v,U)$ the number of neighbors of v in U. Observe that every collection of a linear number of vertices of H (and hence of $G_{H,m}$) contains many subsets of size $t-1$ (and also t), which are typical. Indeed, let $U \subset V(H)$ be of size at least αn. Since the minimum degree of H is at least $dn/2$ we have that

$$\sum_{v \in V(H)} deg_H(v,U) = \sum_{u \in U} deg_H(u) \geq \frac{dn}{2}|U| \geq \frac{\alpha d}{2} n^2.$$

Therefore

$$\sum_{\substack{V_0\subset U,\\ |V_0|=t-1}} |N^*(V_0)| = \sum_{v\in V(H)} \binom{\deg_H(v,U)}{t-1} \geq n\binom{\sum_{v\in V(H)}\frac{\deg_H(v,U)}{n}}{t-1} = n\binom{\alpha dn/2}{t-1}.$$

As every $|N^*(V_0)| \leq n$, we conclude that $|N^*(V_0)| \geq c_2 n$ for at least

$$\binom{\alpha dn/2}{t-1} - \binom{\alpha n}{t-1}c_2$$

subsets $V_0 \subset U$ of size $t-1$. Taking $c_2 = c_2(d)$ small enough we ensure that the later quantity is at least $c_2' n^{t-1}$ for some $c_2' > 0$.

In the final step of of the proof, we switch to the equivalent model $G_{H,p}$ with $p = \omega\left(n^{-2/(t-1)}\right)$ and apply Janson's inequality (see Section 27.6).

Due to monotonicity we may also assume that edge probability $p = p(n)$ does not to exceed $n^{-2/(t-1)}$ by much, say $p \leq n^{-2/(t-1)}\log n$.
Let μ_U denote the expected number of c_2-typical copies of K_{t-1} in U after adding R to H. Here we only include a set T of $t-1$ vertices if T is complete using the edges of R only. Let Δ_U denote the correlation term:

$$\Delta_U = \sum_{\substack{T,T'\subset U\\ |T|=|T'|=t-1\\ |T\cap T'|\geq 2}} \mathbb{P}(T \text{ and } T' \text{ each induce a } c_2\text{-typical } K_{t-1} \in G_{H,p}).$$

It is easy to check that

$$\mu_U = \Theta\left(n^{t-1}p^{\binom{t-1}{2}}\right) = \omega(n),$$

and that

$$\Delta_U = O(n^{2t-4}p^{(t-1)(t-2)-1}) = O(n^{2/(t-1)}(\log n)^{(t-1)(t-2)-1}) = o(\mu_U).$$

Using the Janson inequality it now follows that

$$\mathbb{P}(\text{there are no } c_2\text{-typical copies of } K_{t-1} \text{ in } U) \leq e^{-\mu_U+\frac{\Delta_U}{2}} = e^{-\omega(n)}.$$

Therefore with probability $1-2^n e^{-\omega(n)} = 1-o(1)$ every set U of size at least αn contains a c_2-typical copy of K_{t-1}, which completes the proof of (ii).

The proof of statement (i) is obtained along the same lines. Observe that the expected number μ of c_2-typical copies of K_t in $G_{H,p}$ is $\Theta(n^t p^{\binom{t}{2}}) = o(1)$, and the correlation term is

$$\Delta = \sum_{\substack{T,T'\subset U\\ |T|=|T'|=t-1\\ |T\cap T'|\geq 2}} \mathbb{P}(T \text{ and } T' \text{ each induce } K_t \in G_{H,p}) = O\left(n^{2t-2}p^{2\binom{t}{2}-1}\right) = o(1).$$

Once again the assertion (i) follows using Janson inequality.

To prove statement (iii) we need the following proposition, which is directly implied by the results for a random graph $\mathbb{G}_{n,p}$ due to Rödl and Ruciński [771] and Kohayakawa and Kreuter [578] (or see Theorems 3.5 and 3.6 and Corollary 3.7 of [605])

Proposition 14.16. *Let $t \geq 3$ be a fixed integer . For every $a > 0$, there exist C and N_0 such that for $n > N_0$ and $p > Cn^{-(2t-3)/(t(t-1))}$ we have*

$$\mathbb{P}\left(\mathbb{G}_{n,p} \to (K_3, K_t)\right) > 1 - e^{-an^{3/2}}.$$

□

Statement (iii) then follows by trivial combination of the union bound with Proposition 14.16, since the edge probability in our random graph $G_{H,p}$, $p = n^{-\frac{2}{t-1}} \geq n^{-\frac{2(t-2)-3}{(t-2)(t-3)}}$ for $t \geq 5$.

We are now ready to prove Theorem 14.13.

Proof. (of Theorem 14.13)

We start with a proof of statement (ii) of the theorem. Let $H_0 = K_{\frac{n}{2},\frac{n}{2}}$ and color all edges of H Red and the random edges of $G_{H_0,m}$ Blue. Notice that w.h.p. $G_{H_0,m}$ does not contain a Red copy of K_3 nor neither a Blue copy of K_t, $t \geq 3$. The later follows from Theorem 5.3 since w.h.p. the random graph $\mathbb{G}_{n,m}$ does not contain a complete subgraph K_t if $m = o(n^{2-2/(t-1)})$.

Notice that for $t = 3$ the statement (i) of the theorem follows directly from Theorem 5.3 (see Exercise 14.5.6), while the proof for $t \geq 4$ needs much more effort, and is solely based on Lemma 14.15 under the assumption that we can take the constants c_1, c_2, α, β, specified in its statements (i) and (ii), as small as will be required in the proof.

Consider an arbitrary Blue-Red coloring of the edges of $G_{H,p}$ and assume, by the way of contradiction, that this coloring contains neither a Red K_3 nor a Blue K_t for fixed $t \geq 4$. Observe first, that there is a vertex whose Blue degree in edge colored $G_{H,p}$ is $c_1 n/2$. Indeed, by (i) of Lemma14.15 $G_{H,p}$ contains a c_1-typical copy of K_t, and hence $|N^*(V(K_t))| \geq c_1 n$. Since there is no Blue K_t, at least one edge inside $V(K_t)$ must be colored Red, and let (u,v) be such an edge. For every $w \in N^*(u,v)$ at least one of (u,w) or (v,w) is Blue - otherwise we get a Red K_3. Therefore the Blue degree of u or v is at least $c_1 n/2$.

Let v be the vertex with Blue degree at least $c_1 n/2$ and let $N_B(v)$ denote v's set of Blue neighbors. Applying statement (ii) of Lemma 14.15 greedily, with $\alpha = c_1 c_2/48$ and $t = 4$, we see that $N_B(v)$ contains a family of $c_1 n/12$ pairwise disjoint c_2-typical copies of K_3. Each of these must have a Blue edge - otherwise we have a Red K_3 and we are done. Let these Blue edges be $(x_1, y_1), \ldots, (x_s, y_s)$, where $s = c_1n/12$. Denote by $N^*(x_i, y_i)$ the common neighborhood of (x_i, y_i) for $i = 1, \ldots, s$. Note that each $N^*(x_i, y_i)$ is linear in size, since (x_i, y_i) is a pair from from c_2-typical K_3. Clearly

$$\min_i |N^*(x_i, y_i)| \geq c_2n.$$

Suppose first that for some $i \in \{1, \ldots, s\}$,

$$|\{w \in N^*(x_i, y_i) : (x_i, w), (y_i, w) \text{ are both Blue }\}| \geq c_2n/2. \tag{14.5}$$

Then, for such choice of i, we obtain a Blue edge (x_i, y_i) with common Blue neighborhood $N_B^*(x_i, y_i) \subseteq N^*(x_i, y_i)$ of size at least $c_2n/2$. When $t = 4$, property (ii) of Lemma 14.15 guarantees that $N_B^*(x_i, y_i)$ contains a copy of $K_{t-1} = K_3$. If at least one of the edges of this K_3 is blue, then together with the pair (x_i, y_i) we obtain a Blue K_4. If not, then we obtain a Red K_3 yielding a contradiction.

To arrive at a contradiction when $t \geq 5$, we invoke statement (iii) of Lemma 14.15. Since $|N_B^*(x_i, y_i)| \geq c_2n/2 = \beta n$, we have

$$G_{H,p}[N_B^*(x_i, y_i)] \to (K_3, K_{t-2}).$$

Thus together with the Blue edge (x_i, y_i) and the fact that $N_B^*(x_i, y_i)$ is the Blue neighborhood of (x_i, y_i), we have that $G_{H,p} \to (K_3, K_t)$, yielding again a contradiction with our assumption on the coloring of edges of $G_{H,p}$.

To finish the proof, we have to answer the remaining question: what happens if for none of $i \in \{1, \ldots, s\}$ equation (14.5) holds.

Then, for each i, the edge (x_i, y_i) is incident to $c_2n/2$ Red edges. Moreover since each Red edge is counted at most twice as we vary i, we infer that a Blue neighborhood $N_B(v)$ of v is incident to at least $(1/2)(c_1n/12)(c_2n/2) = c_1c_2n^2/48$ Red edges. Hence there exist a vertex $w \in V$ whose Red neighborhood in $N_B(v)$ is of size at least $c_1c_2n/48$. Let $\overline{N}$ be the intersection of the Blue neighborhood of v and the red neighborhood of w, so that $|\overline{N}| \geq c_1c_2n/48$. Once more, according to (ii) of Lemma 14.15, $\overline{N}$ contains a copy of K_{t-1}. If this copy has a Red edge, then this Red edge forms a Red K_3 with w. Otherwise, all of the edges of the above K_{t-1} are Blue, forming a Blue K_t together with v, yielding a contradiction and thus concluding the proof of Theorem 14.13. $\square$

14.5 Exercises

14.5.1 Let Let $G = K_t$ be a complete graph on t vertices, an let $2 \leq r \leq t$. Find $m_r(G)$.

14.5.2 Determine $m_2(G)$ when

(a) G is obtained from a complete bipartite graph $K_{t,t}$ by adding one edge to to each part of the bipartition,

(b) $G = K_{2,t,t}$ is a complete 3-partite graph,

(c) $G = K_{t,t,t,t}$.

14.5.3 Prove that for monotone graph property $\mathscr{P}$, if $G_{H,p} \in \mathscr{P}$ w.h.p. then also does $G_{H,m}$.

14.5.4 Show that (14.3) holds.

14.5.5 Prove statement (i) of Lemma 14.15.

14.5.6 Let $H \in \overline{\mathscr{G}}(n,d)$ and $m = \omega(n)$. Show that it follows directly from Theorem 14.1 that $G_{H,m} \to (K_3, K_3)$ w.h.p.

14.6 Notes

Perturbed dense graphs

In this subsection all of the results presented use the notation introduced in the beginning of this Chapter. Hence, H always stands for an n-vertex (dense) graph belonging to the family $\mathscr{G}(n,d)$ with minimum degree dn, for $d > 0$, while $G_{H,m}$ and $G_{H,p}$ are random graphs being respective unions of H with $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$.

Spanning subgraphs

Böttcher, Montgomery, Parczyk and Person in [187] prove that for every $d > 0$ and $\Delta \geq 5$, and for every n-vertex graph F with maximum degree at most Δ, they show that if $p = \omega(n^{-2/(\Delta+1)})$ then then $G_{H,p}$ contains a copy of F w.h.p.

Krivelevich, Kwan, and Sudakov in [592] show that for any dense graph G and bounded-degree tree T on the same number of vertices, a modest random perturbation of G will typically contain a copy of T. In particular, they prove that if T is an n-vertex tree with maximum degree Δ, then there exist $c = c(d, \Delta)$, such that a random graph $G_{H,m}$, where $m = cn$, w.h.p. contain T. Böttcher, Han, Kohayakawa, Montgomery, Parczyk, and Person [186] proved that the same holds for all spanning trees simultaneously, while Joos and Kim [530] extended results from [592] to spanning trees with unbounded maximum degree.

A graph G is called *universal* for a family of graphs $\mathscr{F}$ if it contains every element $F \in \mathscr{F}$ as a subgraph. Let $\mathscr{F}(n,2)$ be the family of all n-vertex graphs with maximum degree 2. Parczyk [728] proves that for $0 < d < 1$, $p = \omega(n^{-2/3})$, $G_{H,p}$ is $\mathscr{F}(n,2)$-universal w.h.p.

F-factors

Recall that, given a graph F, a graph G has an *F-factor* if it contains $\lfloor \frac{|V(G)|}{|V(H)|} \rfloor$ pairwise vertex-disjoint copies of F.

Balogh, Treglown, and Wagner proved in [70] that if F is a non-empty graph and n is divisible by $|V(F)|$, then for every $d > 0$ there is $c = c(d,F)$ such that $G_{H,p}$ has an F-factor w.h.p. if $p \geq cn^{-1/d^*(F)}$, where $d^*(F) = \max\left\{ \frac{|E(F')|}{(|V(F')|-1)} : F' \subseteq F, |V(F')| \geq 2 \right\}$.

Böttcher, Parczyk, Sgueglia, and Skokan [188] consider the special case when $F = C_l$, where C_l stands for a cycle of length l. They show that for any integer $l \geq 3$, there exist $c > 0$ such that w.h.p. $G_{H,p}$ contains a C_l-factor if $d > 1/l$ and $p \geq c/n$, and also if $d = 1/l$ and $c \geq c \log n/n$. They also show that for any integer $l \geq 3$, there exists $c > 0$ such that w.h.p. we can find in $G_{H,p}$, $\min\{\delta(G_{H,p}), \lfloor n/l \rfloor\}$ pairwise disjoint copies of C_l, provided that $p \geq c \log n/n$. See also the earlier paper [189] of the same authors dealing with a special case for triangles ($l = 3$).

Han, Montgomery and Treglown in [465] consider the case when $F = K_r$ and find a sharp threshold for the existence of a K_r-factor in $G_{H,p}$. Let $2 \leq k \leq r$ be integers. Then given any $1 - \frac{k}{r} < d < 1 - \frac{k-1}{r}$, $G_{H,p}$ has K_r-factor w.h.p. if $p = \omega(n^{-2/k})$ and w.h.p. does not have a K_r-factor if $p = o(n^{-2/k})$.

Extending the perturbation model, Gomez-Leos and Martin in [445] determined the threshold for the existence of $K_{r,r}$-factor, but this time in a randomly perturbed bipartite graph (with linear minimum degree).

Powers of Hamilton cycles

The main result of Bohman, Frieze and Martin [134] on the existence of Hamilton cycle in randomly perturbed dense graphs (see Theorem 18.20) opened a wide avenue of research on the powers of Hamilton cycles in $G_{H,p}$ and $G_{H,m}$.

In a most recent paper [190], Böttcher, Parczyk, Sgueglia and Skokan investigate the appearance of the square of a Hamilton cycle in the random graph $G_{H,p}$, where $H \in \mathscr{G}(n,d)$ for $d \in (0,1)$. They demonstrate that, as d ranges over the interval $(0,1)$, the perturbed threshold performs a countably infinite number of "jumps". This was known earlier when $d > 1/2$. In the range of $d \in (1/2, 2/3)$ the threshold was determined by Dudek, Reiher Ruciński and Schacht [314] while for $d \geq 2/3$ it was established in [584]. Hence [190] completely settles the question of determining the perturbed threshold for the square of a Hamilton cycle for the whole range of d.

Antoniuk, Dudek, Reiher, Ruciński and Schacht in [46] investigate the existence of powers of Hamiltonian cycles in $G_{H,m}$. For all integers $k \geq 1$, $r \geq 0$, and $l \geq (r+1)r$, and for any $d \geq k/(k+1)$ they show that if $m = O(n^{2-2/l})$ then $G_{H,m}$ w.h.p. contains $(kl+r)$-th power of a Hamiltonian cycle. In particular, for $r = 1$

and $l = 2$ this implies that adding $m = O(n)$ random edges to a graph $H \in \mathscr{G}(n,d)$ already ensures the $(2k+1)$-st power of a Hamiltonian cycle w.h.p, which was proved independently by Nenadov and Trujic [715].

Antoniuk, Dudekk and Ruciński in [47] consider a specific case when H is an n-vertex graph with minimum degree at least $(1/2+\varepsilon)n$ and for every integer $k \geq 2$ precisely estimate the threshold probability for the event that random $G_{H,p}$ contains the k-th power of a Hamiltonian cycle.

In [187] it is proved that, for every integer $k \geq 2$ and $d > 0$, there is some $\eta > 0$ for which the k-th power of a Hamilton cycle w.h.p. appears in $G_{H,p}$ when $p = \omega(n^{-1/k-\eta})$.

For similar results on Hamilton cycles and existence of their k-th powers in a related model, when a dense n-vertex graph is perturbed by a random geometric graph or a random regular graph, see papers [342], [343] and [344], respectively.

Ramsey properties

Krivelevich, Sudakov and Tetali in [605] initiated the study of *edge* Ramsey properties of randomly perturbed graphs. Theorem 14.13, proved in Section 14.4, shows when $G_{H,m} \to (K_3, K_t)$ w.h.p. for $t \geq 3$. Combining their work with results of Das and Treglown [279] and of Powierski [728], it is known that $G_{H,m} \to (K_s, K_t)$ w.h.p. for all values of (s,t), except for the case when $s = 4$ and $t \geq 5$. See [279] for other results on this topic.

Das, Morris and Treglown in [278] focus on *vertex* Ramsey properties of randomly perturbed graphs. In particular they resolve the $(F, K_r)_v$-Ramsey problem for $r \geq 2$ and and arbitrary graph F.

Perturbed sparse graphs

Aigner-Horev, Hefetz and Krivelevich [9] analyse the asymptotic behaviour of the order of the largest complete minor, the order of the largest complete topological minor, the vertex-connectivity, and the diameter of (possibly sparse) graphs that are randomly perturbed using the binomial random graph.

Hahn-Klimroth, Maesaka, Mogge, Mohr and Parczyk [460] consider sparse n-vertex graph H^* with minimum vertex degree at most dn, where $d = o(1)$ and a random perturbed graph $G_{H^*,p}$ being a union of H^* with a random binomial graph $\mathbb{G}_{n,p}$. In particular, they prove the following extension of Theorem 18.20. Let $d = d(n) : \mathbb{N} \mapsto (0,1), b = b(d)$ and $p = b/n$. Then $G_{H^*,p}$ w.h.p. contains a Hamilton cycle. They also discuss embeddings of bounded degree trees and other spanning structures in this model.

Krivelevich , Reichman and Samotij [600] consider as a starting point all n-vertex connected graphs. Given such graph H_c they take its union with $\mathbb{G}_{n,p}$, where $p = \varepsilon/n$ and $\varepsilon > 0$ is a small positive constant. Pointing out to the well known facts that connected graphs can be bad expanders, can have very large diameter, and possibly contain no long paths, they in contrast show that w.h.p. $G_{H_c,p}$ has edge expansion $\Omega(1/\log n)$, diameter $O(\log n)$, vertex expansion $\Omega(1/\log n)$, and contains a path of length $\Omega(n)$, where for the last two properties one has to additionally assume that H_c has bounded maximum degree.

Perturbed digraphs

Bohman, FriezeFrieze, A., Krivelevich and Martin in [135] proved the analogous result to Theorm 18.20 for consistently oriented Hamilton cycles in the randomly perturbed digraph model. Subsequently, Krivelevich , Kwanwan, M. and Sudakov [591] extended this result to consistently oriented cycles of any length between 2 and n, while Araujo, Balogh, Krueger, Piga and Treglown in [53], under the same conditions, proved that the randomly perturbed digraph w.h.p. contains *every* orientation of a cycle of every length between 2 and n.

Perturbed hypergraphs

We next present results on randomly perturbed dense k-uniform hypergraphs (briefly, *k-graphs*) $H = (V, \mathscr{E})$, where V is an n-vertex set and the edge set $\mathscr{E}$ is a family of k-vertex subsets of V obtained by taking the union of H with a random binomial hypergraph $\mathbb{H}_{n,p;k}$, defined in Chapter 12. In particular, these results deal with *l-overlapping* Hamilton cycles, briefly (Hamilton l-cycle), for $1 \le l \le k-1$. In particular, a Hamilton $(k-1)$-cycle is usually called a *tight* Hamilton cycle while a Hamilton 1-cycle is called a *loose* Hamilton cycle (see also Chapter 12).

In the paper [591] Krivelevich, Kwan and Sudakov extended Theorem 18.20 to hypergraphs. They proved that if $d > 0$ and H is a k-graph on $n \in (k-1)\mathbb{N}$ vertices with $\delta_{k-1}(H) \ge dn$, then there exists a function $c_k = c_k(d)$ such that for $p = c_k n^{-(k-1)}$, $H \cup \mathbb{H}_{n,p;k}$ w.h.p. contains a Hamilton 1-cycle (loose Hamilton cycle). Here δ_t stands for the minimum t-degree of H, defined for $1 \le t \le k-1$ as follows. Given a k-graph H with a set S of t vertices (where $1 \le t \le k-1$), let $N_H(S)$ be the collection of $(k-t)$-sets T such that $S \cup T \in \mathscr{E}(H)$, and let $deg_H(S) := |N_H(S)|$. The minimum t-degree $\delta_t(H)$ is the minimum of $deg_H(S)$ over all t-vertex sets S in H.

McDowell and Mycroft [670] partially answered the question posed in [591] whether the above result can be extended for any l and t, $1 \le l, t \le k-1$, proving it for $t \ge \max\{l, k-l\}$. Finally, Han and Zao in [464] solved the problem completely. Since the minimum 1-degree condition is the weakest among t-degree conditions for all $t \ge 1$, they stated and proved their result for $t = 1$ only. They show that for integers $k \ge 3, 1 \le l \le k-1$ and $d > 0$ there exist $\varepsilon > 0$ and an integer $C > 0$ such that the following holds for sufficiently large $n \in (k-1)\mathbb{N}$. Suppose H is a k-graph on n vertices with $\delta_1(H) \geq dn^{k-1}$ and $p \geq n^{-(k-l)-\varepsilon}$ for $l \geq 2$, and $p \geq Cn^{-(k-1)}$ for $l = 1$. Then $H \cup \mathbb{H}_{n,p;k}$ has a Hamilton l-cycle w.h.p.

Recently, Chang, Han and Thoma in [212] extended the above result to r-th power of Hamilton cycle. They proved that if $k \geq 3, r \geq 2$ and $d > 0$, there exists $\varepsilon > 0$ such that if H is an n-vertex k-graph with minimum codegree $\delta_{k-1}(H) \geq dn$ and $p \geq n^{-\binom{k+r-2}{k-1}^{-1}-\varepsilon}$, then the union $H \cup \mathbb{H}_{n,p;k}$ w.h.p. contains the r-th power of a tight Hamilton cycle. Earlier, BedenknechtHan, Kohayakawa and Oliveira Mota in [135] proved this result for $d \geq 1 - n^{-\binom{k+r-2}{k-1}^{-1}}$.

Finally, let us mention that Ramsey properties for randomly perturbed dense 3-uniform hypergrahs have been considered by Aigner-Horev, Hefetz and Schacht in [12].

Part III

Other models

Chapter 15

Trees

The properties of various kinds of trees are one of the main objects of study in graph theory mainly due to their wide range of application in various areas of science. Here we concentrate our attention on the "average" properties of two important classes of trees: labeled and recursive. The first class plays an important role in both the sub-critical and super-critical phase of the evolution of random graphs. On the other hand random recursive trees serve as an example of the very popular random *preferential attachment* models. In particular we will point out, an often overlooked fact, that the first demonstration of a *power law* for the degree distribution in the preferential attachment model was shown in a special class of inhomogeneous random recursive trees.

The families of random trees, whose properties are analyzed in this chapter, fall into two major categories according to the order of their heights: they are either of square root (labeled trees) or logarithmic (recursive trees) height. While most of square-root-trees appear in probability context, most log-trees are encountered in algorithmic applications.

15.1 Labeled Trees

Consider the family $\mathscr{T}_n$ of all n^{n-2} labeled trees on vertex set $[n] = \{1,2,\ldots,n\}$. Let us choose a tree T_n uniformly at random from the family $\mathscr{T}_n$. The tree T_n is called a *random tree* (*random Cayley tree*).

The Prüfer code [754] establishes a bijection between labeled trees on vertex set $[n]$ and the set of sequences $[n]^{n-2}$ of length $n-2$ with items in $[n]$. Such a coding also implies that there is a one-to-one correspondence between the number of labeled trees on n vertices with a given degree sequence $d_1, d_2, \ldots, d_n$ and the number of ways in which one can distribute $n-2$ particles into n cells, such that ith cell contains exactly $d_i - 1$ particles.

If the positive integers $d_i,\ i = 1,2,\ldots,n$ satisfy

$$d_1 + d_2 + \cdots + d_n = 2(n-1),$$

then there exist

$$\binom{n-2}{d_1-1, d_2-1, \ldots, d_n-1} \tag{15.1}$$

trees with n labeled vertices, the ith vertex having degree d_i.

The following observation is a simple consequence of the Prüfer bijection. Namely, there are

$$\binom{n-2}{i-1}(n-1)^{n-i-1} \tag{15.2}$$

trees with n labeled vertices in which the degree of a fixed vertex v is equal to i.

Let X_v be the degree of the vertex v in a random tree T_n, and let $X_v^* = X_v - 1$. Dividing the above formula by n^{n-2}, it follows that, for every i, X_i^* has the $\mathrm{Bin}(n-2, 1/n)$ distribution, which means that the asymptotic distribution of X_i^* tends to the Poisson distribution with mean one.

This observation allows us to obtain an immediate answer to the question of the limiting behavior of the maximum degree of a random tree. Indeed, the proof of Theorem 3.4 yields:

Theorem 15.1. *Denote by $\Delta = \Delta(T_n)$ the maximum degree of a random tree. Then w.h.p.*

$$\Delta(T_n) \approx \frac{\log n}{\log\log n}.$$

□

The classical approach to the study of the properties of labeled trees chosen at random from the family of all labeled trees was purely combinatorial, i.e., via counting trees with certain properties. In this way, Rényi and Szekeres [756], using complex analysis, found the height of a random labeled tree on n vertices (see also Stepanov [818], while for a general probabilistic context of their result, see a survey paper by Biane, Pitman and Yor [111]).

Assume that a tree with vertex set $V = [n]$ is rooted at vertex 1. Then there is a unique path connecting the root with any other vertex of the tree. The height of a tree is the length of the longest path from the root to any pendant vertex of the tree. Pendant vertices are the vertices of degree one.

Theorem 15.2. *Let $h(T_n)$ be the height of a random tree T_n. Then*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{h(T_n)}{\sqrt{2n}} < x\right) = \eta(x),$$

where

$$\eta(x) = \frac{4\pi^{5/2}}{x^3} \sum_{k=1}^{\infty} k^2 e^{-(k\pi/x)^2}.$$

Moreover,

$$\mathbb{E}\, h(T_n) \approx \sqrt{2\pi n} \ \ \textit{and} \quad \operatorname{Var} h(T_n) \approx \frac{\pi(\pi-3)}{3} n.$$

We will now introduce a useful relationship between certain characteristics of random trees and branching processes. Consider a Galton-Watson branching process $\mu(t),\ t = 0,1,\ldots,$ starting with M particles, i.e., with $\mu(0) = M$, in which the number of offspring of a single particle is equal to r with probability $p_r, \sum_{r=0}^{\infty} p_r = 1$. Denote by Z_M the total number of offspring in the process $\mu(t)$. Dwass [319] (see also Viskov [841]) proved the following relationship.

Lemma 15.3. *Let $Y_1, Y_2, \ldots, Y_N$ be a sequence of independent identically distributed random variables, such that*

$$\mathbb{P}(Y_1 = r) = p_r \ \textit{for} \ \ r = 1,2,\ldots,N.$$

Then

$$\mathbb{P}(Z_M = N) = \frac{M}{N} \mathbb{P}(Y_1 + Y_2 + \ldots + Y_N = N - M).$$

Now, instead of a random tree T_n chosen from the family of all labeled trees $\mathscr{T}_n$ on n vertices, consider a tree chosen at random from the family of all $(n+1)^{n-1}$ trees on $n+1$ vertices, with the root labeled 0 and all other vertices labeled from 1 to n. In such a random tree, with a natural orientation of the edges from the root to pendant vertices, denote by V_t the set of vertices at distance t from the root 0. Let the number of outgoing edges from a given vertex be called its *out-degree* and $X_{r,t}^{+}$ be the number of vertices of out-degree r in V_t. For our branching process, choose the probabilities p_r, for $r = 0,1,\ldots$, as equal to

$$p_r = \frac{\lambda^r}{r!} e^{-\lambda},$$

i.e., assume that the number of offspring has the Poisson distribution with mean $\lambda > 0$. Note that λ is arbitrary here.

Let $Z_{r,t}$ be the number of particles in the tth generation of the process, having exactly r offspring. Next let $\underline{X} = [m_{r,t}]$, $r,t = 0,1,\ldots,n$ be a matrix of non-negative integers. Let $s_t = \sum_{r=0}^{n} m_{r,t}$ and suppose that the matrix $\underline{X}$ satisfies the following conditions:

(i) $s_0 = 1$,

$$s_t = m_{1,t-1} + 2m_{2,t-1} + \ldots nm_{n,t-1} \quad \text{for} \quad t = 1,2,\ldots n.$$

(ii) $s_t = 0$ implies that $s_{t+1} = \ldots = s_n = 0$.

(iii) $s_0 + s_1 + \ldots + s_n = n+1$.

Then, as proved by Kolchin [581], the following relationship holds between the out-degrees of vertices in a random rooted tree and the number of offspring in the Poisson process starting with a single particle.

Theorem 15.4.

$$\mathbb{P}([X_{r,t}^{+}] = \underline{X}) = \mathbb{P}([Z_{r,t}] = \underline{X} | Z = n+1).$$

Proof. In Lemma 15.3 let $M = 1$ and $N = n+1$. Then,

$$\begin{aligned}
\mathbb{P}(Z_1 = n+1) &= \frac{1}{n+1}\,\mathbb{P}(Y_1 + Y_2 + \ldots + Y_{n+1} = n) \\
&= \frac{1}{n+1} \sum_{r_1+\ldots+r_{n+1}=n} \prod_{i=1}^{n+1} \frac{\lambda^{r_i}}{r_i!} e^{-\lambda} \\
&= \frac{(n+1)^n \lambda^n e^{-\lambda(n+1)}}{(n+1)!}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{P}([Z_{r,t}] = \underline{X} | Z = n+1) = \\
&= \frac{\prod_{t=0}^{n} \binom{s_t}{m_{0,t},\ldots,m_{n,t}} p_0^{m_{0,t}} \ldots p_n^{m_{n,t}}}{\mathbb{P}(Z = n+1)} \\
&= \frac{(n+1)!\prod_{t=0}^{n} \frac{s_t!}{m_{0,t}!\; m_{1,t}!\; \ldots\; m_{n,t}!} \prod_{r=0}^{n} \left(\frac{\lambda^r}{r!} e^{-\lambda}\right)^{m_{r,t}}}{(n+1)^n \lambda^n e^{-\lambda(n+1)}} \\
&= \frac{(n+1)!\; s_1!\; s_2!\; \ldots\; s_n!}{(n+1)^n} \prod_{t=0}^{n} \prod_{r=0}^{n} \frac{1}{m_{r,t}!\; (r!)^{m_{r,t}}}.
\end{aligned} \tag{15.3}$$

On the other hand, one can construct all rooted trees such that $[X_{r,t}^+] = \underline{X}$ in the following manner. We first layout an unlabelled tree in the plane. We choose a single point $(0,0)$ for the root and then points $S_t = \{(i,t) : i = 1,2,\ldots,s_t\}$ for $t = 1,2,\ldots,n$. Then for each t,r we choose $m_{r,t}$ points of S_t that will be joined to r points in S_{t+1}. Then, for $t = 0,1,\ldots,n-1$ we add edges. Note that S_n, if non-empty, has a single point corresponding to a leaf. We go through S_t in increasing order of the first component. Suppose that we have reached (i,t) and this has been assigned out-degree r. Then we join (i,t) to the first r vertices of S_{t+1} that have not yet been joined by an edge to a point in S_t. Having put in these edges, we assign labels $1,2,\ldots,n$ to $\bigcup_{t=1}^n S_t$. The number of ways of doing this is

$$\prod_{t=1}^{n} \frac{s_t!}{\prod_{r=1}^n m_{r,t}!} \times n!.$$

The factor $n!$ is an over count. As a set of edges, each tree with $[X_{r,t}^+] = \underline{X}$ appears exactly $\prod_{t=0}^n \prod_{r=0}^n (r!)^{m_{r,t}}$ times, due to permutations of the trees below each vertex. Summarising, the total number of tree with out-degrees given by the matrix $\underline{X}$ is

$$n!\, s_1!\, s_2! \,\ldots\, s_n! \prod_{t=0}^{n} \prod_{r=0}^{n} \frac{1}{m_{r,t}!\, (r!)^{m_{r,t}}},$$

which, after division by the total number of labeled trees on $n+1$ vertices, i.e., by $(n+1)^{n-1}$, results in an identical formula to that given for the random matrix $[X_+^{r,t}]$ in the case of $[Z_{r,t}]$, see (15.3). To complete the proof one has to notice that for those matrices $\underline{X}$ which do not satisfy conditions (i) to (iii) both probabilities in question are equal to zero. $\square$

Hence, roughly speaking, a random rooted labeled tree on n vertices has asymptotically the same shape as a branching process with Poisson, parameter one in terms of family sizes. Grimmett [452] uses this probabilistic representation to deduce the asymptotic distribution of the distance from the root to the nearest pendant vertex in a random labeled tree T_n, $n \geq 2$. Denote this random variable by $d(T_n)$.

Theorem 15.5. *As $n \to \infty$,*

$$\mathbb{P}(d(T_n) \geq k) \to \exp\left\{\sum_{i=1}^{k-1} \alpha_i\right\},$$

where the α_i are given recursively by

$$\alpha_0 = 0, \quad \alpha_{i+1} = e^{\alpha_i} - e^{-1} - 1.$$

Proof. Let k be a positive integer and consider the sub-tree of T_n induced by the vertices at distance at most k from the root. Within any level (strata) of T_n, order the vertices in increasing lexicographic order, and then delete all labels, excluding that of the root. Denote the resulting tree by T_n^k.

Now consider the following branching process constructed recursively according to the following rules:

(i) Start with one particle (the unique member of generation zero).

(ii) For $k \geq 0$, the $(k+1)$th generation A_{k+1} is the union of the families of descendants of the kth generation together with one additional member which is allocated at random to one of these families, each of the $|A_k|$ families having equal probability of being chosen for this allocation. As in Theorem 15.4, all family sizes are independent of each other and the past, and are Poisson distributed with mean one.

Lemma 15.6. *As $n \to \infty$ the numerical characteristics of T_n^k have the same distribution as the corresponding characteristics of the tree defined by the first k generations of the branching process described above.*

Proof. For a proof of Lemma 15.6, see the proof Theorem 3 of [452].

Let Y_k be the size of the kth generation of our branching process and let N_k be the number of members of the kth generation with no offspring. Let $\mathbf{i} = (i_1, i_2, \ldots, i_k)$ be a sequence of positive integers, and let

$$A_j = \{N_j = 0\} \quad \text{and} \quad B_j = \{Y_j = i_j\} \quad \text{for} \quad j = 1, 2, \ldots, k.$$

Then, by Lemma 15.6, as $n \to \infty$,

$$\mathbb{P}(d(T_n) \geq k) \to \mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k) \; .$$

Now,

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \sum_{\mathbf{i}} \prod_{j=1}^{k} \mathbb{P}(A_j | A_1 \cap \ldots \cap A_{j-1} \cap B_1 \cap \ldots B_j)$$
$$\times \mathbb{P}(B_j | A_1 \cap \ldots \cap A_{j-1} \cap B_1 \cap \ldots B_{j-1}),$$

Using the Markov property,

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \sum_{\mathbf{i}} \prod_{j=1}^{k} \mathbb{P}(A_j | B_j) \, \mathbb{P}(B_j | A_{j-1} \cap B_{j-1})$$

$$= \sum_{\mathbf{i}} \prod_{j=1}^{k} \left(1 - e^{-1}\right)^{i_j - 1} C_j(i_j), \tag{15.4}$$

where $C_j(i_j) = \mathbb{P}(B_j | A_{j-1} \cap B_{j-1})$ is the coefficient of x^{i_j} in the probability generating function $D_j(x)$ of Y_j conditional upon $Y_{j-1} = i_{j-1}$ and $N_j = 0$. Thus

$$Y_j = 1 + Z + R_1 + \ldots + R_{i_{j-1}-1},$$

where Z has the Poisson distribution and the R_i are independent random variables with Poisson distribution conditioned on being non-zero. Hence

$$D_j(x) = xe^{x-1} \left(\frac{e^x - 1}{e - 1} \right)^{i_{j-1}-1}.$$

Now,

$$\sum_{i_k=1}^{\infty} (1 - e^{-1})^{i_k - 1} C_k(i_k) = \frac{D_k(1 - e^{-1})}{1 - e^{-1}}.$$

We can use this to eliminate i_k in (15.4) and give

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \sum_{(i_1, \ldots, i_{k-1})} \prod_{j=1}^{k-1} \beta_1^{i_j - 1} C_j(i_j) e^{\beta_1 - 1} \left(\frac{e^{\beta_1} - 1}{e - 1} \right)^{i_{k-1}-1}, \tag{15.5}$$

where $\beta_1 = 1 - e^{-1}$. Eliminating i_{k-1} from (15.5) we get

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \sum_{(i_1, \ldots, i_{k-2})} \prod_{j=1}^{k-2} \beta_1^{i_j - 1} C_j(i_j) e^{\beta_1 + \beta_2 - 2} \left(\frac{e^{\beta_2} - 1}{e - 1} \right)^{i_{k-2}-1},$$

where $\beta_2 = (e^{\beta_1} - 1)$. Continuing we see that, for $k \geq 1$,

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \exp\left\{ \sum_{i=1}^{k} (\beta_i - 1) \right\} = \exp\left\{ \sum_{i=1}^{k} \alpha_i \right\},$$

where $\beta_0, \beta_1, \ldots$ are given by the recurrence

$$\beta_0 = 1, \quad \beta_{i+1} = \left(e^{\beta_i} - 1\right) e^{-1},$$

and $\alpha_i = \beta_i - 1$. One can easily check that β_i remains positive and decreases monotonically as $i \to \infty$, and so $\alpha_i \to -1$. □

Another consequence of Lemma 15.3 is that, for a given N, one can associated with the sequence $Y_1, Y_2, \ldots, Y_N$, a generalized occupancy scheme of distributing n particles into N cells (see [581]). In such scheme, the joint distribution of the number of particles in each cell $(\nu_1, \nu_2, \ldots, \nu_N)$ is given, for $r = 1, 2, \ldots, N$ by

$$\mathbb{P}(\nu_r = k_r) = \mathbb{P}\left(Y_r = k_r \,\middle|\, \sum_{r=1}^{N} Y_r = n\right). \tag{15.6}$$

Now, denote by $X_r^+ = \sum_{t=0}^{n} X_{r,t}^+$ the number of vertices of out-degree r in a random tree on $n+1$ vertices, rooted at a vertex labeled 0. Denote by $Z^{(r)} = \sum_{t=0}^{n} Z_{r,t}$, the number of particles with exactly r offspring in the Poisson process $\mu(t)$. Then by Theorem 15.4,

$$\mathbb{P}(X_r^+ = k_r,\ r = 0, 1, \ldots, n) = \mathbb{P}(Z^{(r)} = k_r,\ r = 0, 1, \ldots, n | Z_1 = n+1).$$

Hence by equation (15.1), the fact that we can choose $\lambda = 1$ in the process $\mu(t)$ and (15.6), the joint distribution of out-degrees of a random tree coincides with the joint distribution of the number of cells containing the given number of particles in the classical model of distributing n particles into $n+1$ cells, where each choice of a cell by a particle is equally likely.

The above relationship, allows us to determine the asymptotic behavior of the expectation of the number X_r of vertices of degree r in a random labeled tree T_n.

Corollary 15.7.

$$\mathbb{E}X_r \approx \frac{n}{(r-1)!\, e}.$$

15.2 Recursive Trees

We call a tree on n vertices labeled $1, 2, \ldots, n$ a *recursive tree* (or *increasing tree*) if the tree is rooted at vertex 1 and, for $2 \le i \le n$, the labels on the unique path from the root to vertex i form an increasing sequence. It is not difficult to see that any such tree can be constructed "recursively": Starting with the vertex labeled 1 and assuming that vertices "arrive" in order of their labels, and connect themselves by an edge to one of the vertices which "arrived" earlier. So the number of recursive (increasing) trees on n vertices is equal to $(n-1)!$.

A *random recursive tree* is a tree chosen uniformly at random from the family of all $(n-1)!$ recursive trees. Or equivalently, it can be generated by a recursive procedure in which each new vertex chooses a neighbor at random from previously arrived vertices. We assume that our tree is rooted at vertex 1 and all edges are directed from the root to the leaves.

Let T_n be a random recursive tree and let $D^+_{n,i}$ be the out-degree of the vertex with label i, i.e the number of "children" of vertex i. We start with the exact probability distribution of these random variables.

Theorem 15.8. *For $i = 1,2,\dots,n$ and $r = 1,2,\dots,n-1$,*

$$\mathbb{P}(D^+_{n,i} = r) = \frac{(i-1)!}{(n-1)!}\sum_{k=r}^{n-i}\binom{k}{r}(i-1)^{k-r}|s(n-i,k)| \tag{15.7}$$

where $s(n-i,k)$ is the Stirling number of the first kind.

Proof. Conditioning on tree T_{n-1} we see that, for $r \geq 1$,

$$\mathbb{P}(D^+_{n,i} = r) = \frac{n-2}{n-1}\mathbb{P}(D^+_{n-1,i} = r) + \frac{1}{n-1}\mathbb{P}(D^+_{n-1,i} = r-1). \tag{15.8}$$

Fix i and let

$$\Phi_{n,i}(z) = \sum_{r=0}^{n-i}\mathbb{P}(D^+_{n,i} = r)z^r$$

be the probability generating function of $D^+_{n,i}$.
Multiplying (15.8) by z^r and then summing over $r \geq 1$ we see that

$$\Phi_{n,i}(z) - \mathbb{P}(D^+_{n,i} = 0) =$$
$$\frac{n-2}{n-1}\Big(\Phi_{n-1,i}(z) - \mathbb{P}(D^+_{n-1,i} = 0)\Big) + \frac{z}{n-1}\Phi_{n-1,i}(z).$$

Notice, that the probability that vertex i is a leaf equals

$$\mathbb{P}(D^+_{n,i} = 0) = \prod_{j=i}^{n-1}\left(1 - \frac{1}{j}\right) = \frac{i-1}{n-1}. \tag{15.9}$$

Therefore

$$\Phi_{n,i}(z) = \frac{n-2}{n-1}\Phi_{n-1,i}(z) + \frac{z}{n-1}\Phi_{n-1,i}(z).$$

With the boundary condition,

$$\Phi_{i,i}(z) = \mathbb{P}(D^+_{i,i} = 0) = 1.$$

One can verify inductively that

$$\Phi_{n,i}(z) = \prod_{k=1}^{n-i}\left(\frac{z+i+k-2}{i+k-1}\right)$$

$$= \frac{(i-1)!}{(n-1)!}(z+i-1)(z+i)\dots(z+n-2). \tag{15.10}$$

Recall the definition of Stirling numbers of the first kind $s(n,k)$. For non-negative integers n and k

$$z(z-1)\dots(z-n+1) = \sum_{k=1}^{n} s(n,k)z^k.$$

Hence

$$\begin{aligned}
\Phi_{n,i}(z) &= \frac{(i-1)!}{(n-1)!}\sum_{k=1}^{n-i} |s(n-i,k)|(z+i-1)^k \\
&= \frac{(i-1)!}{(n-1)!}\sum_{k=1}^{n-i}\sum_{r=0}^{k}\binom{k}{r} z^r(i-1)^{k-r}|s(n-i,k)| \\
&= \sum_{r=0}^{n-i}\left(\frac{(i-1)!}{(n-1)!}\sum_{k=r}^{n-i}\binom{k}{r}(i-1)^{k-r}|s(n-i,k)|\right) z^r.
\end{aligned}$$

$\square$

It follows from (15.10), by putting $z=0$, that the expected number of vertices of out-degree zero is

$$\sum_{i=1}^{n}\frac{i-1}{n-1} = \frac{n}{2}.$$

Then (15.8) with $i=r=1$ implies that $\mathbb{P}(D^+_{n,1}=1) = 1/(n-1)$. Hence, if L_n is the number of leaves in T_n, then

$$\mathbb{E}L_n = \frac{n}{2} + \frac{1}{n-1}. \tag{15.11}$$

For a positive integer n, let $\zeta_n(s) = \sum_{k=1}^{n} k^{-s}$ be the incomplete Riemann zeta function, and let $H_n = \zeta(1) = \sum_{k=1}^{n} k^{-1}$ be the nth harmonic number, and let $\delta_{n,k}$ denote the Kronecker function $1_{n=k}$.

Theorem 15.9. *For $1 \le i \le n$, let $D_{n,i}$ be the degree of vertex i in a random recursive tree T_n. Then*

$$\mathbb{E}D_{n,i} = H_{n-1} - H_{i-1} + 1 - \delta_{1,i},$$

while

$$\operatorname{Var} D_{n,i} = H_{n-1} - H_{i-1} - \zeta_{n-1}(2) + \zeta_{i-1}(2).$$

Proof. Let N_j be the label of that vertex among vertices $1,2,\dots j-1$ which is the parent of vertex j. Then for $j \geq 1$ and $1 \leq i < j$

$$D_{n,i} = \sum_{j=i+1}^{n} \delta_{N_j,i}. \tag{15.12}$$

By definition $N_2, N_3, \dots, N_n$ are independent random variables and for all i, j,

$$\mathbb{P}(N_j = i) = \frac{1}{j-1}. \tag{15.13}$$

The expected value of $D_{n,i}$ follows immediately from (15.12) and (15.13). To compute the variance observe that

$$\operatorname{Var} D_{n,i} = \sum_{j=i+1}^{n} \frac{1}{j-1}\left(1 - \frac{1}{j-1}\right).$$

$\square$

From the above theorem it follows that $\operatorname{Var} D_{n,i} \leq \mathbb{E} D_{n,i}$. Moreover, for fixed i and n large, $\mathbb{E} D_{n,i} \approx \log n$, while for i growing with n the expectation $\mathbb{E} D_{n,i} \approx \log n - \log i$. The following theorem, see Kuba and Panholzer [611], shows a standard limit behavior of the distribution of $D_{n,i}$.

Theorem 15.10. *Let $i \geq 1$ be fixed and $n \to \infty$. Then*

$$(D_{n,i} - \log n)/\sqrt{\log n} \xrightarrow{d} N(0,1).$$

Now, let $i = i(n) \to \infty$ as $n \to \infty$. If

(i) i=o(n), then

$$(D_{n,i} - (\log n - \log i))/\sqrt{\log n - \log i} \xrightarrow{d} N(0,1),$$

(ii) $i = cn$, $0 < c < 1$, then

$$D_{n,i} \xrightarrow{d} \operatorname{Po}(-\log c),$$

(iii) $n - i = o(n)$, then

$$\mathbb{P}(D^+_{n,i} = 0) \to 1.$$

Now, consider another parameter of a random recursive tree.

Theorem 15.11. *Let $r \geq 1$ be fixed and let $X_{n,r}$ be the number of vertices of degree r in a random recursive tree T_n. Then, w.h.p.*

$$X_{n,r} \approx n/2^r,$$

and

$$\frac{X_{n,r} - n/2^r}{\sqrt{n}} \xrightarrow{d} Y_r,$$

as $n \to \infty$, where Y_r has the $N(0, \sigma_r^2)$ distribution.

□

In place of proving the above theorem we will give a simple proof of its immediate implication, i.e., the asymptotic behavior of the expectation of the random variable $X_{n,r}$. The proof of asymptotic normality of suitably normalized $X_{n,r}$ is due to Janson and can be found in [501]. (In fact, in [501] a stronger statement is proved, namely, that, asymptotically, for all $r \geq 1$, random variables $X_{n,r}$ are jointly Normally distributed.)

Corollary 15.12. *Let $r \geq 1$ be fixed. Then*

$$\mathbb{E} X_{n,r} \approx n/2^r.$$

Proof. Let us introduce a random variable $Y_{n,r}$ counting the number of vertices of degree at least r in T_n. Obviously,

$$X_{n,r} = Y_{n,r} - Y_{n,r+1}. \tag{15.14}$$

Moreover, using a similar argument to that given for formula (15.7), we see that for $2 \leq r \leq n$,

$$\mathbb{E}[Y_{n,r}|T_{n-1}] = \frac{n-2}{n-1} Y_{n-1,r} + \frac{1}{n-1} Y_{n-1,r-1} \tag{15.15}$$

Notice, that the boundary condition for the recursive formula (15.15) is, trivially given by

$$\mathbb{E} Y_{n,1} = n.$$

We will show, that $\mathbb{E} Y_{n,r}/n \to 2^{-r+1}$ which, by (15.14), will imply the theorem. Set

$$a_{n,r} := n2^{-r+1} - \mathbb{E} Y_{n,r}. \tag{15.16}$$

$\mathbb{E} Y_{n,1} = n$ implies that $a_{n,1} = 0$. We see from (15.11) that the expected number of leaves in a random recursive tree on n vertices is given by

$$\mathbb{E} X_{n,1} = \frac{n}{2} + \frac{1}{n-1}.$$

Hence $a_{n,2} = 1/(n-1)$ as $\mathbb{E}Y_{n,2} = n - \mathbb{E}X_{n,1}$.
Now we show that,

$$0 < a_{n,1} < a_{n,2} < \cdots < a_{n,n-1}. \tag{15.17}$$

From the relationships (15.15) and (15.16) we get

$$a_{n,r} = \frac{n-2}{n-1} a_{n-1,r} + \frac{1}{n-1} a_{n-1,r-1}. \tag{15.18}$$

Inductively assume that (15.17) holds for some $n \geq 3$. Now, by (15.18), we get

$$a_{n,r} > \frac{n-2}{n-1} a_{n-1,r-1} + \frac{1}{n-1} a_{n-1,r-1} = a_{n-1,r-1}.$$

Finally, notice that

$$a_{n,n-1} = n2^{2-n} - \frac{2}{(n-1)!},$$

since there are only two recursive trees with n vertices and a vertex of degree $n-1$. So, we conclude that $a(n,r) \to 0$ as $n \to \infty$, for every r, and our theorem follows. □

Finally, consider the maximum degree $\Delta_n = \Delta_n(T_n)$ of a random recursive tree T_n. It is easy to see that for large n, its expected value should exceed $\log n$, since it is as large as the expected degree of the vertex 1, which by Theorem 15.9 equals $H_{n-1} \approx \log n$. Szymański [825] proved that the upper bound is $O(\log_2 n)$ (see Goh and Schmutz [442] for a strengthening of his result). Finally, Devroye and Lu (see [290]) have shown that in fact $\Delta_n \approx \log_2 n$. This is somewhat surprising. While each vertex in $[1, n^{1-o(1)}]$ only has a small chance of having such a degree, there are enough of these vertices to guarantee one w.h.p..

Theorem 15.13. *In a random recursive tree T_n, w.h.p.*

$$\Delta_n \approx \log_2 n.$$

The next theorem was originally proved by Devroye [285] and Pittel [740]. Both proofs were based on an analysis of certain branching processes. The proof below is related to [285].

Theorem 15.14. *Let $h(T_n)$ be the height of a random recursive tree T_n. Then w.h.p.*

$$h(T_n) \approx e \log n.$$

Proof.
Upper Bound: For the upper bound we simply estimate the number ν_1 of vertices at height $h_1 = (1+\varepsilon)e\log n$ where $\varepsilon = o(1)$ but is sufficiently large so that claimed inequalities are valid. Each vertex at this height can be associated with a path $i_0 = 1, i_1, \ldots, i_h$ of length h in T_n. So, if $S = \{i_1, \ldots, i_h\}$ refers to such a path, then

$$\mathbb{E}\,\nu_1 = \sum_{|S|=h_1} \prod_{i\in S} \frac{1}{i-1} \le \frac{1}{h_1!}\left(\sum_{i=1}^{n} \frac{1}{i}\right)^{h_1} \le \left(\frac{(1+\log n)e}{h_1}\right)^{h_1} = o(1), \quad (15.19)$$

assuming that $h_1\varepsilon \to \infty$.

Explanation: If $S = \left\{i_1 = 1, i_2, \ldots, i_{h_1}\right\}$ then the term $\prod_{j=1}^{h_1} 1/i_j$ is the probability that i_j chooses i_{j-1} in the construction of T_n.
Lower Bound: The proof of the lower bound is more involved. We consider a different model of tree construction and relate it to T_n. We consider a *Yule* process. We run the process for a specific time t and construct a tree $Y(t)$. We begin by creating a single particle x_1 at time 0 this will be the root of a tree $Y(t)$. New particles are generated at various times $\tau_1 = 0, \tau_2, \ldots,$. Then at time τ_k there will be k particles $X_k = \{x_1, x_2, \ldots, x_k\}$ and we will have $Y(t) = Y(\tau_k)$ for $\tau_k \le t < \tau_{k+1}$. After x_k has been added to $Y(\tau_k)$, each $x \in X_k$ is associated with an exponential random variable E_x with mean one[1]. If z_k is the particle in X_k that minimizes $E_x, x \in X_k$ then a new particle x_{k+1} is generated at time $\tau_{k+1} = \tau_k + E_{z_k}$ and an edge $\{z_k, x_{k+1}\}$ is added to $Y(\tau_k)$ to create $Y(\tau_{k+1})$. After this we independently generate new random variables $E_x, x \in X_{k+1}$.

Remark 15.15. The *memory-less property* of the exponential random variable, i.e. $\mathbb{P}(Z \ge a+b \mid Z \ge a) = \mathbb{P}(Z \ge b)$, implies that we could equally well think that at time $t \ge \tau_k$ the E_x are independent exponentials conditional on being at least τ_k. In which case the choice of z_k is uniformly random from X_k, even conditional on the processes prior history.

Suppose then that we focus attention on $Y(y; s, t)$, the sub-tree rooted at y containing all descendants of y that are generated after time s and before time t.

We observe three things:

(T1) The tree $Y(\tau_n)$ has the same distribution as T_n. This is because each particle in X_k is equally likely to be z_k.

[1] An exponential random variable Z with mean λ is characterised by $\mathbb{P}(Z \ge x) = e^{-x/\lambda}$.

(T2) If $s < t$ and $y \in Y(s)$ then $Y(y;s,t)$ is distributed as $Y(t-s)$. This follows from Remark 15.15, because when $z_k \notin Y(y;s,t)$ it does not affect any of the the variables $E_x, x \in Y(y;s,t)$.

(T3) If $x,y \in Y(s)$ then $Y(x;s,t)$ and $Y(y;s,t)$ are independent. This also follows from Remark 15.15 for the same reasons as in (T2).

It is not difficult to prove (see Exercise (vii) or Feller [350]) that if $P_n(t)$ is the probability there are exactly n particles at time t then

$$P_n(t) = e^{-t}(1-e^{-t})^{n-1}. \tag{15.20}$$

Next let

$$t_1 = (1-\varepsilon)\log n.$$

Then it follows from (15.20) that if $\nu(t)$ is the number of particles in our Yule process at time t then

$$\mathbb{P}(\nu(t_1) \geq n) \leq \sum_{k \geq n} e^{-t_1}(1-e^{-t_1})^{k-1} = \left(1 - \frac{1}{n^{1-\varepsilon}}\right)^{n-1} = o(1). \tag{15.21}$$

We will show that w.h.p. the tree $T_{\nu(t_1)}$ has height at least

$$h_0 = (1-\varepsilon)et_1$$

and this will complete the proof of the theorem.

We will choose $s \to \infty$, $s = O(\log t_1)$. It follows from (15.20) that if $\nu_0 = \varepsilon e^s$ then

$$\mathbb{P}(\nu(s) \leq \nu_0) = \sum_{k=0}^{\nu_0} e^{-s}(1-e^{-s})^{k-1} \leq \varepsilon = o(1). \tag{15.22}$$

Suppose now that $\nu(s) \geq \nu_0$ and that the vertices of $T_{1;0,s}$ are $\{x_1, x_2, \ldots, x_{\nu(s)}\}$. Let $\sigma = \nu_0^{1/2}$ and consider the sub-trees $A_j, j = 1,2,\ldots,\tau$ of $T_{1;0,t_1}$ rooted at $x_j, j = 1,2,\ldots,\nu(s)$. We will show that

$$\mathbb{P}(T_{x_1;s,t_1} \text{ has height at least } (1-\varepsilon)^3 e\log n) \geq \frac{1}{2\sigma\log\sigma}. \tag{15.23}$$

Assuming that (15.23) holds, since the trees $A_1, A_2, \ldots, A_\tau$ are independent, by T3, we have

$$\mathbb{P}(h(T_n) \leq (1-\varepsilon)^3 e\log n) \leq$$
$$o(1) + \mathbb{P}(h(T_{1;0,t_1}) \leq (1-\varepsilon)^3 e\log n) \leq o(1) + \left(1 - \frac{1}{2\sigma\log\sigma}\right)^{\nu_0} = o(1).$$

To prove all this we will associate a Galton-Watson branching process with each of $x_1, x_2, \dots, x_\tau$. Consider for example $x = x_1$ and let $\tau_0 = \log \sigma$. The vertex x will be the root of a branching process Π, which we now define. We will consider the construction of $Y(x; s, t)$ at times $\tau_i = s + i\tau_0$ for $i = 1, 2, \dots, i_0 = (t_1 - s)/\tau_0$. The children of x in Π are the vertices at depth at least $(1-\varepsilon)e\tau_0$ in $Y(x; s, \tau_1)$. In general, the particles in generation i will correspond to particles at depth at least $(1-\varepsilon)e\tau_0$ in the tree $Y(\xi; \tau_{i-1}, \tau_i)$ where ξ is a particle of $Y(x; s, t)$ included in generation $i-1$ of Π.

If the process Π does not ultimately become extinct then generation i_0 corresponds to vertices in $Y(t)$ that are at depth

$$i_0 \times (1-\varepsilon)e\tau_0 = (1-\varepsilon)e(t_1 - s) \geq (1-\varepsilon)^3 e \log n.$$

We will prove that

$$\mathbb{P}(\Pi \text{ does not become extinct}) \geq \frac{1}{2\sigma \log \sigma}, \tag{15.24}$$

and this implies (15.23) and the theorem.

To prove (15.24) we first show that μ, the expected number of progeny of a particle in Π satisfies $\mu > 1$ and after that we prove (15.24).

Let $D(h, m)$ denote the expected number of vertices at depth h in the tree T_m. Then for any $\xi \in \Pi$,

$$\mu \geq D((1-\varepsilon)e\tau_0, \sigma) \times \mathbb{P}(\nu(\tau_0) \geq \sigma). \tag{15.25}$$

It follows from (15.20) and $\sigma = e^{\tau_0}$ that

$$\mathbb{P}(|Y(\xi, 0, \tau_0)| \geq \sigma) = \sum_{k=\sigma}^{\infty} e^{-\tau_0}(1 - e^{-\tau_0})^k$$
$$= (1 - e^{-\tau_0})^\sigma \geq \frac{1}{2e}. \tag{15.26}$$

We show next that for $m \gg h$ we have

$$D(h, m) \geq \frac{(\log m - \log h - 1)^h}{h!}. \tag{15.27}$$

To prove this, we go back to (15.19) and write

$$D(h, m) = \frac{1}{h} \sum_{i=2}^{m} \frac{1}{i-1} \sum_{S \in \binom{[2,m] \setminus \{i\}}{h-1}} \prod_{j \in S \setminus \{i\}} \frac{1}{j-1}$$

$$= \frac{1}{h} \sum_{S \in \binom{[2,m]}{h-1}} \prod_{j \in S} \frac{1}{j-1} \sum_{1 \neq k \notin S}^{m} \frac{1}{k-1} \geq \frac{1}{h} \sum_{S \in \binom{[2,m]}{h-1}} \prod_{j \in S} \frac{1}{j-1} \sum_{k=h+1}^{m} \frac{1}{k}$$
$$\geq \frac{\log m - \log h - 1}{h} D(h-1,m). \quad (15.28)$$

Equation (15.27) follows by induction since $D(1,m) \geq \log m$.

Explanation of (15.28)**:** We choose a path of length h by first choosing a vertex i and then choosing $S \subseteq [2,m] \setminus \{i\}$. We divide by h because each h-set arises h times in this way. Each choice will contribute $\prod_{j \in S \cup \{i\}} \frac{1}{j-1}$. We change the order of summation i, S and then lower bound $\sum_{1 \neq k \notin S}^{m} \frac{1}{k-1}$ by $\sum_{k=h+1}^{m} \frac{1}{k}$.

We now see from (15.25), (15.26) and (15.27) that

$$\mu \geq \frac{(\tau_0 - \log((1-\varepsilon)e\tau_0) - 1)^{(1-\varepsilon)e\tau_0}}{((1-\varepsilon)e\tau_0)!} \times \frac{1}{2e} \geq$$
$$\frac{1}{2e\sqrt{2\pi}} \times \frac{1}{(1-\varepsilon/2)^{(1-\varepsilon)e\tau_0}} \gg 1,$$

if we take $\varepsilon\tau_0 / \log \tau_0 \to \infty$.

We are left to prove (15.24). Let $G(z)$ be the probability generating function for the random variable Z equal to the number of descendants of a single particle. We first observe that for any $\theta \geq 1$,

$$\mathbb{P}(Z \geq \theta\sigma) \leq \mathbb{P}(|Y(\xi,0,\tau_0)| \geq \theta\sigma) = \sum_{k=\theta\sigma}^{\infty} e^{-\tau_0}(1-e^{-\tau_0})^k \leq e^{-\theta}.$$

Note that for $0 \leq x \leq 1$, any $k \geq 0$ and $a \geq k$ it holds that

$$\left(1 - \frac{k}{a}\right) + \frac{k}{a} x^a \geq x^k.$$

We then write for $0 \leq x \leq 1$,

$$G(x) \leq \sum_{k=0}^{\theta\sigma} p_k x^k + \mathbb{P}(Z \geq \theta\sigma) \leq \sum_{k=0}^{\theta\sigma} p_k x^k + e^{-\theta}$$
$$\leq \sum_{k=0}^{\theta\sigma} \left(\left(1 - \frac{k}{\theta\sigma}\right) p_k + \frac{k}{\theta\sigma} p_k x^{\theta\sigma} \right) + e^{-\theta}$$
$$\leq \sum_{k=0}^{\infty} \left(\left(1 - \frac{k}{\theta\sigma}\right) p_k + \frac{k}{\theta\sigma} p_k x^{\theta\sigma} \right) + e^{-\theta}$$
$$= H(x) = 1 - \frac{\mu}{\theta\sigma} + \frac{\mu}{\theta\sigma} x^{\theta\sigma} + e^{-\theta}.$$

The function H is monotone increasing in x and so $\rho = \mathbb{P}(\Pi \text{ becomes extinct})$ being the smallest non-negative solution to $x = G(x)$ (see Theorem 29.1) implies that ρ is at most the smallest non-negative solution q to $x = H(x)$. The convexity of H and the fact that $H(0) > 0$ implies that q is at most the value ζ satisfying $H'(\zeta) = 1$ or

$$q \leq \zeta = \frac{1}{\mu^{1/(\theta\sigma-1)}} < 1.$$

But $\rho = G(\rho) \leq G(q) \leq H(q)$ and so

$$1-\rho \geq \frac{\mu}{\theta\sigma}\left(1-\frac{1}{\mu^{\theta\sigma/(\theta\sigma-1)}}\right)-e^{-\theta} \geq \frac{\mu-1}{\theta\sigma}-e^{-\theta} \geq \frac{1}{2\sigma\log\sigma},$$

after putting $\theta = 2\log\sigma$ and using $\mu \gg 1$. □

Devroye, Fawzi and Fraiman [286] give another proof of the above theorem that works for a wider class of random trees called *scaled attachment random recursive trees*, where each vertex i attaches to the random vertex $\lfloor iX_i \rfloor$ and $X_0, \ldots, X_n$ is a sequence of independent identically distributed random variables taking values in $[0,1)$.

15.3 Inhomogeneous Recursive Trees

Plane-oriented recursive trees

This section is devoted to the study of the properties of a class of *inhomogeneous recursive trees* that are closely related to the Barabási-Albert "preferential attachment model", see [72]. Bollobás, Riordan, Spencer and Tusnády gave a proper definition of this model and showed how to reduce it to random *plane-oriented recursive trees*, see [177]. In this section we present some results that preceded [72] and created a solid mathematical ground for the further development of general preferential attachment models, which will be discussed later in the book (see Chapter 18).

Suppose that we build a recursive tree in the following way. We start as before with a single vertex labeled 1 and add $n-1$ vertices labeled $2,3,\ldots,n$, one by one. We assume that the children of each vertex are ordered (say, from left to right). At each step a new vertex added to the tree is placed in a position "in between" old vertices. A tree built in this way is called a *plane-oriented recursive tree*. To study this model it is convenient to introduce an *extension* of a plane-oriented recursive tree: given a plane-oriented recursive tree we connect each vertex with external nodes, representing a possible insertion position for an incoming new vertex. See Figure 15.3 for a diagram of all plane-oriented recursive trees on $n = 3$ vertices, together with their extensions.

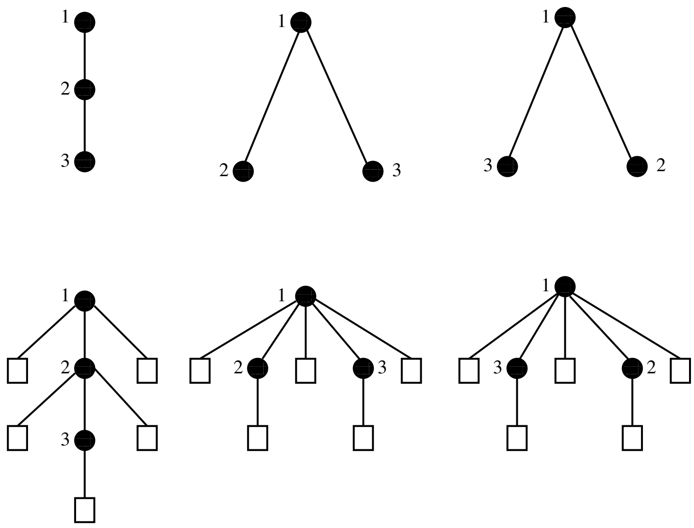


Figure 15.1: Plane-oriented recursive trees and their extensions, $n = 3$

Assume now, as before that all the edges of a tree are directed toward the leaves, and denote the out-degree of a vertex v by $d^+(v)$. Then the total number of extensions of an plane-oriented recursive tree on n vertices is equal to

$$\sum_{v \in V} (d^+(v) + 1) = 2n - 1.$$

So a new vertex can choose one those those $2n - 1$ places to join the tree and create a tree on $n + 1$ vertices. If we assume that this choice in each step is made uniformly at random then a tree constructed this way is called a *random plane-oriented recursive tree*. Notice that the probability that the vertex labeled $n + 1$ is attached to vertex v is equal to $\frac{d^+(v)+1}{2n-1}$ i.e., it is proportional to the degree of v. Such random trees, called *plane-oriented* because of the above *geometric* interpretation, were introduced by Szymański [824] under the name of *non-uniform recursive trees*. Earlier, Prodinger and Urbanek [753] described plane-oriented recursive trees combinatorially, as labeled ordered (or plane) trees with the property that labels along any path down from the root are increasing. Such trees are also known in the literature as *heap-ordered trees* (see Chen and Ni [217], Prodinger [752], Morris, Panholzer and Prodinger [696]) or, more recently, as *scale-free trees*. So, random plane-oriented recursive trees are the simplest example of random preferential attachment graphs.

Denote by a_n the number of plane-oriented recursive trees on n vertices. This number, for $n \geq 2$ satisfies an obvious recurrence relation

$$a_{n+1} = (2n - 1)a_n.$$

Solving this equation we get that

$$a_n = 1 \cdot 3 \cdot 5 \cdots (2n-3) = (2n-3)!!.$$

This is also the number of Stirling permutations, introduced by Gessel and Stanley [432], i.e. the number of permutations of the multiset $\{1,1,2,2,3,3,\ldots,n,n\}$, with the additional property that, for each value of $1 \le i \le n$, the values lying between the two copies of i are greater than i.

There is a one-to-one correspondence between such permutations and plane-oriented recursive trees, given by Koganov [575] and, independently, by Janson [503]. To see this relationship consider a plane-oriented recursive tree on $n+1$ vertices labelled $0,1,2,\ldots,n$, where the vertex with label 0 is the root of the tree and is connected to the vertex labeled 1 only, and the edges of the tree are oriented in the direction from the root. Now, perform a *depth first search* of the tree in which we start from the root. Next we go to the leftmost child of the root, explore that branch recursively, go to the next child in order etc., until we stop at the root. Notice that every edge in such a walk is traversed twice. If every edge of the tree gets a label equal to the label of its end-vertex furthest from the root, then the depth first search encodes each tree by a string of length $2n$, where each label $1,2,\ldots,n$ appears twice. So the unique code of each tree is a unique permutation of the multiset $\{1,1,2,2,3,3,\ldots,n,n\}$ with additional property described above. Note also that the insertion of a pair $(n+1,n+1)$ into one of the $2n-1$ gaps between labels of the permutation of this multiset, corresponds to the insertion of the vertex labeled $n+1$ into a plane-oriented recursive tree on n vertices.

Let us start with exact formulas for probability distribution of the out-degree $D^+_{n,i}$ of a vertex with label i, $i = 1,2,\ldots,n$ in a random plane-oriented recursive tree. Kuba and Panholzer [611] proved the following theorem.

Theorem 15.16. *For $i = 1,2,\ldots,n$ and $r = 1,2,\ldots,n-1$,*

$$\mathbb{P}(D^+_{n,i} = r) = \sum_{k=0}^{r} \binom{r}{k} (-1)^k \frac{\Gamma(n-3/2)\Gamma(i-1/2)}{\Gamma(i-1-k/2)\Gamma(n-1/2)},$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ is the Gamma function. Moreover,

$$\mathbb{E}(D^+_{n,i}) = \frac{\binom{2i-2}{i-1}4^{n-i}}{\binom{2n-2}{n-1}} - 1 \tag{15.29}$$

For simplicity, we show below that the formula (15.29) holds for $i = 1$, i.e., the expected value of the out-degree of the root of a random plane-oriented recursive tree, and investigate its behavior as $n \to \infty$. It is then interesting to compare the latter with the asymptotic behavior of the degree of the root of a random recursive tree. Recall that for large n this is roughly $\log n$ (see Theorem 15.10).

The result below was proved by Mahmoud, Smythe and Szymański [651].

Corollary 15.17. *For $n \geq 2$ the expected value of the degree of the root of a random plane-oriented recursive tree is*

$$\mathbb{E}(D^+_{n,1}) = \frac{4^{n-1}}{\binom{2n-2}{n-1}} - 1,$$

and,

$$\mathbb{E}(D^+_{n,1}) \approx \sqrt{\pi n}.$$

Proof. Denote by

$$u_n = \frac{4^n}{\binom{2n}{n}} = \prod_{i=1}^{n} \frac{2i}{2i-1} = \frac{(2n)!!}{(2n-1)!!}.$$

Hence, in terms of u_n, we want to prove that $\mathbb{E}(D^+_{n,1}) = u_{n-1} - 1$.
It is easy to see that the claim holds for $n = 1, 2$ and that

$$\mathbb{P}(D^+_{n,1} = 1) = \prod_{i=1}^{n-1} \left(1 - \frac{2}{2i-1}\right) = \frac{1}{2n-3},$$

while, for $r > 1$ and $n \geq 1$,

$$\mathbb{P}(D^+_{n+1,1} = r) = \\ \left(1 - \frac{r+1}{2n-1}\right) \mathbb{P}(D^+_{n,1} = r) + \frac{r}{2n-1} \mathbb{P}(D^+_{n,1} = r-1).$$

Hence

$$\begin{aligned}
&\mathbb{E}(D^+_{n+1,1}) \\
&= \sum_{r=1}^{n} r \left(\frac{2n-r-2}{2n-1} \mathbb{P}(D^+_{n,1} = r) + \frac{r}{2n-1} \mathbb{P}(D^+_{n,1} = r-1)\right) = \\
&\frac{1}{2n-1} \left(\sum_{r=1}^{n-1} r(2n-r-2) \mathbb{P}(D^+_{n,1} = r) + \sum_{r=1}^{n-1} (r+1)^2 \mathbb{P}(D^+_{n,1} = r)\right) \\
&= \frac{1}{2n-1} \sum_{r=1}^{n} (2nr+1) \mathbb{P}(D^+_{n,1} = r).
\end{aligned}$$

So, we get the following recurrence relation

$$\mathbb{E}(D^+_{n+1,1}) = \frac{2n}{2n-1}\,\mathbb{E}(D^+_{n,1}) + \frac{1}{2n-1}$$

and the first part of the theorem follows by induction.

To see that the second part also holds one has to use the Stirling approximation to check that

$$u_n = \sqrt{\pi n} - 1 + \frac{3}{8}\sqrt{\pi/n} + \cdots .$$

$\square$

The next theorem, due to Kuba and Panholzer [611], summarizes the asymptotic behavior of the suitably normalized random variable $D^+_{n,i}$.

Theorem 15.18. *Let $i \geq 1$ be fixed and let $n \to \infty$. If*

(i) $i = 1$, *then*

$$n^{-1/2} D^+_{n,1} \xrightarrow{d} D_1, \text{ with density } \quad f_{D_1}(x) = (x/2)e^{-x^2/2},$$

i.e., is asymptotically Rayleigh distributed with parameter $\sigma = \sqrt{2}$,

(ii) $i \geq 2$, *then $n^{-1/2} D^+_{n,i} \xrightarrow{d} D_i$, with density*

$$f_{D_i}(x) = \frac{2i-3}{2^{2i-1}(i-2)!} \int_x^\infty (t-x)^{2i-4} e^{-t^2/4} dt.$$

Let $i = i(n) \to \infty$ as $n \to \infty$. If

(i) $i = o(n)$, *then the normalized random variable $(n/i)^{-1/2} D^+_{n,i}$ is asymptotically Gamma distributed $\gamma(\alpha, \beta)$, with parameters $\alpha = -1/2$ and $\beta = 1$,*

(ii) $i = cn,\ 0 < c < 1$, *then the random variable $D^+_{n,i}$ is asymptotically negative binomial distributed NegBinom(r, p) with parameters $r = 1$ an $p = \sqrt{c}$,*

(iii) $n - i = o(n)$, *then* $\mathbb{P}(D^+_{n,i} = 0) \to 1$, *as $n \to \infty$.*

We now turn our attention to the number of vertices of a given out-degree. The next theorem shows a characteristic feature of random graphs built by preferential attachment rule where every new vertex prefers to attach to a vertex with high degree (*rich get richer* rule). The proportion of vertices with degree r in such a random graph with n vertices grows like n/r^α, for some constant $\alpha > 0$, i.e., its distribution obeys a so called *power law*. The next result was proved by Szymański [824] (see also [651] and [826]) and it indicates such a behavior for the degrees of the vertices of a random plane-oriented recursive tree, where $\alpha = 3$.

Theorem 15.19. *Let r be fixed and denote by $X^+_{n,r}$ the number of vertices of out-degree r in a random plane-oriented recursive tree T_n. Then,*

$$\mathbb{E}X^+_{n,r} = \frac{4n}{(r+1)(r+2)(r+3)} + O\left(\frac{1}{r}\right).$$

Proof. Observe first that conditional on T_n,

$$\mathbb{E}(X^+_{n+1,r}|T_n) = X^+_{n,r} - \frac{r+1}{2n-1}X^+_{n,r} + \frac{r}{2n-1}X^+_{n,r-1} + 1_{r=0}, \tag{15.30}$$

which gives

$$\mathbb{E}X^+_{n+1,r} = \frac{2n-r-2}{2n-1}\mathbb{E}X^+_{n,r} + \frac{r}{2n-1}\mathbb{E}X^+_{n,r-1} + 1_{r=0} \tag{15.31}$$

for $r \geq 1$, ($X^+_{n,-1} = 0$).

We will show that the difference

$$a_{n,r} \stackrel{def}{=} \mathbb{E}X^+_{n,r} - \frac{4n}{(r+1)(r+2)(r+3)}.$$

is asymptotically negligible with respect to the leading term in the statement of the theorem. Substitute $a_{n,r}$ in the equation (15.31) to get that for $r \geq 1$,

$$a_{n+1,r} = \frac{2n-r-2}{2n-1}a_{n,r} + \frac{r}{2n-1}a_{n,r-1} - \frac{1}{2n-1}. \tag{15.32}$$

We want to show that $|a_{n,r}| \leq \frac{2}{\max\{r,1\}}$, for all $n \geq 1, r \geq 0$. Note that this is true for all n and $r = 0,1$, since from (15.31) it follows (inductively) that for $n \geq 2$

$$\mathbb{E}X^+_{n,0} = \frac{2n-1}{3} \text{ and so } a_{n,0} = -\frac{1}{3}.$$

For $n \geq 2$,

$$\mathbb{E}X^+_{n,1} = \frac{n}{6} - \frac{1}{12} + \frac{3}{4(2n-3)} \text{ and so } a_{n,1} = -\frac{1}{12} + \frac{3}{4(2n-3)}.$$

We proceed by induction on r. By definition

$$a_{r,r} = -\frac{4r}{(r+1)(r+2)(r+3)},$$

and so,

$$|a_{r,r}| < \frac{2}{r}.$$

We then see from (15.32) that for and $r \geq 2$ and $n \geq r$ that

$$\begin{aligned} |a_{n+1,r}| &\leq \frac{2n-r-2}{2n-1} \cdot \frac{2}{r} + \frac{r}{2n-1} \cdot \frac{2}{r-1} - \frac{1}{2n-1}. \\ &= \frac{2}{r} - \frac{2}{(2n-1)r}\left(r+1-\frac{r^2}{r-1}-\frac{r}{2}\right) \\ &\leq \frac{2}{r}, \end{aligned}$$

which completes the induction and the proof of the theorem. □

In fact much more can be proved.

Theorem 15.20. *Let $\varepsilon > 0$ and r be fixed. Then, w.h.p.*

$$(1-\varepsilon)a_r \leq \frac{X^+_{n,r}}{n} \leq (1+\varepsilon)a_r, \tag{15.33}$$

where

$$a_r = \frac{4}{(r+1)(r+2)(r+3)}.$$

Moreover,

$$\frac{\left(X^+_{n,r} - na_r\right)}{\sqrt{n}} \xrightarrow{d} Y_r, \tag{15.34}$$

as $n \to \infty$, jointly for all $r \geq 0$, where the Y_r are jointly Normally distributed with expectations $\mathbb{E} Y_r = 0$ and covariances $\sigma_{rs} = \operatorname{Cov}(Y_r, Y_s)$ given by

$$\sigma_{rs} = 2\sum_{k=0}^{r}\sum_{l=0}^{s} \frac{(-1)^{k+l}}{k+l+4}\binom{r}{k}\binom{s}{l}\left(\frac{2(k+l+4)!}{(k+3)!(l+3)!} - 1 - \frac{(k+1)(l+1)}{(k+3)(l+3)}\right).$$

Proof. For the proof of asymptotic normality of a suitably normalized random variable $X^+_{n,r}$, i.e., for the proof of statement (15.34)) see Janson [501]. We will give a short proof of the first statement (15.33), due to Bollobás, Riordan, Spencer and Tusnády [177] (see also Mori [694]).

Consider a random plane-oriented recursive tree T_n as an element of a process $(T_t)_{t=o}^{\infty}$. Fix $n \geq 1$ and $r \geq 0$ and for $0 \leq t \leq n$ define the martingale

$$Y_t = \mathbb{E}(X^+_{n,r}|T_t) \quad \text{where} \quad Y_0 = \mathbb{E}(X^+_{n,r}) \ \text{ and } \ Y_n = X^+_{n,r}.$$

One sees that the differences

$$|Y_{t+1} - Y_t| \leq 2.$$

For a proof of this, see the proof of Theorem 18.3. Applying the Hoeffding- Azuma inequality (see Theorem 27.16) we get, for any fixed r,

$$\mathbb{P}(|X^+_{n,r} - \mathbb{E}X^+_{n,r}| \geq \sqrt{n\log n}) \leq e^{-(1/8)\log n} = o(1).$$

But Theorem 15.19 shows that for any fixed r, $\mathbb{E}X^+_{n,r} \gg \sqrt{n\log n}$ and (15.33) follows. □

Similarly, as for uniform random recursive trees, Pittel [740] established the asymptotic behavior of the height of a random plane-oriented recursive tree.

Theorem 15.21. *Let h^*_n be the height of a random plane-oriented recursive tree. Then w.h.p.*

$$h^*_n \approx \frac{\log n}{2\gamma},$$

where γ is the unique solution of the equation

$$\gamma e^{\gamma+1} = 1,$$

i.e., $\gamma = 0.27846..$, so $\frac{1}{2\gamma} = 1.79556....$

Inhomogeneous recursive trees: a general model

As before, consider a tree that grows randomly in time. Each time a new vertex appears, it chooses exactly one of the existing vertices and attaches to it. This way we build a tree T_n of order n with $n+1$ vertices labeled $\{0,1,\dots,n\}$, where the vertex labeled 0 is the root. Now assume that for every $n \geq 0$ there is a probability distribution

$$P^{(n)} = (p_0, p_1, \dots, p_n), \quad \sum_{j=0}^{n} p_j = 1.$$

Suppose that T_n has been constructed for some $n \geq 1$. Given T_n we add an edge connecting one of its vertices with a new vertex labeled $n+1$ and thus forming a tree T_{n+1}. A vertex $v_n \in \{0,1,2,\dots,n\}$ is chosen to be a neighbor of the incoming vertex with probability

$$\mathbb{P}(v_n = j|T_n) = p_j, \quad \text{for} \quad j = 0,1,\dots,n.$$

Note that for the uniform random recursive tree we have

$$p_j = \frac{1}{n+1}, \qquad \text{for } 0 \leq j \leq n.$$

We say that a random recursive tree is *inhomogeneous* if the attachment rule of new vertices is determined by a non-uniform probability distribution. Most often the probability that a new vertex chooses a vertex $j \in \{0,1,\ldots,n\}$ is proportional to $w(d_n(j))$, the value of a weight function w applied to the degree $d_n(j)$ of vertex j after n-th step. Then the probability distribution $P^{(n)}$ is defined

$$p_j = \frac{w(d_n(j))}{\sum_{k=0}^{n} w(d_n(k))}.$$

Consider a special case when the weight function is linear and, for $0 \le j \le n$,

$$w(d_n(j)) = d_n(j) + \beta,\ \beta > -1, \tag{15.35}$$

so that the total weight

$$w_n = \sum_{k=0}^{n} (d_n(k) + \beta) = 2n + (n+1)\beta. \tag{15.36}$$

Obviously the model with such probability distribution is only a small generalisation of plane-oriented random recursive trees and we obtain the latter when we put $\beta = 0$ in (15.35). Inhomogeneous random recursive trees of this type are known in the literature as either *scale free random trees* or *Barabási-Albert random trees*. For obvious reasons, we will call such graphs *generalized random plane-oriented recursive trees*.

Let us focus the attention on the asymptotic behavior of the maximum degree of such random trees. We start with some useful notation and observations.

Let $X_{n,j}$ denote the weight of vertex j in a generalized plane-oriented random recursive tree, with initial values $X_{1,0} = X_{j,j} = 1 + \beta$ for $j > 0$. Let

$$c_{n,k} = \frac{\Gamma\left(n + \frac{\beta}{\beta+2}\right)}{\Gamma\left(n + \frac{\beta+k}{\beta+2}\right)},\ \ n \ge 1,\ k \ge 0,$$

be a double sequence of normalising constants. Note that

$$\frac{c_{n+1,k}}{c_{n,k}} = \frac{w_n}{w_n + k}, \tag{15.37}$$

and, for any fixed k,

$$c_{n,k} = n^{-k/(\beta+2)}(1 + O(n^{-1})).$$

Let k be a positive integer and

$$X_{n,j;k} = c_{n,k}\binom{X_{n,j} + k - 1}{k}.$$

Lemma 15.22. *Let $\mathscr{F}_n$ be the σ-field generated by the first n steps. If $n \geq \max\{1, j\}$, then $\left(X_{n,j;k}, \mathscr{F}_n\right)$ is a martingale.*

Proof. Because $X_{n+1,j} - X_{n,j} \in \{0,1\}$, we see that

$$\begin{aligned}
&\binom{X_{n+1,j}+k-1}{k} \\
&= \binom{X_{n,j}+k-1}{k} + \binom{X_{n,j}+k-1}{k-1}\binom{X_{n+1,j}-X_{n,j}}{1} \\
&= \binom{X_{n,j}+k-1}{k}\left(1+\frac{k(X_{n+1,j}-X_{n,j})}{X_{n,j}}\right).
\end{aligned}$$

Hence, noting that

$$\mathbb{P}(X_{n+1,j} - X_{n,j} = 1 | \mathscr{F}_n) = \frac{X_{n,j}}{w_n},$$

and applying (15.37)

$$\mathbb{E}(X_{n+1,j;k}|\mathscr{F}_n) = X_{n,j;k}\frac{c_{n+1,k}}{c_{n,k}}\left(1+\frac{k}{w_n}\right) = X_{n,j;k},$$

we arrive at the lemma. □

Thus, the random variable $X_{n,j;k}$, as a non-negative martingale, is bounded in L_1 and it almost surely converges to $X_j^k/k!$, where X_j is the limit of $X_{n,j;1}$. Since $X_{n,j;k} \leq cX_{n,j;2k}$, where the constant c does not depend on n, it is also bounded in L_2, which implies that it converges in L_1. Therefore we can determine all moments of the random variable X_j. Namely, for $j \geq 1$,

$$\frac{X_j^k}{k!} = \lim_{n\to\infty} \mathbb{E}X_{n,j;k} = X_{j,j;k} = c_{j,k}\binom{\beta+k}{k}. \tag{15.38}$$

Let Δ_n be the maximum degree in a generalized random plane-oriented recursive tree T_n and let, for $j \leq n$,

$$\Delta_{n,j} = \max_{0\leq i\leq j} X_{n,i;1} = \max_{0\leq i\leq j} c_{n,1}X_{n,i}.$$

Note that since $X_{n,i}$ is the weight of vertex i, i.e., its degree plus β, we find that $\Delta_{n,n} = c_{n,1}(\Delta_n + \beta)$. Define

$$\xi_j = \max_{0\leq i\leq j} X_i \quad \text{and} \quad \xi = \xi_\infty = \sup_{j\geq 0} X_j. \tag{15.39}$$

Now we are ready to prove the following result, due to Móri [695].

Theorem 15.23.

$$\mathbb{P}\left(\lim_{n\to\infty} n^{-1/(\beta+2)}\Delta_n = \xi\right) = 1.$$

The limiting random variable ξ is almost surely finite and positive and it has an absolutely continuous distribution. The convergence also holds in L_p, for all p, $1 \leq p < \infty$.

Proof. In the proof we skip the part dealing with the positivity of ξ and the absolute continuity of its distribution.

By Lemma 15.22, $\Delta_{n,n}$ is the maximum of martingales, therefore $(\Delta_{n,n}|\mathscr{F})$ is a non-negative sub-martingale, and so

$$\mathbb{E}\,\Delta_{n,n}^k \leq \sum_{j=0}^{n} \mathbb{E}\,X_{n,j;1}^k \leq \sum_{j=0}^{\infty} \mathbb{E}\,X_j^k = k!\binom{\beta+k}{k}\sum_{j=0}^{\infty} c_{j,k} < \infty,$$

if $k > \beta+2$. (Note $c_{0,k}$ is defined here as equal to $c_{1,k}$). Hence $(\Delta_{n,n}|\mathscr{F})$ is bounded in L_k, for every positive integer k, which implies both almost sure convergence and convergence in L_p, for any $p \geq 1$.

Assume that $k > \beta+2$ is fixed. Then, for $n \geq k$,

$$\mathbb{E}(\Delta_{n,n} - \Delta_{n,j})^k \leq \sum_{i=j+1}^{n} \mathbb{E}\,X_{n,i;1}.$$

Take the limit as $n \to \infty$ of both sides of the above inequality. Applying (15.39) and (15.38), we get

$$\mathbb{E}\left(\lim_{n\to\infty} n^{-1/(\beta+2)}\Delta_n - \xi_j\right)^k \leq \sum_{i=j+1}^{\infty} \mathbb{E}\,\xi_i^k = k!\binom{\beta+k}{k}\sum_{i=j+1}^{\infty} c_{j,k}.$$

The right-hand side tends to 0 as $j \to \infty$, which implies that $n^{-1/(\beta+2)}\Delta_n$ tends to ξ, as claimed. $\square$

To conclude this section, setting $\beta = 0$ in Theorem 15.23, one can obtain the asymptotic behavior of the maximum degree of a plane-oriented random recursive tree.

15.4 Exercises

(i) Use the Prüfer code to show that there is one-to-one correspondence between the family of all labeled trees with vertex set $[n]$ and the family of all ordered sequences of length $n-2$ consisting of elements of $[n]$.

(ii) Prove Theorem 15.1.

(iii) Let Δ be the maximum degree of a random labeled tree on n vertices. Use (15.1) to show that for every $\varepsilon > 0$, $\mathbb{P}(\Delta > (1+\varepsilon)\log n/\log\log n)$ tends to 0 as $n \to \infty$.

(iv) Let Δ be defined as in the previous exercise and let $t(n,k)$ be the number of labeled trees on n vertices with maximum degree at most k. Knowing that $t(n,k) < (n-2)!\left(1+1+\frac{1}{2!}+\ldots+\frac{1}{(k-1)!}\right)^n$, show that for every $\varepsilon > 0$, $\mathbb{P}(\Delta < (1-\varepsilon)\log n/\log\log n)$ tends to 0 as $n \to \infty$.

(v) Determine a one-to-one correspondence between the family of permutations on $\{2,3,\ldots,n\}$ and the family of recursive trees on the set $[n]$.

(vi) Let L_n denote the number of leaves of a random recursive tree with n vertices. Show that $\mathbb{E}L_n = n/2$ and $\operatorname{Var}L_n = n/12$.

(vii) Prove (15.20).

(viii) Show that $\Phi_{n,i}(z)$ given in Theorem 15.8 is the probability generating function of the convolution of $n-i$ independent Bernoulli random variables with success probabilities equal to $1/(i+k-1)$ for $k = 1,2,\ldots,n-i$.

(ix) Let L_n^* denotes the number of leaves of a random plane-oriented recursive tree with n vertices. Show that

$$\mathbb{E}L_n^* = \frac{2n-1}{3} \quad \text{and} \quad \operatorname{Var}L_n^* = \frac{2n(n-2)}{9(2n-3)}.$$

(x) Prove that L_n^*/n (defined above) converges in probability, to 2/3.

15.5 Notes

Labeled trees

The literature on random labeled trees and their generalizations is very extensive. For a comprehensive list of publications in this broad area we refer the reader to a recent book of Drmota [303], to a chapter of Bollobás's book [155] on random graphs, as well as to the book by Kolchin [583]. For a review of some classical results, including the most important contributions, forming the foundation of the research on random trees, mainly due to Meir and Moon (see, for example : [676], [677]and [679]), one may also consult a survey by Karoński [547].

Recursive trees

Recursive trees have been introduced as probability models for system generation (Na and Rapoport [703]), spread of infection (Meir and Moon [678]), pyramid schemes (Gastwirth [426]) and stemma construction in philology (Najock and Heyde [707]). Most likely, the first place that such trees were introduced in the literature, is the paper by Tapia and Myers [829], presented there under the name "concave node-weighted trees". Systematic studies of random recursive trees were initiated by Meir and Moon ([678] and [693]) who investigated distances between vertices as well as the process of cutting down such random trees. Observe that there is a bijection between families of recursive trees and binary search trees, and this has opened many interesting directions of research, as shown in a survey by Mahmoud and Smythe [650] and the book by Mahmoud [648].

Early papers on random recursive trees (see, for example, [703], [426] and [302]) were focused on the distribution of the degree of a given vertex and of the number of vertices of a given degree. Later, these studies were extended to the distribution of the number of vertices at each level, which is referred to as the *profile*. Recall, that in a rooted tree, a *level* (*strata*) consists of all those vertices that are at the same distance from the root.

The profile of a random recursive tree is analysed in many papers. For example, Drmota and Hwang [304] derive asymptotic approximations to the correlation coefficients of two level sizes in random recursive trees and binary search trees. These coefficients undergo sharp sign-changes when one level is fixed and the other is varying. They also propose a new means of deriving an asymptotic estimate for the expected *width*, which is the number of nodes at the most abundant level.

Devroye and Hwang [288] propose a new, direct, correlation-free approach based on central moments of profiles to the asymptotics of width in a class of random trees of logarithmic height. This class includes random recursive trees.

Fuchs, Hwang, Neininger [420] prove convergence in distribution for the profile, normalized by its mean, of random recursive trees when the limit ratio α of the level and the logarithm of tree size lies in $[0,e)$. Convergence of all moments is shown to hold only for $\alpha \in (0,1)$ (with only convergence of finite moments when $\alpha \in (1,e)$).

van der Hofstadt, Hooghiemstra and Van Mieghem [483] study the covariance structure of the number of nodes k and l steps away from the root in random recursive trees and give an analytic expression valid for all k,l and tree sizes n.

For an arbitrary positive integer $i \leq i_n \leq n-1$, a function of n, Su, Liu and Feng [821] demonstrate the distance between nodes i and n in random recursive trees T_n, is asymptotically normal as $n \to \infty$ by using the classical limit theory method.

Holmgren and Janson [488] proved limit theorems for the sums of functions of sub-trees of binary search trees and random recursive trees. In particular, they give new simple proofs of the fact that the number of fringe trees of size $k = k_n$ in a binary search tree and the random recursive tree (of total size n) asymptotically has a Poisson distribution if $k \to \infty$, and that the distribution is asymptotically normal for $k = o(\sqrt{n})$. Recall that a *fringe tree* is a sub-tree consisting of some vertex of a tree and all its descendants (see Aldous [18]). For other results on that topic see Devroye and Janson [289].

Feng, Mahmoud and Panholzer [351] study the variety of sub-trees lying on the fringe of recursive trees and binary search trees by analysing the distributional behavior of $X_{n,k}$, which counts the number of sub-trees of size k in a random tree of size n, with $k = k(n)$. Using analytic methods, they characterise for both tree families the phase change behavior of $X_{n,k}$.

One should also notice interesting applications of random recursive trees. For example, Mehrabian [675] presents a new technique for proving logarithmic upper bounds for diameters of evolving random graph models, which is based on defining a coupling between random graphs and variants of random recursive trees. Goldschmidt and Martin [443] describe a representation of the Bolthausen-Sznitman coalescent in terms of the cutting of random recursive trees.

Bergeron, Flajolet, Salvy [102] have defined and studied a wide class of *random increasing trees*. A a tree with vertices labeled $\{1,2,\ldots,n\}$ is *increasing* if the sequence of labels along any branch starting at the root is increasing. Obviously, recursive trees and binary search trees (as well as the general class of inhomogeneous trees, including plane-oriented trees) are increasing. Such a general model, which has been intensively studied, yields many important results for random trees discussed in this chapter. Here we will restrict ourselves to pointing out just a few papers dealing with random increasing trees authored by Dobrow and Smythe [301], Kuba and Panholzer [611] and Panholzer and Prodinger [725], as well as with their generalisations, i.e., *random increasing k-trees*, published by Zhang, Rong, and Comellas [865], Panholzer and Seitz [726] and Darrasse, Hwang and Soria [277],

Inhomogeneous recursive trees

Plane-oriented recursive trees

As we already mentioned in Section 15.3, Prodinger and Urbanek [753], and, independently, Szymański [824] introduced the concept of plane-oriented random trees (more precisely, this notion was introduced in an unpublished paper by Dondajewski and Szymański [302]), and studied the vertex degrees of such random trees. Mahmoud, Smythe and Szymański [651], using Pólya urn models, investigated the exact and limiting distributions of the size and the number of leaves in the branches of the tree (see [503] for a follow up). Lu and Feng [629] considered the strong convergence of the number of vertices of given degree as well as of the degree of a fixed vertex (see also [650]). In Janson's [501] paper, the distribution of vertex degrees in random recursive trees and random plane recursive trees are shown to be asymptotically normal. Brightwell and Luczak [194] investigate the number $D_{n,k}$ of vertices of each degree k at each time n, focusing particularly on the case where $k = k(n)$ is a growing function of n. They show that $D_{n,k}$ is concentrated around its mean, which is approximately $4n/k^3$, for all $k \leq (n \log n)^{-1/3}$, which is best possible up to a logarithmic factor.

Hwang [490] derives several limit results for the profile of random plane-oriented recursive trees. These include the limit distribution of the normalized profile, asymptotic bimodality of the variance, asymptotic approximation to the expected width and the correlation coefficients of two level sizes.

Fuchs [419] outlines how to derive limit theorems for the number of sub-trees of size k on the fringe of random plane-oriented recursive trees.

Finally, Janson, Kuba and Panholzer [505] consider generalized Stirling permutations and relate them with certain families of generalized plane recursive trees.

Generalized recursive trees

Móri [694] proves the strong law of large numbers and central limit theorem for the number of vertices of low degree in a generalized random plane-oriented recursive tree. Szymański [826] gives the rate of concentration of the number of vertices with given degree in such trees. Móri [695] studies maximum degree of a scale-free trees. Zs. Katona [560] shows that the degree distribution is the same on every sufficiently high level of the tree and in [559] investigates the width of scale-free trees.

Rudas, Toth, Valko [780], using results from the theory of general branching processes, give the asymptotic degree distribution for a wide range of weight functions. Backhausz and Móri [59] present sufficient conditions for the almost sure existence of an asymptotic degree distribution constrained to the set of selected vertices and describe that distribution.

Bertoin, Bravo [103] consider Bernoulli bond percolation on a large scale-free tree in the super-critical regime, i.e., when there exists a giant cluster with high probability. They obtain a weak limit theorem for the sizes of the next largest clusters, extending a result in Bertoin [105] for large random recursive trees.

Devroye, Fawzi, Fraiman [286] study depth properties of a general class of random recursive trees called *attachment random recursive trees*. They prove that the height of such tree is asymptotically given by $\alpha_{max} \log n$ where α_{max} is a constant. This gives a new elementary proof for the height of uniform random recursive trees that does not use branching random walk. For further generalisations of random recursive trees see Mahmoud [649].

Chapter 16

Mappings

In the evolution of the random graph $\mathbb{G}_{n,p}$, during its sub-critical phase, tree components and components with exactly one cycle, i.e. graphs with the same number of vertices and edges, are w.h.p. the only elements of its structure. Similarly, they are the only graphs outside the giant component after the phase transition, until the random graph becomes connected w.h.p. In the previous chapter we studied the properties of random trees. Now we focus our attention on random mappings of a finite set into itself. Such mappings can be represented as digraphs with the same number of vertices and edges. So the study of their "average" properties may help us to better understand the typical structure of classical random graphs. We start the chapter with a short look at the basic properties of random permutations (one-to-one mappings) and then continue to the general theory of random mappings.

16.1 Permutations

Let f be chosen uniformly at random from the set of all $n!$ permutations on the set $[n]$, i.e., from the set of all one-to-one functions $[n] \to [n]$. In this section we will concentrate our attention on the properties of a functional digraph representing a random permutation.

Let D_f be the functional digraph $([n], (i, f(i)))$. The digraph D_f consists of vertex disjoint cycles of any length $1, 2, \ldots, n$. Loops represent fixed points, see Figure 16.1.

Let $X_{n,t}$ be the number of cycles of length t, $t = 1, 2, \ldots, n$ in the digraph D_f. Thus $X_{n,1}$ counts the number of fixed points of a random permutation. One can easily check that

$$\mathbb{P}(X_{n,t} = k) = \frac{1}{k!t^k} \sum_{i=0}^{\lfloor n/t \rfloor} \frac{(-1)^i}{t^i i!} \to \frac{e^{-1/t}}{t^k k!} \text{ as } n \to \infty, \tag{16.1}$$

x	1	2	3	4	5	6	7	8	9	10
f(x)	3	10	1	4	2	6	8	9	7	5

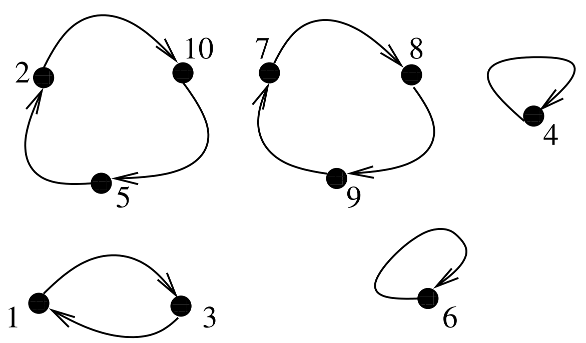


Figure 16.1: A permutation digraph example

for $k = 0, 1, 2, \ldots, n$. Indeed, convergence in (16.1) follows directly from Lemma 26.10 and the fact that

$$B_i = \mathbb{E}\binom{X_{n,t}}{i} = \frac{1}{n!} \cdot \frac{n!}{(t!)^i (n-ti)!} \frac{((t-1)!)^i (n-ti)!}{i!} = \frac{1}{t^i i!}.$$

This means that $X_{n,t}$ converges in distribution to a random variable with Poisson distribution with mean $1/t$.

Moreover, direct computation gives

$$\begin{aligned}
&\mathbb{P}(X_{n,1} = j_1, X_{n,2} = j_2, \ldots, X_{n,n} = j_n) \\
&= \frac{1}{n!} \frac{n!}{\prod_{t=1}^{n} j_t! (t!)^{j_t}} \prod_{t=1}^{n} ((t-1)!)^{j_t} \\
&= \prod_{t=1}^{n} \left(\frac{1}{t}\right)^{j_t} \frac{1}{j_t!},
\end{aligned}$$

for non-negative integers $j_1, j_2, \ldots, j_n$ satisfying $\sum_{t=1}^{n} t j_t = n$.

Hence, asymptotically, the random variables $X_{n,t}$ have independent Poisson distributions with expectations $1/t$, respectively (see Goncharov [447] and Kolchin [580]).

Next, consider the random variable $X_n = \sum_{j=1}^{n} X_{n,j}$ counting the total number of cycles in a functional digraph D_f of a random permutation. It is not difficult to show that X_n has the following probability distribution.

Theorem 16.1. *For $k = 1, 2, \ldots, n$,*

$$\mathbb{P}(X_n = k) = \frac{|s(n,k)|}{n!},$$

where the $s(n,k)$ are Stirling numbers of the first kind, i.e., numbers satisfying the following relation:

$$x(x-1)\cdots(x-n+1) = \sum_{k=0}^{n} s(n,k)x^k \ .$$

Moreover,

$$\mathbb{E}X_n = H_n = \sum_{j=1}^{n} \frac{1}{j}, \qquad \mathrm{Var}\, X_n = H_n - \sum_{j=1}^{n} \frac{1}{j^2} \ .$$

Proof. Denote by $c(n,k)$ the number of digraphs D_f (permutations) on n vertices and with exactly k cycles. Consider a vertex n in D_f. It either has a loop (belongs to a unit cycle) or it doesn't. If it does, then D_f is composed of a loop in n and a cyclic digraph (permutation) on $n-1$ vertices with exactly $k-1$ cycles. and there are $c(n-1,k-1)$ such digraphs (permutations). Otherwise, the vertex n can be thought as dividing (lying on) one of the $n-1$ arcs which belongs to cyclic digraph on $n-1$ vertices with k cycles and there are $(n-1)c(n-1,k)$ such permutations (digraphs) of the set $[n]$. Hence

$$c(n,k) = c(n-1,k-1) + (n-1)c(n-1,k) \ .$$

Now, multiplying both sides by x^k, dividing by $n!$ and summing up over all k, we get

$$G_n(x) = (x+n-1)G_{n-1}(x).$$

where $G_n(x)$ is the probability generating function of X_n. But $G_1(x) = x$, so

$$G_n(x) = \frac{x(x+1)\cdots(x+n-1)}{n!},$$

and the first part of the theorem follows. Note that

$$G_n(x) = \binom{x+n-1}{n} = \frac{\Gamma(x+n)}{\Gamma(x)\Gamma(n+1)},$$

where Γ is the Gamma function.

The results for the expectation and variance of X_n can be obtained by calculating the first two derivatives of $G_n(x)$ and evaluating them at $x = 1$ in a standard way but one can also show them using only the fact that the cycles of functional digraphs must be disjoint. Notice, for example, that

$$\begin{aligned}\mathbb{E}X_n &= \sum_{\emptyset \neq S \subset [n]} \mathbb{P}(S \text{ induces a cycle}) \\ &= \sum_{k=1}^{n} \binom{n}{k} \frac{(k-1)!(n-k)!}{n!} = H_n.\end{aligned}$$

Similarly one can derive the second factorial moment of X_n counting ordered pairs of cycles (see Exercises 16.3.2 and 16.3.3) which implies the formula for the variance. □

Goncharov [447] proved a Central Limit Theorem for the number X_n of cycles.

Theorem 16.2.

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_n - \log n}{\sqrt{\log n}} \leq x\right) = \int_{-\infty}^{x} e^{-t^2/2} dt,$$

i.e., the standardized random variable X_n converges in distribution to the standard Normal random variable.

Another numerical characteristic of a digraph D_f is the length L_n of its longest cycle. Shepp and Lloyd [803] established the asymptotic behavior of the expected value of L_n.

Theorem 16.3.

$$\lim_{n\to\infty} \frac{\mathbb{E}L_n}{n} = \int_0^{\infty} \exp\left\{-x - \int_x^{\infty} \frac{1}{y} e^{-y} dy\right\} dx = 0.62432965....$$

16.2 Mappings

Let f be chosen uniformly at random from the set of all n^n mappings from $[n] \to [n]$. Let D_f be the functional digraph $([n], (i, f(i)))$ and let G_f be the graph obtained from D_f by ignoring orientation. In general, D_f has unicyclic components only, where each component consists of a directed cycle C with trees rooted at vertices of C, see the Figure 16.2.

Therefore the study of functional digraphs is based on results for permutations of the set of cyclical vertices (these lying on cycles) and results for forests consisting of trees rooted at these cyclical vertices (we allow also trivial one vertex trees). For example, to show our first result on the connectivity of G_f we will need the following enumerative result for the forests.

x	1	2	3	4	5	6	7	8	9	10
f(x)	3	10	5	4	2	5	8	9	7	5

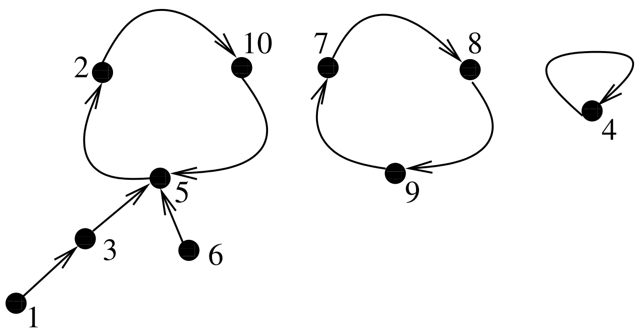


Figure 16.2: A mapping digraph example

Lemma 16.4. *Let $T(n,k)$ denote the number of forests with vertex set $[n]$, consisting of k trees rooted at the vertices $1,2,\ldots,k$. Then,*

$$T(n,k) = kn^{n-k-1}.$$

Proof. Observe first that by (15.2) there are $\binom{n-1}{k-1}n^{n-k}$ trees with $n+1$ labelled vertices in which the degree of a vertex $n+1$ is equal to k. Hence there are

$$\binom{n-1}{k-1}n^{n-k}\Big/\binom{n}{k} = kn^{n-k-1}$$

trees with $n+1$ labeled vertices in which the set of neighbors of the vertex $n+1$ is exactly $[k]$. An obvious bijection (obtained by removing the vertex $n+1$ from the tree) between such trees and the considered forests leads directly to the lemma. □

Theorem 16.5.

$$\mathbb{P}(G_f \textit{ is connected }) = \frac{1}{n}\sum_{k=1}^{n}\frac{(n)_k}{n^k} \approx \sqrt{\frac{\pi}{2n}}.$$

Proof. If G_f is connected then there is a cycle with k vertices say such that after removing the cycle we have a forest consisting of k trees rooted at the vertices of the cycle. Hence,

$$\begin{aligned}\mathbb{P}(G_f \text{ is connected }) &= n^{-n}\sum_{k=1}^{n}\binom{n}{k}(k-1)!\;T(n,k)\\ &= \frac{1}{n}\sum_{k=1}^{n}\frac{(n)_k}{n^k} = \frac{1}{n}\sum_{k=1}^{n}\prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right)\\ &= \frac{1}{n}\sum_{k=1}^{n}u_k.\end{aligned}$$

If $k \geq n^{3/5}$, then

$$u_k \leq \exp\left\{-\frac{k(k-1)}{2n}\right\} \leq \exp\left\{-\frac{1}{3}n^{1/6}\right\},$$

while, if $k < n^{3/5}$,

$$u_k = \exp\left\{-\frac{k^2}{2n} + O\left(\frac{k^3}{n^2}\right)\right\}.$$

So

$$\begin{aligned}\mathbb{P}(G_f \text{ is connected }) &= \frac{1+o(1)}{n}\sum_{k=1}^{n^{3/5}} e^{-k^2/2n} + O\left(ne^{-n^{1/6}/3}\right)\\ &= \frac{1+o(1)}{n}\int_0^\infty e^{-x^2/2n}dx + O\left(ne^{-n^{1/6}/3}\right)\\ &= \frac{1+o(1)}{\sqrt{n}}\int_0^\infty e^{-y^2/2}dy + O\left(ne^{-n^{1/6}/3}\right)\\ &\approx \sqrt{\frac{\pi}{2n}}.\end{aligned}$$

□

Let Z_k denote the number of cycles of length k in a random mapping. Then

$$\mathbb{E}Z_k = \binom{n}{k}(k-1)!\;n^{-k} = \frac{1}{k}\prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right) = \frac{u_k}{k}.$$

If $Z = Z_1 + Z_2 + \cdots + Z_n$, then

$$\mathbb{E}Z = \sum_{k=1}^{n}\frac{u_k}{k} \approx \int_{x=1}^{\infty}\frac{1}{x}e^{-x^2/2n}dx = \frac{1}{2}\int_{s=1}^{\infty}\frac{1}{s}e^{-s/2n}ds \approx \frac{1}{2}\log n.$$

(To estimate the integral $\frac{1}{2}\int_{s=1}^{\infty}\frac{1}{s}e^{-s/2n}$ we break it into $I_1+I_2+I_3$ where $I_1 = \int_{s=1}^{n/\omega}\cdots ds \approx \log n$, $\omega = \log n$, $I_2 = \int_{s=n/\omega}^{\omega n}\cdots ds \le \log\left(\frac{\omega n}{n/\omega}\right) = o(\log n)$ and $I_3 = \int_{s=\omega n}^{\infty}\cdots ds = o(1)$.)

Moreover the expected number of vertices of cycles in a random mapping is equal to

$$\mathbb{E}\left(\sum_{k=1}^{n} kZ_k\right) = \sum_{k=1}^{n} u_k \approx \sqrt{\frac{\pi n}{2}}.$$

□

Note that the functional digraph of a random mapping can be interpreted as a representation of a process in which vertex $i \in [n]$ chooses its image independently with probability $1/n$. So, it is natural to consider a general model of a random mapping $\hat{f} : [n] \to [n]$ where, independently for all $i \in [n]$,

$$\mathbb{P}\left(\hat{f}(i) = j\right) = p_j,\ j = 1,2,\ldots,n, \tag{16.2}$$

and

$$p_1 + p_2 + \ldots + p_n = 1.$$

This model was introduced (in a slightly more general form) independently by Burtin [206] and Ross [774]. We will first prove a generalisation of Theorem 16.5.

Theorem 16.6.

$$\mathbb{P}(G_{\hat{f}} \text{ is connected }) =$$
$$= \sum_i p_i^2 \left(1 + \sum_{j\ne i} p_j + \sum_{j\ne i}\sum_{k\ne i,j} p_j p_k + \sum_{j\ne i}\sum_{k\ne i,j}\sum_{l\ne i,j,k} p_j p_k p_l + \cdots\right).$$

To prove this theorem we use the powerful "Burtin–Ross Lemma". The short and elegant proof of this lemma given here is due to Jaworski [517] (His general approach can be applied to study other characteristics of a random mappings, not only their connectedness).

Lemma 16.7 (Burtin–Ross Lemma)**.** *Let $\hat{f}$ be a generalized random mapping defined above and let $G_{\hat{f}}[U]$ be the subgraph of $G_{\hat{f}}$ induced by $U \subset [n]$. Then*

$$\mathbb{P}(G_{\hat{f}}[U] \text{ does not contain a cycle}) = \sum_{k\in[n]\setminus U} p_k.$$

Proof. The proof is by induction on $r = |U|$. For $r = 0$ and $r = 1$ it is obvious. Assume that the result holds for all values less than r, $r \geq 2$. Let $\emptyset \neq S \subset U$ and denote by $\mathscr{A}$ the event that $G_{\hat{f}}[S]$ is the union of disjoint cycles and by $\mathscr{B}$ the event that $G_{\hat{f}}[U \setminus S]$ does not contain a cycle . Notice that events $\mathscr{A}$ and $\mathscr{B}$ are independent, since the first one depends on choices of vertices from S, only, while the second depends on choices of vertices from $U \setminus S$. Hence

$$\mathbb{P}(G_{\hat{f}}[U] \text{ contains a cycle }) = \sum_{\emptyset \neq S \subset U} \mathbb{P}(\mathscr{A})\,\mathbb{P}(\mathscr{B}).$$

But if $\mathscr{A}$ holds then $\hat{f}$ restricted to S defines a permutation on S. So,

$$\mathbb{P}(\mathscr{A}) = |S|! \prod_{j \in S} p_j.$$

Since $|U \setminus S| < r$, by the induction assumption we obtain

$$\begin{aligned}
&\mathbb{P}(G_{\hat{f}}[U] \text{ contains a cycle }) = \\
&= \sum_{\emptyset \neq S \subset U} |S|! \prod_{j \in S} p_j \sum_{k \in [n] \setminus (U \setminus S)} p_k \\
&= \sum_{\emptyset \neq S \subset U} |S|! \prod_{j \in S} p_j \left(1 - \sum_{k \in (U \setminus S)} p_k\right) \\
&= \sum_{S \subset U,\ |S| \geq 1} |S|! \prod_{k \in S} p_k - \sum_{S \subset U,\ |S| \geq 2} |S|! \prod_{k \in S} p_k \\
&= \sum_{k \in U} p_k,
\end{aligned}$$

completing the induction. $\square$

Before we prove Theorem 16.6 we will point out that Lemma 16.4 can be immediately derived from the above result. To see this, in Lemma 16.7 choose $p_j = 1/n$, for $j = 1, 2, \cdots n$, and U such that $|U| = r = n - k$. Then, on one hand,

$$\mathbb{P}(G_f[U] \text{ does not contain a cycle}) = \sum_{i \in [n] \setminus U} \frac{1}{n} = \frac{k}{n}.$$

On the other hand,

$$\mathbb{P}(G_f[U] \text{ does not contain a cycle}) = \frac{T(n,k)}{n^{n-k}},$$

where $T(n,k)$ is the number of forests on $[n]$ with k trees rooted in vertices from the set $[n] \setminus U$. Comparing both sides we immediately get the result of Lemma 16.4, i.e., that

$$T(n,k) = kn^{n-k-1}.$$

Proof (of Theorem 16.6*).* Notice that $G_{\hat{f}}$ is connected if and only if there is a subset $U \subseteq [n]$ such that U spans a single cycle while there is no cycle on $[n] \setminus U$. Moreover, the events "$U \subseteq [n]$ spans a cycle" and "there is no cycle on $[n] \setminus U$" are independent. Hence, by Lemma 16.7,

$$\begin{aligned}
&Pr(G_{\hat{f}} \text{ is connected}) = \\
&= \sum_{\emptyset \neq U \subseteq [n]} \mathbb{P}(U \subset [n] \text{ spans a cycle}) \mathbb{P}(\text{there is no cycle on } [n] \setminus U) \\
&= \sum_{\emptyset \neq U \subset [n]} (|U|-1)! \prod_{j \in U} p_j \sum_{k \in U} p_k \qquad (16.3) \\
&= \sum_i p_i^2 \left(1 + \sum_{j \neq i} p_j + \sum_{j \neq i} \sum_{k \neq i,j} p_j p_k + \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} p_j p_k p_l + \cdots \right).
\end{aligned}$$

□

Using the same reasoning as in the above proof, one can show the following result due to Jaworski [517].

Theorem 16.8. *Let X be the number of components in $G_{\hat{f}}$ and Y be the number of its cyclic vertices (vertices belonging to a cycle). Then for $k = 1, 2, \ldots, n$,*

$$\mathbb{P}(X = k) = \sum_{\substack{U \subset [n] \\ |U| \geq k}} \prod_{j \in U} p_j |s(|U|, k)| - \sum_{\substack{U \subset [n] \\ |U| \geq k+1}} \prod_{j \in U} p_j |s(|U|-1, k)| |U|,$$

where $s(\cdot, \cdot)$ is the Stirling number of the first kind. On the other hand,

$$\mathbb{P}(Y = k) = k! \sum_{\substack{U \subset [n] \\ |U| = k}} \prod_{j \in U} p_j - (k+1)! \sum_{\substack{U \subset [n] \\ |U| = k+1}} \prod_{j \in U} p_j.$$

The Burtin–Ross Lemma has another formulation which we present below.

Lemma 16.9 (Burtin-Ross Lemma - the second version)**.** *Let $\hat{g} : [n] \to [n] \cup \{0\}$ be a random mapping from the set $[n]$ to the set $[n] \cup \{0\}$, where, independently for all $i \in [n]$,*

$$\mathbb{P}\left(\hat{g}(i) = j\right) = q_j, \; j = 0, 1, 2, \ldots, n,$$

and

$$q_0 + q_1 + q_2 + \ldots + q_n = 1.$$

Let $D_{\hat{g}}$ be the random directed graph on the vertex set $[n] \cup \{0\}$, generated by the mapping $\hat{g}$ and let $G_{\hat{g}}$ denote its underlying simple graph. Then

$$\mathbb{P}(G_{\hat{g}} \text{ is connected }) = q_0.$$

Notice that the event that $G_{\hat{g}}$ is connected is equivalent to the event that $D_{\hat{g}}$ is a (directed) tree, rooted at vertex $\{0\}$, i.e., there are no cycles in $G_{\hat{g}}[[n]]$.

We will use this result and Lemma 16.9 to prove the next theorem (for more general results, see [518]).

Theorem 16.10. *Let $D_{\hat{f}}$ be the functional digraph of a mapping $\hat{f}$ defined in* (16.2) *and let Z_R be the number of predecessors of a set $R \subset [n]$, $|R| = r$, $r \geq 1$, of vertices of $D_{\hat{f}}$, i.e.,*

$$Z_R = |\{j \in [n] : \textit{for some non-negative integer } k,\ \hat{f}^{(k)}(j) \in R\}|,$$

where $\hat{f}^{(0)}(j) = j$ and for $k \geq 1$, $\hat{f}^{(k)}(j) = \hat{f}(\hat{f}^{(k-1)}(j))$.
Then, for $k = 0, 1, 2, \ldots, n - r$,

$$\mathbb{P}(Z_R = k + r) = \Sigma_R \sum_{\substack{U \subset [n] \setminus R \\ |U| = k}} (\Sigma_{U \cup R})^{k-1} (1 - \Sigma_{U \cup R})^{n-k},$$

where for $A \subseteq [n]$, $\Sigma_A = \sum_{j \in A} p_j$.

Proof. The distribution of Z_R follows immediately from the next observation and the application of Lemma 16.9. Denote by $\mathscr{A}$ the event that there is a forest spanned on the set $W = U \cup R$, where $U \subset [n] \setminus R$, composed of r (directed) trees rooted at vertices of R. Then

$$\mathbb{P}(Z_R = k + r) = \sum_{\substack{U \subset [n] \setminus R \\ |U| = k}} \mathbb{P}(\mathscr{A} | \mathscr{B} \cap \mathscr{C})\, \mathbb{P}(\mathscr{B})\, \mathbb{P}(\mathscr{C}), \tag{16.4}$$

where $\mathscr{B}$ is the event that all edges that begin in U end in W, while $\mathscr{C}$ denotes the event that all edges that begin in $[n] \setminus W$ end in $[n] \setminus W$. Now notice that

$$\mathbb{P}(\mathscr{B}) = (\Sigma_W)^k, \text{ while } \mathbb{P}(\mathscr{C}) = (1 - \Sigma_W)^{n-k}.$$

Furthermore,

$$\mathbb{P}(\mathscr{A} | \mathscr{B} \cap C) = \mathbb{P}(G_{\hat{g}} \text{ is connected }),$$

where $\hat{g} : U \to U \cup \{0\}$, where $\{0\}$ stands for the set R collapsed to a single vertex, is such that for all $u \in U$ independently,

$$q_j = \mathbb{P}(\hat{g}(u) = j) = \frac{p_j}{\Sigma_W}, \text{ for } j \in U, \text{ while } q_0 = \frac{\Sigma_R}{\Sigma_W}.$$

So, applying Lemma 16.9, we arrive at the thesis. □

We will finish this section by stating the central limit theorem for the number of components of G_f, where f is a uniform random mapping $f : [n] \to [n]$ (see Stepanov [819]). It is an analogous result to Theorem 16.2 for random permutations.

Theorem 16.11.

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_n - \frac{1}{2}\log n}{\sqrt{\frac{1}{2}\log n}} \le x\right) = \int_{-\infty}^{x} e^{-t^2/2} dt,$$

the standardized random variable X_n converges in distribution to the standard Normal random variable.

16.3 Exercises

16.3.1 Prove directly that if $X_{n,t}$ is the number of cycles of length t in a random permutation then $\mathbb{E} X_{n,t} = 1/t$.

16.3.2 Find the expectation and the variance of the number X_n of cycles in a random permutation using fact that the rth derivative of the gamma function equals $\frac{d^r}{(dx)^r}\Gamma(x) = \int_0^\infty (\log t)^r t^{x-1} e^{-t} dt$,

16.3.3 Determine the variance of the number X_n of cycles in a random permutation (start with computation of the second factorial moment of X_n, counting ordered pairs of cycles).

16.3.4 Find the probability distribution for the length of a typical cycle in a random permutation, i.e., the cycle that contains a given vertex (say vertex 1). Determine the expectation and variance of this characteristic.

16.3.5 Find the probability distribution of the number of components in a functional digraph D_f of a uniform random mapping $f : [n] \to [n]$.

16.3.6 Show that the length of the cycle containing item 1 in a random permutation is uniformly distributed in $[n]$.

16.3.7 Show that if X denotes the number of cycles in a random permutation of $[n]$ then $\mathbb{P}(X \ge t) \le \mathbb{P}(\mathrm{Bin}(t, 1/2) \le \lceil \log_2 n \rceil)$. Deduce that for every constant $L > 0$, there exists a constant $K > 0$, such that $\mathbb{P}(X \ge K \log n) \le n^{-L}$.

16.3.8 Now let X denote the number of cycles in the digraph D_f of a random mapping $f:[n]\to[n]$. Show that for every constant $L>0$, there exists a constant $K>0$, such that $\mathbb{P}(X\geq K\log n)\leq n^{-L}$.

16.3.9 Determine the expectation and variance of the number of components in a functional digraph $D_{\hat{f}}$ of a generalized random mapping $\hat{f}$ (see Theorem 16.8)

16.3.10 Find the expectation and variance of the number of cyclic vertices in a functional digraph $D_{\hat{f}}$ of a generalized random mapping $\hat{f}$ (see Theorem 16.8).

16.3.11 Prove Theorem 16.8.

16.3.12 Show that Lemmas 16.7 and 16.9 are equivalent.

16.3.13 Prove the Burtin-Ross Lemma for a bipartite random mapping, i.e. a mapping with bipartition $([n],[m])$, where each vertex $i\in[n]$ chooses its unique image in $[m]$ independently with probability $1/m$, and, similarly, each vertex $j\in[m]$ selects its image in $[n]$ with probability $1/n$.

16.3.14 Consider an evolutionary model of a random mapping (see [519],[520]), i.e., a mapping $\hat{f}_q[n]\to[n]$, such that for $i,j\in[n]$, $\mathbb{P}(\hat{f}_q(i)=j)=q$ if $i=j$ while, $\mathbb{P}(\hat{f}_q(i)=j)=(1-q)/(n-1)$ if $i\neq j$, where $0\leq q\leq 1$. Find the probability that $\hat{f}_q$ is connected.

16.3.15 Show that there is one-to-one correspondence between the family of n^n mappings $f:[n]\to[n]$ and the family of all doubly-rooted trees on the vertex set $[n]$ (Joyal bijection)

16.4 Notes

Permutations

Systematic studies of the properties of random permutations of n objects were initiated by Goncharov in [446] and [447]. Golomb [444] showed that the expected length of the longest cycle of D_f, divided by n is monotone decreasing and gave a numerical value for the limit, while Shepp and Lloyd in [803] found the closed form for this limit (see Theorem 16.3). They also gave the corresponding result for kth moment of the rth longest cycle, for $k,r=1,2,\ldots$ and showed the limiting distribution for the length of the rth longest cycle.

Kingman [568] and, independently, Vershik and Schmidt [839], proved that for a random permutation of n objects, as $n\to\infty$, the process giving the proportion of elements in the longest cycle, the second longest cycle, and so on, converges in distribution to the Poisson-Dirichlet process with parameter 1 (for further results in this direction see Arratia, Barbour and Tavaré [54]). Arratia and Tavaré [55] provide explicit bounds on the total variation distance between the process which counts the sizes of cycles in a random permutations and a process of independent Poisson random variables.

For other results, not necessarily of a "graphical" nature, such as, for example, the order of a random permutation, the number of derangements, or the number of monotone sub-sequences, we refer the reader to the respective sections of books by Feller [350], Bollobás [156] and Sachkov [788] or, in the case of monotone sub-sequences, to a recent monograph by Romik [773].

Mappings

Uniform random mappings were introduced in the mid 1950's by Rubin and Sitgraves [775], Katz [561] and by Folkert [369]. More recently, much attention has been focused on their usefulness as a model for epidemic processes, see for example the papers of Gertsbakh [431], Ball, Mollison and Scalia-Tomba [68], Berg [100], Mutafchiev [702], Pittel [737] and Jaworski [520]. The component structure of a random functional digraph D_f has been studied by Aldous [16]. He has shown, that the joint distribution of the normalized order statistics for the component sizes of D_f converges to the Poisson-Dirichlet distribution with parameter $1/2$. For more results on uniform random mappings we refer the reader to Kolchin's monograph [582], or a chapter of Bollobás' [156].

The general model of a random mapping $\hat{f}$, introduced by Burtin [206] and Ross [774], has been intensively studied by many authors. The crucial Burtin-Ross Lemma (see Lemmas: 16.7 and 16.9) has many alternative proofs (see [45]) but the most useful seems to be the one used in this chapter, due to Jaworski [517]. His approach can also be applied to derive the distribution of many other characteristics of a random digraph D_f, as well as it can be used to prove generalisations of the Burtin-Ross Lemma for models of random mappings with independent choices of images. (For an extensive review of results in that direction see [518]). Aldous, Miermont, Pitman ([22],[23]) study the asymptotic structure of $D_{\hat{f}}$ using an ingenious coding of the random mapping $\hat{f}$ as a stochastic process on the interval $[0,1]$ (see also the related work of Pitman [735], exploring the relationship between random mappings and random forests).

Hansen and Jaworski (see [466], [467]) introduce a random mapping $f^D : [n] \to [n]$ with an in-degree sequence, which is a collection of exchangeable random variables $(D_1, D_2, \ldots, D_n)$. In particular, they study predecessors and successors of a given set of vertices, and apply their results to random mappings with preferential and anti-preferential attachment.

Chapter 17

k-out

Several interesting graph properties require that the minimum degree of a graph be at least a certain amount. E.g. having a Hamilton cycle requires that the minimum degree is at least two. In Chapter 6 we saw that $\mathbb{G}_{n,m}$ being Hamiltonian and having minimum degree at least two happen at the same time w.h.p. One is therefore interested in models of a random graph which guarantee a certain minimum degree. We have already seen d-regular graphs in Chapter 9. In this chapter we consider another simple and quite natural model $\mathbb{G}_{k-out}$ that generalises random mappings. It seems to have first appeared in print as Problem 38 of "The Scottish Book" [657]. We discuss the connectivity of this model and then matchings and Hamilton cycles. We also consider a related model of "Nearest Neighbor Graphs".

17.1 Connectivity

For an integer k, $1 \leq k \leq n-1$, let $\vec{\mathbb{G}}_{k-out}$ be a random digraph on vertex set $V = \{1,2,\ldots,n\}$ with arcs (directed edges) generated independently for each $v \in V$ by a random choice of k distinct arcs (v,w), where $w \in V \setminus \{v\}$, so that each of the $\binom{n-1}{k}$ possible sets of arcs is equally likely to be chosen. Let $\mathbb{G}_{k-out}$ be the random graph(multigraph) obtained from $\vec{\mathbb{G}}_{k-out}$ by ignoring the orientation of its arcs, but retaining all edges.
Note that $\vec{\mathbb{G}}_{1-out}$ is a functional digraph of a random mapping $f : [n] \to [n]$, with a restriction that loops (fixed points) are not allowed. So for $k = 1$ the following result holds.

Theorem 17.1.

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{1-out} \text{ is connected }) = 0.$$

The situation changes when each vertex is allowed to choose more than one neighbor. Denote by $\kappa(G)$ and $\lambda(G)$ the vertex and edge connectivity of a graph G respectively, i.e., the minimum number of vertices (respectively edges) the deletion of which disconnects G. Let $\delta(G)$ be the minimum degree of G. The well known Whitney's Theorem states that, for any graph G,

$$\kappa(G) \le \lambda(G) \le \delta(G).$$

In the next theorem we show that for random $k-out$ graphs these parameters are equal w.h.p. It is taken from Fenner and Frieze [358]. The Scottish Book [657] contains a proof that $\mathbb{G}_{k-out}$ is connected for $k \ge 2$.

Theorem 17.2. *Let $\kappa = \kappa(\mathbb{G}_{k-out}), \lambda = \lambda(\mathbb{G}_{k-out})$ and $\delta = \delta(\mathbb{G}_{k-out})$. Then, for $2 \le k = O(1)$,*

$$\lim_{n\to\infty} \mathbb{P}(\kappa = \lambda = \delta = k) = 1.$$

Proof. In the light of Whitney's Theorem, to prove our theorem we have to show that the following two statements hold:

$$\lim_{n\to\infty} \mathbb{P}(\kappa(\mathbb{G}_{k-out}) \ge k) = 1, \tag{17.1}$$

and

$$\lim_{n\to\infty} \mathbb{P}(\delta(\mathbb{G}_{k-out}) \le k) = 1. \tag{17.2}$$

Then, w.h.p.

$$k \le \kappa \le \lambda \le \delta \le k,$$

and the theorem follows.

To prove statement (17.1) consider the deletion of r vertices from the random graph $\mathbb{G}_{k-out}$, where $1 \le r \le k-1$. If $\mathbb{G}_{k-out}$ can be disconnected by deleting r vertices, then there exists a partition (R,S,T) of the vertex set V, with $\mid R \mid = r$, $\mid S \mid = s$ and $\mid T \mid = t = n-r-s$, with $k-r+1 \le s \le n-k-1$, such that $\mathbb{G}_{k-out}$ has no edge joining a vertex in S with a vertex in T. The probability of such an event, for an arbitrary partition given above, is equal to

$$\left(\frac{\binom{r+s-1}{k}}{\binom{n-1}{k}}\right)^s \left(\frac{\binom{n-s-1}{k}}{\binom{n-1}{k}}\right)^{n-r-s} \le \left(\frac{r+s}{n}\right)^{sk} \left(\frac{n-s}{n}\right)^{(n-r-s)k}$$

Thus

$$\mathbb{P}(\kappa(\mathbb{G}_{k-out}) \le r) \le \sum_{s=k-r+1}^{\lfloor (n-r)/2 \rfloor} \frac{n!}{s!r!(n-r-s)!} \left(\frac{r+s}{n}\right)^{sk} \left(\frac{n-s}{n}\right)^{(n-r-s)k}$$

We have replaced $n-k-1$ by $\lfloor (n-r)/2 \rfloor$ because we can always interchange S and T so that $|S| \le |T|$.

But, by Stirling's formula,

$$\frac{n!}{s!r!(n-r-s)!} \le \alpha_s \frac{n^n}{s^s(n-r-s)^{n-r-s}}$$

where

$$\alpha_s = \alpha(s,n,r) \le c\left(\frac{n}{s(n-r-s)}\right)^{1/2} \le \frac{2c}{s^{1/2}},$$

for some absolute constant $c > 0$.

Thus

$$\mathbb{P}(\kappa(\mathbb{G}_{k-out}) \le r) \le 2c \sum_{s=k-r+1}^{\lfloor (n-r)/2 \rfloor} \frac{1}{s^{1/2}} \left(\frac{r+s}{s}\right)^s \left(\frac{n-s}{n-r-s}\right)^{(n-r-s)} u_s$$

where

$$u_s = (r+s)^{(k-1)s}(n-s)^{(k-1)(n-r-s)} n^{n-k(n-r)}.$$

Now,

$$\left(\frac{r+s}{s}\right)^s \left(\frac{n-s}{n-r-s}\right)^{n-r-s} \le e^{2r},$$

and

$$(r+s)^s(n-s)^{n-r-s}$$

decreases monotonically, with increasing s, for $s \le (n-r)/2$. Furthermore, if $s \le n/4$ then the decrease is by a factor of at least 2.
Therefore

$$\begin{aligned}\mathbb{P}(\kappa(\mathbb{G}_{k-out}) \le r) &\le 2ce^{2r}\left(\sum_{s=k-r+1}^{n/4} 2^{-(k-1)(s-k+r-1)} + \frac{2}{n^{1/2}} \cdot \frac{n}{4}\right) u_{k-r+1} \\ &\le 5ce^{2r} n^{1/2} u_{k-r+1} \le 5ce^{2r} a n^{3/2-k(k-r)},\end{aligned}$$

where

$$a = (k+1)^{(k-1)(k-r+1)}.$$

It follows that

$$\lim_{n\to\infty} \mathbb{P}(\kappa(\mathbb{G}_{k-out}) \le r) = \lim_{n\to\infty} \mathbb{P}(\kappa(\mathbb{G}_{k-out}) \le k-1) = 0,$$

which implies that

$$\lim_{n\to\infty} \mathbb{P}(\kappa(\mathbb{G}_{k-out}) \ge k) = 1,$$

i.e., that equation (17.1) holds.
To complete the proof we have to show that equation (17.2) holds, i.e., that

$$\mathbb{P}(\delta(\mathbb{G}_{k-out}) = k) \to 1 \text{ as } n \to \infty.$$

Since $\delta \geq k$ in $\mathbb{G}_{k-out}$, we have to show that w.h.p. there is a vertex of degree k in $\mathbb{G}_{k-out}$.
Let $\mathscr{E}_v$ be the event that vertex v has indegree zero in $\vec{\mathbb{G}}_{k-out}$. Thus the degree of v in $\mathbb{G}_{k-out}$ is k if and only if $\mathscr{E}_v$ occurs. Now

$$\mathbb{P}(\mathscr{E}_v) = \left(\frac{\binom{n-2}{k}}{\binom{n-1}{k}} \right)^{n-1} = \left(1 - \frac{k}{n-1}\right)^{n-1} \to e^{-k}.$$

Let Z denote the number of vertices of degree k in $\mathbb{G}_{k-out}$. Then we have shown that $\mathbb{E}(Z) \approx ne^{-k}$. Now the random variable Z is determined by kn independent random choices. Changing one of these choices can change the value of Z by at most one. Applying the Azuma-Hoeffding concentration inequality – see Section 27.7, in particular Lemma 27.17 we see that for any $t > 0$

$$\mathbb{P}(Z \leq \mathbb{E}(Z) - t) \leq \exp\left\{ -\frac{2t^2}{kn} \right\}.$$

Putting $t = ne^{-k}/2$ we see that $Z > 0$ w.h.p. and the theorem follows. □

17.2 Perfect Matchings

Non-bipartite graphs

Assuming that the number of vertices n of a random graph $\mathbb{G}_{k-out}$ is even, Frieze [385] proved the following result.

Theorem 17.3.

$$\lim_{\substack{n\to\infty \\ n \text{ even}}} \mathbb{P}(\mathbb{G}_{k-out} \textit{ has a perfect matching}) = \begin{cases} 0 & \textit{if } k = 1 \\ 1 & \textit{if } k \geq 2. \end{cases}$$

We will only prove a weakening of the above result to where $k \geq 15$. We follow the ideas of Section 6.1. So, we begin by examining the expansion properties of $G = \mathbb{G}_{a-out}, a \geq 3$.

Lemma 17.4. *W.h.p.* $|N_G(S)| \geq |S|$ *for all* $S \subseteq [n], |S| \leq \kappa_a n$ *where* $\kappa_a = \frac{1}{2}\left(\frac{1}{30}\right)^{1/(a-2)}$.

Proof. The probability there exists a set S with insufficient expansion is at most

$$\sum_{s=3}^{\kappa_a n} \binom{n}{s} \binom{n}{s-1} \left(\frac{2s}{n}\right)^{as} \leq \sum_{s=3}^{\kappa_a n} \left(\frac{ne}{s}\right)^{2s} \left(\frac{2s}{n}\right)^{as}$$

$$= \sum_{s=3}^{\kappa_a n} \left(\left(\frac{s}{n}\right)^{a-2} e^2 2^a\right)^s = o(1). \quad (17.3)$$

Lemma 17.5. *Let $b = \lceil (1+\kappa_a^{-2})/2 \rceil$. Then as $n \to \infty$, n even, $G_{(a+b)-out}$ has a perfect matching w.h.p.*

Proof. First note that $G_{(a+b)-out}$ contains $H = \mathbb{G}_{a-out} \cup \mathbb{G}_{b-out}$ in the following sense. Start the construction of $\mathbb{G}_{(a+b)-out}$ with H. If there is a $v \in [n]$ that chooses edge $\{v,w\}$ in both $\mathbb{G}_{a-out}$ and $\mathbb{G}_{b-out}$ then add another random choice for v.

Let us show that H has a perfect matching w.h.p. Enumerate the edges of $\mathbb{G}_{b-out}$ as $e_1, e_2, \ldots, e_{bn}$. Here $e_{(i-1)n+j}$ is the ith edge chosen by vertex j. Let $\mathbb{G}_0 = \mathbb{G}_{a-out}$ and let $\mathbb{G}_i = \mathbb{G}_0 + \{e_1, e_2, \ldots, e_i\}$. If $\mathbb{G}_i$ does not have a perfect matching, consider the sets $A, A(x), x \in A$ defined prior to (6.6). It follows from Lemma 17.4 that w.h.p. all of these sets are of size at least $\kappa_a n$. Thus, $\mathbb{P}(y \in A(x)) \geq \frac{\kappa_a n - b}{n}$. We subtract b to account for the previously inspected edges associated with x's choices.

It follows that

$$\begin{aligned}
&\mathbb{P}(\mathbb{G}_{(a+b)-out} \text{ does not have a perfect matching}) \\
&\leq \mathbb{P}(H \text{ does not have a perfect matching}) \\
&\leq \mathbb{P}(\mathrm{Bin}(b\kappa_a n, \kappa_a - b/n) \leq n/2) = o(1).
\end{aligned}$$

□

Putting $a = 8$ gives $b = 7$ and a proof that $\mathbb{G}_{15-out}$, n even, has a perfect matching w.h.p.

Bipartite graphs

We now consider the related problem of the existence of a perfect matching in a random k-out bipartite graph.

Let $U = \{u_1, u_2, \ldots, u_n\}, V = \{v_1, v_2, \ldots, v_n\}$ and let each vertex from U choose independently and without repetition, k neighbors in V, and let each vertex from V choose independently and without repetition k neighbors in U. Denote by $\vec{\mathbb{B}}_{k-out}$ the digraphs generated by the above procedure and let $\mathbb{B}_{k-out}$ be its underlying simple bipartite graph.

Theorem 17.6.

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{B}_{k-out} \text{ has a perfect matching}) = \begin{cases} 0 & \text{if } k=1 \\ 1 & \text{if } k \geq 2. \end{cases}$$

We will give two different proofs. The first one - existential- of a combinatorial nature is due to Walkup [844]. The second one - constructive- of an algorithmic nature, is due to Karp, Rinnooy-Kan and Vohra [557]. We start with the combinatorial approach.

Existence proof

Let X denote the number of perfect matchings in $\mathbb{B}_{k-out}$. Then

$$\mathbb{P}(X>0) \leq \mathbb{E}(X) \leq n!\, 2^n (k/n)^n.$$

The above bound follows from the following observations. There are $n!$ ways of pairing the vertices of U with the vertices of V. For each such pairing there are 2^n ways to assign directions for the connecting edges, and then each possible matching has probability $(k/n)^n$ of appearing in $\mathbb{B}_{k-out}$.
So, by Stirling's formula,

$$\mathbb{P}(X>0) \leq 3n^{1/2}(2k/e)^n,$$

which, for $k=1$ tends to 0 as $n\to\infty$, and the first statement of our theorem follows.
To show that $\mathbb{B}_{k-out}$ has a perfect matching w.h.p. notice that since this is an increasing graph property, it is enough to show that it is true for $k=2$. Note also, that if there is no perfect matching in $\mathbb{B}_{k-out}$, then there must exist a set $R\subset U$ (or $R\subset V$) such that the cardinality of neighborhood of $S=N(R)$ of R in U (respectively, in V) is smaller than the cardinality of the set R itself, i.e., $|S|<|R|$. We will call such a pair (R,S) a *bad pair*, and, in particular, we will restrict our attention to the "minimal bad pairs", i.e., such that there is no $R'\subset R$ for which $(R',N(R'))$ is bad.

If (R,S) is a bad pair with $R\subseteq U$ then $(V\setminus S, U\setminus R)$ is also a bad pair. Given this, we can concentrate on showing that w.h.p. there are no bad pairs (R,S) with $2\leq |R|\leq (n+1)/2$.

Every minimal bad pair has to have the following two properties:

(i) $|S|=|R|-1$,

(ii) every vertex in S has at least two neighbors in R.

The first property is obvious. To see why property (ii) holds, suppose that there is a vertex $v \in S$ with at most one neighbor u in R. Then the pair $(R \setminus \{u\}, S \setminus \{v\})$ is also "bad pair" and so the pair (R,S) is not minimal.

Let $r \in [2,(n+1)/2]$ and let Y_r be the number of minimal bad pairs (R,S), with $|R| = r$ in $\mathbb{B}_{k-out}$. To complete the proof of the theorem we have to show that $\sum_r \mathbb{E}Y_r \to 0$ as $n \to \infty$. By symmetry, choose (R,S), such that $R = \{u_1,u_2,\dots u_r\} \subset U$ and $S = \{v_1,v_2,\dots v_{r-1}\} \subset V$ is a minimal "bad pair". Then

$$\mathbb{E}Y_r = 2\binom{n}{r}\binom{n}{r-1}P_r Q_r, \tag{17.4}$$

where

$$P_r = \mathbb{P}((R,S) \text{ is bad})$$

and

$$Q_r = \mathbb{P}((R,S) \text{ is minimal } \mid (R,S) \text{ is bad}).$$

We observe that, for any fixed k,

$$P_r = \left(\frac{\binom{r-1}{k}}{\binom{n}{k}}\right)^r \left(\frac{\binom{n-r}{k}}{\binom{n}{k}}\right)^{n-r+1}.$$

Hence, for $k = 2$,

$$P_r \le \left(\frac{r}{n}\right)^{2r} \left(\frac{n-r}{n}\right)^{2(n-r)}. \tag{17.5}$$

Then we use Stirling's formula to show,

$$\begin{aligned}\binom{n}{r}\binom{n}{r-1} &= \frac{r}{n-r+1}\binom{n}{r}^2 \\ &\le \frac{r}{n-r+1}\frac{n}{r(n-r)}\left(\frac{n}{r}\right)^{2r}\left(\frac{n}{n-r}\right)^{2(n-r)}. \end{aligned} \tag{17.6}$$

To estimate Q_r we have to consider condition (ii) which a minimal bad pair has to satisfy. This implies that a vertex $v \in S = N(R)$ is chosen by at least one vertex from R (denote this event by A_v), or it chooses both its neighbors in R (denote this event by B_v). Then the events $A_v, v \in S$ are negatively correlated (see Section 27.2) and the events $B_v, v \in S$ are independent of other events in this collection. Let $S = \{v_1,v_2,\dots,v_{r-1}\}$. Then we can write

$$Q_r \le \mathbb{P}\left(\bigcap_{i=1}^{r-1}(A_{v_i} \cup B_{v_i})\right)$$

$$
\begin{aligned}
&= \prod_{i=1}^{r-1} \mathbb{P}\left(A_{v_i} \cup B_{v_i} \,\middle|\, \bigcap_{j=1}^{i-1} (A_{v_j} \cup B_{v_j})\right) \\
&\leq \prod_{i=1}^{r-1} \mathbb{P}\left(A_{v_i} \cup B_{v_i}\right) \\
&= \left(1 - \mathbb{P}(A^c_{v_1})\,\mathbb{P}(B^c_{v_1})\right)^{r-1} \\
&\leq \left(1 - \left(\frac{r-2}{r-1}\right)^{2r} \left(1 - \frac{\binom{r}{2}}{\binom{n}{2}}\right)\right)^{r-1} \\
&\leq \eta^{r-1} \qquad (17.7)
\end{aligned}
$$

for some absolute constant $0 < \eta < 1$ when $r \leq (n+1)/2$.

Going back to (17.4), and using (17.5), (17.6), (17.7)

$$
\sum_{r=2}^{(n+1)/2} \mathbb{E} X_r \leq 2 \sum_{r=2}^{(n+1)/2} \frac{\eta^{r-1} n}{(n-r)(n-r+1)} = o(1).
$$

Hence $\sum_r \mathbb{E} X_r \to 0$ as $n \to \infty$, which means that w.h.p. there are no bad pairs, implying that $\mathbb{B}_{k-out}$ has a perfect matching w.h.p. □

Frieze and Melsted [408] considered the related question. Suppose that M,N are disjoint sets of size m,n and that each $v \in M$ chooses $d \geq 3$ neighbors in N. Suppose that we condition on each vertex in N being chosen at least twice. They show that w.h.p. there is a matching of size equal to $\min\{m,n\}$. Fountoulakis and Panagiotou [372] proved a slightly weaker result, in the same vein.

Algorithmic Proof

We will now give a rather elegant algorithmic proof of Theorem 17.6. It is due to Karp, Rinnooy-Kan and Vohra [557]. We do this for two reasons. First, because it is a lovely proof and second this proof is the basis of the proof that 2-in,2-out is Hamiltonian in [252]. In particular, this latter example shows that constructive proofs can sometimes be used to achieve results not obtainable through existence proofs alone.

Start with the random digraph $\vec{\mathbb{B}}_{2-out}$ and consider two multigraphs, G_U and G_V with labeled vertices and edges, generated by $\vec{\mathbb{B}}_{2-out}$ on the sets of the bipartition (U,V) in the following way. The vertex set of the graph G_U is U and two vertices, u and u', are connected by an edge, labeled v, if a vertex $v \in V$ chooses u and u' as its two neighbors in U. Similarly, the graph G_V has vertex set V and we put an edge labeled u between two vertices v and v', if a vertex $u \in U$ chooses v and v' as its two neighbors in V. Hence graphs G_U and G_V are random multigraphs with exactly n labeled vertices and n labeled edges.

We will describe below, a randomized algorithm which w.h.p. finds a perfect matching in $\mathbb{B}_{2-out}$ in $O(n)$ expected number of steps.

Algorithm PAIR

- **Step 0**. Set $H_U = G_U$ and let H_V be empty graph on vertex set V. Initially all vertices in H_U are *unmarked* and all vertices in G_V are *unchecked*. Let $CORE$ denote the set of edges of G_U that lie on cycles in G_U i.e. the edges of the 2-core of G_U.
- **Step 1**. If every isolated tree in H_U contains a marked vertex, go to Step 5. Otherwise, select any isolated tree T in H_U in which all vertices are unmarked. Pick a random vertex u in T and mark it.
- **Step 2**. Add the edge $\{x,y\}, x,y \in V$ that has label u to the graph H_V.
- **Step 3**. Let C_x, C_y be the components of H_V just before the edge labeled u is added. Let $C = C_x \cup C_y$. If all vertices in C are checked, go to Step 6. Otherwise, select an unchecked vertex v in C. If possible, select an unchecked vertex v for which the edge labeled v in H_U belongs to $CORE$.
- **Step 4**. Delete the edge labeled v from H_U, return to Step 1.
- **Step 5**. STOP and declare success.
- **Step 6**. STOP and declare failure.

We next argue that Algorithm PAIR, when it finishes at Step 5, does indeed produce a perfect matching in $\mathbb{B}_{2-out}$. There are two simple invariants of this process that explain this:

(I1) The number of marked vertices plus the number of edges in H_U is equal to n.

(I2) The number of checked vertices is equal to the number of edges in H_V.

For **I1**, we observe that each round marks one vertex and deletes one edge of H_U. Similarly, for **I2**, we observe that each round checks one vertex and adds one edge to H_V.

Lemma 17.7. *Up until (possible) failure in Step 6, the components of H_V are either trees with a unique unchecked vertex or are unicyclic components with all vertices checked. Also, failure in Step 6 means that PAIR tries to add an edge to a unicyclic component.*

Proof. This is true initially, as initially H_V has no edges and all vertices are unchecked. Assume this to be the case when we add an edge $\{x,y\}$ to H_V. If $C_x \neq C_y$ are both trees then we will have a choice of two unchecked vertices in $C = C_x \cup C_y$ and C will be a tree. After checking one vertex, our claim will still hold. The other possibilities are that C_x is a tree and C_y is unicyclic. In this case there is one unchecked vertex and this will be checked and C will be unicyclic. The other possibility is that $C = C_x = C_y$ is a tree. Again there is only one unchecked vertex and adding $\{x,y\}$ will make C unicyclic. □

Lemma 17.8. *If H_U consists of trees and unicyclic components then all the trees in H_U contain a marked vertex.*

Proof. Suppose that H_U contains k trees with marked vertices and ℓ trees with no marked vertices and that the rest of the components are unicyclic. It follows that H_U contains $n-k-\ell$ edges and then (**I1**) implies that $\ell = 0$. □

Lemma 17.9. *If the algorithm stops in Step 5, then we can extract a perfect matching from H_U, H_V.*

Proof. Suppose that we arrive at Step 5 after k rounds. Suppose that there are k trees with a marked vertex. Let the component sizes in H_U be $n_1, n_2, \dots, n_k$ for the trees and $m_1, m_2, \dots, m_\ell$ for the remaining components. Then,

$$n_1 + n_2 + \cdots + n_k + m_1 + m_2 + \cdots + m_\ell = |V(H_U)| = n.$$
$$|E(H_U)| = n - k,$$

from **I1** and so

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) + \\ (\geq m_1) + (\geq m_2) + (\geq m_\ell) = n - k.$$

It follows that the components of H_U that are not trees with a marked vertex have as many edges as vertices and so are unicyclic.

We now show, given that H_U, H_V only contain trees and unicyclic components, that we can extract a perfect matching. The edges of H_U define a matching of $\mathbb{B}_{2-out}$ of size $n-k$. Consider a tree T component with marked vertex ρ. Orient the edges of T away from ρ. Now consider an edge $\{x,y\}$ of T, oriented from x to y. Suppose that this edge has label $z \in V$. We add the edge $\{y,z\}$ to M_1. These edges are disjoint: z appears as the label of exactly one edge and y is the head of exactly one oriented edge.

For the unicyclic components, we orient the unique cycle $C = (u_1, u_2, \dots, u_s)$ arbitrarily in one of two ways. We then consider the trees attached to each of the u_i and orient them away from the u_i. An oriented edge $\{x,y\}$ with label z yields a matching edge $\{y,z\}$ as before.

The remaining k edges needed for a perfect matching come from H_V. We extract a set of k matching edges out of H_V in the same way we extracted $n-k$ edges from H_U. We only need to check that these k edges are disjoint from those chosen from H_U. Let $\{y,z\}$ be such an edge, obtained from the edge $\{x,y\}$ of H_V, which has label z. z is marked in H_U and so is the root of a tree and does not appear in any matching edge of M_1. y is a checked vertex and so the edge labelled y has been deleted from H_U and this prevents y appearing in a matching edge of M_1. □

Lemma 17.10. *W.h.p. Algorithm PAIR cannot reach Step 6 in fewer than* $0.49n$ *iterations.*

Proof. It follows from Lemma 2.10 that w.h.p. after $\leq 0.499n$ rounds, H_V only contains trees and unicyclic components. The lemma now follows from Lemma 17.7. □

To complete our analysis, it only remains to show

Lemma 17.11. *W.h.p., at most* $0.49n$ *rounds are needed to make* H_U *the union of trees and unicyclic components.*

Proof. Recall that each edge of H_U corresponds to an unchecked vertex of H_V, the edges corresponding to checked vertices having been deleted. Moreover, each tree component T of H_V has one unchecked vertex, u_T say. If u_T is the label of an edge of H_U belonging to *CORE* then due to the choice rule for vertex checking in Step 3, every vertex of T must be the label of an edge of *CORE*. Hence the number of edges left in *CORE*, after a given iteration of the algorithm, is equal to the number of tree components of H_V, where every vertex labels an edge of *CORE*. We use this to estimate the number of edges of CORE that remain in H_U after $.49n$ iterations.

Let $xe^{-x} = 2e^{-2}$, where $0 < x < 1$. One can easily check that $0.40 < x < 0.41$. It follows from Lemma 2.16 that w.h.p. $|CORE| \approx \left(1-\frac{x}{2}\right)^2 n$, which implies, that $0.63\, n \leq |CORE| \leq 0.64\, n$.

Let Z be the number of tree components in H_V made up of vertices which are the labels of edges belong to *CORE*. Then, after at most $0.49n$ rounds,

$$\mathbb{E}Z \leq o(1) + \sum_{k=1}^{(\log n)^2} \binom{n}{k} k^{k-2} \binom{0.49n}{k-1} (k-1)! \frac{(0.64)^k}{\binom{n}{2}^{k-1}} \times \\ \times \left(1 - \frac{k(n-k)}{\binom{n}{2}}\right)^{.49n-(k-1)} \tag{17.8}$$

$$\leq (1+o(1))n \sum_{k=1}^{(\log n)^2} \frac{k^{k-2}}{k!} (0.64)^k e^{-0.98k}$$

$$\leq (1+o(1))n \times$$
$$\left[\left(0.64\theta + \frac{(0.64\theta)^2}{2} + \frac{(0.64\theta)^3}{2} + \frac{2(0.64\theta)^4}{3}\right) + \sum_{k=5}^{\infty} \frac{\left((0.64)e^{.02}\right)^k}{2k^{5/2}}\right]$$

where $\theta = e^{-0.98}$

$$\leq (1+o(1))n\left[0.279 + \frac{1}{2 \times 5^{5/2}\left(1-(0.64)e^{.02}\right)}\right]$$
$$\leq (1+o(1))n\,[0.279+0.026]$$
$$\leq (0.305)n.$$

Explanation of (17.8)**;** The $o(1)$ term corresponds to components of size greater than $(\log n)^2$ and w.h.p. there are none of these. For the summand, we choose k vertices and a tree on these k vertices in $\binom{n}{k}k^{k-2}$ ways. The term $\binom{0.49n}{k-1}(k-1)!$ gives the number of sequences of edge choices that lead to a given tree. The term $\binom{n}{2}^{-(k-1)}$ is the probability that these edges exist and $(0.64)^k$ bounds the probability that the vertices of the tree correspond to edges in $CORE$. The final term is the probability that the tree is actually a component.

So after $0.49n$ rounds, in expectation, the number of edges left in $CORE$, is at most $\frac{0.305}{0.63} < 0.485$ of its original size, and the Chebyshev inequality (applied to Z) can be used to show that w.h.p. it is at most 0.49 of its original size. However, randomly deleting approximately 0.51 fraction of the edges of $CORE$ will w.h.p. leave just trees and unicyclic components in H_U. To see this, observe that if we delete $0.505n$ random edges from G_U then we will have a random graph in the sub-critical stage and so w.h.p. it will consist of trees and unicyclic components. But deleting $0.505n$ random edges will w.h.p. delete less than a 0.51 fraction of CORE. □

This completes the proof that w.h.p. Algorithm PAIR finishes before $0.49n$ rounds with a perfect matching. In summary,

Theorem 17.12. *W.h.p. the algorithm* **PAIR** *finds a perfect matching in the random graph $\mathbb{B}_{2-out}$ in at most $.49n$ steps.*

One can ask whether one can w.h.p. secure a perfect matching in a bipartite random graph having more edges then $\mathbb{B}_{1-out}$, but less than $\mathbb{B}_{2-out}$. To see that it is possible, consider the following two-round procedure. In the first round assume that each vertex from the set U chooses exactly one neighbor in V and, likewise, every vertex from the set V chooses exactly one neighbor in U. In the next round, only those vertices from U and V which have not been selected in the first round get a second chance to make yet another random selection. It is easy to see that, for large n, such a second chance is, on the average, given to approximately n/e vertices on each side. I.e, that the average out-degree of vertices in U and V is approximately $1+1/e$. Therefore the underlying simple graph is denoted as $\mathbb{B}_{(1+1/e)-out}$, and Karoński and Pittel [550] proved that the following result holds.

Theorem 17.13. *With probability* $1-O(n^{-1/2})$ *a random graph* $\mathbb{B}_{(1+1/e)-out}$ *contains a perfect matching.*

17.3 Hamilton Cycles

Bohman and Frieze [138] proved the following:

Theorem 17.14.

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{k-out} \text{ has a Hamiltonian Cycle}) = \begin{cases} 0 & \text{if } k \le 2 \\ 1 & \text{if } k \ge 3. \end{cases}$$

To see that this result is best possible note that one can show that w.h.p. the random graph $\mathbb{G}_{2-out}$ contains a vertex adjacent to three vertices of degree two, which prevents the existence of a Hamiltonian Cycle. The proof that $\mathbb{G}_{3-out}$ w.h.p. contains a Hamiltonian Cycle is long and complicated, we will therefore prove the weaker result given below which has a straightforward proof, using the ideas of Section 6.2. It is taken from Frieze and Łuczak [403].

Theorem 17.15.

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{k-out} \text{ has a Hamiltonian Cycle}) = 1, \text{ if } k \ge 5.$$

Proof. Let $H = G_0 \cup G_1 \cup G_2$ where $G_i = \mathbb{G}_{k_i-out}$, where (i) $k_0 = 1, k_1 = k_2 = 2$ and (ii) G_0, G_1, G_2 are generated independently of each other. Then we can couple the construction of H and $\mathbb{G}_{5-out}$ so that $H \subseteq \mathbb{G}_{5-out}$. This is because in the construction of H, some random choices in the construction of the associated digraphs might be repeated. In which case, having constructed H, we can give $\mathbb{G}_{5-out}$ some more edges.

It follows from Theorem 17.3 that w.h.p. $G_i, i = 1,2$ contain perfect matchings $M_i, i = 1,2$. Here we allow n to be odd and so a perfect matching may leave one vertex isolated. By symmetry M_1, M_2 are uniform random
matchings. Let $M = M_1 \cup M_2$. The components of M are cycles. There could be

degenerate 2-cycles consisting of two copies of the same edge and there may be a path in the case n is odd.

Lemma 17.16. *Let X be the number of components of M. Then w.h.p.*

$$X \leq 3\log n.$$

Proof. Let C be the cycle containing vertex 1. We show that

$$\mathbb{P}\left(|C| \geq \frac{n}{2}\right) \geq \frac{1}{2}. \tag{17.9}$$

To see this note that

$$\mathbb{P}(|C| = 2k) = \prod_{i=1}^{k-1}\left(\frac{n-2i}{n-2i+1}\right)\frac{1}{n-2k+1} < \frac{1}{n-2k+1}.$$

Indeed, consider the M_1-edge $\{1 = i_1, i_2\} \in C$ containing vertex 1. Let $\{i_2, i_3\} \in C$ be the M_2-edge containing i_2, Now, $\mathbb{P}(i_3 \neq 1) = (n-2)/(n-1)$. Assume that $i_3 \neq 1$ and let $\{i_3, i_4\} \in C$ be the M_1-edge containing i_3. Let $\{i_4, i_5\} \in C$ be the M_2-edge containing i_4. Then $\mathbb{P}(i_5 \neq 1) = (n-4)/(n-3)$, and so on until we close the cycle with probability $1/(n-2k+1)$. Hence

$$\mathbb{P}\left(|C| < \frac{n}{2}\right) < \sum_{k=1}^{\lfloor n/4 \rfloor} \frac{1}{n-2k+1} < \frac{1}{2},$$

and the bound given in (17.9) follows.
Consider next the following experiment. Choose the size s of the cycle containing vertex 1. Next choose the size of the cycle containing a particular vertex from the remaining $n-s$ vertices. Continue until the cycle chosen contains all remaining vertices. Observe now, that deleting any cycle from M leaves a random pair of matchings of the remaining vertices. So, by this observation and the fact that the bound (17.9) holds, whatever the currently chosen cycle sizes, with probability at least $1/2$, the size of the remaining vertex set halves, at least. Thus,

$$\mathbb{P}(X \geq 3\log n) \leq \mathbb{P}(\mathrm{Bin}(3\log n, 1/2) \leq \log_2 n) = o(1).$$

$\square$

We use rotations as in Section 6.2. Lemma 17.16 enables us to argue that we only need to add random edges trying to find x, y where $y \in END(x)$, at most $O(\log n)$ times. We show next that $H_1 = G_1 \cup G_2$ has sufficient expansion.

Lemma 17.17. *W.h.p.* $S \subseteq [n], |S| \leq n/1000$ *implies that* $|N_{H_1}(S)| \geq 2|S|$.

Proof. Let X be the number of vertex sets that violate the claim. Then,

$$\begin{aligned}
\mathbb{E}X &\leq \sum_{k=1}^{n/1000} \binom{n}{k}\binom{n}{2k}\left(\left(\frac{\binom{3k}{2}}{\binom{n-1}{2}}\right)^2\right)^k \\
&\leq \sum_{k=1}^{n/1000} \left(\frac{e^3n^3}{4k^3}\frac{81k^4}{n^4}\right)^k \\
&= \sum_{k=1}^{n/1000} \left(\frac{81e^3k}{4n}\right)^k \\
&= o(1).
\end{aligned}$$

□

If n is even then we begin our search for a Hamilton cycle by choosing a cycle of H_1 and removing an edge. This will give us our current path P. If n is odd we use the path P joining the two vertices of degree one in $M_1 \cup M_2$. We can ignore the case where the isolated vertex is the same in M_1 and M_2 because this only happens with probability $1/n$. We run Algorithm Pósa of Section 6.2 and observe the following: At each point of the algorithm we will have a path P plus a collection of vertex disjoint cycles spanning the vertices not in P. This is because in Step (d) the edge $\{u,v\}$ will join two cycles, one will be the newly closed cycle and the other will be a cycle of M. It follows that w.h.p. we will only need to execute Step (d) at most $3\log n$ times.

We now estimate the probability that we reach the start of Step (d) and fail to close a cycle. Let the edges of G_0 be $\{e_1, e_2, \dots, e_n\}$ where e_i is the edge chosen by vertex i. Suppose that at the beginning of Step (d) we have identified END. We can go through the vertices of END until we find $x \in END$ such that $e_x = \{x,y\}$ where $y \in END(x)$. Because G_0 and H_1 are independent, we see by Lemma 17.17 that we can assume $\mathbb{P}(y \in END(x)) \geq 1/1000$. Here we use the fact that adding edges to H_1 will not decrease the size of neighborhoods. It follows that with probability $1 - o(1/n)$ we will examine fewer than $(\log n)^2$ edges of G_0 before we succeed in closing a cycle.

Now we tryclosing cycles $O(\log n)$ times and w.h.p. each time we look at $O((\log n)^2)$ edges of G_0. So, if we only examine an edge of G_0 once, we will w.h.p. still always have $n/1000 - O((\log n)^3)$ edges to try. The probability we fail to find a Hamilton cycle this way, given that H_1 has sufficient expansion, can therefore be bounded by $\mathbb{P}(\mathrm{Bin}(n/1000 - O((\log n)^3), 1/1000) \leq 3\log n) = o(1)$. □

17.4 Nearest Neighbor Graphs

Consider the complete graph K_n, on vertex set $V = \{1,2,,\ldots,n\}$, in which each edge is assigned a cost $C_{i,j}, i \neq j$, and the costs are independent identically distributed continuous random variables. Color an edge *green* if it is one of the k shortest edges incident to either end vertex, and color it *blue* otherwise. The graph made up of the *green* edges only is called the *k-th nearest neighbor graph* and is denoted by $G_{k-\text{nearest}}$. Note that in the random graph $G_{k-\text{nearest}}$ the edges are no longer independent, as in the case of $G_{k-\text{out}}$ or in the classical model $\mathbb{G}_{n,p}$. Assume without loss of generality that the $C_{i,j}$ are exponential random variables of mean one. Cooper and Frieze [251] proved

Theorem 17.18.

$$\lim_{n\to\infty} \mathbb{P}(G_{k-nearest} \text{ is connected }) = \begin{cases} 0 & \text{if } k = 1, \\ \gamma & \text{if } k = 2, \\ 1 & \text{if } k \geq 3, \end{cases}$$

where $0.99081 \leq \gamma \leq 0.99586$.

A similar result holds for a random *bipartite k-th nearest neighbor graph*, generated in a similar way as $G_{k-\text{nearest}}$ but starting with the complete bipartite graph $K_{n,n}$ with vertex sets $V_1, V_2 = \{1,2,,\ldots,n\}$, and denoted by $B_{k-\text{nearest}}$. The following result is from Pittel and Weishar [744].

Theorem 17.19.

$$\lim_{n\to\infty} \mathbb{P}(B_{k-nearest} \text{ is connected }) = \begin{cases} 0 & \text{if } k = 1, \\ \gamma & \text{if } k = 2, \\ 1 & \text{if } k \geq 3, \end{cases}$$

where $0.996636 \leq \gamma$.

The paper [744] contains an explicit formula for γ.

Consider the related problem of the existence of a perfect matching in the bipartite k-th nearest neighbor graph $B_{k-\text{nearest}}$. For convenience, to simplify computations, we will assume here that the $C_{i,j}$ are iid exponential random variables with rate $1/n$. Coppersmith and Sorkin [274] showed that the expected size of the largest matching in $B_{1-\text{nearest}}$ (which itself is a forest) is w.h.p. asymptotic to

$$\left(2 - e^{-e^{-1}} - e^{-e^{-e^{-1}}}\right) n \approx 0.807n.$$

The same expression was obtained independently in [744]. Also, w.h.p., $B_{2-\text{n}earest}$ does not have a perfect matching. Moreover, w.h.p., in a maximal matching there are at least $\frac{2\log n}{13\log\log n}$ unmatched vertices, see [744].
The situation changes when each vertex chooses three, instead of one or two, of its "green" edges. Then the following theorem was proved in [744]:

Theorem 17.20. *$B_{3-nearest}$ has a perfect matching, w.h.p.*

Proof. The proof is analogous to the proof of Theorem 17.6 and uses Hall's Theorem. We use the same terminology. We can, as in Theorem 17.6, consider only bad pairs of "size" $k \le n/2$. Consider first the case when $k < \varepsilon n$, where $\varepsilon < 1/(2e^2)$, i.e., "small" bad pairs. Notice, that in a bad pair, each of the k vertices from V_1 must choose its neighbors from the set of $k-1$ vertices from V_2. Let A_k be the number of such sets. Then,

$$\mathbb{E}A_k \le 2\binom{n}{k}\binom{n}{k-1}\left(\frac{k}{n}\right)^{3k} \le 2\frac{n^{2k}}{(k!)^2}\left(\frac{k}{n}\right)^{3k} \le 2\left(\frac{ke^2}{n}\right)^k.$$

(The factor 2 arises from allowing R to be chosen from V_1 or V_2.)

Let P_k be the probability that there is a bad pair of size k in $B_{3-\text{n}earest}$. Then the probability that $B_{3-\text{n}earest}$ contains a bad pair of size less than $t = \lfloor \varepsilon n \rfloor$ is, letting $l = \lfloor (\log n)^2 \rfloor$, at most

$$\begin{aligned}
\sum_{k=4}^{t} P_k &\le 2\sum_{k=4}^{t}\left(\frac{ke^2}{n}\right)^k \\
&= 2\sum_{k=4}^{l}\left(\frac{ke^2}{n}\right)^k + 2\sum_{k=l+1}^{t}\left(\frac{ke^2}{n}\right)^k \\
&\le 2\sum_{k=4}^{l}\left(\frac{le^2}{n}\right)^k + 2\sum_{k=l+1}^{t}\left(\varepsilon e^2\right)^k \\
&\le \frac{2l^2e^8}{n^4} + \left(\varepsilon e^2\right)^l.
\end{aligned}$$

So, if $\varepsilon < 1/(2e^2)$, then

$$\sum_{k=4}^{\lfloor \varepsilon n \rfloor} P_k \to 0.$$

It suffices to show that

$$\sum_{k=\lfloor \varepsilon n \rfloor+1}^{n/2} P_k \to 0.$$

To prove that there are no "large" bad pairs, note that for a pair to be bad it must be the case that there is a set of $n-k+1$ vertices of V_2 that do not choose any of

the k vertices from V_1. Let $R \subset V_1, |R| = k$ and $S \subset V_2, |S| = k-1$. Without loss of generality, assume that $R = \{1,2,\dots k\}, S = \{1,2,\dots k-1\}$. Then let $Y_i, i = 1,2,\dots k$ be the smallest weight in $K_{n,n}$ among the weights of edges connecting vertex $i \in R$ with vertices from $V_2 \setminus S$, and let $Z_j, j = k, k+1, \dots n$ be the smallest weight among the weights of edges connecting vertex $j \in V_2 \setminus S$ with vertices from R. Then, each Y_i has an exponential distribution with rate $(n-k+1)/n$ and each Z_j has the exponential distribution with rate k/n.
Notice that in order for there not to be any edge in $B_{3-\mathrm{n}earest}$ between respective sets R and $V_2 \setminus S$ the following property should be satisfied: each vertex $i \in R$ has at least three neighbors in $K_{n,n}$ with weights smaller than Y_i and each vertex $j \in V_2 \setminus S$ has at least three neighbors in $K_{n,n}$ with weights smaller than the corresponding Z_j. If we condition on the value $Y_i = y$, then the probability that vertex i has at least three neighbors with respective edge weight smaller than Y_i, is given by

$$P_{n,k}(y) = 1 - \left(e^{-y/n}\right)^{k-1} - (k-1)\left(1 - e^{-y/n}\right)\left(e^{-y/n}\right)^{k-2} - \binom{k-1}{2}\left(1 - e^{-y/n}\right)^2\left(e^{-y/n}\right)^{k-3}$$

Putting $a = k/n$

$$P_{n,k}(y) \approx f(a,y) = 1 - e^{-ay} - aye^{-ay} - \frac{1}{2}a^2y^2e^{-ay}.$$

Similarly, the probability that there are three neighbors of vertex $j \in V_2 \setminus S$ with edge weights smaller than Z_j is $\approx f(1-a, Z_j)$.
So, the probability that there is a bad pair in $B_{3-\mathrm{n}earest}$ can be bounded by

$$P_k \leq 2\binom{n}{k}\binom{n}{k-1}E_k,$$

where, by the Cauchy-Schwarz inequality and independence separately of $Y_1, \dots, Y_n$ and $Z_1, \dots, Z_n$,

$$\begin{aligned} E_k &= \mathbb{E}\left(\prod_{i=1}^{k} f(a,Y_i) \prod_{j=k}^{n} f(1-a,Z_j)\right) \\ &\leq \left(\mathbb{E}\left(\prod_{i=1}^{k} f^2(a,Y_i)\right)\right)^{1/2} \left(\left(\prod_{j=k}^{n} f^2(1-a,Z_j)\right)\right)^{1/2} \\ &= \prod_{i=1}^{k} \mathbb{E}(f^2(a,Y_i))^{1/2} \prod_{j=k}^{n} \mathbb{E}(f^2(1-a,Z_j))^{1/2} \\ &= \mathbb{E}(f^2(a,Y_1))^{k/2}\,\mathbb{E}(f^2(1-a,Z_n))^{(n-k+1)/2}. \end{aligned}$$

Asymptotically, Y_1 has an exponential $(1-a)$ distribution, so

$$\begin{aligned}
&\mathbb{E}(f^2(a,Y_1)) \\
&\approx \int_0^\infty \left(1 - e^{-ay} - aye^{-ay} - \frac{1}{2}a^2y^2e^{-ay}\right)^2 (1-a)e^{-(1-a)y}dy \\
&= (1-a)\int_0^\infty (e^{-(1-a)y} - 2e^{y} - 2aye^{-y} - a^2y^2e^{-y} + e^{-(1+a)y} \\
&+ 2aye^{-(1+a)y} + 2a^2y^2e^{-(1+a)y} + a^3y^3e^{-(1+a)y} + \frac{1}{4}a^4y^4e^{-(1+a)y})dy.
\end{aligned}$$

Now using

$$\int_0^\infty y^i e^{-cy}dy = \frac{i!}{c^{i+1}},$$

we obtain

$$\begin{aligned}
\mathbb{E}(f^2(a,Y_1)) = (1-a)\Bigg(&\frac{1}{1-a} - 2 - 2a - 2a^2 + \frac{1}{1+a} \\
&+ \frac{2a}{(1+a)^2} + \frac{4a^2}{(1+a)^3} + \frac{6a^3}{(1+a)^4} + \frac{6a^4}{(1+a)^5}\Bigg) \\
&= \frac{2a^6(10+5a+a^2)}{(1+a)^5}.
\end{aligned}$$

Letting

$$g(a) = \mathbb{E}(f^2(a,Y_1))^{a/2},$$

we have

$$P_k \le 2\binom{n}{k}\binom{n}{k-1}(g(a)g(1-a))^n \approx 2\left(\frac{g(a)g(1-a)}{a^{2a}(1-a)^{2(1-a)}}\right)^n = 2h(a)^n.$$

Numerical examination of the function $h(a)$ shows that it is bounded below 1 for a in the interval $[\delta, 0.5]$, which implies that the expected number of bad pairs is exponentially small for any $k > \delta n$, with $k \le n/2$. Taking $\delta < \varepsilon < 1/(2e^2)$, we conclude that, w.h.p., there are no bad pairs in $B_{3-\text{n}earest}$, and so we arrive at the theorem. □

17.5 Exercises

17.5.1 Let $p = \frac{\log n + (m-1)\log\log n + \omega}{n}$ where $\omega \to \infty$. Show that w.h.p. it is possible to orient the edges of $\mathbb{G}_{n,p}$ to obtain a digraph D such that the minimum out-degree $\delta^+(D) \ge m$.

17.5.2 The random digraph $D_{k-in,\ell-out}$ is defined as follows: each vertex $v \in [n]$ independently randomly chooses k in-neighbors and ℓ out-neighbors. Show that w.h.p. $D_{m-in,m-out}$ is m-strongly connected for $m \geq 2$ i.e. to destroy strong connectivity, one must delete at least m vertices.

17.5.3 Show that w.h.p. the diameter of G_{k-out} is asymptotically equal to $\log_{2k} n$ for $k \geq 2$.

17.5.4 For a graph $G = (V,E)$ let $f : V \to V$ be a G-mapping if $(v, f(v))$ is an edge of G for all $v \in V$. Let G be a connected graph with maximum degree d. Let $H = \bigcup_{i=0}^{k} H_i$ where (i) $k \geq 1$, (ii) H_0 is an arbitrary spanning tree of G and (iii) $H_1, H_2, \ldots, H_k$ are independent uniform random G-mappings. Let $\theta_k = 1 - \left(1 - \frac{1}{d}\right)^{2k}$ and let $\alpha = 16/\theta_k$. Show that w.h.p. for every $A \subset V$, we have

$$|e_H(A)| \geq \frac{\theta_k}{16 \log n} \cdot |e_G(A)|.$$

where $e_G(A)$ (resp. $e_H(A)$) is the number of edges of G (resp. H) with exactly one endpoint in A.

17.5.5 Let G be a graph with n vertices and minimum degree $(\frac{1}{2} + \varepsilon)n$ for some fixed $\varepsilon > 0$. Let $H = \bigcup_{i=1}^{k} H_i$ where (i) $k \geq 2$ and (ii) $H_1, H_2, \ldots, H_k$ are independent uniform random G-mappings. Show that w.h.p. H is connected.

17.5.6 Show that w.h.p. G_{k-out} contains k edge disjoint spanning trees. (Hint: Use the Nash-Williams condition [708] – see Frieze and Łuczak [404]).

17.5.7 In the random digraph $G_{k-in,k-out}$ each $v \in [n]$ independently chooses k uniformly random in-neighbors and k uniformly random out-neighbors. Show that $G_{k-in,k-out}$ is k-strongly connected for $k \geq 2, k = O(1)$.

17.6 Notes

k-out process

Jaworski and Łuczak [521] studied the following process that generates G_{k-out} along the way. Starting with the empty graph, a vertex v is chosen uniformly at random from the set of vertices of minimum out-degree. We then add the arc (v,w) where w is chosen uniformly at random from the set of vertices that are not out-neighbors of v. After kn steps the digraph in question is precisely $\vec{\mathbb{G}}_{k-out}$. Ignoring orientation, we denote the graph obtained after m steps by $U(n,m)$. The paper [521] studied the structure of $U(n,m)$ for $n \le m \le 2m$. These graphs sit between random mappings and G_{2-out}.

Nearest neighbor graphs

There has been some considerable research on the nearest neighbor graph generated by n points $\mathscr{X} = \{X_1, X_2, \dots, X_n\}$ chosen randomly in the unit square. Given a positive integer k we define the k-nearest neighbor graph $\mathbb{G}_{\mathscr{X},k}$ by joining vertex $X \in \mathscr{X}$ to its k nearest neighbors in Euclidean distance. We first consider the existence of a giant component. Teng and Yao [830] showed that if $k \ge 213$ then there is a giant component w.h.p. Balister and Bollobás [65] reduced this number to 11. Now consider connectivity. Balister, Bollobás, Sarkar and Walters [67] proved that there exists a critical constant c^* such that if $k \le c \log n$ and $c < c^*$ then w.h.p. $\mathbb{G}_{\mathscr{X},k}$ is not connected and if $k \ge c \log n$ and if $c > c^*$ then w.h.p. $\mathbb{G}_{\mathscr{X},k}$ is connected. The best estimates for c^* are given in Balister, Bollobás, Sarkar and Walters [66] i.e. $0.3043 < c^* < 0.5139$.

When distances are independently generated then the situation is much clearer. Cooper and Frieze [251] proved that if $k = 1$ then the k-nearest neighbor graph $\mathscr{O}_1$ is not connected; the graph $\mathscr{O}_2$ is connected with probability $\gamma \in [.99081, .99586]$; for $k \ge 2$, the graph $\mathscr{O}_k$ is k-connected w.h.p.

Directed k-in, ℓ-out

There is a natural directed version of G_{k-out} called $D_{k-in,\ell-out}$ where each vertex randomly chooses k in-neighbors and ℓ out-neighbors.
Cooper and Frieze [249] studied the connectivity of such graphs. They prove for example that if $1 \le k, \ell \le 2$ then

$$\lim_{n\to\infty} \mathbb{P}(D_{k-in,\ell-out} \text{ is strongly connected}) = (1-(2-k)e^{-\ell})(1-(2-\ell)e^{-k}).$$

In this result, one can in a natural way allow $k, \ell \in [1,2]$. Hamiltonicity was discussed in [252] where it was shown that w.h.p. $D_{2-in,2-out}$ is Hamiltonian.

The random digraph $\mathbb{D}_{n,p}$ as well as $\vec{\mathbb{G}}_{k-out}$ are special cases of a random digraph where each vertex, independently of others, first chooses its out-degree d according to some probability distribution and then the set of its images - uniformly from all d-element subsets of the vertex set. If d is chosen according to the binomial distribution then it is $\mathbb{D}_{n,p}$ while if d equals k with probability 1, then it is $\vec{\mathbb{G}}_{k-out}$. Basic properties of the model (monotone properties, k-connectivity), were studied in Jaworski and Smit [523] and in Jaworski and Palka [522] .

k-out subgraphs of large graphs

Just as in Section 6.5, we can consider replacing the host graph K_n by graphs of large degree. Let an n vertex graph G be *strongly Dirac* if its minimum degree is at least cn for some constant $c > 1/2$. Frieze and Johansson [397] consider the subgraph G_k obtained from G by letting each vertex independently choose k neighbors in G. They show that w.h.p. G_k is k-connected for $k \geq 2$ and that G_k is Hamiltonian for k sufficiently large. The paper by Frieze, Goyal, Rademacher and Vempala [392] shows the use of G_k as a cut-sparsifier.

k-out with preferential attachment

Peterson and Pittel [734] considered the following model: Vertices $1, 2, \ldots, n$ in this order, each choose k random out-neighbors one at a time, subject to a "preferential attachment" rule: the current vertex selects vertex i with probability proportional to a given parameter $\alpha = \alpha(n)$ plus the number of times i has already been selected. Intuitively, the larger α gets, the closer the resulting k-out mapping is to the uniformly random k-out mapping. They prove that $\alpha = \Theta(n^{1/2})$ is the threshold for α growing "fast enough" to make the random digraph approach the uniformly random digraph in terms of the total variation distance. They also determine an exact limit of this distance for $\alpha = \beta n^{1/2}$.

Chapter 18

Real World Networks

There has recently been an increased interest in the networks that we see around us in our every day lives. Most prominent are the Internet or the World Wide Web or social networks like Facebook and Linked In. The networks are constructed by some random process. At least we do not properly understand their construction. It is natural to model such networks by random graphs. When first studying so-called “real world networks”, it was observed that often the degree sequence exhibits a tail that decays polynomially, as opposed to classical random graphs, whose tails decay exponentially. See, for example, Faloutsos, Faloutsos and Faloutsos [348]. This has led to the development of other models of random graphs such as the ones described below.

18.1 Preferential Attachment Graph

Fix an integer $m > 0$, constant and define a sequence of graphs $G_1, G_2, \ldots, G_t$. The graph G_t has vertex set $[t]$ and G_1 consists of m loops on vertex 1. Suppose we have constructed G_t. To obtain G_{t+1} we apply the following rule. We add vertex $t+1$ and connect it to m randomly chosen vertices $y_1, y_2, \ldots, y_m \in [t]$ in such a way that for $i = 1, 2, \ldots, m$,

$$\mathbb{P}(y_i = w) = \frac{\deg(w, G_t)}{2mt},$$

where $deg(w, G_t)$ is the degree of w in G_t.

In this way, G_{t+1} is obtained from G_t by adding vertex $t+1$ and m randomly chosen edges, in such a way that the neighbors of $t+1$ are biased towards higher degree vertices.

When $m = 1$, G_t is a tree and this is basically a plane-oriented recursive tree as considered in Section 15.5.

This model was considered by Barabási and Albert [72]. This was followed by a rigorous analysis of a marginally different model in Bollobás, Riordan, Spencer and Tusnády [177].

Expected Degree Sequence: Power Law

Fix t and let $V_k(t)$ denote the set of vertices of degree k in G_t, where $m \le k = \tilde{O}(t^{1/2})$. Let $D_k(t) = |V_k(t)|$ and $\bar{D}_k(t) = \mathbb{E}(D_k(t))$. Then (compare with (15.30) when $m = 1$)

$$\mathbb{E}(D_k(t+1)|G_t) = \\ D_k(t) + m\left(\frac{(k-1)D_{k-1}(t)}{2mt} - \frac{kD_k(t)}{2mt}\right) + 1_{k=m} + \varepsilon(k,t). \quad (18.1)$$

Explanation of (18.1)**:** The total degree of G_t is $2mt$ and so $\frac{(k-1)D_{k-1}(t)}{2mt}$ is the probability that y_i is a vertex of degree $k-1$, creating a new vertex of degree k. Similarly, $\frac{kD_k(t)}{2mt}$ is the probability that y_i is a vertex of degree k, destroying a vertex of degree k. At this point $t+1$ has degree m and this accounts for the term $1_{k=m}$. The term $\varepsilon(k,t)$ is an error term that accounts for the possibility that $y_i = y_j$ for some $i \neq j$.

Thus

$$\varepsilon(k,t) = O\left(\binom{m}{2}\frac{k}{mt}\right) = \tilde{O}(t^{-1/2}). \quad (18.2)$$

Taking expectations over G_t, we obtain

$$\bar{D}_k(t+1) = \bar{D}_k(t) + 1_{k=m} + m\left(\frac{(k-1)\bar{D}_{k-1}(t)}{2mt} - \frac{k\bar{D}_k(t)}{2mt}\right) + \varepsilon(k,t). \quad (18.3)$$

Under the assumption $\bar{D}_k(t) \approx d_k t$ (justified below) we are led to consider the recurrence

$$d_k = \begin{cases} 1_{k=m} + \frac{(k-1)d_{k-1} - kd_k}{2} & \text{if } k \ge m, \\ 0 & \text{if } k < m, \end{cases} \quad (18.4)$$

or

$$d_k = \begin{cases} \frac{k-1}{k+2}d_{k-1} + \frac{2\cdot 1_{k=m}}{k+2} & \text{if } k \ge m, \\ 0 & \text{if } k < m. \end{cases}$$

Therefore

$$d_m = \frac{2}{m+2}$$

$$d_k = d_m \prod_{l=m+1}^{k} \frac{l-1}{l+2} = \frac{2m(m+1)}{k(k+1)(k+2)}. \tag{18.5}$$

So for large k, under our assumption $\bar{D}_k(t) \approx d_k t$, we see that

$$\bar{D}_k(t) \approx \frac{2m(m+1)}{k^3} t.$$

We now show that the assumption $\bar{D}_k(t) \approx d_k t$ can be justified. Note that the following theorem is vacuous for $k \gg t^{1/6}$.

Theorem 18.1.

$$|\bar{D}_k(t) - d_k t| = \tilde{O}(t^{1/2}) \text{ for } k = \tilde{O}(t^{1/2}).$$

Proof. Let

$$\Delta_k(t) = \bar{D}_k(t) - d_k t.$$

Then, replacing $\bar{D}_k(t)$ by $\Delta_k(t) + d_k t$ in (18.3) and using (18.2) and (18.4) we get

$$\Delta_k(t+1) = \frac{k-1}{2t}\Delta_{k-1}(t) + \left(1 - \frac{k}{2t}\right)\Delta_k(t) + \tilde{O}(t^{-1/2}). \tag{18.6}$$

Now assume inductively on t that for every $k \geq 0$

$$|\Delta_k(t)| \leq At^{1/2}(\log t)^{\beta},$$

where $(\log t)^{\beta}$ is the hidden power of logarithm in $\tilde{O}(t^{-1/2})$ of (18.6) and A is an unspecified constant.

This is trivially true for $k < m$ also for small t if we make A large enough. So, replacing $\tilde{O}(t^{-1/2})$ in (18.6) by the more explicit $\alpha t^{-1/2}(\log t)^{\beta}$ we get

$$\begin{aligned}
&\Delta_k(t+1) \leq \\
&\leq \left|\frac{k-1}{2t}\Delta_{k-1}(t)\right| + \left|\left(1 - \frac{k}{2t}\right)\Delta_k(t)\right| + \alpha t^{-1/2}(\log t)^{\beta} \\
&\leq \frac{k-1}{2t}At^{1/2}(\log t)^{\beta} + \left(1 - \frac{k}{2t}\right)At^{1/2}(\log t)^{\beta} + \alpha t^{-1/2}(\log t)^{\beta} \\
&\leq (\log t)^{\beta}(At^{1/2} + \alpha t^{-1/2}).
\end{aligned}$$

Note that if t is sufficiently large then

$$(t+1)^{1/2} = t^{1/2}\left(1 + \frac{1}{t}\right)^{1/2} \geq t^{1/2} + \frac{1}{3t^{1/2}},$$

and so

$$\Delta_k(t+1) \leq (\log(t+1))^\beta \left(A\left[(t+1)^{1/2} - \frac{1}{3t^{1/2}}\right] + \frac{\alpha}{t^{1/2}}\right)$$
$$\leq A\left(\log(t+1)\right)^\beta (t+1)^{1/2},$$

assuming that $A \geq 3\alpha$. □

In the next section, we will justify our bound of $\tilde{O}(t^{1/2})$ for vertex degrees. After that we will prove concentration of the number of vertices of degree k, for small k.

Maximum Degree

Fix $s \leq t$ and let X_l be the degree of vertex s in G_l for $s \leq l \leq t$. We prove the following high probability upper bound on the degree of vertex s.

Lemma 18.2.

$$\mathbb{P}(X_t \geq Ae^m (t/s)^{1/2} (\log(t+1))^2) = O(t^{-A}).$$

Proof. Note first that $X_s = m$. If $0 < \lambda < \varepsilon_t = \frac{1}{\log(t+1)}$ then,

$$\mathbb{E}\left(e^{\lambda X_{l+1}} | X_l\right)$$
$$= e^{\lambda X_l} \sum_{k=0}^{m} \binom{m}{k} \left(\frac{X_l}{2ml}\right)^k \left(1 - \frac{X_l}{2ml}\right)^{m-k} e^{\lambda k}$$
$$\leq e^{\lambda X_l} \sum_{k=0}^{m} \binom{m}{k} \left(\frac{X_l}{2ml}\right)^k \left(1 - \frac{X_l}{2ml}\right)^{m-k} (1 + k\lambda(1+k\lambda))$$
$$= e^{\lambda X_l} \left(1 + \frac{\lambda(1+\lambda)X_l}{2l} + \frac{(m-1)\lambda^2 X_l^2}{4ml^2}\right)$$
$$\leq e^{\lambda X_l} \left(1 + \frac{\lambda X_l}{2l}(1+m\lambda)\right), \qquad \text{since } X_l \leq 2ml,$$
$$\leq e^{\lambda \left(1 + \frac{(1+m\lambda)}{2l}\right) X_l}.$$

We define a sequence $\lambda = (\lambda_s, \lambda_{s+1}, \ldots, \lambda_t)$ where

$$\lambda_{j+1} = \left(1 + \frac{1+m\lambda_j}{2j}\right)\lambda_j < \varepsilon_t.$$

Here our only choice will be λ_s. We show below that we can find a suitable value for this, but first observe that if we manage this then

$$\mathbb{E}\left(e^{\lambda_s X_t}\right) \leq \mathbb{E}\left(e^{\lambda_{s+1} X_{t-1}}\right) \cdots \leq \mathbb{E}\left(e^{\lambda_t X_s}\right) \leq 1 + o(1).$$

Now

$$\lambda_{j+1} \leq \left(1 + \frac{1 + m\varepsilon_t}{2j}\right)\lambda_j,$$

implies that

$$\lambda_t = \lambda_s \prod_{j=s}^{t}\left(1 + \frac{1 + m\varepsilon_t}{2j}\right) \leq \lambda_s \exp\left\{\sum_{j=s}^{t} \frac{1 + m\varepsilon_t}{2j}\right\} \leq e^m \left(\frac{t}{s}\right)^{1/2} \lambda_s.$$

So a suitable choice for $\lambda = \lambda_s$ is

$$\lambda_s = e^{-m}\varepsilon_t \left(\frac{s}{t}\right)^{1/2}.$$

This gives

$$\mathbb{E}\left(\exp\left\{e^{-m}\varepsilon_t (s/t)^{1/2} X_t\right\}\right) \leq 1 + o(1).$$

So,

$$\mathbb{P}\left(X_t \geq Ae^m (t/s)^{1/2} (\log(t+1))^2)\right) \leq \\ e^{-\lambda_s Ae^m (t/s)^{1/2} (\log(t+1)^2)}\, \mathbb{E}\left(e^{\lambda_s X_t}\right) = O(t^{-A}).$$

□

Thus with probability $1 - o(1)$ as $t \to \infty$ we have that the maximum degree in G_t is $O(t^{1/2}(\log t)^2)$. This is not best possible. One can prove that w.h.p. the maximum degree is $O(t^{1/2}\omega(t))$ and $\Omega(t^{1/2}/\omega(t))$ for any $\omega(t) \to \infty$, see for example Flaxman, Frieze and Fenner [364].

Concentration of Degree Sequence

Fix a value k for a vertex degree. We show that $D_k(t)$ is concentrated around its mean $\bar{D}_k(t)$.

Theorem 18.3.

$$\mathbb{P}(|D_k(t) - \bar{D}_k(t)| \geq u) \leq 2\exp\left\{-\frac{u^2}{8mt}\right\}. \tag{18.7}$$

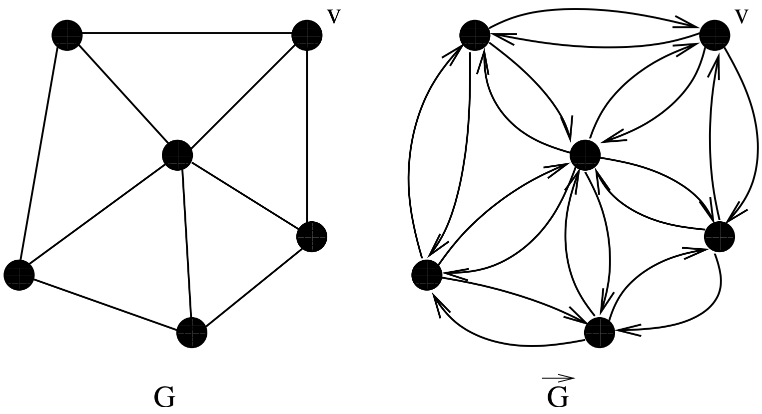


Figure 18.1: Constructing $\vec{G}$ from G.

Proof. Let $Y_1,Y_2,\dots,Y_{mt}$ be the sequence of edge choices made in the construction of G_t, and for $Y_1,Y_2,\dots,Y_i$ let

$$Z_i = Z_i(Y_1,Y_2,\dots,Y_i) = \mathbb{E}(D_k(t) \mid Y_1,Y_2,\dots,Y_i). \tag{18.8}$$

We will prove next that $|Z_i - Z_{i-1}| \le 4$ and then (18.7) follows directly from the Azuma-Hoeffding inequality, see Section 27.7. Fix $Y_1,Y_2,\dots,Y_i$ and $\hat{Y}_i \neq Y_i$. We define a map (measure preserving projection) φ of

$$Y_1,Y_2,\dots,Y_{i-1},Y_i,Y_{i+1},\dots,Y_{mt}$$

to

$$Y_1,Y_2,\dots,Y_{i-1},\hat{Y}_i,\hat{Y}_{i+1},\dots,\hat{Y}_{mt}$$

such that

$$|Z_i(Y_1,Y_2,\dots,Y_i) - Z_i(Y_1,Y_2,\dots,\hat{Y}_i)| \le 4. \tag{18.9}$$

In the preferential attachment model we can view vertex choices in the graph G as random choices of arcs in a digraph $\vec{G}$, which is obtained by replacing every edge of G by a directed 2-cycle (see Figure 18.1).

Indeed, if we choose a random arc and choose its head then v will be chosen with probability proportional to the number of arcs with v as head i.e. its degree. Hence $Y_1,Y_2,\dots$ can be viewed as a sequence of arc choices. Let

$$Y_i = (x,y) \text{ where } x > y$$
$$\hat{Y}_i = (\hat{x},\hat{y}) \text{ where } \hat{x} > \hat{y}.$$

Note that $x = \hat{x}$ if $i \mod m \neq 1$.
Now suppose $j > i$ and $Y_j = (u,v)$ arises from choosing (w,v). Then we define

$$\varphi(Y_j) = \begin{cases} Y_j & (w,v) \neq Y_i \\ (w,\hat{y}) & (w,v) = Y_i \end{cases} \tag{18.10}$$

This map is measure preserving since each sequence $\varphi(Y_1,Y_2,\dots,Y_t)$ occurs with probability $\prod_{j=i+1}^{tm} j^{-1}$. Only $x,\hat{x},y,\hat{y}$ change degree under the map φ so $D_k(t)$ changes by at most four. $\square$

We will now study the degrees of early vertices.

Degrees of early vertices

Let $d_t(s)$ denote the degree of vertex s at time t. Then we have $d_s(s) = m$ and

$$\mathbb{E}(d_{t+1}(s)|G_t) = d_t(s) + \frac{md_t(s)}{2mt} = d_t(s)\left(1+\frac{1}{2t}\right).$$

So, because

$$\frac{2^{2s+1}s!(s-1)!}{(2s)!} \approx 2\left(\frac{\pi}{s}\right)^{1/2} \text{ for large } s,$$

we have

$$\mathbb{E}(d_s(t)) = m\prod_{i=s}^{t-1}\frac{2i+1}{2i} = mA(s)\frac{(2t-1)!}{2^{2t}((t-1)!)^2} \approx m\left(\frac{t}{s}\right)^{1/2} \text{ for large } s. \quad (18.11)$$

For random variables X,Y and a sequence of random variables $\mathscr{Z} = Z_1,Z_2,\dots,Z_k$ taking discrete values, we write

$$X \succ Y \text{ to mean that } \Pr(X \geq a) \geq \Pr(Y \geq a)$$

and

$$X\mid_{\mathscr{Z}} \succ Y\mid_{\mathscr{Z}} \text{ to mean that } \Pr(X \geq a \mid Z_l = z_l, l = 1,\dots,k) \geq \Pr(Y \geq a \mid Z_l = z_l, l = 1,\dots,k),$$

for all choices of a,z.

Fix $i \leq j-2$ and let $X = d_{j-1}(t), Y = d_j(t)$ and $Z_l = d_l(t), l = i,\dots,j-2$.

Lemma 18.4. $X\mid_{\mathscr{Z}} \succ Y\mid_{\mathscr{Z}}$.

Proof. Consider the construction of $G_{j+1},G_{j+2},\dots,G_t$. We condition on those edge choices of $j+1,j+2,\dots,t$ that have one end in $i,i+1,\dots,j-2$. Now if vertex j does not choose an edge $(j-1,j)$ then the conditional distributions of $d_{j-1}(t),d_j(t)$ are identical. If vertex j does choose edge $(j-1,j)$ and we do not include this edge in the value of the degree of $j-1$ at times $j+1$ onwards, then the conditional distributions of $d_{j-1}(t),d_j(t)$ are again identical. Ignoring this edge will only reduce the chances of $j-1$ being selected at any stage and the lemma follows. $\square$

Corollary 18.5. *If $j \geq i-2$ then $d_i(t) \succ (d_{i+1}(t)+\cdots+d_j(t))/(j-i)$.*

Proof. Fix $i \leq l \leq j$ and then we argue by induction that

$$d_{i+1}(t)+\cdots+d_l(t)+(j-l)d_{l+1}(t) \prec d_{i+1}(t)+\cdots+(j-l+1)d_l(t). \quad (18.12)$$

This is trivial for $j=l$ as the LHS is then the same as the RHS. Also, if true for $l=i$ then

$$d_{i+1}(t)+\cdots+d_j(t) \prec (j-i)d_{i+1}(t) \prec (j-i)d_i(t)$$

where the second inequality follows from Lemma 18.4 with $j=i+1$.

Putting $\mathscr{X}=d_{i+1}(t),\ldots,d_{l-1}(t)$ we see that (18.12) is implied by

$$d_l(t)+(j-l)d_{l+1}(t)\mid_{\mathscr{X}} \prec (j-l+1)d_l(t)\mid_{\mathscr{X}} \text{ or } d_{l+1}(t)\mid_{\mathscr{X}} \prec d_l(t)\mid_{\mathscr{X}},$$

after subtracting $(j-l)d_{l+1}(t)$. But the latter follows from Lemma 18.4. □

Lemma 18.6. *Fix $1 \leq s = O(1)$ and let $\omega = \log^2 t$ and let $D_s(t) = \sum_{i=s+1}^{s+\omega} d_s(t)$. Then w.h.p. $D_s(t) \approx 2m(\omega t)^{1/2}$.*

Proof.
We have from (18.11) that

$$\mathbb{E}(D_s(t)) \approx m \sum_{i=s+1}^{s+\omega} \left(\frac{t}{i}\right)^{1/2} \approx 2m(\omega t)^{1/2}.$$

Going back to the proof of Theorem 18.3 we consider the map φ as defined in (18.10). Unfortunately, (18.9) does not hold here. But we can replace 4 by $10\log t$, *most of the time*. So we let $Y_1,Y_2,\ldots,Y_{mt}$ be as in Theorem 18.3. Then let ψ_i denote the number of times that $(w,v)=Y_i$ in equation (18.10). Now ψ_j is the sum of $mt-j$ independent Bernouilli random variables and $\mathbb{E}(\psi_i) \leq \sum_{j=i+1}^{mt} 1/mj \leq m^{-1}\log mt$. It follows from Hoeffding's inequality that $\Pr(\psi_i \geq 10\log t) \leq t^{-10}$. Given this, we define a new random variable $\widehat{d_s}(t)$ and let $\widehat{D}_s(t) = \sum_{j=1}^{\omega} \widehat{d}_{s+j}(t)$. Here $\widehat{d}_{s+j}(t) = d_{s+j}(t)$ for $j=1,2,\ldots,\omega$ unless there exists i such that $\psi_i \geq 10\log t$. If there is an i such that $\psi_i \geq 10\log t$ then assuming that i is the first such we let $\widehat{D}_s(t) = Z_i(Y_1,Y_2,\ldots,Y_i)$ where Z_i is as defined in (18.8), with $D_k(t)$ replaced by $\widehat{D}_s(t)$. In summary we have

$$\Pr(\widehat{D}_s(t) \neq D_s(t)) \leq t^{-10}. \quad (18.13)$$

So,

$$|\mathbb{E}(\widehat{D}_s(t)) - \mathbb{E}(D_s(t)| \leq t^{-9}.$$

And finally,

$$|Z_i - Z_{i-1}| \leq 20\log t.$$

This is because each $Y_i, \hat{Y}_i$ concerns at most two of the vertices $s+1, s+2, \ldots, s+\omega$. So,

$$\Pr(|\widehat{D}_s(t) - \mathbb{E}(\widehat{D}_s(t))| \geq u) \leq \exp\left\{-\frac{u^2}{800mt\log^2 t}\right\}. \tag{18.14}$$

Putting $u = \omega^{3/4}t^{1/2}$ into (18.14) yields the claim. □

Combining Corollary 18.5 and Lemma 18.6 we have the following theorem.

Theorem 18.7. *Fix $1 \leq s = O(1)$ and let $\omega = \log^2 t$. Then w.h.p. $d_i(t) \geq mt^{1/2}/\omega^{1/2}$ for $i = 1, 2, \ldots, s$.*

Proof. Corollary 18.5 and (18.13) implies that $d_i(t) \succ D_i(t)/\omega$. Now apply Lemma 18.6. □

18.2 Spatial Preferential Attachment

The Spatial Preferential Attachment (SPA) model, was introduced by Aiello, Bonato, Cooper, Janssen and Prałat in [8]. This model combines preferential attachment with geometry by introducing ”spheres of influence” of vertices, whose volumes depend on their in-degrees.

We first fix parameters of the model. Let $m \in \mathbb{N}$ be the dimension of space $\mathbb{R}^m$, $p \in [0,1]$ be the link (arc) probability and fix two additional parameters A_1, A_2, where $A_1 < 1/p$ while $A_2 > 0$. Let S be the unit hypercube in $\mathbb{R}^m$, with the torus metric $d(\cdot,\cdot)$ derived from the L^∞ metric. In particular, for any two points x and y in S,

$$d(x,y) = \min\{\|x - y + u\|_\infty : u \in \{-1,0,1\}^m\}. \tag{18.15}$$

For each positive real number $\alpha < 1$, and $u \in S$, define the ball around u with volume α as

$$B_\alpha(u) = \{x \in S : d(u,x) \leq r_\alpha\},$$

where $r_\alpha = \alpha^{1/m}/2$, so that r_α is chosen such that B_α has volume α.

The SPA model generates a stochastic sequence of directed graphs $\{G_t\}$, where $G_t = (V_t, E_t)$ and $V_t \subset S$, i.e., all vertices are placed in the m-dimensional hypercube $S = [0,1]^m$.

Let $deg^-(v;t)$ be the in-degree of the vertex v in G_t, and $deg^+(v;t)$ its out-degree. Then, the *sphere of influence* $S(v;t)$ of the vertex v at time $t \geq 1$ is the ball centered at v with the following volume:

$$|S(v,t)| = \min\left\{\frac{A_1 deg^-(v;t) + A_2}{t}, 1\right\}. \tag{18.16}$$

In order to construct a sequence of graphs we start at $t = 0$ with G_0 being the null graph. At each time step t we construct G_t from G_{t-1} by, first, choosing a new vertex v_t uniformly at random (**uar**) from the cube S and adding it to V_{t-1} to create V_t. Then, independently, for each vertex $u \in V_{t-1}$ such that $v_t \in S(u,t-1)$, a directed link (v_t,u) is created with probability p. Thus,the probability that a link (v_t,u) is added in time-step t equals $p|S(u,t-1)|$.

Power law and vertex in-degrees

Theorem 18.8. *Let $N_{i,n}$ be the number of vertices of in-degree i in the SPA graph G_t at time $t = n$, where $n \geq 0$ is an integer. Fix $p \in (0,1]$. Then for any $i \geq 0$,*

$$\mathbb{E}(N_{i,n}) = (1+o(1))c_i n, \tag{18.17}$$

where

$$c_0 = \frac{1}{1+pA_2}, \tag{18.18}$$

and for $1 \leq i \leq n$,

$$c_i = \frac{p^i}{1+pA_2+ipA_1} \prod_{j=0}^{i-1} \frac{jA_1+A_2}{1+pA_2+jpA_1}. \tag{18.19}$$

In [8] a stronger result is proved which indicates that fraction $N_{i,n}/n$ follows a power law. It is shown that for $i = 0,1,\ldots,i_f$, where $i_f = (n/\log^8 n)^{pA_1/(4pA_1+2)}$, w.h.p.

$$N_{i,n} = (1+o(1))c_i n.$$

Since, for some constant c,

$$c_i = (1+o(1))ci^{-(1+1/pA_1)},$$

it shows that for large i the expected proportion $N_{i,n}/n$ follows a power law with exponent $1+\frac{1}{pA_1}$, and concentration for all values of i up to i_f.

To prove Theorem 18.8 we need the following result of Chung and Lu (see [224], Lemma 3.1) on real sequeances

Lemma 18.9. *If $\{\alpha_t\},\{\beta_t\}$and$\{\gamma_t\}$ are real sequences satisfying the relation*

$$\alpha_{t+1} = \left(1-\frac{\beta_t}{t}\right)\alpha_t + \gamma_t.$$

Furthermore, suppose $\lim_{t\to\infty}\beta_t = \beta > 0$ and $\lim_{t\to\infty}\gamma_t = \gamma$. Then $\lim_{t\to\infty}\frac{\alpha_t}{t}$ exists and

$$\lim_{t\to\infty}\frac{\alpha_t}{t} = \frac{\gamma}{1+\beta}$$

□

Proof of Theorem 18.8

The equations relating the random variables $N_{i,t}$ are described as follows. Since G_1 consists of one isolated node, $N_{0,1} = 1$, and $N_{i,1} = 0$ for $i > 0$. For all $t > 0$, we derive that

$$\mathbb{E}(N_{0,t+1} - N_{0,t}|G_t) = 1 - pN_{0,t}\frac{A_2}{t}, \tag{18.20}$$

while

$$\mathbb{E}(N_{i,t+1} - N_{i,t}|G_t) = pN_{i-1,t}\frac{A_1 i + A_2}{t} - pN_{i,t}\frac{A_1(i-1)+A_2}{t}. \tag{18.21}$$

Now applying Lemma 18.9 to (18.20) with

$$\alpha_t = \mathbb{E}(N_{0,t}) \;\; \beta_t = pA_2 \;\; \text{and} \;\; \gamma_t = 1,$$

we get that

$$\mathbb{E}(N_{0,t}) = c_0 + o(t),$$

where c_0 as in (18.18).

For $i > 0$ Lemma 18.9 can be inductively applied with

$$\alpha_t = \mathbb{E}(N_{i,t}), \;\; \beta_t = p(A_1 i + A_2) \;\; \text{and} \;\; \gamma_t = \mathbb{E}(N_{i-1,t})\frac{A_1(i-1)+A_2}{t},$$

to show that

$$\mathbb{E}(N_{i,t}) = c_i + o(t),$$

where

$$c_i = pc_{i-1}\frac{A_1(i-1)+A_2}{1+p(A_1 i + A_2)}.$$

One can easily verify that the expressions for c_0, and c_i, $i \geq 1$, given in (18.18) and (18.19), satisfy the respective recurrence relations derived above. □

Knowing the expected in-degree of a node, given its age, can be used to analyze geometric properties of the SPA graph G_t. Let us note also that the result below for $i \gg 1$ was proved in [515] and extended to all $i \geq 1$ in [270]. As before, let v_i be the node added at time i.

Theorem 18.10. *Suppose that* $i = i(t) \gg 1$ *as* $t \to \infty$. *Then,*

$$\mathbb{E}(\deg^-(v_i,t)) \;\; = \;\; (1+o(1))\frac{A_2}{A_1}\left(\frac{t}{i}\right)^{pA_1} - \frac{A_2}{A_1}, \tag{18.22}$$

$$\mathbb{E}(|S(v_i,t)|) \quad = \quad (1+o(1))A_2 t^{pA_1-1} i^{-pA_1}.$$

Moreover, for all $i \geq 1$,

$$\begin{aligned} \mathbb{E}(\deg^-(v_i,t)) &\leq \frac{eA_2}{A_1}\left(\frac{t}{i}\right)^{pA_1} - \frac{A_2}{A_1}, \\ \mathbb{E}(|S(v_i,t)|) &\leq (1+o(1))eA_2 t^{pA_1-1} i^{-pA_1}. \end{aligned} \tag{18.23}$$

Proof. In order to simplify calculations, we make the following substitution:

$$X(v_i,t) = \deg^-(v_i,t) + \frac{A_2}{A_1}. \tag{18.24}$$

It follows from the definition of the process that

$$X(v_i,t+1) = \begin{cases} X(v_i,t)+1, & \text{with probability } \frac{pA_1X(v_i,t)}{t} \\ X(v_i,t), & \text{otherwise.} \end{cases}$$

We then have,

$$\begin{aligned} \mathbb{E}(X(v_i,t+1) \mid X(v_i,t)) &= (X(v_i,t)+1)\frac{pA_1X(v_i,t)}{t} + X(v_i,t)\left(1 - \frac{pA_1X(v_i,t)}{t}\right) \\ &= X(v_i,t)\left(1+\frac{pA_1}{t}\right). \end{aligned}$$

Taking expectations over $X(v_i,t)$, we get

$$\mathbb{E}(X(v_i,t+1)) = \mathbb{E}(X(v_i,t))\left(1+\frac{pA_1}{t}\right).$$

Since all nodes start with in-degree zero, $X(v_i,i) = \frac{A_2}{A_1}$. Note that, for $0<x<1$, $\log(1+x) = x - O(x^2)$. If $i \gg 1$, one can use this to get

$$\mathbb{E}(X(v_i,t)) = \frac{A_2}{A_1}\prod_{j=i}^{t-1}\left(1+\frac{pA_1}{j}\right) = (1+o(1))\frac{A_2}{A_1}\exp\left(\sum_{j=i}^{t-1}\frac{pA_1}{j}\right),$$

and in all cases $i \geq 1$,

$$\mathbb{E}(X(v_i,t)) \leq \frac{A_2}{A_1}\exp\left(\sum_{j=i}^{t-1}\frac{pA_1}{j}\right).$$

Therefore, when $i \gg 1$,

$$\mathbb{E}(X(v_i,t)) = (1+o(1))\frac{A_2}{A_1}\exp\left(pA_1\log\left(\frac{t}{i}\right)\right) = (1+o(1))\frac{A_2}{A_1}\left(\frac{t}{i}\right)^{pA_1},$$

and (18.22) follows from (18.24) and (18.16). Moreover, for any $i \geq 1$

$$\mathbb{E}(X(v_i,t)) \leq \frac{A_2}{A_1}\exp\left(pA_1\left(\log\left(\frac{t}{i}\right)+1/i\right)\right) \leq \frac{eA_2}{A_1}\left(\frac{t}{i}\right)^{pA_1},$$

and (18.23) follows from (18.24) and (18.16) as before, which completes the proof. $\square$

Directed diameter

Consider the graph G_t produced by the SPA model. For a given pair of vertices $v_i, v_j \in V_t$ ($1 \leq i < j \leq t$), let $l(v_i,v_j)$ denote the length of the shortest directed path from v_j to v_i if such a path exists, and let $l(v_i,v_j) = 0$ otherwise. The directed diameter of a graph G_t is defined as

$$D(G_t) = \max_{1\leq i<j\leq t} l(v_i,v_j).$$

We next prove the following upper bound on $D(G_t)$ (see [270]):

Theorem 18.11. *Consider the SPA model. There exists an absolute constant $c_1 > 0$ such that w.h.p.*

$$D(G_t) \leq c_1 \log t.$$

Proof. Let $C = 18\max(A_2,1)$. We prove that with probability $1-o(t^{-2})$ we have that for any $1 \leq i < j \leq t$, G_t does not contain a directed (v_i,v_j)-path of length exceeding $k^* = C\log t$. As there are at most t^2 pairs v_i, v_j, Theorem 18.11 will follow.

In order to simplify the notation, we use v to denote the vertex added at step $v \leq t$. Let vPu be a directed (v,u)-path of length given by $vPu = (v, t_{k-1}, t_{k-2}, \dots, t_1, u)$, let $t_0 = u, t_k = v$.

$$\Pr(vPu \text{ exists}) = \prod_{i=1}^{k} p\left(\frac{A_1\deg^-(t_{i-1},t_i)+A_2}{t_i}\right).$$

Let $N(v,u,k)$ be the number of directed (v,u)-paths of length k, then

$$\mathbb{E}(N(v,u,k)) = \sum_{u<t_1<\cdots<t_{k-1}<v} p^k\,\mathbb{E}\left(\prod_{i=1}^{k}\left(\frac{A_1\deg^-(t_{i-1},t_i)+A_2}{t_i}\right)\right).$$

We first consider the case where u tends to infinity together with t. It follows from Theorem 18.10 that

$$\mathbb{E}(\deg^-(t_{i-1},t_i)) = (1+o(1))\frac{A_2}{A_1}\left(\frac{t_i}{t_{i-1}}\right)^{pA_1} - \frac{A_2}{A_1}.$$

Thus

$$\begin{aligned}
\mathbb{E}(N(v,u,k)) &= \sum_{u<t_1<\cdots<t_{k-1}<v} p^k \prod_{i=1}^{k} \frac{1}{t_i}\left(A_1\,\mathbb{E}(\deg^-(t_{i-1},t_i)) + A_2\right) \\
&= \sum_{u<t_1<\cdots<t_{k-1}<v} (1+o(1))^k (A_2 p)^k \prod_{i=1}^{k} \frac{1}{t_i}\left(\frac{t_i}{t_{i-1}}\right)^{pA_1} \\
&= (1+o(1))^k (A_2 p)^k \left(\frac{v}{u}\right)^{pA_1} \frac{1}{v} \sum_{u<t_1<\cdots<t_{k-1}<v} \prod_{i=1}^{k-1} \frac{1}{t_i}.
\end{aligned}$$

However

$$\begin{aligned}
\sum_{u<t_1<\cdots<t_{k-1}<v} \prod_{i=1}^{k-1} \frac{1}{t_i} &\le \frac{1}{(k-1)!}\left(\sum_{u<s<v} \frac{1}{s}\right)^{k-1} \\
&\le \frac{1}{(k-1)!}\left(\log v/u + 1/u\right)^{k-1} \\
&\le \left(\frac{e(\log v/u + 1/u)}{k-1}\right)^{k-1}.
\end{aligned}$$

Let $k^* = C\log t$, where $C = 18\max(1, A_2)$. Assuming t sufficiently large, and recalling that $pA_1 < 1$, we have

$$\begin{aligned}
\sum_{k>k^*} \mathbb{E}(N(v,u,k)) &\le 2A_2 \sum_{k>k^*} \left(\frac{(1+o(1))A_2 pe(\log v/u + 1/u)}{k-1}\right)^{k-1} \\
&\le 2A_2 \left(\frac{(1+o(1))A_2 e(\log v/u + 1/u)}{C\log t}\right)^{k^*} \frac{1}{1-3A_2/C} \qquad (18.25) \\
&= O(6^{-18\log t}) \qquad (18.26) \\
&= o(t^{-4}).
\end{aligned}$$

The result follows for u tending to infinity. In the case where u is a constant, it follows from Theorem 18.10 that a multiplicative correction of e can be used in $\mathbb{E}(\deg^-(t_{i-1},t_i))$, leading to replacing e by e^2 in (18.25) and then 6 in (18.26) by 2, giving a bound of $O(2^{-18\log t}) = o(t^{-4})$ as before. This finishes the proof of the theorem. □

18.3 Preferential Attachment with Deletion

In this section we consider models where edges or vertices are deleted as well as added as the process continues. Random vertex deletions were considereed in Chung and Lu [223] and by Cooper, Frieze and Vera [273]. These papers show that power laws for the degree sequence persist, assuming that vertices arrive at a greater rate than vertices are deleted. The arguments are based on the analysis of equations like (18.1). This is complicated by the fact that the number of edges is now a random variable.

Random Edge Deletion

The model we consider here is as follows: suppose that $\alpha < 1$ is such that αm is an integer. Suppose that after adding vertex $t+1$ and its m incident edges, we randomly delete αm edges from the current graph. The effect of this is to replace (18.1) by

$$\mathbb{E}(D_k(t+1)|G_t) = \\ D_k(t) + m\left(\frac{(k-1)D_{k-1}(t)}{2m(1-\alpha)t} - \frac{kD_k(t)}{2m(1-\alpha)t}\right) + 1_{k=m} + \varepsilon(k,t). \quad (18.27)$$

Here the error $\varepsilon(k,t)$ has also to absorb the possibility that we delete and edge incident with $t+1$. This is $O(1/t)$ and so is negligible. Given this, we follow the subsequent analysis and obtain

$$\bar{D}_k(t) \approx \frac{2(1-\alpha)m((1-\alpha)m+1)}{k^3}t. \quad (18.28)$$

We repeat that random vertex deletion is more difficult to analyse. Deletion of a vertex results in the deletion of a random number of edges and we have to handle the distribution of the inverse of the number of edges.

Adversarial Vertex Deletion

One can also consider *adversarial* deletions as studied in Flaxman, Frieze and Vera [367] and this will be the topic of this section. We will consider the process $\mathscr{P}$, which generates a sequence of graphs $G_t = (V_t, E_t)$, for $t = 1, 2, \ldots, n$. It is follows the construction of G_t in Section 18.1 except that after each addition of a vertex, an adversary is allowed to delete a (possibly empty) set of at most δn vertices.

Theorem 18.12. *For any sufficiently small constant δ there exists a sufficiently large constant $m = m(\delta)$ and a constant $\theta = \theta(\delta, m)$ such that w.h.p. G_n has a "giant" connected component with size at least θn.*

The proof of this is quite complicated, but it does illustrate some new ideas over and above what we have seen so far in this book.

In the theorem above, the constants are phrased to indicate the suspected relationship, although we do not attempt to optimize them. Our unoptimized calculations work for $\delta \leq 1/50$ and $m \geq \delta^{-2} \times 10^8$ and $\theta = 1/30$.

The proof of Theorem 18.12 is based on an idea developed by Bollobas and Riordan in [176]. There they couple the graph G_n with $G(n,p)$, the Bernoulli random graph, which has vertex set $[n]$ and each pair of vertices appears as an edge independently with probability p. We couple a carefully chosen induced subgraph of G_n with $G(n',p)$.

To describe the induced subgraph in our coupling, we now make a few definitions. We say that a vertex v of G_t is *good* if it was created after time $t/2$ and the number of its *original edges* that remain undeleted exceeds $m/6$. By original edges of v, we mean the m edges that were created when v was added. Let Γ_t denote the set of good vertices of G_t and $\gamma_t = |\Gamma_t|$. We say that a vertex of G_t is *bad* if it is not good. Notice that once a vertex becomes bad it remains bad for the rest of the process. On the other hand, a vertex that was good at time t_1 can become bad at a later time t_2, simply because it was created at a time before $t_2/2$.

Let
$$p = \frac{m}{1500n}$$
and let $\sim$ denote "has the same distribution as".

Theorem 18.13. *For any sufficiently small constant δ there exists a sufficiently large constant $m = m(\delta)$ such that we can couple the construction of G_n and a random graph H_n, with vertex set Γ_n, such that $H_n \sim G(\gamma_n, p)$ and w.h.p. $|E(H_n) \setminus E(G_n)| \leq 10^{-3} e^{-\delta^2 m/10^7} mn$.*

In Section 18.3 we prove Theorem 18.13. In Section 18.3 we prove a lower bound on the number of good vertices, a key ingredient for the proof of Theorem 18.12, given in section 18.3.

Proof of Theorem 18.12.

We will prove the following two lemmas in Section 18.3.

Lemma 18.14. *Let G obtained by deleting fewer than $n/100$ edges from a realization of $G_{n,c/n}$. If $c \geq 10$ then w.h.p. G has a component of size at least $n/3$.*

Lemma 18.15. W.h.p., *for all t with $n/2 < t \leq n$ we have $\gamma_t \geq t/10$.*

With these lemmas, the proof of Theorem 18.12 is only a few lines:

Let $G = G_n$ and $H = G(\gamma_n, p)$ be the graphs constructed in Theorem 18.13. Let $G' = G \cap H$. Then $E(H) \setminus E(G') = E(H) \setminus E(G)$ and so w.h.p. $|E(H) \setminus E(G')| \leq 10^{-3} e^{-\delta^2 m/10^7} mn$. By Lemma 18.15,$|G'| = \gamma_n \geq n/10$ w.h.p. Since m is large enough, $p = m/1500n > 10/\gamma_n$ and $10^{-3} e^{-\delta^2 m/10^7} mn < n/1000 \leq \gamma_n/100$. Then, by Lemma 18.14, w.h.p. G' (and therefore G) has a component of size at least $|G'|/3 \geq n/30$. □

The Coupling: Proof of Theorem 18.13.

We construct $G[k] \sim G_k$ and $H[k] \sim G(\gamma_k, p)$ for $k \geq n/2$ inductively. $G[k]$ will be constructed by following the definition of the preferential attachment process $\mathscr{P}$ and $H[k]$ will be constructed by coupling its construction with the construction of $G[k]$.

For $k \leq n/2$, we only make the size of $H[k]$ correct and do not try to make the edge structure look like $G[k]$. Thus we just take $H[n/2]$ to be an independent copy of $G(\gamma_{n/2}, p)$ with vertex set $\Gamma_{n/2}$.

For $k > n/2$, having constructed $G[k]$ and $H[k]$ with $G[k] \sim G_k$ and $H[k] \sim G(\gamma_k, p)$, we construct $G[k+1]$ and $H[k+1]$ as follows: Let $G[k] = (V_k, E_k)$, and let $\nu_k = |V_k|$, $\eta_k = |E_k|$ and recall that the number of good vertices is denoted $\gamma_k = |\Gamma_k|$.

If $\gamma_k < k/10$ then we call this a *failure of type 0* and generate $G[n]$ and $H[n]$ independently. (By Lemma 18.15 the probability of occurrence of this event is $o(1)$.)

Otherwise we have $\gamma_k \geq k/10$. In this case, to construct $G[k+1]$ process $\mathscr{P}$ adds vertex x_{k+1} to $G[k]$ and links it to vertices $t_1, \ldots, t_m \in V_k$ chosen according to the preferential attachment rule. To construct $H[k+1]$, let $\{t_1, \ldots, t_r\} = \{t_1, \ldots, t_m\} \cap \Gamma_k$ be the subset of selected vertices that are good at time k. Let $\varepsilon_0 = 1/120$. If r, the number of good vertices selected, is less than $(1-\delta)\varepsilon_0 m$ then we call this a *type 1 failure* and generate $H[k+1]$ by joining x_{k+1} to each vertex in $H[k]$ independently with probability p.

Since the number of good vertices $|\Gamma_k| = \gamma_k \geq k/10$ and any $v \in \Gamma_k$ is still incident to at least $m/6$ of its original edges and $\eta_k \leq mk$, we have

$$\Pr(t_i \in \Gamma_k) = \sum_{v \in \Gamma_k} \frac{\deg_{G[k]}(v)}{2\eta_k} \geq \frac{k}{10} \frac{m}{6} \frac{1}{2mk} = \varepsilon_0.$$

So, by comparing r with a Binomial random variable, we obtain an exponential upper bound on the probability of a type 1 failure:

$$\Pr(r \le m\varepsilon_0(1-\delta/2)) \le \Pr(\text{Bin}(m,\varepsilon_0) \le (1-\delta/2)m\varepsilon_0) \\ \le e^{-\delta^2\varepsilon_0 m/8} = e^{-\delta^2 m/960}.$$

Now for every $i = 1,\dots,m$ and for every $v \in \Gamma_k$,

$$\Pr(t_i = v) = \frac{\deg_{G[k]}(v)}{2\eta_k} \ge \frac{m}{12mk} = \frac{1}{12k}$$

Let $\perp$ be a new symbol. For each $i = 1,\dots,r$ we choose $s_i \in \Gamma_k \cup \{\perp\}$ such that for each $v \in \Gamma_k$ we have $\Pr(s_i = v) = \frac{1}{12k}$. We couple the selection of the s_i's with the selection of the t_i's such that if $s_i \ne \perp$ then $s_i = t_i$. Let $S = \{s_i : i = 1,\dots,r\} \setminus \{\perp\}$ and $X = |S|$. Let $Y \sim \text{Bin}(\gamma_k, p)$. If $r \ge m\varepsilon_0(1-\delta/2)$ then

$$\mathbb{E}(X) \ge r\frac{\gamma_k}{12k} - \binom{m}{2}\frac{1}{\gamma_k} \ge (1-\delta/2)\varepsilon_0 m\frac{\gamma_k}{12n} - \frac{200m^2}{n^2} \\ \ge (1+\delta)\gamma_k p = (1+\delta)\,\mathbb{E}(Y).$$

Since $\mathbb{E}(X) \ge (1+\delta)\,\mathbb{E}(Y)$, the probability that $(1+\delta/2)Y > X$ is at most the probability that X or Y deviates from its mean by a factor of $\delta/5$. And, since

$$\mathbb{E}(X) \ge \mathbb{E}(Y) = \gamma_k p \ge \frac{k}{10}\frac{m}{1500n} \ge \frac{m}{30000}.$$

By Chernoff's bound, $\Pr(Y \ge (1+\delta/5)\,\mathbb{E}[Y])$ is at most $e^{-\delta^2 m/10^7}$.
It follows from Azuma's inequality that for any $u > 0$, $\Pr(|X - E(X)| > u) \le e^{-u^2/(2r)}$. This is because X is determined by r independent trials and changing the outcome of a single trial can only change X by at most 1. Putting $u = \delta\,\mathbb{E}(X)/5$ we get

$$\Pr(X \le (1-\delta/5)\,\mathbb{E}(X)) \le e^{-\delta^2 r/50} \le e^{-\delta^2 m/12000}.$$

We say we have a *type 2 failure* if $Y > X$, so we have a type 2 failure with probability at most $2e^{-\delta^2 m/10^7}$. In which case we generate $H[k+1]$ by joining x_{k+1} to each vertex in $H[k]$ independently with probability p.

Conditioning on X, the s_i's form a subset S of Γ_k of size X chosen uniformly at random from all of these subsets. We choose S_1 uniformly at random between all the subsets of Γ_k of size Y, coupling the selection of S_1 to the selection of S such that $S_1 \subseteq S$ when $Y \le X$. Now, to generate $H[k+1]$, we join x_{k+1} to every vertex in S_1 (deterministically).

After the adversary deletes a (possible empty) set of vertices in $G[k]$, we delete all the vertices $H[k]$ that don't belong to Γ_{k+1}, possibly including $x_{\lfloor (k+1)/2 \rfloor}$, simply because of its age.

For $k \geq n/2$ this process yields an $H[k]$ with vertex set Γ_k and identically distributed with $G(\gamma_k, p)$, so we have $H[n] \sim G(\gamma_n, p)$.

We call an edge e in $H[n]$ *misplaced* if e is not an edge of $G[n]$. We are interested in bounding the number of misplaced edges. Misplaced edges can only be created when we have a failure. The probability of having a type 1 or 2 failure at step k is at most $3e^{-\delta^2 m/10^7}$. Let M_k denote the number of misplaced edges created between good vertices when we have a failure of type 1 or 2 at step k. Then M_k is stochastically smaller than $Y_k \sim \mathrm{Bin}(\gamma_k, p)$ and thus stochastically dominated by $Z_k \sim \mathrm{Bin}(n, p)$, a binomial random variable with $\mathbb{E}[Z_k] = np = m/1500$.

Let M denote the total number of misplaced edges at time n. Let θ_k be the indicator for an error of type 1 or 2 at step k. Thus,

$$M \leq \sum_{k=n/2}^{n} M_k \leq \sum_{k=n/2}^{n} Z_k \theta_k.$$

Note that Z_k is independent of θ_k and $\Pr(\theta_k = 1) \leq \rho = 3e^{-\delta^2 m/10^7}$, regardless of the value of $\theta_{k'}$, $k' \neq k$. Thus M is stochastically dominated by

$$M^* = \sum_{k=n/2}^{n} Z_k \zeta_k$$

where the ζ_k are independent Bernouilli random variables with $\Pr(\zeta_k = 1) = \rho$.

$$\mathbb{E}(M^*) \leq \sum_{k=n/2}^{n} 3e^{-\delta^2 m/10^7} m/1500 = \frac{\rho mn}{3000}.$$

and

$$\begin{aligned}
&\Pr\left(M^* > \frac{(1+\delta)\rho mn}{3000}\right) \\
&\leq \Pr\left(M^* > \frac{(1+\delta)\rho mn}{3000} \;\middle|\; \sum_{k=n/2}^{n} \zeta_k \leq \frac{n}{2}\rho\left(1+\frac{\delta}{3}\right)\right) + \Pr\left(\sum_{k=n/2}^{n} \zeta_k > \frac{n}{2}\rho\left(1+\frac{\delta}{3}\right)\right) \\
&\leq \Pr(\mathrm{Bin}\left(\frac{n^2}{2}\rho\left(1+\frac{\delta}{3}\right), p\right) > \left(1+\frac{\delta}{3}\right)^2 \frac{n^2}{2}\rho p) + \Pr\left(\mathrm{Bin}\left(\frac{n}{2}, \rho\right) > \frac{n}{2}\rho\left(1+\frac{\delta}{3}\right)\right) \\
&\leq \exp\left(-\frac{\delta^2 n^2 \rho\left(1+\frac{\delta}{3}\right)p}{54}\right) + \exp\left(-\frac{n\rho\delta^2}{54}\right) \\
&= \exp\left(-\frac{\delta^2 nm\rho\left(1+\frac{\delta}{3}\right)}{90000}\right) + \exp\left(-\frac{n\rho\delta^2}{54}\right)
\end{aligned}$$

$$\leq \exp\left(-10^{-5}\delta^2\left(1+\frac{\delta}{3}\right)nm\rho\right)$$

$\square$

Bounding the number of good vertices: Proof of Lemma 18.15.

We now prove Lemma 18.15, which is restated here for convenience.

Lemma 18.16. W.h.p., *for all t with $n/2 < t \leq n$ we have $\gamma_t \geq t/10$.*

Proof. Let z_t denote the total number of edges created after time $t/2$ that have been deleted by the adversary, up to time t. Let ν_t' and η_t' be the number of vertices and edges respectively in G_t that were created after time $t/2$. Notice that $\eta_t' = mt/2 - z_t$ and $\nu_t' \leq t/2$. Also, since each vertex contributes at most m edges, and bad vertices (not in Γ_t) contribute at most $m/6$ edges, we have $\eta_t' \leq m\gamma_t + m(\nu_t' - \gamma_t)/6$. So

$$\gamma_t \geq \frac{6\eta_t' - m\nu_t'}{5m} \geq \frac{3mt - 6z_t - mt/2}{5m} = \frac{t}{2} - \frac{6z_t}{5m},$$

So if $z_t \leq mt/3$ then $\gamma_t \geq t/10$. Thus, to prove the lemma, it is sufficient to show that

$$\Pr\left(z_t \geq \frac{mt}{3}\right) \leq e^{-\delta^2 mn/10}. \tag{18.29}$$

To show that inequality (18.29) holds, we will compare our process with another process $\mathscr{P}^\star$ in which the adversary deletes no vertices until time t and then deletes the same set of vertices as in $\mathscr{P}$.

Fix $t \geq n/2$. We begin by showing that we can couple the $\mathscr{P}$ and $\mathscr{P}^\star$ in such way that for

$$t_0 = 1000\delta n,$$

$$\Pr(z_t(\mathscr{P}) \geq z_t(\mathscr{P}^\star) + mt_0) = O\left(ne^{-\delta^2 mn/7}\right). \tag{18.30}$$

(The reason for this choice of t_0 is inequality (18.32) in Lemma 18.17).

Generate G_s for $s = 1, \ldots, t$ by process $\mathscr{P}$. Let $D_1, D_2, \ldots$ be the sequence of vertex sets deleted by the adversary in this realization of $\mathscr{P}$. Let $D = \bigcup_{\tau=1}^{t} D_\tau$ denote the set of vertices deleted by the adversary by time t.

We define $G_s^\star$ inductively. To begin, generate $G_{t_0}^\star$ according to preferential attachment (with no adversary). For every s with $t_0 \leq s < t$ let $G_s = (V_s, E_s)$ and $G_s^\star = (V_s^\star, E_s^\star)$. Define $X_s = E_s^\star \setminus E_s$, the set of edges that have been deleted by the adversary's moves.

Selecting a vertex by preferential attachment is equivalent to choosing an edge uniformly at random and then randomly selecting one of the end points of the edge. So we can view the transition from G_s to G_{s+1} as adding x_{s+1} to G_s, choosing m edges $e_1, \ldots, e_m$ (here with replacement), and for each i, selecting a random endpoint y_i of e_i and adding an edge between x_{s+1} and y_i.

To construct $G^\star_{s+1}$, we first add x_{s+1} to $G^\star_s$. To choose $y^\star_1, \ldots, y^\star_m$ we apply the following procedure, for each i:

- With probability $1 - |X_s|/(ms)$ we set $e^\star_i = e_i$ and $y^\star_i = y_i$
- With probability $|X_s|/(ms)$, we choose $e^\star_i$ uniformly at random from X_s. Notice that $e^\star_i$ has already been deleted from G_s by the adversary and therefore it is incident to at least one deleted vertex, $v_i \in D$. Now, we randomly choose $y^\star_i$ from the two end points of $e^\star_i$. If the total degree T_s of the vertices $V_s \cap D$ that $\mathscr{P}$ will delete in the future is at most $ms/2$ then $\Pr(y_i \in D) \leq 1/2$ and we can couple the (random) decisions in such way that if y_i is going to be deleted by time t then $y^\star_i = v_i$. Otherwise we say we have a *failure* and choose $y^\star_i$ independently of y_i.

In the coupling, after time t_0 and before the first failure, an edge incident with x_{s+1} and destined for deletion in $\mathscr{P}$ is matched with an edge incident with x_{s+1} and destined for deletion in $\mathscr{P}^\star$. So, until the first failure, T_s is bounded by $T^\star_s$, the corresponding total degree of $V_s \cap D$ in $G^\star_s$. In Lemma 18.17 below, we prove that $\Pr(T^\star_s > sm/2) = O\left(e^{-\delta^2 mn/6}\right)$ and therefore the probability of having a failure is $O\left(ne^{-\delta^2 mn/6}\right) = O\left(e^{-\delta^2 mn/7}\right)$.

To repeat, if there is no failure and if e_i is deleted in $\mathscr{P}$ before time t we have two possibilities: x_{s+1} is deleted or y_i is deleted. In either case, x_{s+1} or $y^\star_i$ will be deleted by time t in $\mathscr{P}^\star$ and therefore $e^\star_i$ will be deleted, and Equation (18.30) follows.

We will show that

$$\Pr\left(z_t(\mathscr{P}^\star) \geq \frac{mt}{4}\right) \leq O(e^{-\delta^2 mn}), \tag{18.31}$$

and then Inequality (18.29) follows from Equation (18.30).

To prove Inequality (18.31) let s be such that $t/2 \leq s \leq t$ and $x_s \notin D$. We want to upper bound the probability in the process $\mathscr{P}^\star$ that an edge created at time s chooses its end point in D. For $i = 1, \ldots, m$,

$$\Pr(y^\star_i \in D \mid T^\star_s) = \frac{T^\star_s}{2ms}.$$

By Lemma 18.17 (below), we have $\Pr(T^\star_s \geq mt/2) \leq O(e^{-\delta^2 mn})$ so

$$\Pr(y^*_i \in D) \leq \frac{1}{4} + o(1).$$

Therefore $z_t(\mathscr{P}^\star)$ is stochastically dominated by Bin$\left(\frac{mt}{2}, \frac{1}{4} + o(1)\right)$. Inequality (18.31) now follows from Chernoff's bound. This completes the proof of Lemma 18.16. $\square$

Lemma 18.17. *Let $A \subset \{x_1, \dots, x_t\}$, with $|A| \le \delta n$. Let $t \ge 1000\delta n$ and let G_t be a graph generated by preferential attachment (i.e. the process $\mathscr{P}$, but without an adversary). Let T_A denote the total degree of the vertices in A. Then*

$$\Pr(\exists A : \; T_A \ge mt/2) = O\left(e^{-\delta^2 mn}\right).$$

Proof. Let $A' = \{x_1, \dots, x_{\delta n}\}$ be the set of the oldest δn vertices. We can couple the construction of G_t with G'_t, another graph generated by preferential attachment, such that $T_{A'} \ge T_A$. Therefore $\Pr(T_A \ge mt) \le \Pr(T_{A'} \ge mt)$, and we can assume $A = A'$.

Now we consider the process $\mathscr{P}$ in δ^{-1} rounds, Each round consisting of δn steps. Let T_i be the total degree of A at the end of the ith round. Notice that $T_1 = 2\delta mn$ and $T_2 \le 3\delta mn$. For $i \ge 2$, fix s with $i\delta n < s \le (i+1)\delta n$. Then the probability that x_s chooses a vertex in A is at most $\frac{T_i + \delta mn}{2i\delta mn}$. So given T_i, the difference $T_{i+1} - T_i$ is stochastically dominated by $Y_i \sim$ Bin$\left(\delta mn, \frac{T_i + \delta mn}{2i\delta mn}\right)$.

Therefore, for $i \ge 2$,

$$\begin{aligned}
\Pr(T_{i+1} \ge 3i^{2/3}\delta mn) &\le \Pr(T_{i+1} \ge 3i^{2/3}\delta mn \mid T_i \le 3(i-1)^{2/3}\delta mn) \\
&\qquad + \Pr(T_i \ge 3(i-1)^{2/3}\delta mn) \\
&\le \Pr(T_{i+1} \ge 3i^{2/3}\delta mn \mid T_i = 3(i-1)^{2/3}\delta mn) \\
&\qquad + \Pr(T_i \ge 3(i-1)^{2/3}\delta mn).
\end{aligned}$$

Now, for $i \ge 2$, we have $3(i^{2/3} - (i-1)^{2/3}) \le 2i^{-1/3}$ and then

$$\begin{aligned}
&\Pr(T_{i+1} \ge 3i^{2/3}\delta mn \mid T_i = 3(i-1)^{2/3}\delta mn) \\
&\quad \le \Pr(Y_i \ge 3(i^{2/3} - (i-1)^{2/3})\delta mn \mid T_i = 3(i-1)^{2/3}\delta mn) \\
&\quad \le \Pr(Y_i \ge 2i^{-1/3}\delta mn \mid T_i = 3(i-1)^{2/3}\delta mn)
\end{aligned}$$

As $Y_i \sim$ Bin$\left(\delta mn, \frac{T_i + \delta mn}{2i\delta mn}\right)$

$$\mathbb{E}[Y_i | T_i = 3(i-1)^{2/3}\delta mn] = \left(\frac{3(i-1)^{2/3}+1}{2i}\right)\delta mn \le \frac{4}{3} i^{-1/3}\delta mn.$$

Since and $i \le \delta^{-1}$, by Chernoff's bound we have

$$\Pr(T_{i+1} \ge 3i^{2/3}\delta mn \mid T_i = 3(i-1)^{2/3}\delta mn) \le e^{-\delta^{4/3}mn/9}.$$

Hence, for any $k \leq \delta^{-1}$,

$$\Pr(T_k > 3(k-1)^{2/3}\delta mn) \leq \sum_{i=2}^{k-2} e^{-\delta^{4/3}mn/9} \leq e^{-2\delta^2 mn}.$$

Now, if $t \geq t_0$ then

$$k = \left\lfloor \frac{t}{\delta n} \right\rfloor \geq 10^3 \tag{18.32}$$

and so

$$3(k-1)^{2/3}\delta mn \leq tm/2.$$

Thus

$$\Pr(T_t \geq tm/2) \leq e^{-2\delta^2 mn}.$$

We inflate the above by $\binom{n}{\delta n}$ to get the bound in the lemma. □

Proof of Lemma 18.14 If after deleting $n/100$ edges the maximum component size is at most $n/3$ then $G_{n,c/n}$ contains a set S of size $n/3 \leq s \leq n/2$ such that there are at most $n/100$ edges joining S to $V \setminus S$. The expected number of edges across this cut is $s(n-s)c/n$ so when $1-\varepsilon = \frac{9}{200c}$ we have $n/100 \leq (1-\varepsilon)s(n-s)c/n$ and by applying the union bound and Chernoff's bound we have

$$\begin{aligned}\Pr(\exists S) &\leq \sum_{s=n/3}^{n/2} \binom{n}{s} e^{-\varepsilon^2 s(n-s)c/(2n)} \\ &\leq \sum_{s=n/3}^{n/2} \left(\frac{ne}{s} e^{-\varepsilon^2 (n-s)c/(2n)}\right)^s \\ &= o(1).\end{aligned}$$

□

18.4 Bootstrap Percolation

This is a simplified mathematical model of the spread of a disease through a graph/network $G = (V,E)$. Initially a set A_0 of vertices are considered to be infected. This is considered to be round 0. Then in round $t > 0$ any vertex that has at least r neighbors in A_{t-1} will become infected. No-one recovers in this model. The main question is as to how many vertices eventually end up getting infected. There is a large literature on this subject with a variety of graphs G and ways of defining A_0.

Here we will assume that each vertex s is placed in A_0 with probability p, independent of other vertices. The proof of the following theorem relies on the fact that with high probability all of the early vertices of G_t become infected during the first round. Subsequently, the connectivity of the random graph is enough to spread the infection to the remaining vertices. The following is a simplified version of Theorem 1 of Abdulla and Fountoulakis [1].

Theorem 18.18. *If $r \leq m$ and $\omega = \log^2 t$ and $p \geq \omega t^{-1/2}$ then w.h.p. all vertices in G_t get infected.*

Proof. Given Theorem 18.7 we can assume that $d_s(t) \geq mt^{1/2}/\omega^{1/2}$ for $1 \leq s \leq m$. In which case, the probability that vertex $s \leq m$ is not infected in round 1 is at most

$$\sum_{i=1}^{m-1} \binom{mt^{1/2}/\omega^{1/2}}{i} p^i (1-p)^{mt^{1/2}/\omega^{1/2}-i} \leq \sum_{i=1}^{m-1} \omega^{i/2} e^{-(1-o(1))m\omega^{1/2}} = o(1).$$

So, w.h.p. $1, 2, \ldots, m$ are infected in round 1. After this we use induction and the fact that every vertex $i > s$ has m neighbors $j < i$. $\square$

18.5 A General Model of Web Graphs

In the model presented in the previous section a new vertex is added at time t and this vertex chooses m random neighbors, with probability proportional to their current degree. Cooper and Frieze [254] generalise this in the following ways: they allow (a) new edges to be inserted between existing vertices, (b) a variable number of edges to be added at each step and (c) endpoint vertices to be chosen by a mixture of uniform selection and copying. This results in a large number of parameters, which will be described below. We first give a precise description of the process.

Initially, at step $t = 0$, there is a single vertex v_0. At any step $t = 1, 2, \ldots, T, \ldots$, there is a birth process in which either new vertices or new edges are added. Specifically, either a procedure NEW is followed with probability $1 - \alpha$, or a procedure OLD is followed with probability α. In procedure NEW, a new vertex v is added to G_{t-1} with one or more edges added between v and G_{t-1}. In procedure OLD, an existing vertex v is selected and extra edges are added at v.

The recipe for adding edges at step t typically permits the choice of initial vertex v (in the case of OLD) and of terminal vertices (in both cases) to be made from G_{t-1} either u.a.r (uniformly at random) or according to vertex degree, or a mixture of these two based on further sampling. The number of edges added to vertex v at step t by the procedures (NEW, OLD) is given by distributions specific to the procedure.

Notice that the edges have an intrinsic direction, arising from the way they are inserted, which one can ignore or not. Here the undirected model is considered with a sampling procedure based on vertex degree. The process allows multiple edges, and self-loops can arise from the OLD procedure. The NEW procedure, as described, does not generate self-loops, although this could easily be modified.

Sampling parameters, notation and main properties

Our undirected model G_t has sampling parameters $\alpha, \beta, \gamma, \delta, p, q$ whose meaning is given below:

Choice of procedure at step t.

α: Probability that an OLD vertex generates edges.

$1-\alpha$: Probability that a NEW vertex is created.

Procedure NEW

$p = (p_i : i \geq 1)$: Probability that the new node generates i new edges.

β: Probability that choices of terminal vertices are made uniformly.

$1-\beta$: Probability that choices of terminal vertices are made according to degree.

Procedure OLD

$q = (q_i : i \geq 1)$: Probability that the old node generates i new edges.

δ: Probability that the initial node is selected uniformly.

$1-\delta$: Probability that the initial node is selected according to degree.

γ: Probability that choices of terminal vertices are made uniformly.

$1-\gamma$: Probability that choices of terminal vertices are made according to degree.

The models require $\alpha < 1$ and $p_0 = q_0 = 0$. It is convenient to assume a *finiteness* condition for the distributions $\{p_j\}$, $\{q_j\}$. This means that there exist j_0, j_1 such that $p_j = 0,\ j > j_0$ and $q_j = 0,\ j > j_1$. Imposing the finiteness condition helps simplify the difference equations used in the analysis.

The model creates edges in the following way: An initial vertex v is selected. If the terminal vertex w is chosen u.a.r, we say v is *assigned uniformly* to w. If the terminal vertex w is chosen according to its vertex degree, we say v is *copied* to w. In either case the edge has an intrinsic direction (v,w), which we may choose to ignore. Note that sampling according to vertex degree is equivalent to selecting an edge u.a.r and then selecting an endpoint u.a.r.

Let

$$\mu_p = \sum_{j=1}^{j_0} jp_j,\ \mu_q = \sum_{j=1}^{j_1} jq_j$$

be the expected number of edges added by NEW or OLD and let

$$\theta = 2((1-\alpha)\mu_p + \alpha\mu_q).$$

To simplify subsequent notation, we introduce new parameters as follows:

$$\begin{aligned}
a &= 1+\beta\mu_p+\frac{\alpha\gamma\mu_q}{1-\alpha}+\frac{\alpha\delta}{1-\alpha},\\
b &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta}+\frac{\alpha(1-\gamma)\mu_q}{\theta}+\frac{\alpha(1-\delta)}{\theta},\\
c &= \beta\mu_p+\frac{\alpha\gamma\mu_q}{1-\alpha},\\
d &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta}+\frac{\alpha(1-\gamma)\mu_q}{\theta},\\
e &= \frac{\alpha\delta}{1-\alpha},\\
f &= \frac{\alpha(1-\delta)}{\theta}.
\end{aligned}$$

We note that

$$c+e=a-1 \text{ and } b=d+f. \tag{18.33}$$

Now define the sequence $(d_0,d_1,\dots,d_k,\dots)$ by $d_0=0$, and for $k\geq 1$

$$d_k(a+bk)=(1-\alpha)p_k+(c+d(k-1))d_{k-1}+\sum_{j=1}^{k-1}(e+f(k-j))q_jd_{k-j}. \tag{18.34}$$

For convenience we define $d_k=0$ for $k<0$. Since $a\geq 1$, this system of equations has a unique solution.

The main quantity we study is the random variable $D_k(t)$, the number of vertices of degree k at step t. Cooper and Frieze [254] prove that, as $t\to\infty$, for small k, $\overline{D}_k(t)\approx d_kt$.

Theorem 18.19. *There exists a constant $M>0$ such that almost surely for all t, $k\geq 1$*

$$|\overline{D}_k(t)-td_k|\leq Mt^{1/2}\log t.$$

This will be proved in Section 18.5.

It is shown in (18.35), that the number of vertices $\nu(t)$ at step t is w.h.p. asymptotic to $(1-\alpha)t$. It follows that the proportion of vertices of degree k is w.h.p. asymptotic to

$$\bar{d}_k=\frac{d_k}{1-\alpha}.$$

The next theorem summarises what is known about the sequence (d_k) defined by (18.34).

Theorem 18.20. *There exist constants $C_1, C_2, C_3, C_4 > 0$ such that*

(i) $C_1 k^{-\zeta} \leq d_k \leq C_2 \min\{k^{-1}, k^{-\zeta/j_1}\}$ *where* $\zeta = (1+d+f\mu_q)/(d+f)$.

(ii) *If* $j_1 = 1$ *then* $d_k \approx C_3 k^{-(1+1/(d+f))}$.

(iii) *If* $f = 0$ *then* $d_k \approx C_4 k^{-(1+1/d)}$.

Evolution of the degree sequence of G_t

Let $\nu(t) = |V(t)|$ be the number of vertices and let $\eta(t) = |2E(t)|$ be the total degree of the graph at the end of step t. $\mathbb{E}\,\nu(t) = (1-\alpha)t$ and $\mathbb{E}\,\eta(t) = \theta t$. The random variables $\nu(t)$, $\eta(t)$ are sharply concentrated provided $t \to \infty$. Indeed $\nu(t)$ has distribution $\mathrm{Bin}(t, 1-\alpha)$ and so by Theorem 27.6 and its corollaries,

$$\mathbb{P}(|\nu(t) - (1-\alpha)t| \geq t^{1/2}\log t) = O(t^{-K}) \tag{18.35}$$

for any constant $K > 0$.

Similarly, $\eta(t)$ has expectation θt and is the sum of t independent random variables, each bounded by $\max\{j_0, j_1\}$. Hence, by Theorem 27.6 and its corollaries,

$$\mathbb{P}(|\eta(t) - \theta t| \geq t^{1/2}\log t) = O(t^{-K}) \tag{18.36}$$

for any constant $K > 0$.

These results are almost sure in the sense that they hold for all $t \geq t_0$ with probability $1 - O(t_0^{-K+1})$. Thus we can focus on processes such that this is true.

We remind the reader that $D_k(t)$ is the number of vertices of degree k at step t and that $\overline{D}_k(t)$ is its expectation. Here $\overline{D}_j(t) = 0$ for all $j \leq 0, t \geq 0$, $\overline{D}_1(0) = 1$, $\overline{D}_k(0) = 0$, $k \geq 2$.

Using (18.35) and (18.36) we see that

$$\overline{D}_k(t+1) = \overline{D}_k(t) + (1-\alpha)p_k + O(t^{-1/2}\log t) \tag{18.37}$$

$$+ (1-\alpha)\sum_{j=1}^{j_0} p_j \left(\beta\left(\frac{j\overline{D}_{k-1}(t)}{(1-\alpha)t} - \frac{j\overline{D}_k(t)}{(1-\alpha)t}\right)\right.$$

$$\left. + (1-\beta)\left(\frac{j(k-1)\overline{D}_{k-1}(t)}{\theta t} - \frac{jk\overline{D}_k(t)}{\theta t}\right)\right) \tag{18.38}$$

$$- \alpha\left(\frac{\delta\overline{D}_k(t)}{(1-\alpha)t} + \frac{(1-\delta)k\overline{D}_k(t)}{\theta t}\right)$$

$$+ \alpha\sum_{j=1}^{j_1} q_j \left(\frac{\delta\overline{D}_{k-j}(t)}{(1-\alpha)t} + \frac{(1-\delta)(k-j)\overline{D}_{k-j}(t)}{\theta t}\right) \tag{18.39}$$

$$+\alpha\sum_{j=1}^{j_1} jq_j\left(\gamma\left(\frac{\overline{D}_{k-1}(t)}{(1-\alpha)t}-\frac{\overline{D}_k(t)}{(1-\alpha)t}\right)+\right.$$
$$\left.(1-\gamma)\left(\frac{(k-1)\overline{D}_{k-1}(t)}{\theta t}-\frac{k\overline{D}_k(t)}{\theta t}\right)\right). \tag{18.40}$$

Here (18.38), (18.39), (18.40) are (respectively) the main terms of the change in the expected number of vertices of degree k due to the effect on: terminal vertices in NEW, the initial vertex in OLD and the terminal vertices in OLD. Rearranging the right hand side, we find:

$$\begin{aligned}
&\overline{D}_k(t+1)=\overline{D}_k(t)+(1-\alpha)p_k+O(t^{-1/2}\log t)\\
&-\frac{\overline{D}_k(t)}{t}\left(\beta\mu_p+\frac{\alpha\gamma\mu_q}{1-\alpha}+\frac{\alpha\delta}{1-\alpha}+\frac{(1-\alpha)(1-\beta)\mu_p k}{\theta}+\right.\\
&\qquad\qquad\left.+\frac{\alpha(1-\gamma)\mu_q k}{\theta}+\frac{\alpha(1-\delta)k}{\theta}\right)\\
&+\frac{\overline{D}_{k-1(t)}}{t}\left(\beta\mu_p+\frac{\alpha\gamma\mu_q}{1-\alpha}+\frac{(1-\alpha)(1-\beta)\mu_p(k-1)}{\theta}+\right.\\
&\qquad\qquad\left.+\frac{\alpha(1-\gamma)\mu_q(k-1)}{\theta}\right)\\
&+\sum_{j=1}^{j_1} q_j\frac{\overline{D}_{k-j}(t)}{t}\left(\frac{\alpha\delta}{1-\alpha}+\frac{\alpha(1-\delta)(k-j)}{\theta}\right).
\end{aligned}$$

Thus for all $k\geq 1$ and almost surely for all $t\geq 1$,

$$\begin{aligned}
\overline{D}_k(t+1)&=\overline{D}_k(t)+(1-\alpha)p_k+O(t^{-1/2}\log t)\\
&+\frac{1}{t}((1-(a+bk))\overline{D}_k(t)+(c+d(k-1))\overline{D}_{k-1}(t)\\
&\qquad+\sum_{j=1}^{j_1} q_j(e+f(k-j))\overline{D}_{k-j}(t)).
\end{aligned} \tag{18.41}$$

The following Lemma establishes an upper bound on d_k given in Theorem 18.20(i).

Lemma 18.21. *The solution of* (18.34) *satisfies* $d_k\leq\frac{C_2}{k}$.

Proof. Assume that $k>k_0$ where k_0 is sufficiently large, and thus $p_k=0$. Smaller values of k can be dealt with by adjusting C_2. We proceed by induction on k. From

(18.34),

$$\begin{aligned}
(a+bk)d_k &\leq (c+d(k-1))\frac{C_2}{k-1}+\sum_{j=1}^{j_1}(e+f(k-j))q_j\frac{C_2}{k-j} \\
&\leq C_2(d+f)+\frac{C_2(c+e)}{k-j_1} \\
&= C_2b+\frac{C_2(a-1)}{k-j_1},
\end{aligned}$$

from (18.33). So

$$\begin{aligned}
d_k-\frac{C_2}{k} &\leq \frac{C_2b}{a+bk}+\frac{C_2(a-1)}{(k-j_1)(a+bk)}-\frac{C_2}{k} \\
&= \frac{C_2(a-1)}{(k-j_1)(a+bk)}-\frac{C_2a}{k(a+bk)} \\
&\leq 0,
\end{aligned}$$

for $k \geq j_1 a$. $\square$

We can now prove Theorem 18.19, which is restated here for convenience.

Theorem 18.22. *There exists a constant $M > 0$ such that almost surely for $t,\ k \geq 1$,*

$$|\overline{D}_k(t)-td_k| \leq Mt^{1/2}\log t. \tag{18.42}$$

Proof. Let $\Delta_k(t) = \overline{D}_k(t)-td_k$. It follows from (18.34) and (18.41) that

$$\begin{aligned}
\Delta_k(t+1) = \Delta_k(t)\left(1-\frac{a+bk-1}{t}\right)+O(t^{-1/2}\log t) \\
+\frac{1}{t}\left((c+d(k-1))\Delta_{k-1}(t)+\sum_{j=1}^{j_1}(e+f(k-j))q_j\Delta_{k-j}(t)\right).
\end{aligned} \tag{18.43}$$

Let L denote the hidden constant in $O(t^{-1/2}\log t)$. We can adjust M to deal with small values of t, so we assume that t is sufficiently large. Let $k_0(t) = \left\lfloor \frac{t+1-b}{a} \right\rfloor$. If $k > k_0(t)$ then we observe that (i) $D_k(t) \leq \frac{t\max\{j_0,j_1\}}{k_0(t)} = O(1)$ and (ii) $td_k \leq t\frac{C_2}{k_0(t)} = O(1)$ follows from Lemma 18.21, and so (18.42) holds trivially.

Assume inductively that $\Delta_\kappa(\tau) \leq M\tau^{1/2}\log\tau$ for $\kappa+\tau \leq k+t$ and that $k \leq k_0(t)$. Then (18.43) and $k \leq k_0$ implies that for M large,

$$|\Delta_k(t+1)| \leq L\frac{\log t}{t^{1/2}}+Mt^{1/2}\log t\ \times$$

$$\left(1+\frac{1}{t}\left(c+dk+\sum_{j=1}^{j_1}(e+fk)q_j-(a+bk-1)\right)\right)$$
$$=L\frac{\log t}{t^{1/2}}+Mt^{1/2}\log t$$
$$\leq M(t+1)^{1/2}\log(t+1)$$

provided $M \gg 2L$. We have used (18.33) to obtain the second line.

This completes the proof by induction. □

A general power law bound for d_k

The following lemma completes the proof of Theorem 18.20(i).

Lemma 18.23. *For $k > j_0$ we have,*

(i) $d_k > 0$.

(ii) $C_1 k^{-(1+d+f\mu_q)/b} \leq d_k \leq C_2 k^{-(1+d+f\mu_q)/bj_1}$.

Proof. **(i)** Let κ be the first index such that $p_\kappa > 0$, so that, from (18.34), $d_\kappa > 0$. It is not possible for both c and d to be zero. Therefore the coefficient of d_{k-1} in (18.34) is non-zero and thus $d_k > 0$ for $k \geq \kappa$.

(ii) Re-writing (18.34) we see that for $k > j_0$, $p_k = 0$ and then d_k satisfies

$$d_k = d_{k-1}\frac{c+d(k-1)}{a+bk}+\sum_{j=1}^{j_1} d_{k-j}q_j\frac{e+f(k-j)}{a+bk}, \tag{18.44}$$

which is a linear difference equation with rational coefficients (see [683]).

We let $d_i = 0$ for $i < 0$ to handle the cases where $k-j < 0$ in the above sum.

Let $y = 1+d+f\mu_q$, then

$$\frac{c+d(k-1)}{a+bk}+\sum_{j=1}^{j_1} q_j\frac{e+f(k-j)}{a+bk} = 1-\frac{y}{a+bk} \geq 0$$

and thus

$$\left(1-\frac{y}{a+bk}\right)\min\{d_{k-1},\dots,d_{k-j_1}\} \leq d_k \leq \left(1-\frac{y}{a+bk}\right)\max\{d_{k-1},\dots,d_{k-j_1}\}. \tag{18.45}$$

It follows that

$$d_{j_0} \prod_{j=j_0+1}^{k} \left(1 - \frac{y}{a+bj}\right) \leq d_k \leq$$
$$\max\{d_1, d_2, \ldots, d_{j_0}\} \prod_{s=0}^{\lfloor (k-j_0)/j_1 \rfloor} \left(1 - \frac{y}{a+b(k-sj_1)}\right). \quad (18.46)$$

The lower bound in (18.46) is proved by induction on k. It is trivial for $k = j_0$ and for the inductive step we have

$$\begin{aligned} d_k &\geq d_{j_0} \left(1 - \frac{y}{a+bk}\right) \min_{i=j_0,\ldots,k-1} \left\{ \prod_{j=j_0+1}^{i} \left(1 - \frac{y}{a+bj}\right) \right\} \\ &= d_{j_0} \prod_{j=j_0+1}^{k} \left(1 - \frac{y}{a+bj}\right). \end{aligned}$$

The upper bound in (18.46) is proved as follows: Let $d_{i_1} = \max\{d_{k-1}, \ldots, d_{k-j_1}\}$, and in general, let $d_{i_{t+1}} = \max\{d_{i_t-1}, \ldots, d_{i_t-j_1}\}$. Using (18.45) we see there is a sequence $k-1 \geq i_1 > i_2 > \cdots > i_p > j_0 \geq i_{p+1}$ such that $|i_t - i_{t-1}| \leq j_1$ for all t, and $p \geq \lfloor (k-j_0)/j_1 \rfloor$. Thus

$$d_k \leq d_{i_{p+1}} \prod_{t=0}^{p} \left(1 - \frac{y}{a+bi_t}\right),$$

and the RHS of (18.46) now follows.

Now consider the product in the LHS of (18.46).

$$\begin{aligned} &\prod_{j=j_0+1}^{k} \left(1 - \frac{y}{a+bj}\right) \\ &= \exp\left\{ \sum_{j=j_0+1}^{k} \left(-\frac{y}{a+bj} - \frac{1}{2}\left(\frac{y}{a+bj}\right)^2 - \cdots \right) \right\} \\ &= \exp\left\{ O(1) - \sum_{j=j_0+1}^{k} \frac{y}{a+bj} \right\} \\ &= C_1 k^{-y/b}. \end{aligned}$$

This establishes the lower bound of the lemma. The upper bound follows similarly, from the upper bound in (18.46). □

The case $j_1 = 1$

We prove Theorem 18.20(ii). When $q_1 = 1$, $p_j = 0, j > j_0 = \Theta(1)$, the general value of d_k, $k > j_0$ can be found directly, by iterating the recurrence (18.34). Thus

$$d_k = \frac{1}{a+bk} \left(d_{k-1}\left((a-1) + b(k-1)\right)\right)$$

$$\begin{aligned} &= d_{k-1}\left(1-\frac{1+b}{a+bk}\right) \\ &= d_{j_0}\prod_{j=j_0+1}^{k}\left(1-\frac{1+b}{a+jb}\right). \end{aligned}$$

Thus, for some constant $C_6 > 0$,

$$d_k \approx C_6(a+bk)^{-x}$$

where

$$x = 1+\frac{1}{b} = 1+\frac{2}{\alpha(1-\delta)+(1-\alpha)(1-\beta)+\alpha(1-\gamma)}.$$

The case $f = 0$

We prove Theorem 18.20(iii). The case $(f = 0)$ arises in two ways. Firstly if $\alpha = 0$ so that a new vertex is added at each step. Secondly, if $\alpha \neq 0$ but $\delta = 1$ so that the initial vertex of an OLD choice is sampled u.a.r.

Observe that $b = d$ now, see (18.33).

We first prove that for a sufficiently large absolute constant $A_2 > 0$ and for all sufficiently large k, that

$$\frac{d_k}{d_{k-1}} = 1-\frac{1+d}{a+dk}+\frac{\xi(k)}{k^2} \tag{18.47}$$

where $|\xi(k)| \leq A_2$.

We first re-write (18.34) as

$$\frac{d_k}{d_{k-1}} = \frac{c+d(k-1)}{a+dk}+\sum_{j=1}^{j_1}\frac{eq_j}{a+dk}\prod_{t=k-j+1}^{k-1}\frac{d_{t-1}}{d_t}. \tag{18.48}$$

(We assume here that $k > j_0$, so that $p_k = 0$.)

Now use induction to write

$$\prod_{t=k-j+1}^{k-1}\frac{d_{t-1}}{d_t} = 1+(j-1)\frac{d+1}{a+dk}+\frac{\xi^*(j,k)}{k^2} \tag{18.49}$$

where $|\xi^*(j,k)| \leq A_3$ for some constant $A_3 > 0$. (We use the fact that j_1 is constant here.)

Substituting (18.49) into (18.48) gives

$$\frac{d_k}{d_{k-1}} = \frac{c+d(k-1)}{a+dk}+\frac{e}{a+dk}+\frac{e(\mu_q-1)(d+1)}{(a+dk)^2}+\frac{\xi^{**}(k)}{(a+dk)k^2}$$

where $|\xi^{**}(k)| \leq eA_3$.

Equation (18.47) follows immediately from this and $c+e=a-1$. On iterating (18.47) we see that for some constant $C_7 > 0$,

$$d_k \approx C_7 k^{-\left(1+\frac{1}{a}\right)}.$$

$\square$]

18.6 Small World

In an influential paper Milgram [682] describes the following experiment. He chose a person X to receive mail and then randomly chose a person Y to send it. If Y did not know X then Y was to send the mail to someone he/she thought more likely to know X and so on. Surprisingly, the mail got through in 64 out of 296 attempts and the number of links in the chain was relatively small, between 5 and 6. More recently, Kleinberg [572] described a model that attempts to explain this phenomenon.

Watts-Strogatz Model

Milgram's experiment suggests that large real-world networks although being globally sparse, in terms of the number edges, have their nodes/vertices connected by relatively short short paths. In addition, such networks are locally dense, i.e. vertices lying in a small neighborhood of a given vertex are connected by many edges. This observation is called the "small world" phenomenon and it has generated many attempts, both theoretical and experimental to build and study appropriate models of small world networks. Unfortunately, for many reasons, the classical Erdős-Rényi- Gilbert random graph $\mathbb{G}_{n,p}$ is missing many important characteristics of such networks. The first attempt to build more realistic model was introduced by Watts and Strogatz in 1998 in *Nature* (see [847]).
The Watts-Strogatz model starts with a kth power of a n-vertex cycle, denoted here as C_n^k. To construct it fix n and k, $n \geq k \geq 1$ and take the vertex set as $V=[n]=\{1,2,\ldots,n\}$ and edge-set $E=\{\{i,j\}: i+1 \leq j \leq i+k\}$, where the additions are taken modulo n.
In particular, $C_n^1 = C_n$ is a cycle on n vertices. For an example of a square C_n^2 of C_n see Figure 18.6 below.

Notice, that for $n > 2k$ graph C_n^k is $2k$-regular and has nk edges. Now choose each of nk edges of C_n^k, one by one, and independently with small probability p

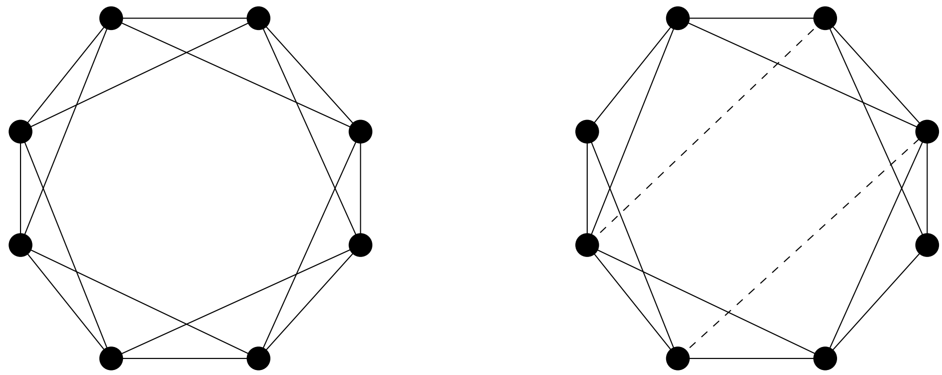


Figure 18.2: C_8 and C_8 after two re-wirings

decide to "rewire" it or leave it unchanged. The procedure goes as follows. We start, say, at vertex labeled 1, and move clockwise k times around the cycle. At the ith pasage of the cycle, at each visted vertex, we take the edge connecting it to its neighbour at distance i to the right and decide, with probability p, if its other endpoint should be replaced by a uniformly random vertex of the cycle. Not however allowing the creation a double edges. Notice that after this k round procedure is completed the number of edges of the Watts-Strogatz random graph is kn, i.e., the same as in "starting" graph C_n^k. To study properties the original Watts-Strogatz model on a formal mathematical ground has proved rather difficult. Therefore Newman and Watts (see [716]) proposed a modified version, where instead of rewiring the edges of C_n^k each of $\binom{n}{2} - nk$ edges not in C_n^k is added independently probability p. In fact this modification, when $k = 1$ was introduced earlier by Ball, Mollison and Scalia-Tomba in [68] as "the great circle" epidemic model. For a rigorous results on typical distances in such random graph see the seminal papers of Barbour and Reinert [78] and [79].

Much earlier Bollobás and Chung in [160] took a similar approach to introducing "shortcuts" in C_n. Namely, let C_n be a cycle with n vertices labeled clockwise $1, 2, \ldots, n$, so that vertex i is adjacent to vertex $i+1$ for $1 \leq i \leq n-1$. Consider the graph G_n obtained by adding a randomly chosen perfect matching to C_n. (We will assume that n is even. For odd n one can add a random near prefect matching.) Note that the graphs generated by this procedure are 3-regular (see Figure 18.3 below).

It is easy to see that a cycle C_n itself has diameter $n/2$. Bollobás and Chung proved that the diameter drops dramatically after adding to C_n such system of random "shortcuts".

Theorem 18.24. *Let G_n be formed by adding a random perfect matching M to an*

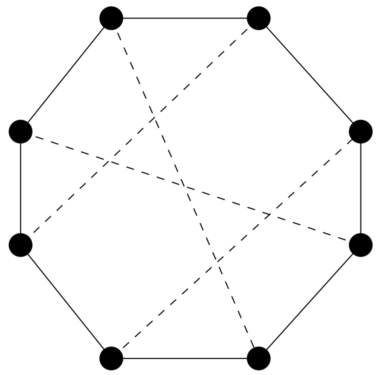


Figure 18.3: $C_8 \cup M$

n-cycle C_n. Then w.h.p.

$$diam(G_n) \leq \log_2 n + \log_2 \log n + 10.$$

Proof. For a vertex u of G_n define sets

$$S_i(u) = \{v : dist(u,v) = i\} \quad \text{and} \quad S_{\leq i}(u) = \bigcup_{j \leq i} S_j(u),$$

where $dist(u,v) = dist_{G_n}(u,v)$ denotes the length of a shortest path between u and v in G_n.
Now define the following process for generating sets $S_i(u)$ and $S_{\leq i}(u)$ in G_n, Start with a fixed vertex u and "uncover" the chord (edge of M) incident to vertex u. This determines set $S_1(u)$. Then we add the neighbours of $S_1(u)$ one by one to determine $S_2(u)$ and proceed to determine $S_i(u)$.
A chord incident to a vertex in $S_i(u)$ is called *"inessential at level i"* if the other vertex in $S_i(u)$ is within distance $3\log_2 n$ in C_n of the vertices determined so far. Notice that $|S_{\leq i}(u) \leq 3 \cdot 2^i$ and so

$$\mathbb{P}(\text{a chord is inessential at level } i \mid S_{\leq i-1}(u)) \leq \frac{18 \cdot 2^{i+1} \log_2 n}{n}. \tag{18.50}$$

Denote by $\mathscr{A}$ the event that for every vertex u at most one of the chords chosen in $S_{\leq i}(u)$ is inessential and suppose that $i \leq \frac{1}{5} \log_2 n$. Then

$$\begin{aligned}\mathbb{P}(\mathscr{A}^c) &= \mathbb{P}(\exists u : \text{at least two of the chords chosen in } S_{\leq i}(u) \text{ are inessential}) \\ &\leq n \binom{3 \cdot 2^{i+1}}{2} \left(\frac{18 \cdot 2^{i+1} \log_2 n}{n} \right)^2 = O\left(n^{-1/5} (\log n)^2\right).\end{aligned}$$

For a fixed vertex u, consider those vertices v in $S_i(u)$ for which there is a unique path from u to v of length i, say $u = u_0, u_1, \ldots, u_{i-1}, u_i = v$, such that

(i) if u_{i-1} is adjacent to v on the cycle C_n then $S_{\leq i}(u)$ contains no vertex on C_n within distance $3\log_2 n$ on the opposite side to v (denote the set of such vertices v by $C_i(u)$),

(ii) if $\{u_{i-1}, v\}$ is a chord then $S_{\leq i}(u) \setminus \{v\}$ contains no vertex within distance, $3\log_2 n$ both to the left and to the right of v (denote the set of such verices by $D_i(u)$).

Obviously,

$$C_i(u) \cup D_i(u) \subseteq S_i(u).$$

Notice that if the event $\mathscr{A}$ holds then, for $i \leq \frac{1}{5}\log_2 n$,

$$|C_i(u)| \geq 2^{i-2} \quad \text{and} \quad |D_i(u)| \geq 2^{i-3}. \tag{18.51}$$

Let $\frac{1}{5}\log_2 n \leq i \leq \frac{3}{5}\log_2 n$. Denote by $\mathscr{B}$ the event that for every vertex u, at most $2^i n^{-1/10}$ inessential chords leave $S_i(u)$. There are at most 2^i chords leaving $S_i(u)$ for such i's and so by (18.50), for large n,

$$\begin{aligned}
\mathbb{P}(\mathscr{B}^c) &= \mathbb{P}(\exists u: \text{ at least } 2^i n^{-1/10} \text{ inessntial chords leave } S_i(u)) \\
&\leq n\binom{2^i}{2^i\, n^{-1/10}}\left(n^{-1/6}\right)^{2^i\, n^{-1/10}} \leq n\left(en^{1/10}n^{-1/6}\right)^{2^i\, n^{-1/10}} \\
&\leq n\left(en^{-1/15}\right)^{n^{1/10}} = O\left(n^{-2}\right).
\end{aligned}$$

For $v \in C_i(u)$ a new neighbor of v in C_n is a potential element of $C_{i+1}(u)$ and a new neighbor, which is the end-vertex of the chord from v, is a potential element of $D_{i+1}(u)$. Also if $v \in D_i(u)$, then the two neigbors of v in C_n are potential elements of $C_{i+1}(u)$. Here "potential" means that the vertices in question become elements of $C_{i+1}(u)$ and $D_{i+1}(u)$ unless the corresponding edge is inessential.

Assuming that the events $\mathscr{A}$ and $\mathscr{B}$ both hold and $\frac{1}{5}\log_2 n \leq i \leq \frac{3}{5}\log_2 n$, then

$$\begin{aligned}
|C_{i+1}(u)| &\geq |C_i(u)| + 2|D_i(u)| - 2^{i+1}n^{-1/10} \\
|D_{i+1}(u)| &\geq |C_i(u)| - 2^{i+1}n^{-1/10},
\end{aligned}$$

while for $i \leq \frac{1}{5}\log_2 n$ the bounds given in (18.51) hold. Hence for all $3 \leq i \leq \frac{3}{5}\log_2 n$ we have

$$|C_i(u)| \geq 2^{i-3} \quad \text{and} \quad |D_i(u)| \geq 2^{i-4}.$$

To finish the proof set

$$i_0 = \left\lceil \frac{\log_2 n + \log_2 \log n + c}{2} \right\rceil,$$

where $c \geq 9$ is a constant.
Let us choose chords leaving $C_{i_0}(u)$ one by one. At each choice the probability of not selecting the other end-vertex in $C_{i_0}(u)$ is at most $1-(2^{i_0-3}/n)$. Since we have to make at least $|C_{i_0}(u)|/2 \geq 2^{i_0-4}$ such choices, we have

$$\mathbb{P}(dist(u,v) > 2i_0+1|\mathscr{A}\cap\mathscr{B}) \leq \left(1-\frac{2^{i_0-3}}{n}\right)^{2^{i_0-4}} \leq \exp(-2^{2i_0-7}/n)$$
$$\leq \exp(-(\log n)2^{c-7}) \leq n^{-4}.$$

Hence,

$$\mathbb{P}(diam(G_n) > 2i_0+1) \leq \mathbb{P}(\mathscr{A}^c)+\mathbb{P}(\mathscr{B}^c)+\sum_{u,v}\mathbb{P}(dist(u,v) > 2i_0+1|\mathscr{A}\cap\mathscr{B})$$
$$\leq c_1(n^{-1/5}(\log n)^2)+c_2n^{-2}+n^{-2} = o(1).$$

Therefore w.h.p. the random graph G_n has diameter at most

$$2\left\lceil\frac{\log_2 n+\log_2 n\log n+9}{2}\right\rceil \leq \log_2 n+\log_2 n\log n+10,$$

which completes the proof of Theorem 18.24. □

In fact, based on the contiguity of a random 3-regular graph and graph G_n defined above, one can prove more precise bounds, showing (see Wormald [859]), that w.h.p. $diam(G_n)$ is highly concentrated, i.e., that

$$\log_2 n+\log_2 n\log n-4 \leq diam(G_n) \leq \log_2 n+\log_2 n\log n+4.$$

Kleinberg's Model

The model can be generalized significantly, but to be specific we consider the following. We start with the $n\times n$ grid G_0 which has vertex set $[n]^2$ and where (i,j) is adjacent to (i',j') iff $d((i,j),(i',j')) = 1$ where $d((i,j),(k,\ell)) = |i-k|+|j-\ell|$. In addition, each vertex $u=(i,j)$ will choose another random neighbor $\varphi(u)$ where

$$\mathbb{P}(\varphi(u) = v = (k,\ell)) = \frac{d(u,v)^{-2}}{D_u}$$

where

$$D_x = \sum_{y\neq x} d(x,y)^{-2}.$$

The random neighbors model "long range contacts". Let the grid G_0 plus the extra random edges be denoted by G.

It is not difficult to show that w.h.p. these random contacts reduce the diameter of G to order $\log n$. This however, would not explain Milgram's success. Instead, Kleinberg proposed the following decentralized algorithm $\mathscr{A}$ for finding a path from an initial vertex $u_0 = (i_0, j_0)$ to a target vertex $u_\tau = (i_\tau, j_\tau)$: when at u move to the neighbor closest in distance to u_τ.

Theorem 18.25. *Algorithm $\mathscr{A}$ finds a path from initial to target vertex of order $O((\log n)^2)$, in expectation.*

Proof. Note that each step of $\mathscr{A}$ finds a node closer to the target than the current node and so the algorithm must terminate with a path.

Observe next that for any vertex x of G we have

$$D_x \leq \sum_{j=1}^{2n-2} 4j \times j^{-2} = 4 \sum_{j=1}^{2n-2} j^{-1} \leq 4\log(3n).$$

As a consequence, v is the long range contact of vertex u, with probability at least $(4\log(3n)d(u,v)^2)^{-1}$.

For $0 < j \leq \log_2 n$, we say that the execution of $\mathscr{A}$ is in Phase j if the distance of the current vertex u to the target is greater than 2^j, but at most 2^{j+1}. We say that $\mathscr{A}$ is in Phase 0 if the distance from u to the target is at most 2.

Let B_j denote the set of nodes at distance 2^j or less from the target. Then

$$|B_j| \geq 1 + \sum_{i=1}^{2^j} i > 2^{2j-1}.$$

Note that by the triangle inequality, each member of B_j is within distance $2^{j+1} + 2^j < 2^{2j+2}$ of u.

Let $X_j \leq 2^{j+1}$ be the time spent in Phase j. Assume first that $\log_2 \log_2 n \leq j \leq \log_2 n$. Phase j will end if the long range contact of the current vertex lies in B_j. The probability of this is at least

$$\frac{2^{2j-1}}{4\log(3n)2^{2j+4}} = \frac{1}{128\log(3n)}.$$

We can reveal the long range contacts as the algorithm progresses. In this way, the long range contact of the current vertex will be independent of the previous contacts of the path. Thus

$$\mathbb{E} X_j = \sum_{i=1}^{\infty} \mathbb{P}(X_j \geq i) \leq \sum_{i=1}^{\infty} \left(1 - \frac{1}{128\log(3n)}\right)^i < 128\log(3n).$$

Now if $0 \leq j \leq \log_2 \log_2 n$ then $X_j \leq 2^{j+1} \leq 2\log_2 n$. Thus the expected length of the path found by $\mathscr{A}$ is at most $128\log(3n) \times \log_2 n$. $\square$

In the same paper, Kleinberg showed that replacing $d(u,v)^{-2}$ by $d(u,v)^{-r}$ for $r \neq 2$ led to non-polylogarithmic path length.

18.7 Exercises

18.7.1 Show that w.h.p. the Preferential Attachment Graph of Section 18.1 has diameter $O(\log n)$. (Hint: Using the idea that vertex t chooses a random edge of the current graph, observe that half of these edges appeared at time $t/2$ or less).

18.7.2 For the next few questions we modify the Preferential Attachment Graph of Section 18.1 in the following way: First let $m = 1$ and preferentially generate a sequence of graphs $\Gamma_1, \Gamma_2, \ldots, \Gamma_{mn}$. Then if the edges of Γ_{mn} are $(u_i, v_i), i = 1, 2, \ldots, mn$ let the edges of $\mathbb{G}_n$ be $(u_{\lceil i/m \rceil}, v_{\lceil i/m \rceil}), i = 1, 2, \ldots, mn$. Show that (18.1) continues to hold.

18.7.3 Show that $\mathbb{G}_n$ of the previous question can also be generated in the following way:

(a) Let π be a random permutation of $[2mn]$. Let $X = \{(a_i, b_i), i = 1, 2, \ldots, mn\}$ where $a_i = \min\{\pi(2i-1), \pi(2i)\}$ and $b_i = \max\{\pi(2i-1), \pi(2i)\}$.

(b) Let the edges of $\mathbb{G}_n$ be $(a_{\lceil i/m \rceil}, b_{\lceil i/m \rceil}), i = 1, 2, \ldots, mn$.

This model was introduced in [177].

18.7.4 Show that the edges of the graph in the previous question can be generated as follows:

(a) Let $\zeta_1, \zeta_2, \ldots, \zeta_{2mn}$ be independent uniform $[0,1]$ random variables. Let $\{x_i < y_i\} = \{\zeta_{2i-1}, \zeta_{2i}\}$ for $i = 1, 2, \ldots, mn$. Sort the y_i in increasing order $R_1 < R_2 < \cdots < R_{mn}$ and let $R_0 = 0$. Then let

$$W_j = R_{mj} \text{ and } I_j = (W_{j-1}, W_j] \text{ for } j = 1, 2, \ldots, n.$$

This model was introduced in [176].

(b) The edges of $\mathbb{G}_n$ are $(u_i, v_i), i = 1, 2, \ldots, mn$ where $x_i \in I_{u_i}, y_i \in I_{v_i}$.

18.7.5 Prove that $(R_1, R_2, \dots, R_{mn})$ can be generated as

$$R_i = \left(\frac{\Upsilon_i}{\Upsilon_{mn+1}}\right)^{1/2}$$

where $\Upsilon_N = \xi_1 + \xi_2 + \cdots + \xi_N$ for $N \geq 1$ and $\xi_1, \xi_2, \dots, \xi_{mn+1}$ are independent exponential copies of $EXP(1)$.

18.7.6 Let L be a large constant and let $\omega = \omega(n) \to \infty$ arbitrarily slowly. Then let $\mathscr{E}$ be the event that

$$\Upsilon_k \approx k \text{ for } \quad \frac{k}{m} \in [\omega, n] \text{ or } k = mn + 1.$$

Show that

(a) $\mathbb{P}(\neg\mathscr{E}) = o(1)$.

(b) Let $\eta_i = \xi_{(i-1)m+1} + \xi_{(i-1)m+2} + \cdots + \xi_{im}$. If $\mathscr{E}$ occurs then

(1) $W_i \approx \left(\frac{i}{n}\right)^{1/2}$ for $\omega \leq i \leq n$, and

(2) $w_i = W_i - W_{i-1} \approx \dfrac{\eta_i}{2m(in)^{1/2}}$ for $\omega \leq i \leq n$.

(c) $\eta_i \leq \log n$ for $i \in [n]$ w.h.p.

(d) $\eta_i \leq \log\log n$ for $i \in [(\log n)^{10}]$ w.h.p.

(e) If $\omega \leq i < j \leq n$ then $\mathbb{P}(\text{edge } ij \text{ exists}) \approx \frac{\eta_i}{2(ij)^{1/2}}$.

(f) $\eta_i \geq \frac{1}{\log\log n}$ and $i \leq \frac{n}{\omega(\log n)^3}$ implies the degree $d_n(i) \approx \eta_i \left(\frac{n}{i}\right)^{1/2}$.

18.8 Notes

There are by now a vast number of papers on different models of "Real World Networks". We point out a few additional results in the area. The books by Durrett [316] and Bollobás, Kozma and Miklós [173] cover the area. See also van der Hofstadt [481].

Preferential Attachment Graph

Perhaps the most striking result is due to Bollobás and Riordan [176]. There they prove that the diameter of $\mathbb{G}_n$ is asymptotic to $\frac{\log n}{\log\log n}$ w.h.p. To prove this they introduced the model in question 4 above. Cooper [246] and Peköz, Röllin and

Ross [731] discuss the degree distribution of $\mathbb{G}_n$ in some detail. Flaxman, Frieze and Fenner [364] show that the if Δ_k, λ_k are the kth largest degree, eigenvalue respectively, then $k \approx \Delta_k^{1/2}$ for $k = O(1)$. The proof follows ideas from Mihail and Papadimitriou [681] and Chung, Lu and Vu [225], [226].

Cooper and Frieze [254] discussed the likely proportion of vertices visited by a random walk on a growing preferential attachment graph. They show that w.h.p. this is just over 40% at all times. Borgs, Brautbar, Chayes, Khanna and Lucier [180] discuss "local algorithms" for finding a specific vertex or the largest degree vertex. Frieze and Pegden [411] describe an algorithm for the same problem, but with reduced storage requirements.

Geometric models

Some real world graphs have a geometric constraint. Flaxman, Frieze and Vera [365], [366] considered a geometric version of the preferential attachment model. Here the vertices $X_1, X_2, \ldots, X_n$ are randomly chosen points on the unit sphere in $\mathbb{R}^3$. X_{i+1} chooses m neighbors and these vertices are chosen with probability $P(deg, dist)$ dependent on (i) their current degree and (ii) their distance from X_{i+1}. van den Esker [345] added *fitness* to the models in [365] and [366]. Jordan [531] considered more general spaces than $\mathbb{R}^3$. Jordan and Wade [532] considered the case $m = 1$ and a variety of definitions of P that enable one to interpolate between the preferential attachment graph and the on-line nearest neighbor graph.

The SPA model was introduced by Aiello, Bonato, Cooper, Janssen and Pralat [8]. Here the vertices are points in the unit hyper-cube D in $\mathbb{R}^m$, equipped with a toroidal metric. At time t each vertex v has a domain of attraction $S(v,t)$ of volume $\frac{A_1 \deg^-(v,t)+A_2}{t}$. Then at time t we generate a uniform random point X_{t+1} as a new vertex. If the new point lies in the domain $S(v,t)$ then we join X_{t+1} to v by an edge directed to v, with probability p. The paper [8] deals mainly with the degree distribution. The papers by Jannsen, Pralat and Wilson [515], [516] show that for graphs formed according to the SPA model it is possible to infer the metric distance between vertices from the link structure of the graph. The paper Cooper, Frieze and Pralat [270] shows that w.h.p. the directed diameter at time t lies between $\frac{c_1 \log t}{\log \log t}$ and $c_2 \log t$.

Random Apollonian networks were introduced by Zhou, Yan and Wang [867]. Here we build a random triangulation by inserting a vertex into a randomly chosen face. Frieze and Tsourakakis [416] studied their degree sequence and eigenvalue structure. Ebrahimzadeh, Farczadi, Gao, Mehrabian, Sato, Wormald and Zung [325] studied their diameter and length of the longest path. Cooper and Frieze [264] gave an improved longest path estimate and this was further improved by Collevecchio, Mehrabian and Wormald [238].

Interpolating between Erdős-Rényi and Preferential Attachment

Pittel [741] considered the following model: $G_0, G_1, \ldots, G_m$ is a random (multi) graph growth process G_m on a vertex set $[n]$. G_{m+1} is obtained from G_m by inserting a new edge e at random. Specifically, the conditional probability that e joins two currently disjoint vertices, i and j, is proportional to $(d_i+\alpha)(d_j+\alpha)$, where d_i, d_j are the degrees of i, j in G_m, and $\alpha > 0$ is a fixed parameter. The limiting case $\alpha = \infty$ is the Erdős-Rényi graph process. He shows that w.h.p. G_m contains a unique giant component iff $c := 2m/n > c_\alpha = \alpha/(1+\alpha)$, and the size of this giant is asymptotic to $n\left[1-\left(\frac{\alpha+c^*}{\alpha+c}\right)^{\alpha}\right]$, where $c^* < c_\alpha$ is the root of $\frac{c}{(\alpha+c)^{2+\alpha}} = \frac{c^*}{(\alpha+c^*)^{2+\alpha}}$. A phase transition window is proved to be contained, essentially, in $[c_\alpha - An^{-1/3}, c_\alpha + Bn^{-1/4}]$, and he conjectured that $1/4$ may be replaced with $1/3$. For the multigraph version, MG_m, he showed that MG_m is connected w.h.p. iff $m \gg m_n := n^{1+\alpha^{-1}}$. He conjectured that, for $\alpha > 1$, m_n is the threshold for connectedness of G_m itself. Janson and Warnke [514] verified this conjecture.

Chapter 19

Weighted Graphs

There are many cases in which we put weights $X_e, e \in E$ on the edges of a graph or digraph and ask for the minimum or maximum weight object. The optimisation questions that arise from this are the backbone of Combinatorial Optimisation. When the X_e are random variables we can ask for properties of the optimum value, which will be a random variable. In this chapter we consider three of the most basic optimisation problems viz. minimum weight spanning trees; shortest paths and minimum weight matchings in bipartite graphs.

19.1 Minimum Spanning Tree

Let X_e, $e \in E(K_n)$ be a collection of independent uniform $[0,1]$ random variables. Consider X_e to be the length of edge e and let L_n be the length of the minimum spanning tree (MST) of K_n with these edge lengths.

Frieze [383] proved the following theorem. The proof we give utilises the rather lovely integral formula (19.1) due to Janson [498], (see also the related equation (7) from [406])˙

Theorem 19.1.

$$\lim_{n\to\infty} \mathbb{E} L_n = \zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3} = 1.202\cdots$$

Proof. Suppose that $T = T(\{X_e\})$ is the MST, unique with probability one. We use the identity

$$a = \int_0^1 1_{\{x<a\}} dx.$$

Therefore

$$L_n = \sum_{e\in T} X_e$$

$$
\begin{aligned}
&= \sum_{e\in T}\int_{p=0}^{1} 1_{\{p<X_e\}}dp\\
&= \int_{p=0}^{1}\sum_{e\in T} 1_{\{p<X_e\}}dp\\
&= \int_{p=0}^{1}|\{e\in T : X_e > p\}|dp\\
&= \int_{p=0}^{1}(\kappa(G_p)-1)dp,
\end{aligned}
$$

where $\kappa(\mathbb{G}_p)$ denote the number of components of graph $\mathbb{G}_p$. Here $\mathbb{G}_p$ is the graph induced by the edges e with $X_e \leq p$, i.e., $\mathbb{G}_p \equiv \mathbb{G}_{n,p}$. The last line may be considered to be a consequence of the fact that the greedy algorithm solves the minimum spanning tree problem. This algorithm examines edges in increasing order of edge weight. It builds a tree, adding one edge at a time. It adds the edge to the forest F of edges accepted so far, only if the two endpoints lie in distinct components of F. Otherwise it moves onto the next edge. Thus the number of edges to be added given F, is $\kappa(F)-1$ and if the longest edge in $e\in F$ has $X_e = p$ then $\kappa(F) = \kappa(G_p)$, which follows by an easy induction. Hence

$$
\mathbb{E}\,L_n = \int_{p=0}^{1}(\mathbb{E}\,\kappa(\mathbb{G}_p)-1)dp. \tag{19.1}
$$

We therefore estimate $\mathbb{E}\,\kappa(\mathbb{G}_p)$. We observe first that

$$
p \geq \frac{6\log n}{n} \Rightarrow \mathbb{E}\,\kappa(G_p) = 1+o(1).
$$

Indeed, $1 \leq \mathbb{E}\,\kappa(G_p)$ and

$$
\begin{aligned}
\mathbb{E}\,\kappa(G_p) &\leq 1+n\mathbb{P}(G_p \text{ is not connected})\\
&\leq 1+n\sum_{k=1}^{n/2}\binom{n}{k}k^{k-2}p^{k-1}(1-p)^{k(n-k)}\\
&\leq 1+\frac{n}{p}\sum_{k=1}^{n/2}\left(\frac{ne}{k}\frac{6k\log n}{n}\frac{1}{n^3}\right)^k\\
&= 1+o(1).
\end{aligned}
$$

Hence, if $p_0 = \frac{6\log n}{n}$ then

$$
\mathbb{E}\,L_n = \int_{p=0}^{p_0}(\mathbb{E}\,\kappa(G_p)-1)dp + o(1)
$$

$$= \int_{p=0}^{p_0} \mathbb{E}\,\kappa(G_p)dp + o(1).$$

Write

$$\kappa(G_p) = \sum_{k=1}^{(\log n)^2} A_k + \sum_{k=1}^{(\log n)^2} B_k + C,$$

where A_k stands for the number of components which are k vertex trees, B_k is the number of k vertex components which are not trees and, finally, C denotes the number of components on at least $(\log n)^2$ vertices. Then, for $1 \leq k \leq (\log n)^2$ and $p \leq p_0$,

$$\begin{aligned}\mathbb{E}A_k &= \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1} \\ &= (1+o(1)) n^k \frac{k^{k-2}}{k!} p^{k-1} (1-p)^{kn}.\end{aligned}$$

$$\begin{aligned}\mathbb{E}B_k &\leq \binom{n}{k} k^{k-2} \binom{k}{2} p^k (1-p)^{k(n-k)} \\ &\leq (1+o(1))(npe^{1-np})^k \\ &\leq 1+o(1).\end{aligned}$$

$$C \leq \frac{n}{(\log n)^2}.$$

Hence

$$\int_{p=0}^{\frac{6\log n}{n}} \sum_{k=1}^{(\log n)^2} \mathbb{E}B_k dp \leq \frac{6\log n}{n} (\log n)^2 (1+o(1)) = o(1),$$

and

$$\int_{p=0}^{\frac{6\log n}{n}} C dp \leq \frac{6\log n}{n} \frac{n}{(\log n)^2} = o(1).$$

So

$$\mathbb{E}L_n = o(1) + (1+o(1)) \sum_{k=1}^{(\log n)^2} n^k \frac{k^{k-2}}{k!} \int_{p=0}^{\frac{6\log n}{n}} p^{k-1}(1-p)^{kn} dp.$$

But

$$\sum_{k=1}^{(\log n)^2} n^k \frac{k^{k-2}}{k!} \int_{p=\frac{6\log n}{n}}^{1} p^{k-1}(1-p)^{kn} dp$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{(\log n)^2} n^k \frac{k^{k-2}}{k!} \int_{p=\frac{6\log n}{n}}^{1} n^{-6k} dp \\
&= o(1).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mathbb{E} L_n &= o(1) + (1+o(1)) \sum_{k=1}^{(\log n)^2} n^k \frac{k^{k-2}}{k!} \int_{p=0}^{1} p^{k-1}(1-p)^{kn} dp \\
&= o(1) + (1+o(1)) \sum_{k=1}^{(\log n)^2} n^k \frac{k^{k-2}}{k!} \frac{(k-1)!(kn))!}{(k(n+1))!} \\
&= o(1) + (1+o(1)) \sum_{k=1}^{(\log n)^2} n^k k^{k-3} \prod_{i=1}^{k} \frac{1}{kn+i} \\
&= o(1) + (1+o(1)) \sum_{k=1}^{(\log n)^2} \frac{1}{k^3} \\
&= o(1) + (1+o(1)) \sum_{k=1}^{\infty} \frac{1}{k^3}.
\end{aligned}
$$

□

One can obtain the same result if the uniform $[0,1]$ random variable is replaced by any random non-negative random variable with distribution F having a derivative equal to one at the origin, e.g. an exponential variable with mean one, see Steele [816].

19.2 Shortest Paths

Let the edges of the complete graph K_n on $[n]$ be given independent lengths X_e, $e \in [n]^2$. Here X_e is exponentially distributed with mean 1. The following theorem was proved by Janson [500]:

Theorem 19.2. *Let X_{ij} be the distance from vertex i to vertex j in the complete graph with edge weights independent $EXP(1)$ random variables. Then, for every $\varepsilon > 0$, as $n \to \infty$,*

(i) For any fixed i, j,

$$\mathbb{P}\left(\left| \frac{X_{ij}}{\log n/n} - 1 \right| \geq \varepsilon \right) \to 0.$$

(ii) For any fixed i,

$$\mathbb{P}\left(\left| \frac{\max_j X_{ij}}{\log n/n} - 2 \right| \geq \varepsilon \right) \to 0.$$

(iii)

$$\mathbb{P}\left(\left|\frac{\max_{i,j} X_{ij}}{\log n/n} - 3\right| \geq \varepsilon\right) \to 0.$$

Proof. First, recall the following two properties of the exponential X:

(P1) $\mathbb{P}(X > \alpha+\beta | X > \alpha) = \mathbb{P}(X > \beta)$.

(P2) If $X_1, X_2, \ldots, X_m$ are independent $EXP(1)$ exponential random variables then $\min\{X_1, X_2, \ldots, X_m\}$ is an exponential with mean $1/m$.

Suppose that we want to find shortest paths from a vertex s to all other vertices in a digraph with non-negative arc-lengths. Recall Dijkstra's algorithm. After several iterations there is a rooted tree T such that if v is a vertex of T then the tree path from s to v is a shortest path. Let $d(v)$ be its length. For $x \notin T$ let $d(x)$ be the minimum length of a path P that goes from s to v to x where $v \in T$ and the sub-path of P that goes to v is the tree path from s to v. If $d(y) = \min\{d(x) : x \notin T\}$ then $d(y)$ is the length of a shortest path from s to y and y can be added to the tree.

Suppose that vertices are added to the tree in the order $v_1, v_2, \ldots, v_n$ and that $Y_j = dist(v_1, v_j)$ for $j = 1, 2, \ldots, n$. It follows from property P1 that

$$Y_{k+1} = \min_{\substack{i=1,2,\ldots,k \\ v \neq v_1,\ldots,v_k}} [Y_i + X_{v_i,v}] = Y_k + E_k$$

where E_k is exponential with mean $\frac{1}{k(n-k)}$ and is independent of Y_k.

This is because X_{v_i,v_j} is distributed as an independent exponential X conditioned on $X \geq Y_k - Y_i$. Hence

$$\begin{aligned}
\mathbb{E} Y_n &= \sum_{k=1}^{n-1} \frac{1}{k(n-k)} \\
&= \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{1}{k} + \frac{1}{n-k}\right) \\
&= \frac{2}{n} \sum_{k=1}^{n-1} \frac{1}{k} \\
&= \frac{2\log n}{n} + O(n^{-1}).
\end{aligned}$$

Also, from the independence of E_k, Y_k,

$$\operatorname{Var} Y_n = \sum_{k=1}^{n-1} \operatorname{Var} E_k$$

$$
\begin{aligned}
&= \sum_{k=1}^{n-1} \left(\frac{1}{k(n-k)}\right)^2 \\
&\le 2\sum_{k=1}^{n/2} \left(\frac{1}{k(n-k)}\right)^2 \\
&\le \frac{8}{n^2}\sum_{k=1}^{n/2} \frac{1}{k^2} \\
&= O(n^{-2})
\end{aligned}
$$

and we can use the Chebyshev inequality (26.3) to prove (ii).
Now fix $j = 2$. Then if i is defined by $v_i = 2$, we see that i is uniform over $\{2,3,\dots,n\}$. So

$$
\begin{aligned}
\mathbb{E} X_{1,2} &= \frac{1}{n-1}\sum_{i=2}^{n}\sum_{k=1}^{i-1} \frac{1}{k(n-k)} \\
&= \frac{1}{n-1}\sum_{k=1}^{n-1} \frac{n-k}{k(n-k)} \\
&= \frac{1}{n-1}\sum_{k=1}^{n-1} \frac{1}{k} \\
&= \frac{\log n}{n} + O(n^{-1}).
\end{aligned}
$$

For the variance of $X_{1,2}$ we have

$$X_{1,2} = \delta_2 Y_2 + \delta_3 Y_3 + \cdots + \delta_n Y_n,$$

where

$$\delta_i \in \{0,1\}; \;\; \delta_2 + \delta_3 + \cdots + \delta_n = 1; \;\; \mathbb{P}(\delta_i = 1) = \frac{1}{n-1}.$$

$$
\begin{aligned}
\operatorname{Var} X_{1,2} &= \sum_{i=2}^{n} \operatorname{Var}(\delta_i Y_i) + \sum_{i \ne j} \operatorname{Cov}(\delta_i Y_i, \delta_j Y_j) \\
&\le \sum_{i=2}^{n} \operatorname{Var}(\delta_i Y_i).
\end{aligned}
$$

The last inequality holds since

$$\operatorname{Cov}(\delta_i Y_i, \delta_j Y_j) = \mathbb{E}(\delta_i Y_i \delta_j Y_j) - \mathbb{E}(\delta_i Y_i)\,\mathbb{E}(\delta_j Y_j) = -\,\mathbb{E}(\delta_i Y_i)\,\mathbb{E}(\delta_j Y_j) \le 0.$$

So

$$\begin{aligned}\operatorname{Var} X_{1,2} &\leq \sum_{i=2}^{n} \operatorname{Var}(\delta_i Y_i)\\ &\leq \sum_{i=2}^{n} \frac{1}{n-1} \sum_{k=1}^{i-1} \left(\frac{1}{k(n-k)}\right)^2\\ &= O(n^{-2}).\end{aligned}$$

We can now use the Chebyshev inequality.

We turn now to proving (iii). We begin with a lower bound. Let $Y_i = \min\left\{X_{i,j} : i \neq j \in [n]\right\}$. Let $A = \left\{i : Y_i \geq \frac{(1-\varepsilon)\log n}{n}\right\}$. Then we have that for $i \in [n]$,

$$\Pr(i \in A) = \exp\left\{-(n-1)\frac{(1-\varepsilon)\log n}{n}\right\} = n^{-1+\varepsilon+o(1)}. \tag{19.2}$$

An application of the Chebyshev inequality shows that $|A| \approx n^{\varepsilon+o(1)}$ w.h.p. Now the expected number of paths from $a_1 \in A$ to $a_2 \in A$ of length at most $\frac{(3-2\varepsilon)\log n}{n}$ can be bounded by

$$n^{2\varepsilon+o(1)} \times n^2 \times n^{-3\varepsilon+o(1)} \times \frac{2\log^2 n}{n^2} = n^{-\varepsilon+o(1)}. \tag{19.3}$$

Explanation for (19.3)**:** The first factor $n^{2\varepsilon+o(1)}$ is the expected number of pairs of vertices $a_1, a_2 \in A$. The second factor is a bound on the number of choices b_1, b_2 for the neighbors of a_1, a_2 on the path. The third factor F_3 is a bound on the expected number of paths of length at most $\frac{\alpha \log n}{n}$ from b_1 to b_2, $\alpha = 1-3\varepsilon$. This factor comes from

$$F_3 \leq \sum_{\ell \geq 0} n^{\ell} \left(\frac{\alpha \log n}{n}\right)^{\ell+1} \frac{1}{(\ell+1)!}.$$

Here ℓ is the number of internal vertices on the path. There will be at most n^{ℓ} choices for the sequence of vertices on the path. We then use the fact that the exponential mean one random variable stochastically dominates the uniform $[0,1]$ random variable U. The final two factors are the probability that the sum of $\ell+1$ independent copies of U sum to at most $\frac{\alpha \log n}{n}$. Continuing we have

$$F_{3,k} \leq \frac{\alpha \log n}{n} \sum_{\ell \geq 0} \frac{(\alpha \log n)^{\ell}}{\ell!} = \frac{\alpha \log n}{n} \cdot e^{\alpha \log n} = n^{-1+\alpha+o(1)}.$$

The final factor in (19.3) is a bound on the probability that $X_{a_1 b_1} + X_{a_2 b_2} \leq \frac{(2+\varepsilon)\log n}{n}$. For this we use the fact that $X_{a_i b_i}, i = 1,2$ is distributed as $\frac{(1-\varepsilon)\log n}{n} + E_i$ where

E_1, E_2 are independent exponential mean one. Now $\Pr(E_1+E_2 \le t) \le (1-e^{-t})^2 \le t^2$ and taking $t = \frac{(k+2)\varepsilon \log n}{n}$ justifies the final factor of (19.3).

It follows from (19.3) that w.h.p. the shortest distance between a pair of vertices in A is at least $\frac{(3-2\varepsilon)\log n}{n}$ w.h.p., completing our proof of the lower bound in (iii).

We now consider the upper bound. Let now $Y_1 = d_{k_3}$ where $k_3 = n^{1/2}\log n$. For $t < 1 - \frac{1+o(1}{n}$ we have that

$$\mathbb{E}(e^{tnY_1}) = \mathbb{E}\left(\exp\left\{tn\sum_{i=1}^{k_3} EXP\left(\frac{1}{i(n-i)}\right)\right\}\right) = \prod_{i=1}^{k}\left(1 - \frac{(1+o(1))t}{i}\right)^{-1}$$

Then for any $\alpha > 0$ and for we have

$$\Pr\left(Y_1 \ge \frac{\alpha \log n}{n}\right) \le \mathbb{E}(e^{tnY_1 - t\alpha\log n}) \le e^{-t\alpha\log n}\prod_{i=1}^{k_3}\left(1 - \frac{(1+o(1))t}{i}\right)^{-1}$$
$$= e^{-t\alpha\log n}\exp\left\{\sum_{i=1}^{k_3}\frac{(1+o(1))t}{i} + O\left(\frac{1}{i^2}\right)\right\} = O(1)\times\exp\left\{\left(\frac{1}{2}+o(1)-\alpha\right)t\log n\right\}.$$

It follows, on taking $\alpha = 3/2 + o(1)$ that w.h.p.

$$Y_j \le \frac{(3+o(1))\log n}{2n} \text{ for all } j \in [n].$$

Letting T_j be the set corresponding to S_{k_3} when we execute Dijkstra's algoritm starting at j, then we have that for $j \ne k$ where $T_j \cap T_k = \emptyset$,

$$\Pr\left(\nexists e \in (T_j : T_k) : X_e \le \frac{\log^{1/2} n}{n}\right) \le \exp\left\{-\frac{k_3^2 \log^{1/2} n}{n}\right\}$$
$$= e^{-\log^{5/2} n} = o(n^{-2})$$

and this is enough to complete the proof of (iii). □

We can as for Spanning Trees, replace the exponential random variables by random variables that behave like the exponential close to the origin. The paper of Janson [500] allows for any random variable X satisfying $\mathbb{P}(X \le t) = t + o(t)$ as $t \to 0$.

19.3 Minimum Weight Assignment

Consider the complete bipartite graph $K_{n,n}$ and suppose that its edges are assigned independent exponentially distributed weights, with rate 1. (The rate of an exponential variable is one over its mean). Denote the minimum total weight of a perfect matching in $K_{n,n}$ by C_n. Aldous [17], [20] proved that $\lim_{n\to\infty}\mathbb{E}C_n = \zeta(2) = \sum_{k=1}^{\infty}\frac{1}{k^2}$. The following theorem was conjectured by Parisi [729]. It was proved independently by Linusson and Wästlund [628] and Nair, Prabhakar and Sharma [706]. The proof given here is from Wästlund [846].

Theorem 19.3.

$$\mathbb{E}C_n = \sum_{k=1}^{n}\frac{1}{k^2} = 1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\cdots+\frac{1}{n^2} \tag{19.4}$$

From the above theorem we immediately get the following corollary, first proved by Aldous [20].

Corollary 19.4.

$$\lim_{n\to\infty}\mathbb{E}C_n = \zeta(2) = \sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6} = 1.6449\cdots$$

Let $EXP(\lambda)$ denote an exponential random variable of rate λ i.e. $\Pr(EXP(\lambda)\geq x) = e^{-\lambda x}$. Consider the complete bipartite graph $K_{n,n}$, with bipartition (A,B), where $A = \{a_1,a_2,\ldots,a_n\}$ and $B = \{b_1,b_2,\ldots,b_n\}$, and with edge weights which are independent copies of $EXP(1)$. We assume that $a_1,a_2,\ldots,a_n$ is a random permutation of A. So $A_r = \{a_1,a_2,\ldots,a_r\}$ is a random r-subset of A.

We add a special vertex b^* to B, with edges to all n vertices of A. Each edge adjacent to b^* is assigned an $EXP(\lambda)$ weight independently, $\lambda > 0$.

For $r\geq 1$ we let M_r be the minimum weight matching of A_r into B and M_r^* be the minimum weight matching of A_r into $B^* = B\cup\{b^*\}$. (As $\lambda\to 0$ it becomes increasingly unlikely that any of the extra edges are actually used in the minimum weight matching.) We denote this matching by M_r^* and we let B_r^* denote the corresponding set of vertices of B^* that are covered by M_r^*. We let $C(n,r)$ denote the weight of M_r.

Define $P(n,r)$ as the normalized probability that b^* participates in M_r^*, i.e.

$$P(n,r) = \lim_{\lambda\to 0}\frac{\Pr(b^*\in B_r^*)}{\lambda}. \tag{19.5}$$

Its importance lies in the following lemma:

Lemma 19.5.
$$\mathbb{E}(C(n,r)-C(n,r-1)) = \frac{P(n,r)}{r}. \tag{19.6}$$

Proof. Choose i randomly from $[r]$ and let $\widehat{B}_i \subseteq B_r$ be the B-vertices in the minimum weight matching of $(A_r \setminus \{a_i\})$ into B^*. Let $X = C(n,r)$ and let $Y = C(n,r-1)$. Let w_i be the weight of the edge (a_i,b^*), and let I_i denote the indicator variable for the event that the minimum weight of an A_r matching that contains this edge is smaller than the minimum weight of an A_r matching that does not use b^*. We can see that I_i is the indicator variable for the event $\{Y_i + w_i < X\}$, where Y_i is the minimum weight of a matching from $A_r \setminus \{a_i\}$ to B. Indeed, if $(a_i,b^*) \in M_r^*$ then $w_i < X - Y_i$. Conversely, if $w_i < X - Y_i$ and no other edge from b^* has weight smaller than $X - Y_i$, then $(a_i,b^*) \in M_r^*$, and when $\lambda \to 0$, the probability that there are two distinct edges from b^* of weight smaller than $X - Y_i$ is of order $O(\lambda^2)$. Indeed, let $\mathscr{F}$ denote the existence of two distinct edges from b^* of weight smaller than X and let $\mathscr{F}_{i,j}$ denote the event that (a_i,b^*) and $a_j,b^*)$ both have weight smaller than X.

Then,

$$\Pr(\mathscr{F}) \leq n^2 \mathbb{E}_X(\max_{i,j} \Pr(\mathscr{F}_{i,j} \mid X)) = n^2 \mathbb{E}((1-e^{-\lambda X})^2) \leq n^2\lambda^2 \mathbb{E}(X^2), \tag{19.7}$$

and since $\mathbb{E}(X^2)$ is finite and independent of λ, this is $O(\lambda^2)$.

Note that Y and Y_i have the same distribution. They are both equal to the minimum weight of a matching of a random $(r-1)$-set of A into B. As a consequence, $\mathbb{E}(Y) = \mathbb{E}(Y_i) = \frac{1}{r}\sum_{j \in A_r} \mathbb{E}(Y_j)$. Since w_i is $EXP(\lambda)$ distributed, as $\lambda \to 0$ we have from (19.7) that

$$P(n,r) = \lim_{\lambda \to 0}\left(\frac{1}{\lambda}\sum_{j \in A_r} \Pr(w_j < X - Y_j) + O(\lambda)\right) =$$
$$\lim_{\lambda \to 0} \mathbb{E}\left(\frac{1}{\lambda}\sum_{j \in A_r}\left(1 - e^{-\lambda(X-Y_j)}\right)\right) = \sum_{j \in A_r} \mathbb{E}(X - Y_i) = r\mathbb{E}(X-Y).$$

□

We now proceed to estimate $P(n,r)$. Fix r and assume that $b^* \notin B_{r-1}^*$. Suppose that M_r^* is obtained from M_{r-1}^* by finding an augmenting path $P = (a_r,\ldots,a_\sigma,b_\tau)$ from a_r to $B \setminus B_{r-1}$ of minimum additional weight. We condition on (i) σ, (ii) the lengths of all edges other than $(a_\sigma,b_j), b_j \in B \setminus B_{r-1}$ and (iii) $\min\left\{w(a_\sigma,b_j) : b_j \in B \setminus B_{r-1}\right\}$. With this conditioning $M_{r-1} = M_{r-1}^*$ will be fixed and so will $P' = (a_r,\ldots,a_\sigma)$. We can now use the following fact: Let $X_1,X_2,\ldots,X_M$ be independent exponential random variables of rates $\lambda_1,\lambda_2,\ldots,\lambda_M$.

Then the probability that X_i is the smallest of them is $\lambda_i/(\lambda_1+\lambda_2+\cdots+\lambda_M)$. Furthermore, the probability stays the same if we condition on the value of $\min\{X_1,X_2,\ldots,X_M\}$. Indeed, for any $\alpha>0$,

$$\begin{aligned}
&\Pr(X_i=\min\{X_1,X_2,\ldots,X_M\}\mid \min\{X_1,X_2,\ldots,X_M\}=\alpha)\\
&=\frac{\Pr(X_i=\alpha\le\min\{X_1,X_2,\ldots,X_M\})}{\sum_{j=1}^M\Pr(X_j=\alpha\le\min\{X_1,X_2,\ldots,X_M\})}\\
&=\frac{\lambda_i e^{-\lambda_i\alpha}\prod_{j\ne i}e^{-\lambda_j\alpha}}{\sum_{j=1}^M\lambda_j e^{-\lambda_j\alpha}\prod_{l\ne j}e^{-\lambda_l\alpha}}\\
&=\frac{\lambda_i}{\lambda_1+\lambda_2+\cdots+\lambda_M}.
\end{aligned}$$

Thus

$$\Pr(b^*\in B_r^*\mid b^*\notin B_{r-1}^*)=\frac{\lambda}{n-r+1+\lambda}.$$

Lemma 19.6.

$$P(n,r)=\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{n-r+1}. \tag{19.8}$$

Proof.

$$\begin{aligned}
\lim_{\lambda\to0}\lambda^{-1}\Pr(b^*\in B_r^*)&=\lim_{\lambda\to0}\lambda^{-1}\left(1-\frac{n}{n+\lambda}\cdot\frac{n-1}{n-1+\lambda}\cdots\frac{n-r+1}{n-r+1+\lambda}\right)\\
&=\lim_{\lambda\to0}\lambda^{-1}\left(1-\left(1+\frac{\lambda}{n}\right)^{-1}\cdots\left(1+\frac{\lambda}{n-r+1}\right)^{-1}\right)\\
&=\lim_{\lambda\to0}\lambda^{-1}\left(\left(\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{n-r+1}\right)\lambda+O(\lambda^2)\right)\\
&=\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{n-r+1}. \qquad (19.9)
\end{aligned}$$

□

It follows from Lemmas 19.5 and 19.6 that

$$\mathbb{E}C_n=\sum_{r=1}^n\frac{1}{r}\sum_{i=1}^r\frac{1}{n-i+1}.$$

It follows that

$$\mathbb{E}(C_{n+1}-C_n)=$$

$$
\begin{aligned}
&= \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{n-i+2}+\sum_{r=1}^{n}\frac{1}{r}\sum_{i=1}^{r}\left(\frac{1}{n-i+2}-\frac{1}{n-i+1}\right)\\
&= \frac{1}{(n+1)^2}+\sum_{i=2}^{n+1}\frac{1}{(n+1)(n-i+2)}+\sum_{r=1}^{n}\frac{1}{r}\left(\frac{1}{n+1}-\frac{1}{n-r+1}\right),\\
&= \frac{1}{(n+1)^2}. \qquad (19.10)
\end{aligned}
$$

$\mathbb{E}(C_1)=1$ and so (19.4) follows from (19.10). □

19.4 Exercises

19.4.1 Suppose that the edges of the complete digraph $\vec{K}_n$ are given independent uniform $[0,1]$ edge costs. Show that if L_n is the cost of the minimum spanning arborescence, then $\lim_{n\to\infty}\mathbb{E}L_n=1$.
(An arborescence is an oriented spanning tree where each edge is oriented away from the root.)

19.4.2 Suppose that the edges of the complete bipartite graph $K_{n,n}$ are given independent uniform $[0,1]$ edge costs. Show that if $L_n^{(b)}$ is the cost of the minimum spanning tree, then

$$\lim_{n\to\infty}\mathbb{E}L_n^{(b)}=2\zeta(3).$$

19.4.3 Let $G=K_{\alpha n,\beta n}$ be the complete unbalanced bipartite graph with bipartition sizes $\alpha n,\beta n$. Suppose that the edges of G are given independent uniform $[0,1]$ edge costs. Show that if $L_n^{(b)}$ is the length of the minimum spanning tree, then

$$\lim_{n\to\infty}\mathbb{E}L_n^{(b)}=\gamma+\frac{1}{\gamma}+\sum_{i_1\geq 1,i_2\geq 1}\frac{(i_1+i_2-1)!}{i_1!i_2!}\frac{\gamma^{i_1}i_1^{i_2-1}i_2^{i_1-1}}{(i_1+\gamma i_2)^{i_1+i_2}},$$

where $\gamma=\alpha/\beta$.

19.4.4 Tighten Theorem 19.1 and prove that

$$\mathbb{E}L_n=\zeta(3)+O\left(\frac{1}{n}\right).$$

19.4.5 Suppose that the edges of K_n are given independent uniform $[0,1]$ edge costs. Let Z_k denote the minimum total edge cost of the union of k edge-disjoint spanning trees. Show that $\lim_{k\to\infty}Z_k/k^2=1$.

19.4.6 Show that if the edges of the complete bipartite graph $K_{n,n}$ are given i.i.d. costs then the minimum cost perfect matching is uniformly random among all $n!$ perfect matchings.

19.4.7 Suppose that the edges of K_n or $K_{n,n}$ are given independent uniform $[0,1]$ edge costs. Show that w.h.p. no edge in the minimum cost perfect matching has cost more than $\frac{\omega \log n}{n}$ where $\omega \to \infty$.

19.4.8 Show that a random permutation π gives rise to a digraph $D_\pi = ([n], \{(i, \pi(i)) : i \in [n]\})$ that w.h.p. consists of $O(\log n)$ vertex disjoint cycles that cover $[n]$.

19.4.9 Consider the Asymmetric Traveling Salesperson Problem (ATSP) where the costs $C(i,j)$ are independent uniform $[0,1]$. Use the claimed results of the previous two problems to show that Karp's patching algorithm finds a tour that is within $(1+o(1))$ of minimal cost w.h.p.
The ATSP asks for the minimum total cost of a directed Hamilton cycle through the complete digraph $\vec{K}_n$.
Karp,s algorithm starts by solving the assignment problem with costs $C(i,j)$. It interprets the perfect matching as the union of disjoint cycles in $\vec{K}_n$ and then patches them together cheaply.
Given cycles C_1, C_2 and edges $e_i = (x_i, y_i) \in C_i, i = 1,2$, a patch replaces removes e_1, e_2 and replaces them with (x_2, y_1) plus (x_1, y_2) creating a single cycle.

19.4.10 Suppose that the edges of $G_{n,p}$ where $0 < p \leq 1$ is a constant, are given exponentially distributed costs with rate 1. Show that if X_{ij} is the shortest distance from i to j then

(a) For any fixed i, j,

$$\mathbb{P}\left(\left|\frac{X_{ij}}{\log n/n} - \frac{1}{p}\right| \geq \varepsilon\right) \to 0.$$

(b)

$$\mathbb{P}\left(\left|\frac{\max_j X_{ij}}{\log n/n} - \frac{2}{p}\right| \geq \varepsilon\right) \to 0.$$

19.4.11 The *quadratic assignment problem* is to

$$\begin{array}{lll} \textit{Minimise} & & \\ & Z = \sum_{i,j,p,q=1}^{n} a_{ijpq} x_{ip} x_{jq} & \\ \textit{Subject to} & & \\ & \sum_{i=1}^{n} x_{ip} = 1 & p = 1,2,\dots,n \\ & \sum_{p=1}^{n} x_{ip} = 1 & i = 1,2,\dots,n \\ & x_{ip} = 0/1. & \end{array}$$

Suppose now that the a_{ijpq} are independent uniform $[0,1]$ random variables. Show that w.h.p. $Z_{\min} \approx Z_{\max}$ where $Z_{\min}$ (resp. $Z_{\max}$) denotes the minimum (resp. maximum) value of Z, subject to the assignment constraints.

19.4.12 The *0/1 knapsack problem* is to

$$\begin{array}{ll} \textit{Maximise} & \\ & Z = \sum_{i=1}^{n} a_i x_i \\ \textit{Subject to} & \\ & \sum_{i=1}^{n} b_i x_i \leq L \\ & x_i = 0/1 \quad \text{for } i = 1,2,\dots,n. \end{array}$$

Suppose that the (a_i, b_i) are chosen independently and uniformly from $[0,1]^2$ and that $L = \alpha n$. Show that w.h.p. the maximum value of Z, $Z_{\max}$, satisfies

$$Z_{\max} \approx \begin{cases} \frac{\alpha^{1/2} n}{2} & \alpha \leq \frac{1}{4}. \\ \frac{(8\alpha - 8\alpha^2 - 1)n}{2} & \frac{1}{4} \leq \alpha \leq \frac{1}{2} \\ \frac{n}{2} & \alpha \geq \frac{1}{2} \end{cases}.$$

19.4.13 Suppose that $X_1, X_2, \dots, X_n$ are points chosen independently and uniformly at random from $[0,1]^2$. Let Z_n denote the total Euclidean length of the shortest tour (Hamilton cycle) through each point. Show that there exist constants c_1, c_2 such that $c_1 n^{1/2} \leq Z_n \leq c_2 n^{1/2}$ w.h.p.

19.4.14 Prove equation (19.12) below.

19.5 Notes

Shortest paths

There have been some strengthenings and generalisations of Theorem 19.2. For example, Bhamidi and van der Hofstad [108] have found the (random) second-order term in (i), i.e., convergence in distribution with the correct norming. They have also studied the number of edges in the shortest path.

Spanning trees

Beveridge, Frieze and McDiarmid [107] considered the length of the minimum spanning tree in regular graphs other than complete graphs. For graphs G of large degree r they proved that the length $MST(G)$ of an n-vertex randomly edge weighted graph G satisfies $MST(G) = \frac{n}{r}(\zeta(3)+o_r(1))$ w.h.p., provided some mild expansion condition holds. For r regular graphs of large girth g they proved that if

$$c_r = \frac{r}{(r-1)^2}\sum_{k=1}^{\infty}\frac{1}{k(k+\rho)(k+2\rho)},$$

then w.h.p. $|MST(G) - c_r n| \leq \frac{3n}{2g}$.

Frieze, Ruszinko and Thoma [414] replaced expansion in [107] by connectivity and in addition proved that $MST(G) \leq \frac{n}{r}(\zeta(3)+1+o_r(1))$ for any r-regular graph.

Cooper, Frieze, Ince, Janson and Spencer [265] show that Theorem 19.1 can be improved to yield $\mathbb{E} L_n = \zeta(3) + \frac{c_1}{n} + \frac{c_2+o(1)}{n^{4/3}}$ for explicit constants c_1, c_2.

Bollobás, Gamarnik, Riordan and Sudakov [169] considered the Steiner Tree problem on K_n with independent random edge weights, $X_e, e \in E(K_n)$. Here they assume that the X_e have the same distribution $X \geq 0$ where $\mathbb{P}(X \leq x) = x + o(x)$ as $x \to 0$. The main result is that if one fixes $k = o(n)$ vertices then w.h.p. the minimum length W of a sub-tree of K_n that includes these k points satisfies $W \approx \frac{k-1}{n}\log\frac{n}{k}$.

Angel, Flaxman and Wilson [44] considered the minimum length of a spanning tree of K_n that has a fixed root and bounded depth k. The edges weights X_e are independent exponential mean one. They prove that if $k \geq \log_2\log n + \omega(1)$ then w.h.p. the minimum length tends to $\zeta(3)$ as in the unbounded case. On the other hand, if $k \leq \log_2\log n - \omega(1)$ then w.h.p. the weight is doubly exponential in $\log_2\log n - k$. They also considered bounded depth Steiner trees.

Using Talagrand's inequality, McDiarmid [663] proved that for any real $t > 0$ we have $\mathbb{P}(|L_n - \zeta(3)| \geq t) \leq e^{-\delta_1 n}$ where $\delta_1 = \delta_2(t)$. Flaxman [363] proved that $\mathbb{P}(|L_n - \zeta(3)| \leq \varepsilon) \geq e^{-\delta_2 n}$ where $\delta_1 = \delta_2(\varepsilon)$.

Assignment problem

Walkup [845] proved if the weights of edges are independent uniform $[0,1]$ then $\mathbb{E}C_n \leq 3$ (see (19.3)) and later Karp [555] proved that $\mathbb{E}C_n \leq 2$. Dyer, Frieze and McDiarmid [323] adapted Karp's proof to something more general: Let Z be the optimum value to the linear program:

$$\text{Minimise } \sum_{j=1}^{n} c_j x_j, \text{subject to } x \in P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\},$$

where A is an $m \times n$ matrix. As a special case of [323], we have that if $c_1, c_2, \ldots, c_n$ are independent uniform $[0,1]$ random variables and x^* is any member of P, then $\mathbb{E}(Z) \leq m(\max_j x_j^*)$. Karp's result can easily be deduced from this.

The assignment problem can be generalized to multi-dimensional versions: We replace the complete bipartite graph $K_{n,n}$ by the complete k-partite hypergraph $K_n^{(k)}$ with vertex partition $V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_k$ where each V_i is of size n. We give each edge of $K_n^{(k)}$ an independent exponential mean one value. Assume for example that $k = 3$. In one version of the 3-dimensional assignment problem we ask for a minimum weight collection of hyper-edges such that each vertex $v \in V$ appears in exactly one edge. The optimal total weight Z of this collection satisfies

$$Z = \Theta\left(\frac{1}{n}\right) \; w.h.p. \tag{19.11}$$

(The upper bound uses the result of [529], see Section 22).

Frieze and Sorkin [415] give an $O(n^3)$ algorithm that w.h.p. finds a solution of value $\frac{1}{n^{1-o(1)}}$.

In another version of the 3-dimensional assignment problem we ask for a minimum weight collection of hyper-edges such that each pair of vertices $v, w \in V$ from different sets in the partition appear in exactly one edge. The optimal total weight Z of this collection satisfies

$$\Omega(n) \leq Z \leq O(n \log n) \; w.h.p. \tag{19.12}$$

(The upper bound uses the result of [323] to greedily solve a sequence of restricted assignment problems).

Part IV

Further topics

Chapter 20

Resilience

Sudakov and Vu [822] introduced the idea of the *local resilience* of a monotone increasing graph property $\mathscr{P}$. Suppose we delete the edges of some graph H on vertex set $[n]$ from $\mathbb{G}_{n,p}$. Suppose that p is above the threshold for $\mathbb{G}_{n,p}$ to have the property. What can we say about the value Δ so that w.h.p. the graph $G = ([n], E(\mathbb{G}_{n,p}) \setminus E(H))$ has property $\mathscr{P}$ for all H with maximum degree at most Δ? We will denote the maximum Δ by $\Delta_{\mathscr{P}}$

In this chapter we discuss the resilience of various properties. In Section 20.2 we discuss the resilience of having a perfect matching. In Section 20.3 we discuss the resileince of having a Hamilton cycle. In Section 20.4 we discuss the resilience of the chromatic number.

20.1 Connectivity

We begin with a simple result. Let $\mathscr{C}$ denote the property of being connected.

Theorem 20.1. *Suppose that $np \gg \log n$. Then w.h.p. in $G_{n,p}$, $\left(\frac{1}{2} - \varepsilon\right) np \leq \Delta_{\mathscr{C}} \leq \left(\frac{1}{2} + \varepsilon\right) np$ for any positive constant ε.*

Proof. For the upper bound we partition $[n]$ into $A = \{1, 2, \ldots, \lfloor n/2 \rfloor\}$ and $B = [n] \setminus A$. The Chernoff bounds imply that each $a \in A$ has $(1 + o(1))np/2$ neighbors in B. It is therefore possible for H to contain all $A : B$ edges and then G will be disconnected.

For the lower bound observe that w.h.p.

$$e(S : \bar{S}) \geq k(n-k)p(1 - \varepsilon/10) \text{ for every set } S \text{ of } k \leq n/2 \text{ vertices} \tag{20.1}$$

Indeed,

$$\Pr(\neg(20.1)) \leq \sum_{k=1}^{n/2} \binom{n}{k} e^{-k(n-k)p\varepsilon^2/100} \leq \sum_{k=1}^{n/2} \left(\frac{ne^{1-np\varepsilon^2/200}}{k} \right)^k = o(1).$$

Now suppose that H is not connected. Then there exist S with $|S| \leq n/2$ such that all edges between S and $[n] \setminus S$ are contained in H. But this means that

$$|S|(1/2-\varepsilon)np \geq |S|(n-|S|)p(1-\varepsilon/10) \geq |S|(n/2)p(1-\varepsilon/10),$$

contradiction. □

20.2 Perfect Matchings

Sudakov and Vu [822] proved that if $\mathscr{M}$ denotes the property of having a perfect matching then

Theorem 20.2. *Suppose that $n = 2m$ is even and that $np \gg \log n$. Then w.h.p. in $G_{n,p}$, $\left(\frac{1}{2}-\varepsilon\right)np \leq \Delta_{\mathscr{M}} \leq \left(\frac{1}{2}+\varepsilon\right)np$ for any positive constant ε.*

Proof. The upper bound $\Delta_{\mathscr{M}} \leq \left(\frac{1}{2}+\varepsilon\right)np$ is easy to prove. Randomly partition $[n]$ into two subsets X,Y of sizes $m+1$ and $m-1$ respectively. Now delete all edges inside X so that X becomes an independent set. Clearly, the remaining graph contains no perfect matching. The Chernoff bounds, Corollary 27.7, imply that we have deleted $\approx np/2$ edges incident with each vertex.

The lower bound requires a little more work. Theorem 3.4 implies that w.h.p. the minimum degree in G is at least $(1-o(1))\left(\frac{1}{2}+\varepsilon\right)np$. We randomly partition $[n]$ into two sets X,Y of size m. We have that w.h.p.

PM1 $d_Y(x) \gtrsim \left(\frac{1}{4}+\frac{\varepsilon}{2}\right)np$ for all $x \in X$ and $d_X(y) \gtrsim \left(\frac{1}{4}+\frac{\varepsilon}{2}\right)np$ for all $y \in Y$.

PM2 $e(S,T) \leq \left(1+\frac{\varepsilon}{3}\right)\frac{np}{4}|S|$ for all $S \subseteq X, T \subseteq Y$, $|S| = |T| \leq n/4$.

Property **PM1** follows immediately from the Chernoff bounds, Corollary 27.7, and the fact that $d_G(v) \gtrsim \left(\frac{1}{2}+\varepsilon\right)np \gg \log n$.

Property **PM2** is derived as follows:

$$\begin{aligned}\Pr\left(\exists S,T : e(S,T) \geq |S|\left(1+\frac{\varepsilon}{3}\right)\frac{np}{4}\right) &\leq \sum_{s=1}^{n/4}\binom{n}{s}^2 e^{-4\varepsilon^2 snp/27}\\ &\leq \sum_{s=1}^{n/4}\left(\frac{n^2e^{2-4\varepsilon^2 np/27}}{s^2}\right)^s = o(1).\end{aligned}$$

Given, **PM1, PM2**, we see that if there exists $S \subseteq X, |S| \leq n/4$ such that $|N_X(S)| \leq |S|$ then for $T = N_X(S)$,

$$\left(\frac{1}{4}+\frac{\varepsilon}{2}\right)np|S| \lesssim e(S,T) \leq \left(1+\frac{\varepsilon}{3}\right)\frac{np}{4}|S|,$$

contradiction. We finish the proof that Hall's condition holds, i.e. deal with $|S| > n/4$ just as we did for $|S| > n/2$ in Theorem 6.1. □

20.3 Hamilton Cycles

Sudakov and Vu [822] proved that if $np \gg \log^4 n$ and $\mathscr{H}$ denotes the Hamiltonicity property, then w.h.p. $\Delta_{\mathscr{H}} = \left(\frac{1}{2} - o(1)\right) np$. This is the optimal value for $\Delta_{\mathscr{H}}$ but a series of papers culminating in the following theorem due to Lee and Sudakov.

Theorem 20.3. *If $np \gg \log n$, then w.h.p. $\Delta_{\mathscr{H}} = \left(\frac{1}{2} - o(1)\right) np$.*

Going even further, Montgomery [689] and Nenadov, Steger and Trujić [714] have given tight hitting time versions. The proofs in these papers rely on the use of Pósa rotations, as in Chapter 6. Some recent papers have introduced the use of the *absorbing method* from extremal combinatorics to related problems. The method was initiated by Rödl, Ruciński and Szemerédi [770]. Our purpose in this section is to give an example of this important technique. Our exposition closely follows the paper of Ferber, Nenadov, Noever, Peter and Trujić [357]. They consider the resilience of Hamiltonicity in the context of random digraphs, but their proof can be adapted and simplified when considering graphs. Their proof in turn utilises ideas from Montgomery [689].

Theorem 20.4. *Suppose that $p \geq \frac{\log^{10} n}{n}$. Then w.h.p. in $G_{n,p}$,*

$$\left(\frac{1}{2} - \varepsilon\right) np \leq \Delta_{\mathscr{H}} \leq \left(\frac{1}{2} + \varepsilon\right) np$$

for any positive constant ε.

From our previous remarks, we can see that $\log^{10} n$ is not optimal. The proof we give can be tightened, but probably not down to $\log n$. The proof of Theorem 20.4 takes up the remainder of this section.

The proof of the upper bound is essentially the same as for Theorem 20.2. After making X independent, there is no possibility of a Hamilton cycle.

The lower bound requires more work.

A pseudo-random condition

We say that a graph $G = (V,E)$ with $|V| = n$ is (n, α, p)-pseudo-random if

Q1 $d_G(v) \geq \left(\frac{1}{2} + \alpha\right) np$ for all $v \in V(G)$.

Q2 $e_G(S) \leq |S| \log^3 n$ for all $S \subseteq V, |S| \leq \frac{10 \log^2 n}{p}$.

Q3 $e_G(S,T) \leq \left(1 + \frac{\alpha}{4}\right) |S||T|p$ for all disjoint $S,T \subseteq V$, $|S|,|T| \geq \frac{\log^2 n}{p}$.

Lemma 20.5. *Let α be an arbitrary small positive constant. Suppose that $p \geq \frac{\log^{10} n}{n}$. Let H be a subgraph of $\mathbb{G} = \mathbb{G}_{n,p}$ with maximum degree $\left(\frac{1}{2} - 3\alpha\right) np$ and let $G = \mathbb{G} - H$. Then w.h.p. G is (n, α, p)-pseudo-random.*

Proof. **Q1:** This follows from the fact that w.h.p. every vertex of $\mathbb{G}_{n,p}$ has degree $(1+o(1))np$, see Theorem 3.4(ii).

Q2: We show that this is true w.h.p. in $\mathbb{G}$ and hence in G. Indeed,

$$\mathbb{P}\left(\exists S : e_{\mathbb{G}}(S) \geq |S| \log^3 n \text{ and } |S| \leq \frac{10 \log^2 n}{p}\right) \leq$$
$$\sum_{s=\log n}^{10p^{-1}\log^2 n} \binom{n}{s} \binom{\binom{s}{2}}{s\log^3 n} p^{s\log^3 n} \leq \sum_{s=\log n}^{10p^{-1}\log^2 n} \left(\frac{ne}{s}\left(\frac{sep}{2\log^3 n}\right)^{\log^3 n}\right)^s \leq$$
$$\sum_{s=\log n}^{10p^{-1}\log^2 n} \left(\frac{ne}{s}\left(\frac{5e}{\log n}\right)^{\log^3 n}\right)^s = o(1).$$

Q3: We show that this is true w.h.p. in $\mathbb{G}$ and hence in G. We first note that the Chernoff bounds, Corollary 27.7, imply that

$$\mathbb{P}\left(e_{\mathbb{G}}(S,T) \geq \left(1 + \frac{\alpha}{4}\right)|S||T|p\right) \leq e^{-\alpha^2 |S||T|p/50}.$$

So,

$$\mathbb{P}\left(\exists S,T : |S|,|T| \geq \frac{\log^2 n}{p} \text{ and } e_{\mathbb{G}}(S,T) \geq (1+\alpha)|S||T|p\right) \leq$$
$$\sum_{s,t=p^{-1}\log^2 n}^{n} \binom{n}{s}\binom{n}{t} e^{-\alpha^2 stp/50} \leq \sum_{s,t=p^{-1}\log^2 n}^{n} \left(\frac{ne^{1-\alpha^2 tp/100}}{s}\right)^s \left(\frac{ne^{1-\alpha^2 sp/100}}{t}\right)^t \leq$$
$$\sum_{s,t=p^{-1}\log^2 n}^{n} \left(\frac{ne^{1-\alpha^2 \log^2 n/100}}{s}\right)^s \left(\frac{ne^{1-\alpha^2 \log^2 n/100}}{t}\right)^t \leq$$
$$\left(\sum_{s=p^{-1}\log^2 n}^{n} \left(\frac{ne^{1-\alpha^2 \log^2 n/100}}{s}\right)^s\right)^2 = o(1).$$

□

Pseudo-random implies Hamiltonian

The rest of this section is devoted to the proof that if $G = ([n], E)$ is (n, α, p)-pseudo-random then G is Hamiltonian.

We randomly partition $[n]$ into sets let $V_i, i = 1, 2, \dots, 5$ such that

$$|V_1| = \left\lceil \frac{4\log^3 n}{p} \right\rceil, |V_i| = \frac{\alpha n}{5(1+2\alpha)}, i = 2,3,4$$

so that

$$|V_5| = \frac{5+7\alpha}{5+10\alpha} n - O\left(\frac{n}{\log^7 n}\right) \approx \frac{5+7\alpha}{5+10\alpha} n.$$

The number of neighbors of v in V_i is distributed as the binomial $\mathrm{Bin}(d_G(v), |V_i|/n)$. Thus for all $v \in [n]$ we have, using our assumption that $np \geq \log^{10} n$,

$$\mathbb{E}(d_{V_i}(v)) = \frac{|V_i|}{n} d_G(v) \geq \left(\frac{1}{2} + \alpha\right) |V_i| p \geq \left(\frac{1}{2} + \alpha\right) \frac{np}{\log^2 n} \gg \log n$$

and so the Chernoff bounds imply that w.h.p.

$$d_{V_i}(v) \gtrsim \left(\frac{1}{2} + \alpha\right) |V_i| p \text{ for all } v \in [n] \tag{20.2}$$

The proof now rests on two lemmas: the following quantities are fixed for the remainder of the proof:

$$\ell = 12\lceil \log n \rceil + 3 \text{ and } t = \left\lceil \frac{4\log^3 n}{p} \right\rceil \tag{20.3}$$

Lemma 20.6. *[Connecting Lemma] Let $\{a_i, b_i\}, i = 1, 2, \dots, t$ be a family of pairs of vertices from $[n]$ with $a_i \neq a_j$ and $b_i \neq b_j$ for every distinct $i, j \in [t]$, ($a_i = b_i$ is allowed). Let $L = \bigcup_{i=1}^t \{a_i, b_i\}$. Assume that $K \subseteq [n] \setminus L$ is such that*

C1 $|K| \gg \ell t \log t$.

C2 *For every $v \in K \cup L$ we have $|d_K(v)| \gtrsim \left(\frac{1}{2} + \alpha\right) p|K|$.*

Then there exist t internally disjoint paths $P_1, P_2, \dots, P_t$ such that P_i connects a_i to b_i and $V(P_i) \setminus \{a_i, b_i\} \subseteq K$. Furthermore, each path is of length ℓ.

Lemma 20.7. *[Absorbing Lemma] There is a path P^* with $V(P^*) \subseteq V_2 \cup V_3 \cup V_4$ such that for every $W \subseteq V_1$ there is a path P^*_W such that $V(P^*_W) = V(P^*) \cup W$ and such that P^* and P^*_W have the same endpoints.*

With these two lemmas in hand, we can show that G is Hamiltonian. Let P^* be as in Lemma 20.7 and let $U = (V_2 \cup V_3 \cup V_4 \cup V_5) \setminus V(P^*)$. If $v \in U$ then

$$d_U(v) \geq d_{V_5}(v) \gtrsim \left(\frac{1}{2}+\alpha\right)|V_5|p$$
$$\gtrsim \left(\frac{1}{2}+\alpha\right)\frac{5+7\alpha}{5+10\alpha}np \geq \left(\frac{1}{2}+\frac{3\alpha}{4}\right)|U|p. \quad (20.4)$$

Next let $k=\left\lfloor \frac{|U|\log^5 n}{n}\right\rfloor$ and $s=\left\lfloor \frac{n}{\log^5 n}\right\rfloor$. Randomly choose disjoint sets $S_1,S_2,\dots,S_k \subseteq U$ of size s and let $S=\bigcup_{i=1}^k S_i$ and $S'=U\setminus S$. It follows from (20.4) and the Chernoff bounds and the fact that

$$\mathbb{E}(d_{S_i}(v)) = \frac{|S_i|d_U(v)}{|U|} \gtrsim \left(\frac{1}{2}+\frac{3\alpha}{4}\right)|S_i|p \gg \log n$$

that w.h.p.

$$d_{S_i}(v) \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|S_i|p \text{ for all } i\in[k], v\in U. \quad (20.5)$$

***Claim* 6.** *Assuming* (20.5)*, we see that there is a perfect matching M_i between S_i,S_{i+1} for* $1\leq i<k$.

We prove the claim below. The matchings $M_1,M_2,\dots,M_{k-1}$ combine to give us s vertex disjoint paths Q_i from S_1 to S_k for $i=1,2,\dots,s$. Together with a 0-length path for each $v\in|S'|$ we have $t'=s+|S'|\leq 2s$ internally disjoint paths Q_i from x_i to $y_i, i=1,2,\dots,t'$ that cover U. Now let $t=t'+1$ and suppose that x_t,y_t are the endpoints of P^*. Applying Lemma 20.6 with $K=V_1$ and $a_i=y_i, b_i=x_{i+1}$ for $i\in[t]$ we obtain a cycle $C=(Q_1,P_1,Q_2,P_2,\dots,Q_{t'},P^*,P_t)$ (here $x_{t+1}=x_1$) that covers $V_2\cup V_3\cup V_4\cup V_5$, see Figure 20.1. Putting $W=V_1\setminus V(C)$ and using the fact that $P^*\subseteq C$, we can use Lemma 20.7 to extend C to a Hamilton cycle of G.

Proof of Claim 6

Fix i and now consider Hall's theorem. We have to show that $|N(X,S_{i+1})|\geq|X|$ for $X\subseteq S_i$.

Case 1: $|X|\leq 10p^{-1}\log^2 n$**.** Let $Y=N(X,S_{i+1})$ and suppose that $|Y|<|X|$. We now have

$$e_G(X\cup Y) \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|X||S_i|p \gtrsim \frac{|X\cup Y|\log^5 n}{4},$$

which contradicts **Q2**.

Case 2: $10p^{-1}\log^2 n\leq|X|\leq s/2$**.** In this case we have

$$e_G(X,Y) \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|S_i||X|p \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|X||Y|p,$$

which contradicts **Q3**. (If $|Y|<p^{-1}\log^2 n$ then we can add arbitrary vertices from $S_{i+1}\setminus Y$ to Y so that we can apply **Q3**.)

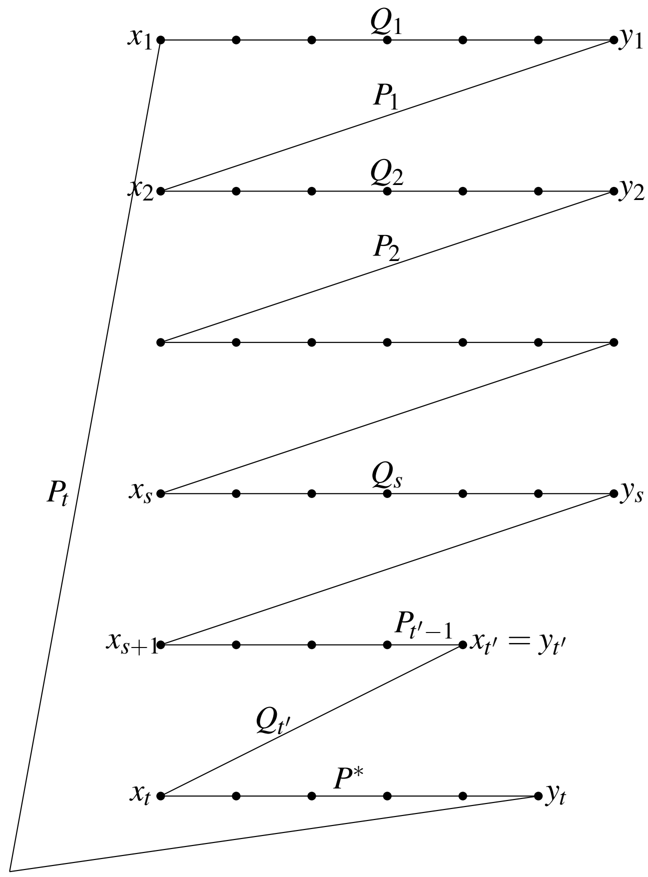


Figure 20.1: Cycle $C = (Q_1, P_1, Q_2, P_2, \ldots, Q_{t'}, P^*, P_t)$

The case $|X| > s/2$ is dealt with just as we did for $|S| > n/2$ in Theorem 6.1. This completes the proof of Claim 6.
End of proof of Claim 6

Proof of Lemma 20.6

We begin with some lemmas on expansion. For sets X,Y and integer ℓ, let $N_G^\ell(X,Y)$ be the set of vertices $y \in Y$ for which there exists $x \in X$ and a path P of length ℓ from x to y such that $V(P) \setminus \{x\} \subseteq Y$. We let $N_G(X,Y) = N_G^1(X,Y)$ denote the set of neighbors of X in Y. The sets X,Y need not be disjoint in this definition.

Lemma 20.8. *Suppose that $X,Y \subseteq [n]$ are (not necessarily disjoint) sets such that*

$|X| = \left\lfloor \frac{\log^2 n}{p} \right\rfloor$, $|Y| \geq \frac{3\log^3 n}{\alpha p}$ *and that* $|N_G(x,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right) p|Y|$ *for all* $x \in X$. *Then*

$$|N_G(X,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{3}\right)|Y|. \tag{20.6}$$

Proof. Let $Z = X \cup N_G(X,Y)$. We have

$$e_G(Z) \geq \sum_{x\in X} N_G(x,Y) - e_G(X) \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right) p|X||Y| - |X|\log^3 n \geq$$
$$\left(\frac{3}{2\alpha}+\frac{3}{2}\right)|X|\log^3 n = \left(\frac{3}{2\alpha}+\frac{3}{2}\right)\frac{|X|}{|Z|}\log^3 n \cdot |Z|$$

If $|Z| < \frac{10\log^2 n}{p}$ then we see that $e_G(Z) \gtrsim \frac{3}{20\alpha}\log^3 n$. This contradicts **Q2** for small α and so we can assume that $|Z| \geq \frac{10\log^2 n}{p}$ and therefore $|N_G(X,Y)| \geq \frac{9\log^2 n}{p}$. On the other hand, if (20.6) fails then from **Q3** we have

$$\left(\frac{1}{2}+\frac{\alpha}{2}\right) p|X||Y| \leq e_G(X, Y\setminus X) + 2e_G(X) \leq$$
$$e_G(X,Y\setminus X) + 2|X|\log^3 n \leq \left(1+\frac{\alpha}{4}\right)|X|\left(\frac{1}{2}+\frac{\alpha}{3}\right)|Y|p + 2|X|\log^3 n,$$

contradiction. □

Lemma 20.9. *Suppose that* $X,Y \subseteq [n]$ *are disjoint sets such that*

D1 $|Y| \geq \frac{3\log^3 n}{\alpha p}$.

D2 $|N_G(X,Y)| \geq \frac{2\log^2 n}{p}$.

D3 $|N_G(S,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y|$ *for all* $S \subseteq Y, |S| \geq \frac{\log^2 n}{p}$.

Then there exists $x \in X$ *such that* $|N_G^\ell(x,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{8}\right)|Y|$

Proof. We first show that there exists $x \in X$ such that $|N_G^{\ell-1}(x,Y)| \geq \frac{2\log^2 n}{p}$. For this we use the following claim:

Claim 7. *Let* $i < \ell$ *and* $A \subseteq X$ *be such that* $|N_G^i(A,Y)| \geq \frac{2\log^2 n}{p}$. *Then there exists* $A' \subseteq A$ *such that* $|A'| \leq \lceil |A|/2 \rceil$ *and* $|N_G^{i+1}(A',Y)| \geq \frac{2\log^2 n}{p}$.

We prove the claim below. Using **D2** and the claim $\ell-2$ times, we obtain a set $X' \subseteq X$ such that $|X'| \le \lceil |X|/2^{\ell-2} \rceil$ and $|N_G^{\ell-1}(X',Y)| \ge \frac{2\log^2 n}{p}$. But $\ell-2 \ge \log_2 n$ and so we have $|X'|=1$. Let $X'=\{x\}$ and $M \subseteq N_G(x,Y)$ be of size $\left\lceil \frac{\log^2 n}{p} \right\rceil$.
By definition, there is a path P_w of length $\ell-1$ from x to each $w \in M$. Let $V^* = (\bigcup_{w\in M} V(P_w)) \setminus \{x\}$. Then **D2** and **D3** imply

$$|N_G^{\ell}(x,Y)| \ge |N_G(M,Y\setminus V^*)| \ge \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y| - \ell|M| \ge \left(\frac{1}{2}+\frac{\alpha}{8}\right)|Y|.$$

□

Proof of Claim 7
First note that if A_1,A_2 is a partition of A with $|A_1| = \lceil |A|/2 \rceil$ then we have

$$|N_G^i(A_1,Y)| + |N_G^i(A_2,Y)| \ge |N_G^i(A,Y)| \ge \frac{2\log^2 n}{p}.$$

We can assume therefore that there exists $A' \subseteq A, |A'| \le \lceil |A|/2 \rceil$ such that $|N_G^i(A',Y)| \ge \frac{\log^2 n}{p}$.
Choose $B \subseteq N_G^i(A',Y), |B| = \left\lceil \frac{\log^2 n}{p} \right\rceil$. Then **D3** implies that $|N_G(B,Y)| \ge \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y|$.
Each $v \in B$ is the endpoint of a path P_v of length i from a vertex in A'. Let $V^* = \bigcup_{v\in B} V(P_v)$. Then,

$$|N_G^{i+1}(A',Y)| \ge |N_G(B,Y)| - |V^*| \ge \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y| - \ell|B| \ge \frac{2\log^2 n}{p}.$$

End of proof of Claim 7
We are now ready to prove an approximate version of Lemma 20.6. Let t,ℓ be as in (20.3).

Lemma 20.10. *Let $\{a_i,b_i\}, i=1,2,\dots,t$ be a family of pairs of vertices from $[n]$ with $a_i \ne a_j$ and $b_i \ne b_j$ for every distinct $i,j \in [t]$. Furthermore, let $R_A,R_B \subseteq [n] \setminus \bigcup_{i=1}^t \{a_i,b_i\}$ be disjoint and such that*

E1 $|R_A|,|R_B| \ge \frac{48t\ell}{\alpha}$.

E2 *For $Z=A,B$, $|N_G(S,R_Z)| \ge \left(\frac{1}{2}+\frac{\alpha}{4}\right)|R_Z|$ for all $S \subseteq R_A \cup R_B \cup \bigcup_{i=1}^t \{a_i,b_i\}$ such that $|S| \ge \frac{\log^2 n}{p}$.*

Then there exists a set $I \subseteq [t], |I| = \lfloor t/2 \rfloor$ and internally disjoint paths $P_i, i \in I$ such that P_i connects a_i to b_i and $V(P_i) \setminus \{a_i,b_i\} \subseteq R_A \cup R_B$.

Proof. We prove this by induction. Assume that we have found $s < \lfloor t/2 \rfloor$ paths P_i from a_i to b_i for $i \in J \subseteq I$, $|J| = s$. Then let

$$R'_A = R_A \setminus \bigcup_{i \in J} V(P_i),\ R'_B = R_B \setminus \bigcup_{i \in J} V(P_i).$$

Choose $h_A, h_B \geq 2\log n$ so that $h_A + h_B + 1 = \ell$.

***Claim* 8.** *There exists $i \in K = [t] \setminus J$ such that*

$$|N_G^{h_A}(a_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_A| \text{ and } |N_G^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_B|.$$

We verify Claim 8 below. Assume its truth for now. Let $S = N_G^{h_A}(a_i, R'_A)$. Then

$$|S| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)(R_A - s\ell) \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)\frac{47t\ell}{\alpha} \geq \frac{\log^2 n}{p}.$$

Now from **E2** we obtain

$$\begin{aligned} |N_G^{h_A+1}(a_i, R'_B)| &\geq |N_G(S, R'_B)| - \ell \geq |N_G(S, R_B)| - t\ell \\ &\geq \left(\frac{1}{2} + \frac{\alpha}{4}\right)|R_B| - t\ell \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R_B| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_B|. \end{aligned}$$

Now from Claim 8 we have that $|N_G^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_B|$ and so $N_G^{h_A+1}(a_i, R'_B) \cap N_G^{h_B}(b_i, R'_B) \neq \emptyset$ and there is a path as claimed.

It only remains to prove Claim 8. Assume inductively that we have found $v_1, v_2, \ldots, v_k \in \{a_i : i \in K\}$ such that $|N_G^{h_A}(v_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{16}\right)|R'_A|$ for $i \in [k]$. The base case is $k = 0$. We apply Lemma 20.9 with $Y = R'_A$ and $X = \{a_i : i \in K\} \setminus \{v_1, v_2, \ldots, v_k\}$. We check that the lemma's conditions are satisfied. $|R'_A| \geq \frac{48t\ell}{\alpha} - t\ell \geq \frac{47\ell \log^3 n}{\alpha p}$ and so **D1** is satisfied. On the other hand **E2** implies that if $S \subseteq R'_A \cup \{a_i : i \in K\}$ is of size at least $\frac{\log^2 n}{p}$ then

$$|N_G(S, R'_A)| \geq |N_G(S, R_A)| - t\ell \geq \left(\frac{1}{2} + \frac{\alpha}{4}\right)|R_A| - t\ell \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R_A| \gg \frac{\log^2 n}{p}.$$

So, **D3** is satisfied and also **D2** if $|X| \geq \frac{\log^2 n}{p}$ i.e. if $k \leq t/2$, completing the induction. So, we obtain $I_A \subseteq [t]$, $|I_A| = \lfloor t/2 \rfloor + 1$ such that

$$|N^{h_A}(a_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_A| \text{ for } i \in I_A.$$

A similar argument proves the existence of $I_B \subseteq [t]$, $|I_B| = \lfloor t/2 \rfloor + 1$ such that $|N^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_B|$ for $i \in I_B$ and the claim follows, since $I_A \cap I_B \neq \emptyset$ and we can therefore choose $i \in I_A \cap I_B$. $\square$

Completing the proof of Lemma 20.6

We now define some parameters that will be used for the remainder of the proof:

$$m = \lceil \log_2 t \rceil + 1, \quad s_i = 2t, i \in [2m], \quad s_{2m+1} = s_{2m+2} = \frac{|K|}{4}, \quad k = 2m+2.$$

We randomly choose disjoint sets $S_i \subseteq K, |S_i| = s_i, i \in [k]$. The Chernoff bounds and **C2** imply that w.h.p.

$$|N_G(v, S_i)| \geq \left(\frac{1}{2} + \frac{\alpha}{2}\right) ps_i \text{ for } v \in K \cup L,\ i \in [k]. \tag{20.7}$$

We prove the following lemma at the end of this section:

Lemma 20.11. *Given sets $[t] = I_1 \supseteq I_2 \supseteq \cdots I_m$ such that $|I_j| = \lceil |I_{j-1}|/2 \rceil$, G contains complete binary trees $T_A(i), T_B(i), i \in [t]$ such that*

F1 *The depth of $T_A(i), T_B(i), i \in I_s$ is $s-1$.*

F2 *$T_A(i)$ is rooted at a_i and $T_B(i)$ is rooted at b_i for $i \in [t]$.*

F3 *The vertices $T_A(i, j)$ at depth $j \in [0, m]$ in $T_A(i)$ are contained in S_j.*

F4 *The vertices $T_B(i, j)$ at depth $j \in [0, m]$ in $T_B(i)$ are contained in S_{m+j}.*

F5 *The trees are vertex disjoint.*

Assuming the truth of the lemma, we proceed as follows. We repeatedly use Lemma 20.10 to find vertex disjoint paths. We first find $\lfloor t/2 \rfloor$ paths $\mathscr{P}_1$ of length ℓ from a_i to b_i for $i \in J_1$. We then let $I_2 = I_1 \setminus J_1$ and construct the trees $T_A(i), T_B(i)$ for $i \in I_2$ and then use Lemma 20.10 once more to find $|I_2|$ vertex disjoint paths $\mathscr{Q}_2$ of length $\ell - 2$ from $T_A(i, 1)$ to $T_B(i, 1)$ for $i \in I_2$. We can now select at least half of the $\mathscr{Q}_2$ to make $\lceil |I_2|/2 \rceil$ paths $\mathscr{P}_2$ from a_i to b_i for $i \in J_2$ and then let $I_3 = I_2 \setminus J_2$. We repeat this process until we have constructed the required set of paths.

We now check that Lemma 20.10 can be applied as claimed. To simplify notation we use the convention that

$$T_A(i, j) = T_B(i, j) = \emptyset \text{ for } i \in \bigcup_{l=1}^{j} I_l.$$

With this convention, we let

$$M_s = K \setminus \left(\left(\bigcup_{i=1}^{t} \bigcup_{j=1}^{s-1} (T_A(i, j) \cup T_B(i, j)) \right) \cup \left(\bigcup_{i=1}^{s-1} \bigcup_{Q \in \mathscr{Q}_i} V(Q) \right) \right)$$

and then in round s we apply Lemma 20.10 with $a_i, b_i, i = 1,2,\ldots,t$ replaced by $x_j, y_j, j = 1,2,\ldots,t$ where the x_j are made up of the $T_A(i,s)$ for $i \in I_s$ and the y_j are made up of the $T_B(i,s)$ for $i \in I_s$ and

$$R_A = S_{2m+1} \setminus M_s \text{ and } R_B = S_{2m+2} \setminus M_s.$$

Thus

$$|R_A| \geq \frac{|K|}{4} - O\left(\sum_{i=1}^{t}\sum_{j=1}^{s-1} \frac{t}{2^j} \cdot 2^j \ell\right) = \frac{|K|}{4} - O(\ell t \log t) \geq \frac{|K|}{5} \gg \ell t \log t,$$

and similarly for $|S_B|$ and so **E1** holds.

Now suppose that $S \subseteq R_A \cup R_B \cup \bigcup_{i=1}^{t} \{x_j, y_j\}$ with $|S| \geq \frac{\log^2 n}{p}$. We can apply Lemma 20.8 with $X = S$ and $Y = R_A$ because **E1** holds and because of Assumption **C2** and (20.7). It follows that

$$|N_G(S, R_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{3}\right)|R_A|$$

and similarly for R_B and so **E2** holds. It now follows from Lemma 20.10 that there are $\lfloor t/2 \rfloor$ indices i for which there is a path from x_j to y_j and these yield at least $\lceil t/2^s \rceil$ indices I_s and a path from a_i to b_i for $i \in I_s$.

It remains only to prove Lemma 20.11.

Proof. Let $V_A(s)$ denote the endpoints of $\mathscr{Q}_s$ in S_s. Because of (20.7), we can reduce this to the following: given a bipartite graph Γ with bipartion $X = \{x_1, x_2, \ldots, x_{\lfloor t/2 \rfloor}\} \subseteq S_s \setminus V_A(s)$, $B = S_{s+1} = \{y_1, y_2, \ldots, y_{2t}\}$ and minimum degree at least $\left(\frac{1}{2} + \frac{\alpha}{2}\right) ps_{i+1} \geq 4\log^3 n$ and such that **Q2, Q3** hold, there exists a partition of B into t pairs $\{z_{i,1}, z_{i,2}\}$, $i \in [t]$ such that both edges $\{x_i, z_{i,l}\}$, $l = 1,2$ exist in Γ. (The reader can check that after each round, there are $t/2$ vertices that are leaves of the current active trees, and need two neighbors to grow the tree. We say that a tree is active if its root is not the endpoint of a path in $\mathscr{Q}_s$.)

We need to verify the following condition:

$$S \subseteq A \text{ implies } |N_\Gamma(S,B)| \geq 2|S|. \tag{20.8}$$

Case 1: $|S| \leq \frac{2\log^2 n}{3p}$**.**

Let $T = N_\Gamma(S,B)$. If (20.8) fails then $|S \cup T| \leq \frac{2\log^2 n}{p}$ and $e_G(S \cup T) > |S|\log^3 n$, violating **Q2**.

Case 2: $|S| > \frac{2\log^2 n}{3p}$**.**

If (20.8) fails then

$$\left(\frac{1}{2} + \frac{\alpha}{2}\right) s_i |S| p \leq e_G(S,T) \leq 2\left(1 + \frac{\alpha}{4}\right)|S|^2 p. \tag{20.9}$$

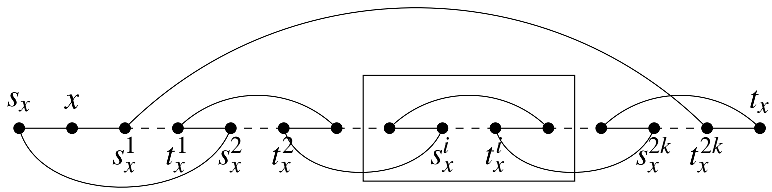


Figure 20.2: The absorber for $k_1 = 3$. The cycle C_x is drawn with solid lines. The dashed lines represent paths. The part inside the rectangle can be repeated to make larger absorbers.

The lower bound is from (20.7) and the upper bound in (20.9) is from **Q3**. Equation (20.9) is a contradiction, because $s_i = 2t \geq 4|S|$. □

Proof of Lemma 20.7

Let ℓ_x be an integer and A_x a graph with $|V(A_x)| = \ell_x + 1$. A_x is called an *absorber* if there are vertices x, s_x, t_x such that A_x contains paths P_x, P_x' of lengths $\ell_x, \ell_x - 1$ from s_x to t_x such that $x \notin V(P_x')$.

Let k, ℓ be integers and consider the graph A_x with $3 + 2k(\ell + 1)$ vertices constructed as follows:

S1 A_x contains a cycle C_x of length $4k + 3$, the solid lines in Figure 20.3.

S2 A_x contains $2k$ pairwise disjoint s_x^i, t_x^i paths, $P_1, P_2, \ldots, P_{2k}$, each of which is of length ℓ, the dashed lines in Figure 20.3.

Lemma 20.12. *A_x is an absorber for vertex x.*

Proof. We take

$$P_x = s_x, x, s_1^x, P_1, t_1^x, s_x^2, P_2, \ldots, t_x^{2k}, t_x.$$
$$P_x' = s_x, s_x^2, P_2, t_x^2, s_x^4, P_4, \ldots, s_x^{2k}, P_{2k}, t_x^{2k}, s_x^1, P_1, t_x^1, s_x^3, P_3, \ldots, t_x^{2k-1}, t_x.$$

□

We first apply Lemma 20.6 to find $C_x, x \in V_1$. We let $L = V_1$ and let $a_i = b_i = x, x \in V_1$. We let $K = V_2$ and $k = 3\lceil \log n \rceil$ so that $\ell = 4k + 3$. The lemma is applicable as $|K| = \frac{\alpha n}{5(1+2\alpha)} \gg \ell t \log t$ and (20.2) implies that **C2** holds.

We next apply Lemma 20.6 to connect $a_i = s_x^i$ to $b_i = t_x^i$ by a path of length ℓ for $i \in [2k], x \in V_1$. We let $K = V_3$ and note that $|K| = \frac{\alpha n}{5(1+2\alpha)} \gg \ell t \log t$ and that (20.2) implies that **C2** holds, once more.
At this point we have paths P_x, P'_x for $x \in V_1$. We finally construct P^*, using Lemma 20.6 to connect the paths $P'_x, x \in V_1$. We let $t = h-1$ where $V_1 = \{x_1, x_2, \dots, x_h\}$. We take $a_i = t_{x_i}$ and $b_i = s_{x_{i+1}}$ for $i \in [h-1]$ and $K = V_4$.
It is easy to see that this construction has the desired property. Where necessary, we can absorb $x \in V_1$ by replacing P'_x by P_x.

20.4 The chromatic number

In this section we consider *adding* the edges of a fixed graph H to $\mathbb{G}_{n,p}$. We examine the case where p is constant and where $\Delta = \Delta(H)$ is sufficiently small. We will see that under some circumstances, we can add quite a few edges without increasing the chromatic number by very much.

Theorem 20.13. *Suppose that H is a graph on vertex set $[n]$ with maximum degree $\Delta = n^{o(1)}$. Let p be constant and let $G = \mathbb{G}_{n,p} + H$. Then w.h.p. $\chi(G) \approx \chi(\mathbb{G}_{n,p})$,* for all choices of H.

Proof. We first observe that

$$\text{W.h.p. every set of } t \leq \frac{n}{10\log^2 n} \text{ vertices, spans fewer than } \frac{2npt}{\log^2 n} \text{ edges of } \mathbb{G}_{n,p}. \quad (20.10)$$

Indeed,

$$\Pr(\exists \text{ a set negating (20.10)}) \leq \sum_{t=2np/\log^2 n}^{n/10\log^2 n} \binom{n}{t} \binom{\binom{t}{2}}{2npt/\log^2 n} p^{2npt/\log^2 n}$$
$$\leq \sum_{t=2np/\log^2 n}^{n/10\log^2 n} \left(\frac{ne}{t}\right)^t \left(\frac{t^2 ep \log^2 n}{4npt}\right)^{2npt/\log^2 n} \leq$$
$$\sum_{t=2np/\log^2 n}^{n/10\log^2 n} \left(\frac{ne}{t} \cdot \left(\frac{te}{4n}\right)^{2np/\log^2 n}\right)^t = o(1).$$

We let $s = 20\Delta \log^2 n$ and randomly partition $[n]$ into s sets $V_1, V_2, \dots, V_s$ of size n/s. Let Y denote the number of edges of H that have endpoints in the same set of

the partition. Then $\mathbb{E}(Y) \leq \frac{|E(H)|}{s} \leq \frac{\Delta n}{2s}$. Therefore, by the Markov inequality, at least one half of partitions have $Y \leq \Delta n/s$. So, if we choose $\log n$ random partitions then w.h.p. at least one will satisfy $Y \leq \Delta n/s$. Furthermore, it follows from (7.21), that w.h.p. the subgraphs G_i of $\mathbb{G}_{n,p}$ induced by each V_i have chromatic number $\approx \frac{n/s}{2\log_b(n/s)}$. Indeed, this will be true for all $\log n$ partitions.
Given $V_1, V_2, \ldots, V_s$, we color G as follows: we color the edges of G_i with $\approx \frac{n/s}{2\log_b(n/s)}$ colors, using different colors for each set and $\approx \frac{n}{2\log_b(n/s)} \approx \frac{n}{2\log_b(n)} \approx \chi(\mathbb{G}_{n,p})$ colors overall. We must of course deal with the at most Y edges that could be improperly colored. Let W denote the endpoints of these edges. Then $|W| \leq \frac{n}{10\log^2 n}$. It follows from (20.10) that we can write $W = \{w_1.w_2, \ldots, w_m\}$ such that w_i has at most $\frac{2np}{\log^2 n} + \Delta$ neighbors in $\{w_1, w_2, \ldots, w_{i-1}\}$ i.e. the *coloring number* of the subgraph of $\mathbb{G}_{n,p}$ induced by W is at most $\frac{2np}{\log^2 n} + \Delta$. It follows that we can re-color the Y badly colored edges using at most $\frac{2np}{\log^2 n} + \Delta + 1 = o(\chi(\mathbb{G}_{n,p}))$ new colors. □

20.5 Exercises

20.5.1 Prove that if $p \geq \frac{(1+\eta)\log n}{n}$ for a postive constant η then $\Delta_{\mathscr{C}} \geq \left(\frac{1}{2} - \varepsilon\right) np$, where $\mathscr{C}$ denotes connectivity. (See Haller and Trujić [461].)

20.5.2 Show that for every $\varepsilon > 0$ there exists $c_\varepsilon > 0$ such that the following is true w.h.p. If $c \geq c_\varepsilon$ and $p = c/n$ and we remove any set of at most $(1-\varepsilon)cn/2$ edges from $\mathbb{G}_{n,p}$, then the remaining graph contains a component of size at least $\varepsilon n/4$.

20.6 Notes

Sudakov and Vu [822] were the first to discuss local resilience in the context of random graphs. Our examples are taken from this paper except that we have given a proof of hamiltonicity that introduces the absorbing method.

Hamiltonicity

Sudakov and Vu proved local resilience for $p \geq \frac{\log^4 n}{n}$ and $\Delta_{\mathscr{H}} = \frac{(1-o(1))np}{2}$. The expression for $\Delta_{\mathscr{H}}$ is best posible, but the needed value for p has been lowered. Frieze and Krivelevich [399] showed that there exist constants K, α such that w.h.p. $\Delta_{\mathscr{H}} \geq \alpha np$ for $p \geq \frac{K\log n}{n}$. Ben-Shimon, Krivelevich and Sudakov [97] improved this to $\alpha \geq \frac{1-\varepsilon}{6}$ holds w.h.p. and then in [98] they obtained a result on resilience for $np - (\log n + \log\log n) \to \infty$, but with K close to $\frac{1}{3}$. (Vertices of degree less than $\frac{np}{100}$ can lose all but two incident edges.) Lee and Sudakov [617] proved the sought after result that for every positive ε there exists $C = C(\varepsilon)$ such that w.h.p. $\Delta_{\mathscr{H}} \geq \frac{(1-\varepsilon)np}{2}$ holds for $p \geq \frac{C\log n}{n}$. Condon, Espuny Díaz, Kim, Kühn and Osthus [239] refined [617]. Let H be a graph with degree sequence $d_1 \geq d_2 \geq \cdots \geq d_n$ where $d_i \leq (n-i)p - \varepsilon np$ for $i < n/2$. They say that G is ε-Pósa-resilient if $G - H$ is Hamiltonian for all such H. Given $\varepsilon > 0$ there is a constant $C = C(\varepsilon)$ such that if $p \geq \frac{C\log n}{n}$ then $G_{n,p}$ is ε-Pósa-resilient w.h.p. The result in [617] has now been improved to give a hitting time result, see Montgomery [689] and Nenadov, Steger and Trujić [714]. The latter paper also proves the optimal resilience of the 2-core when $p = \frac{(1+\varepsilon)\log n}{3n}$.

Fischer, Škorić, Steger and Trujić [361] have shown that there exists $C > 0$ such that if $p \geq \frac{C\log^3 n}{n^{1/2}}$ then not only is there the square of a Hamilton cycle w.h.p., but containing a square is resilient to the deletion of not too many triangles incident with each vertex.

Krivelevich, Lee and Sudakov [593] proved that $G = G_{n,p}, p \gg n^{-1/2}$ remains pancyclic w.h.p. if a subgraph H of maximum degree $(\frac{1}{2} - \varepsilon)np$ is deleted, i.e. pancyclicity is locally resilient. The same is true for random regular graphs when $r \gg n^{1/2}$.

Hefetz, Steger and Sudakov [478] began the study of the resilience of Hamiltonicity for random digraphs. They showed that if $p \gg \frac{\log n}{n^{1/2}}$ then w.h.p. the Hamiltonicity of $D_{n,p}$ is resilient to the deletion of up to $(\frac{1}{2} - o(1))np$ edges incident with each vertex. The value of p was reduced to $p \gg \frac{\log^8 n}{n}$ by Ferber, Nenadov, Noever, Peter and Škorić [357]. Finally, Montgomery [691] proved that in the random digraph process, at the hitting time for Hamiltonicity, the property is resilient w.h.p.

Chapter 21

Extremal Properties

A typical question in extremal combinatorics can be viewed as "how many edges of the complete graph (or hypergraph) on n vertices can a graph have without having some property $\mathscr{P}$". In recent years research has been carried out where the complete graph is replaced by a random graph.

21.1 Containers

Ramsey theory and the Turán problem constitute two of the most important areas in extremal graph theory. For a fixed graph H we can ask how large should n be so that in any r-coloring of the edges of K_n can we be sure of finding a monochromatic copy of H – a basic question in Ramsey theory. Or we can ask for the maximum $\alpha > 0$ such that we take an α proportion of the edges of K_n without including a copy of H – a basic question related to the Turán problem. Both of these questions have analogues where we replace K_n by $\mathbb{G}_{n,p}$.

There have been recent breakthroughs in transferring extremal results to the context of random graphs and hypergraphs. Conlon and Gowers [241], Schacht [793], Balogh, Morris and Samotij [69] and Saxton and Thomason [791] have proved general theorems enabling such transfers. We make use of a new short proof of the *container theorem* due to Nenadov and Pham [710]. Our exposition has been much improved by the comments of Robert Krueger [610].

In this section, we present a special case of Theorem 2.3 of [791] that will enable us to deal with Ramsey and Turán properties of random graphs. For a graph H with $e(H) \geq 2$ we let

$$m_2(H) = \max_{H' \subseteq H, e(H') > 1} \frac{e(H') - 1}{v(H') - 2}. \tag{21.1}$$

Next let

$$\pi(H) = \lim_{n\to\infty} \frac{ex(n,H)}{\binom{n}{2}} \tag{21.2}$$

where as usual, $ex(n,H)$ is the maximum number of edges in an H-free subgraph of K_n.

Theorem 21.1. *For every $\varepsilon > 0$ there exists $h > 0$ such that the following holds: Let H be a graph with $e(H) \geq 2$. Then for every H-free graph I there exist graphs C, F such that $C = C(F)$ and $F \subseteq I \subseteq C$ such that*

(i) $|F| \leq hn^{2-1/m_2(H)}$.

(ii) *For every graph C, the number of copies of H in C is at most $\varepsilon n^{v(H)}$.*

(iii) *For every graph C, $e(C) \leq (\pi(H)+\varepsilon)\binom{n}{2}$.*

(Note that Proposition 21.13 below means that we only need to verify (ii).)

We prove Theorem 21.1 in Section 21.4. But first we give a couple of examples of the use of this theorem.

21.2 Ramsey Properties

The investigation of the Ramsey properties of $\mathbb{G}_{n,p}$ was initiated by Łuczak, Ruciński and Voigt [645]. Later, Rödl and Ruciński [769], [772] proved that the following holds w.h.p. for some constants $0 < c < C$. Here H is some fixed graph containing at least one cycle. Suppose that the edges of $\mathbb{G}_{n,m}$ are colored with r colors. If $m < cn^{2-1/m_2(H)}$ then w.h.p. there exists an r-coloring without a mono-chromatic copy of H, while if $m > Cn^{2-1/m_2(H)}$ then w.h.p. in every r-coloring there is a monochromatic copy of H.

We will give a proof of the 1-statement based on Theorem 21.1. We will closely follow the argument in a recent paper of Nenadov and Steger [711]. The notation $G \to (H)^e_r$ means that in every r-coloring of the edges of G there is a copy of H with all edges the same color. Rödl and Ruciński [772] proved the following

Theorem 21.2. *For any graph H with $e(H) \geq v(H)$ and $r \geq 2$, there exist $c_0, c_1 > 0$ such that*

$$\mathbb{P}(\mathbb{G}_{n,p} \to (H)^e_r) = \begin{cases} o(1) & p \leq c_0 n^{-1/m_2(H)} \\ 1-o(1) & p \geq c_1 n^{-1/m_2(H)} \end{cases}$$

The density $p_0 = n^{-1/m_2(H)}$ is the threshold for every edge of $\mathbb{G}_{n,p}$ to be contained in a copy of H. When $p \leq cp_0$ for small c, the copies of H in $\mathbb{G}_{n,p}$ will be spread out and the associated 0-statement is not so surprising. We will use

Theorem 21.1 to prove the 1-statement for $p \geq c_1 p_0$. The proof of the 0-statement follows [711] and is given in Exercises 21.5.1 to 21.5.6.

We begin with a couple of lemmas:

Lemma 21.3. *For every graph H and $r \geq 2$ there exist constants $\alpha > 0$ and n_0 such that for all $n \geq n_0$ every r-coloring of the edges of K_n contains at least $\alpha n^{v(H)}$ monochromatic copies of H.*

Proof. From Ramsey's theorem we know that there exists $N = N(H,r)$ such that every r-coloring of the edges of K_N contains a monochromatic copy of H. Thus, in any r-coloring of K_n, every N-subset of the vertices of K_n contains at least one monochromatic copy of H. As every copy of H is contained in at most $\binom{n-v(H)}{N-v(H)}$ N-subsets, the theorem follows with $\alpha = 1/N^{v(H)}$. $\square$

From this we get

Corollary 21.4. *For every graph H and every positive integer r there exist constants n_0 and $\delta, \varepsilon > 0$ such that the following is true: If $n \geq n_0$, then for any $E_1, E_2, \dots, E_r \subseteq E(K_n)$ such that for all $1 \leq i \leq r$ the set E_i contains at most $\varepsilon n^{v(H)}$ copies of H, we have*

$$|E(K_n) \setminus (E_1 \cup E_2 \cup \cdots \cup E_r)| \geq \delta n^2.$$

Proof. Let α and n_0 be as given in Lemma 21.3 for H and $r+1$. Further, let $E_{r+1} = E(K_n) \setminus (E_1 \cup E_2 \cup \cdots \cup E_r)$, and consider the coloring $f : E(K_n) \to [r+1]$ given by $f(e) = \min_{i \in [r+1]} \{e \in E_i\}$. By Lemma 21.3 there exist at least $\alpha n^{v(H)}$ monochromatic copies of H under coloring f, and so by our assumption on the sets $E_i, 1 \leq i \leq r$, E_{r+1} must contain at least $\alpha n^{v(H)}$ copies. As every edge is contained in at most $e(H) n^{v(H)-2}$ copies and $E_1 \cup E_2 \cup \cdots E_r$ contains at most $r\varepsilon n^{v(H)}$ copies of H, we see that E_{r+1} contains at least $(\alpha - r\varepsilon e(H)) n^{v(H)}$ copies of H. It follows that $|E_{r+1}| \geq \frac{(\alpha - r\varepsilon e(H)) n^{v(H)}}{e(H) n^{v(H)-2}}$ and so the corollary follows with $\delta = \frac{\alpha - re(H)\varepsilon}{e(H)}$. Here we take $\varepsilon \leq \frac{\alpha}{2re(H)}$. $\square$

Proof. We can now proceed to the proof of the 1-statement of Theorem 21.2. If $\mathbb{G}_{n,p} \not\to (H)^e_r$ then there must exist a coloring $f : E(\mathbb{G}_{n,p}) \to [r]$ such that for all $1 \leq i \leq r$ the set $E_i = f^{-1}(i)$ does not contain a copy of H. By Theorem 21.1 we have that for every such E_i there exists F_i and a container C_i such that $F_i \subseteq E_i \subseteq C_i$. The crucial observation is that $\mathbb{G}_{n,p}$ completely avoids $E_0 = E(K_n) \setminus (C_1 \cup C_2 \cup \cdots \cup C_r)$, which by Corollary 21.4 and a choice of ε has size at least δn^2.

Therefore, we can bound $\mathbb{P}(\mathbb{G}_{n,p} \not\to (H)^e_r)$ by the probability that there exist $\mathscr{F} = \{F_1,\dots,F_r\}$ and $\mathscr{C} = \{C_i = C(T_i) : i = 1,2,\dots,r\}$ such that E_0 is edge-disjoint from $\mathbb{G}_{n,p}$. Thus,

$$\mathbb{P}((\mathbb{G}_{n,p} \not\to (H)^e_r) \le \sum_{F_i, 1\le i\le r} \mathbb{P}(T_i \subseteq \mathbb{G}_{n,p}, 1 \le i \le r \wedge E(\mathbb{G}_{n,p}) \cap E_0 = \emptyset).$$

Note that the two events in the above probability are independent and can thus be bounded by $p^a(1-p)^b$ where $a = |\bigcup_i T_i|$ and $b = \delta n^2$. The sum can be bounded by first deciding on $a \le rhn^{2-1/m_2(H)}$ (h from Theorem 21.1) and then choosing a edges ($\binom{\binom{n}{2}}{a}$ choices) and then deciding for every edge in which T_i it appears (r^a choices). Thus,

$$\begin{aligned}\mathbb{P}((\mathbb{G}_{n,p} \not\to (H)^e_r) &\le e^{-\delta n^2 p} \sum_{a=0}^{rhn^{2-1/m_2(H)}} \binom{\binom{n}{2}}{a} (rp)^a \\ &\le e^{-\delta n^2 p} \sum_{a=0}^{rhn^{2-1/m_2(H)}} \left(\frac{en^2rp}{2a}\right)^a.\end{aligned}$$

Recall that $p = c_1 n^{-1/m_2(H)}$. By choosing c_1 sufficiently large with respect to h we get

$$\begin{aligned}\sum_{a=0}^{rhn^{2-1/m_2(H)}} \left(\frac{en^2rp}{2a}\right)^a &\le n^2 \left(\frac{erc_1}{2rh}\right)^{(rh/c_1)n^2p} \\ &= n^2 \left(\left(\frac{ec_1}{2h}\right)^{2rh/c_1\delta}\right)^{\delta n^2 p/2} \le e^{\delta n^2 p/2},\end{aligned}$$

and thus $\mathbb{P}((\mathbb{G}_{n,p} \not\to (H)^e_r) = o(1)$ as desired. Recall that $(eA/x)^x$ is unimodal with a maximum at $x = A$ and that c_1 is large. This implies that $n^2rp/2 > rhn^{2-1/m_2(H)}$, giving the first inequality and $\left(\frac{ec_1}{2h}\right)^{2rh/c_1\delta} < e$, giving the second inequality. □

21.3 Turán Properties

Early success on the Turán problem for random graphs was achieved by Haxell, Kohayakawa and Łuczak [472], [473], Kohayakawa, Kreuter and Steger [576], Kohayakawa, Łuczak and Rödl [577], Gerke, Prömel, Schickinger and Steger [428], Gerke, Schickinger and Steger [429], Łuczak [641]. It is only recently that Turán's theorem in its full generality has been transferred to $\mathbb{G}_{n,p}$.

From its definition, every H-free graph with n vertices will have $(\pi(H)+o(1))\binom{n}{2}$ edges. In this section we prove a corresponding result for random graphs. Our proof is taken from [791], although Conlon and Gowers [241] gave a proof for 2-balanced H and Schacht [793] gave a proof for general H.

Theorem 21.5. *Suppose that $0 < \gamma < 1$ and H is not a matching. Then there exists $A > 0$ such that if $p \geq An^{-1/m_2(H)}$ and n is sufficiently large then the following event occurs with probability at least $1 - e^{-\gamma^3\binom{n}{2}p/384}$:*

$$\text{Every } H\text{-free subgraph of } \mathbb{G}_{n,p} \text{ has at most } (\pi(H)+\gamma)\binom{n}{2}p \text{ edges.}$$

If H is a matching then $m_2(H) = 1/2$ and then the lower bound on p in the theorem is $O(n^{-2})$ we would not be claiming a high probability result.

To prove the theorem, we first prove the following lemma:

Lemma 21.6. *Given $0 < \eta < 1$ and $h \geq 1$, there is a constant $\varphi = \varphi(\eta, h)$ such that the following holds: Let M be a set, $|M| = N$ and let $\mathscr{I} \subseteq 2^M$. Let $t \geq 1$, $\varphi t/N \leq p \leq 1$ and let $\eta N/2 \leq d \leq N$. Suppose there exists $C : 2^M \to 2^M$ and $\mathscr{T} \subseteq \binom{M}{\leq t}$ such that for each $I \in \mathscr{I}$ there exists $T_I \in \mathscr{T}$ such that $T_I \subseteq I$ and $C_I = C(T_I) \subseteq M$, where $|C_I| \leq d$. Let $X \subseteq M$ be a random subset where each element is chosen independently with probability p. Then*

$$\mathbb{P}(\exists I \in \mathscr{I} : |C_I \cap X| > (1+\eta)pd \text{ and } I \subseteq X) \leq e^{-\eta^2 dp/24}. \tag{21.3}$$

Proof. For $T \in \mathscr{T}$ let E_T be the event that

$$T \subseteq X \text{ and } |C(T) \cap X| \geq (1+\eta)pd.$$

The event E_T is contained in $F_T \cap G_T$ where F_T is the event that $T \subseteq X$ and G_T is the event that $|(C(T) \setminus T) \cap X| \geq (1+\eta)dp - |T|$. Since F_T and G_T are independent, $\mathbb{P}(E_T) \leq \mathbb{P}(F_T)\mathbb{P}(G_T)$. Now $|T| \leq t \leq Np/\varphi \leq 2dp/\varphi\eta \leq \eta dp/2$ if φ is large. So by the Chernoff bound, see Lemma 27.6,

$$\mathbb{P}(G_T) \leq \mathbb{P}(\text{Bin}(d,p) \geq (1+\eta/2)dp) \leq e^{-\eta^2 dp/12}.$$

Note that $\mathbb{P}(F_T) = p^{|T|}$. Let $x = Np/t \geq \varphi$, so that $t \leq Np/x \leq 2dp/\eta x$. If φ is large we may assume that $p(N-t) > t$. So

$$\sum_T \mathbb{P}(F_T) \leq \sum_{i=0}^{t} \binom{N}{i} p^i \leq 2\left(\frac{eNp}{t}\right)^t = 2(xe)^t \leq 2(xe)^{2dp/\eta x} \leq e^{\eta^2 dp/24},$$

if φ, and therefore x, is large. If there exists $I \subseteq X, I \in \mathscr{I}$ with $|C(T_I) \cap X| \geq (1+\eta)dp$ then the event E_T holds. Hence the probability in (21.3) is bounded by

$$\sum_T \mathbb{P}(F_T)\mathbb{P}(G_T) \leq e^{\eta^2 dp/24} e^{-\eta^2 dp/12} = e^{-\eta^2 dp/24}.$$

$\square$

With this lemma in hand, we can complete the proof of Theorem 21.5.

Let $\mathscr{I}$ be the set of H-free graphs on vertex set $[n]$. We take $M = \binom{[n]}{2}$ and $X = E(\mathbb{G}_{n,p})$ and $N = \binom{n}{2}$. For $I \in \mathscr{I}$, let T_I and $h = h(H,\varepsilon)$ be given by Theorem 21.1. Each H-free graph $I \in \mathscr{I}$ is contained in C_I and so if $\mathbb{G}_{n,p}$ contains an H-free subgraph with $(\pi(H)+\gamma)Np$ edges then there exists I such that $|X \cap C_I| \geq (\pi(H)+\gamma)Np$. Our aim is to apply Lemma 21.6 with

$$\eta = \frac{\gamma}{2},\ d = \left(\pi(H) + \frac{\gamma}{4}\right)N,\ t = hn^{2-1/m_2(H)}.$$

The conditions of Lemma 21.6 then hold after noting that $d \geq \eta N/2$ and that $p \geq An^{-1/m_2(H)} \geq \varphi t/N$ if A is large enough. Note also that $|C_I| \leq d$. Now $(1+\eta)dp \leq (\pi(H)+\gamma)Np$, and so the probability that the event in the statement of the theorem fails to occur is bounded by

$$e^{-\eta^2 dp/24} \leq \exp\left\{-\frac{\gamma^3 Np}{384}\right\}$$

completing the proof. □

21.4 Containers and the proof of Theorem 21.1

Let V be a finite set. Given a subset $X \subseteq V$, let $\langle X\rangle = \{S \subseteq V : X \subseteq S\}$. Throughout this section we use $V = V(\mathscr{H})$ and $N = |V|$, where $\mathscr{H}$ is a given hypergraph. If all edges in a hypergraph $\mathscr{H}$ have size at most ℓ, we say that $\mathscr{H}$ is an $(\leq \ell)$-*graph*. In an ℓ-graph or ℓ-uniform hypergraph, all edges are of size exactly ℓ. The following is a re-arrangement of the proof due to Nenadov and Pham [710].

We say a probability measure μ over 2^V is (p,K)-*uniformly-spread* if we have

$$\mu(\langle X\rangle) \leq \frac{Kp^{|X|-1}}{N} \text{ for every non-empty } X \subseteq V. \tag{21.4}$$

Remark 21.7. Note that if μ is the uniform distribution on the edges of $\mathscr{H}$, then $\mu(v) = deg(v)/|\mathscr{H}|$. Requiring this to be at most K/N is equivalent to the maximum degree being at most K times the average degree. Then the higher co-degrees each shrink by a factor of p.

If $S \subseteq V$ then $\mathscr{H}[C]$ is the sub-hypergraph of$\mathscr{H}$ made up of the edges of $\mathscr{H}$ that are contained in S.

Theorem 21.8. *For every $\ell \in \mathbb{N}$ and $K, \varepsilon > 0$ there exists $h > 0$ such that the following holds. Suppose $\mathscr{H}$ is an $(\leq \ell)$-graph, and let μ be (p,K)-uniformly-spread measure over 2^V, supported on $\mathscr{H}$. Then for every independent set $I \subseteq V(\mathscr{H})$ there exist $F \subseteq I \subseteq C = C(F) \subseteq V$ such that* *(a)* $|F| \leq hNp$.

(b) $\mu(\mathscr{H}[C]) < \varepsilon$.

Theorem 21.8 follows by iterated application of the following lemma, known as the *hypergraph container* lemma.

Lemma 21.9. *For every $\ell \in \mathbb{N}$ and $K > 1$ there exists $\delta > 0$ such that the following holds. Suppose $\mathscr{H}$ is an $(\leq \ell)$-graph, and let μ be a (p,K)-uniformly-spread measure over 2^V, supported on $\mathscr{H}$. Then for every independent set $I \subseteq V$ there exists $F \subseteq I \subseteq C = C(F) \subseteq V$ such that*

(a) $|F| \leq \ell Np$.

(b) $\mu(C) \leq (1-\delta)$.

Moreover, C can be unambigously constructed from any $F \subseteq \hat{F} \subseteq I$.

Proof of Lemma 21.9. We prove the lemma by induction on ℓ. For $\ell = 1$, take $F = \emptyset$ and $C \subseteq V$ to be the set of all vertices $v \in V$ with $\mu(v) = 0$. As there are at least N/K vertices with strictly positive measure, the lemma holds for $\delta = 1/K$, since (21.4) implies that $\mu(\langle x \rangle) \leq K/N$. We now prove the lemma for $\ell \geq 2$. Without loss of generality, we may assume $|I| \geq Np$, since otherwise we can take $F = I = C$.

Set $F = \emptyset \subseteq I$, $\mathscr{L} = \emptyset \subseteq 2^V$, and $\mathscr{D}, \mathscr{H}' = \emptyset \subseteq \mathscr{H}$. Repeat the following for Np rounds:

Step 1 Take $v \in I \setminus F$ to be a largest vertex with respect to $\mu(\langle v \rangle \cap \mathscr{R})$, where $\mathscr{R} = \mathscr{H}[V \setminus F] \setminus \mathscr{D}$ (tie-breaking done in some canonical way, e.g. by agreeing on the ordering of V).
The larger the degree, the more vertices are ruled out from C.

Step 2 Add v to F.

Step 3 Set $\mathscr{H}' = \mathscr{H}' \cup (\langle v \rangle \cap \mathscr{R})$.

Step 4 For each $X \in 2^V \setminus \mathscr{L}$ of size $|X| \leq \ell - 1$ such that

$$\mu(\langle X \rangle \cap \mathscr{H}') > \frac{Kp^{|X|}}{N}, \tag{21.5}$$

add X to $\mathscr{L}$ and set $\mathscr{D} = \mathscr{D} \cup (\langle X \rangle \cap \mathscr{R})$.
We remove the edges of $\mathscr{D}$ from consideration so that hypergraph $\mathscr{H}''$ below has a uniformly spread measure in Case 1, below. After whch we can apply induction.

After this we will construct C. There are two cases. Let $\alpha = 2^{-\ell-2}$. If $\mu(\mathscr{H}')$ is large, then we can apply the inductive hypothesis to an appropriate $(\leq \ell-1)$-graph, otherwise we can immediately find a small container C for which $I \subseteq C$.

Case 1: $\mu(\mathscr{H}') \geq \alpha p$.

As μ is (p,K)-uniformly-spread the value $\mu(\langle X\rangle \cap \mathscr{H}')$ increases by at most $\mu(\langle X \cup \{v\}\rangle) \leq Kp^{|X|}/N$ after adding a vertex v to F. Once a subset X satisfies (21.5) no more hyperedges which contain X are added to $\mathscr{H}'$, thus at the end of the process we have

$$\mu(\langle X\rangle \cap \mathscr{H}') \leq \frac{2Kp^{|X|}}{N}, \tag{21.6}$$

for every $X \subseteq V$ of size $|X| \leq \ell-1$. Second, given $F \subseteq \hat{F} \subseteq I$, we can reconstruct F from $\hat{F}$ together with the order in which the vertices were added, thus we can also reconstruct $\mathscr{H}'$ and $\mathscr{R}$.

Let $\mathscr{H}''$ denote the $(\leq \ell-1)$-graph consisting of sets X such that $X = H' \setminus F$ for some $H' \in \mathscr{H}'$. Set μ' to be the probability measure over $2^{V\setminus F}$ given by

$$\mu'(X) \propto \begin{cases} \mu((X \cup 2^F) \cap \mathscr{H}'), & \text{if } X \in \mathscr{H}'', \\ 0, & \text{otherwise.} \end{cases}$$

From (21.6) and $\mu(\mathscr{H}') \geq \alpha p$ we conclude that μ' is $(2K\alpha^{-1}, p)$-uniformly-spread. Also observe that I is an independent set in $\mathscr{H}''$, (if $X'' \subseteq I$ then $X'' \cup F \subseteq I$ and so I contains an edge of $\mathscr{H}' \subseteq \mathscr{H}$) thus by the induction hypothesis there exists $F' \subseteq V$ of size $|F'| \leq (\ell-1)Np$ and $C = C(F')$ such that $|C| \leq (1-\delta)N$ and $I \subseteq C$. Note that we can reconstruct C from $F := F \cup F'$.

Case 2: $\mu(\mathscr{H}') < \alpha p$.

First we show that

$$\mu(\mathscr{D}) \leq \frac{2^\ell \mu(\mathscr{H}')}{p}. \tag{21.7}$$

Indeed, for each $e \in \mathscr{D}$ there exists $X \in \mathscr{L}$ such that $e \in \langle X\rangle$. Thus, $\sum_{X\in\mathscr{L}} \mu(\langle X\rangle) \geq \mu(\mathscr{D})$. On the other hand, we have by (21.5) that

$$\sum_{X\in\mathscr{L}} \mu(\langle X\rangle \cap \mathscr{H}') > \sum_{X\in\mathscr{L}} \frac{Kp^{|X|}}{N} \geq p \sum_{X\in\mathscr{L}} \mu(\langle X\rangle).$$

Here in the last inequality we use that μ is (p,K)-uniformly spread. Furthermore, each edge e in $\mathscr{H}'$ may contribute to at most 2^ℓ terms $\mu(\langle X\rangle \cap \mathscr{H}')$. Hence,

$$\mu(\mathscr{H}') \geq 2^{-\ell} \sum_{X\in\mathscr{L}} \mu(\langle X\rangle \cap \mathscr{H}') > 2^{-\ell} p\mu(\mathscr{D}),$$

as claimed in (21.7).

Next, we show that

$$\mu(\mathscr{H}') \geq (Np) \max_{v \in I \setminus F} \mu(\langle v \rangle \cap \mathscr{R}). \tag{21.8}$$

Let $\mathscr{R}_i$ denote the hypergraph $\mathscr{R}$ at the moment when the i-th vertex v_i was added to F (thus $\mathscr{R} = \mathscr{R}_{|F|}$). We observe that, since $\mathscr{R}$ is non-increasing and by our choice of v in each step,

$$\mu(\mathscr{H}') \geq \sum_{i=1}^{|F|} \mu(\langle v_i \rangle \cap \mathscr{R}_i) \geq \sum_{i=1}^{|F|} \max_{v \in I \setminus F} \mu(\langle v \rangle \cap \mathscr{R}_{|F|}),$$

yielding (21.8).

By (21.7), we have $\mu(\mathscr{D}) < 1/4$ and hence $\mu(\mathscr{R}) \geq \mu(\mathscr{H}) - \mu(\mathscr{H}') - \mu(\mathscr{D}) > 1/2$. By (21.8), for every $v \in I \setminus F$ we have

$$\mu(\langle v \rangle \cap \mathscr{R}) \leq \alpha/N. \tag{21.9}$$

Let now $B \subseteq V \setminus F$ denote the set of all $v \in V \setminus F$ such that $\mu(\langle v \rangle \cap \mathscr{R}) \leq \alpha/N$. By (21.9) we have $I \setminus F \subseteq B$. Furthermore,

$$\mu(\mathscr{R}) \leq \sum_{v \in B} \mu(\langle v \rangle \cap \mathscr{R}) + \sum_{w \in V \setminus (F \cup B)} \mu(\langle w \rangle \cap \mathscr{R}) < \alpha + (N - |B|) \cdot K/N.$$

Hence, $|B| < N - (\mu(\mathscr{R}) - \alpha)N/K < (1-\delta)N$ for $\delta = 1/(4K)$. This concludes the construction of desired F and $C = B \cup F$. $\square$

For the sake of completeness, we derive Theorem 21.8 from Lemma 21.9.

Proof of Theorem 21.8. Let $\delta > 0$ be as given by Lemma 21.9 for ℓ and K/ε (as K). We prove the theorem for $h = \ell \log(K\varepsilon^{-1})/\log(1+\delta)$.

We find a *fingerprint* F and a *container* C as follows. Set $F = \emptyset$ and $C = V$, and as long as $\mu(\mathscr{H}[C \setminus F]) \geq \varepsilon$ do the following: Let F' and C' be as given by Lemma 21.9 applied with μ' being a probability measure over $2^{C \setminus F}$ given by $\mu'(X) \propto \mu(X)$ if $X \in \mathscr{H}[C \setminus F]$, and $\mu'(X) = 0$ otherwise. Set $F := F \cup F'$ and $C := C'$, and proceed to the next iteration.

If $\mu(\mathscr{H}[C \setminus F]) \geq \varepsilon$, then for nonempty $X \subseteq C \setminus F$,

$$\mu'(\langle X \rangle) \leq \frac{\mu(\langle X \rangle)}{\mu(\mathscr{H}[C \setminus F])} \leq \frac{Kp^{|X|-1}/N}{\varepsilon} \leq \frac{Kp^{|X|-1}}{\varepsilon |C \setminus F|},$$

and hence μ' is $(p, K/\varepsilon)$-uniformly-spread each time we apply Lemma 21.9. Furthermore, if $\mu(\mathscr{H}[C]) \geq \varepsilon$, then $|C \setminus F| \geq \varepsilon N/K$. In each iteration the set $C \setminus F$

shrinks by a factor of $1-\delta$, thus we are done after at most $\log(K\varepsilon^{-1})/\log(1+\delta)$ iterations. The set F grows by at most ℓNp in each iteration, which gives an upper bound of hNp on its final size for the above choice of $h = h(K,\varepsilon)$. Due to the last property in Lemma 21.9, the final set C can be unambiguously constructed from F. $\square$

H-free graphs

In this section we prove the following generalisation of Theorem 21.1 to ℓ-graphs.

Theorem 21.10. *For every $\ell \in \mathbb{N}$ and $K,\varepsilon > 0$ there exists $h > 0$ such that the following holds: Let H be an ℓ-graph with $e(H) \geq 2$. Then for every H-free ℓ-graph I there exist ℓ-graphs F,C such that $C = C(F)$ and $F \subseteq I \subseteq C$ such that*

(i) $|F| \leq hn^{\ell-1/m_2(H)}$.

(ii) *For every ℓ-graph C, the number of copies of H in C is at most $\varepsilon n^{v(H)}$, and $e(C) \leq (\pi(H)+\varepsilon)\binom{n}{\ell}$.*

To prove this, we will apply Theorem 21.8 to the following hypergraph, whose independent sets correspond to H-free ℓ-graphs on vertex set $[N]$.

Definition 21.11. Let H be an ℓ-graph. Let $r = e(H)$. The r-graph $\mathbb{G}_H$ has vertex set $V = \binom{[n]}{\ell}$, where $B = \{v_1,\ldots,v_r\} \in \binom{V}{r}$ is an edge whenever B, considered as an ℓ-graph with vertices in $[n]$ and with r edges, is isomorphic to H. Note that $|\mathbb{G}_H| = \alpha_H n^{v(H)}$ where $\alpha_H = \Omega(1)$.

Lemma 21.12. *Let H be an ℓ-graph with $r = e(H) \geq 2$. Let n be sufficiently large. Then, if $p = n^{-1/m_2(H)}$ and $K = v(H)!/\ell!\alpha_H$,*

$$\frac{|\langle X\rangle \cap \mathbb{G}_H|}{|\mathbb{G}_H|} \leq \frac{Kp^{|X|-1}}{N},$$

where $N = |V|$ as usual.

Proof. Consider $X \subseteq [n]^{(\ell)}$ (so X is both a set of vertices of $\mathbb{G}_H$ and an ℓ-graph on vertex set $[n]$). This is the number of ways of extending X to an ℓ-graph isomorphic to H. If X as an ℓ-graph is not isomorphic to any subgraph of H, then clearly $|\langle X\rangle \cap \mathbb{G}_H| = 0$. Otherwise, let $v(X)$ be the number of vertices in X considered as an ℓ-graph, so there exists $W \subseteq V$, $|W| = v(X)$ with $X \subseteq W^{(\ell)}$. Edges of $\mathbb{G}_H$ containing X correspond to copies of H in V containing X, each such copy given by a choice of $v(H)-v(X)$ vertices in $V \setminus W$ and a permutation of the vertices of H. Hence for n sufficiently large,

$$|\langle X\rangle \cap \mathbb{G}_H| \leq v(H)!\binom{n-v(X)}{v(H)-v(X)} \leq v(H)!n^{v(H)-v(X)}$$

Now

$$\frac{N|\langle X\rangle \cap \mathbb{G}_H|}{p^{|X|-1}|\mathbb{G}_H|} \leq \frac{v(H)!n^{v(H)-v(X)}n^{\ell}}{\ell!\alpha_H n^{v(H)-(|X|-1)/m_2(H)}} = \frac{v(H)!n^{-v(X)+\ell+(|X|-1)/m_2(H)}}{\ell!\alpha_H} \leq K.$$

□

A well-known supersaturation theorem bounds the number of edges in containers.

Proposition 21.13 (Erdős and Simonovits [337]). *Let H be an ℓ-graph and let $\varepsilon > 0$. There exists n_0 and $\eta > 0$ such that if C is an ℓ-graph on $n \geq n_0$ vertices containing at most $\eta n^{v(H)}$ copies of H then $e(C) \leq (\pi(H)+\varepsilon)\binom{n}{\ell}$.*

Proof of Theorem 21.10. Let $\eta = \eta(\varepsilon, H)$ be given by Proposition 21.13, and let $\beta = \min\{\varepsilon, \eta\}$. Recall that $r = e(H)$. Apply Theorem 21.8 to $\mathbb{G}_H$ with β playing the role of ε. Lemma 21.12 implies that the uniform distribution over $\mathbb{G}_H$ is (p,K)-uniformly spread. (i) is immediate with h of (a) replaced by $h/\ell!$. (ii) follows from (b) and Proposition 21.13.

□

21.5 Exercises

21.5.1 An edge e of G is H-open if it is contained in at most one copy of H and H-closed otherwise. The H-core $\hat{G}_H$ of G is obtained by repeatedly deleting H-open edges. Show that $G \to (H)^e_2$ implies that $\hat{G}_{H'} \to (H')^e_2$ for every $H' \subseteq H$. (Thus one only needs to prove the 0-statement of Theorem 21.2 for strictly 2-balanced H. A graph H is strictly 2-balanced if $H' = H$ is the unique maximiser in (21.1)).

21.5.2 A subgraph G' of the H-core is H-closed if it contains at least one copy of H and every copy of H in $\hat{G}_H$ is contained in G' or is edge disjoint from G'. Show that the edges of $\hat{G}_H$ can be partitioned into inclusion minimal H-closed subgraphs.

21.5.3 Show that there exists a sufficiently small $c > 0$ and a constant $L = L(H,c)$ such that if H is 2-balanced and $p \leq cn^{-1/m_2(H)}$ then w.h.p. every inclusion minimal H-closed subgraph of $\mathbb{G}_{n,p}$ has size at most L. (Try $c = o(1)$ first here).

21.5.4 Show that if $e(G)/v(G) \leq m_2(H)$ and $m_2(H) > 1$ then $G \not\to (H)^e_2$.

21.5.5 Show that if H is 2-balanced and $p = cn^{-1/m_2(H)}$ then w.h.p. every subgraph G of $\mathbb{G}_{n,p}$ with $v(G) \leq L = O(1)$ satisfies $e(G)/v(G) \leq m_2(H)$.

21.5.6 Prove the 0-statement of Theorem 21.2 for $m_2(H) > 1$.

21.6 Notes

The largest triangle-free subgraph of a random graph

Babai, Simonovits and Spencer [58] proved that if $p \geq 1/2$ then w.h.p. the largest triangle-free subgraph of $\mathbb{G}_{n,p}$ is bipartite. They used Szemerédi's regularity lemma in the proof. Using the sparse version of this lemma, Brightwell, Panagiotou and Steger [195] improved the lower bound on p to n^{-c} for some (unspecified) positive constant c. DeMarco and Kahn [283] improved the lower bound to $p \geq Cn^{-1/2}(\log n)^{1/2}$, which is best possible up to the value of the constant C. And in [284] they extended their result to K_r-free graphs.

Anti-Ramsey Property

Let H be a fixed graph. A copy of H in an edge colored graph G is said to be rainbow colored if all of its edges have a different color. The study of rainbow copies of H was initiated by Erdős, Simonovits and Sós [336]. An edge-coloring of a graph G is said to be b-bounded if no color is used more than b times. A graph is G said to have property $\mathscr{A}(b,H)$ if there is a rainbow copy of H in every b-bounded coloring. Bohman, Frieze, Pikhurko and Smyth [141] studied the threshold for $\mathbb{G}_{n,p}$ to have property $\mathscr{A}(b,H)$. For graphs H containing at least one cycle they prove that there exists b_0 such that if $b \geq b_0$ then there exist $c_1, c_2 > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{A}(b,H)) = \begin{cases} 0 & p \leq c_1 n^{-1/m_2(H)} \\ 1 & p \geq c_2 n^{-1/m_2(H)} \end{cases}. \tag{21.10}$$

A reviewer of this paper pointed out a simple proof of the 1-statement. Given a b-bounded coloring of G, let the edges colored i be denoted $e_{i,1}, e_{i,2}, \ldots, e_{i,b_i}$ where $b_i \leq b$ for all i. Now consider the auxiliary coloring in which edge $e_{i,j}$ is colored with j. At most b colors are used and so in the auxiliary coloring there will be a monochromatic copy of H. The definition of the auxiliary coloring implies that this copy of H is rainbow in the original coloring. So the 1-statement follows directly from the results of Rödl and Ruciński [772], i.e. Theorem 21.2.

Nenadov, Person, Škorić and Steger [709] gave further threshold results on both Ramsey and Anti-Ramsey theory of random graphs. In particular they proved that in many cases $b_0 = 2$ in (21.10).

Chapter 22

Thresholds

In this chapter we describe remarkable results that will fairly easily give us good estimates for the thresholds for various structures.

A hypergraph $\mathscr{H}$ (thought of as a set of edges) is r-bounded if $e \in \mathscr{H}$ implies that $|e| \leq r$. For a set $S \subseteq X$ we let $\langle S \rangle = \{T : S \subseteq T \subseteq X\}$ denote the subsets of X that contain S. We say that a measure μ on the edges of $\mathscr{H}$ is κ-spread if

$$\mu(\{A \in \mathscr{H} : S \subseteq A\}) \leq \frac{1}{\kappa^{|S|}}, \quad \forall S \subseteq X.$$

In particular, we say that $\mathscr{H}$ is κ-spread if the uniform distribution is κ-spread i.e.

$$|\mathscr{H} \cap \langle S \rangle| \leq \frac{|\mathscr{H}|}{\kappa^{|S|}}, \quad \forall S \subseteq X. \tag{22.1}$$

Let X_m denote a random m-subset of X and X_p denote a subset of X where each $x \in X$ is included independently in X_p with probability p. The following theorem is from Frankston, Kahn, Narayanan and Park [376]:

Theorem 22.1. *Let $\mathscr{H}$ be an r-bounded, κ-spread hypergraph and let $X = V(\mathscr{H})$. There is an absolute constant $K > 0$ such that if*

$$p \geq \frac{(K \log r)}{\kappa} \text{ or equivalently } m \geq \frac{(K \log r)|X|}{\kappa} \tag{22.2}$$

then w.h.p. X_p contains an edge of $\mathscr{H}$ or equivalently X_m contains an edge of $\mathscr{H}$. Here w.h.p. assumes that $r \to \infty$.

This is an extremely powerful theorem as we will shortly see. But there is an even stronger theorem. Park and Pham [730] proved the so-called Kahn-Kalai conjecture [537] which implies Theorem 22.1. This requires some preparation, but first let us state some consequences of Theorem 22.1. In the first two examples, $X = \binom{[n]}{r}$ and $H = \mathbb{H}_{n,m;k}$ (not to be confused with $\mathscr{H}$).

Shamir's Problem: This is the name given to the problem of finding the threshold for the existence of a perfect matching in the random hypergraph $\mathbb{H}_{n,m;k}$. A perfect matching of an n-vertex, k-uniform hypergraph H, where $k|n$, is a set of n/k disjoint edges that cover all vertices of H. In this and the problem of loose hamilton cycles, we assume that k is a constant. If $m = cn\log n$ then the expected number of isolated vertices in $\mathbb{H}_{n,m;k}$ is $\approx n^{1-ck}$. If $ck < 1$ then there will be isolated vertices, w.h.p., see Exercise 12.4.1. Suppose now that $\mathscr{H}$ is the hypergraph with vertex set $X = \binom{[n]}{k}$ and an edge of size n/k corresponding to each perfect matching of the complete k-uniform hypergraph $\mathscr{H}_{n,k}$ on vertex set $[n]$., viz. each partition of $[n]$ into n/k sets of size k. Thus $\mathscr{H}$ is n/k-bounded. Now $\mathscr{H}$ has exactly $\frac{n!}{(n/k)!k!^{n/k}}$ edges. We see more generally, that if we choose a set $S = \{e_1, e_2, \dots, e_s\} \subseteq X$, i.e. s subsets of $[n]$ of size k then $E(\mathscr{H}) \cap \langle S\rangle = \emptyset$ unless $e_1, e_2, \dots, e_s$ are disjoint. If they are disjoint then, using Stirling's approximation, we find that for some constant $c_k \geq 1$,

$$\begin{aligned}
\frac{|E(\mathscr{H}) \cap \langle S\rangle|}{|E(\mathscr{H})|} &= \frac{(n-ks)!}{n!} \cdot \frac{(n/k)!k!^s}{(n/k-s)!} \\
&\leq c_k e^{(k-1)s} \frac{(n-ks)^{n-ks}\left(\frac{n}{k}\right)^{n/k}(k!)^s}{n^n\left(\frac{n}{k}-s\right)^{n/k-s}} \\
&= c_k e^{(k-1)s} \frac{(n-ks)^{n-ks-n/k+s} n^{n/k}(k!)^s}{n^n k^s} \\
&\leq c_k e^{(k-1)s} n^{-(k-1)s}(k-1)!^s.
\end{aligned}$$

Thus, $\mathscr{H}$ is $\kappa = \frac{n^{k-1}}{c_k(k-1)!e^{k-1}}$ spread. Applying Theorem 22.1 with this value of κ and $r = n/k$, we see that there exists $K > 0$ such if $m = Kn\log n$ then $\mathbb{H}_{n,m;k}$ has a perfect matching w.h.p. This was first proved in a breakthrough paper by Johansson, Kahn and Vu [529].

Loose Hamilton Cycles: We consider the case of 1-overlapping Hamilton cycles of Section 12.2. In this case $X = \binom{[n]}{k}$ and the edges of $\mathscr{H}$ correspond to the loose Hamilton cycles of the complete k-uniform hypergraph $\mathscr{H}_{n,k}$ on vertex set $[n]$. We see from Exercise 12.4.1 that we need at least $\Omega(n\log n)$ random edges to have a loose Hamilton cycle w.h.p. Now there are $n/(k-1)$ edges in a loose cycle and so we take $r = n/(k-1)$. The number of loose Hamilton cycles in $\mathscr{H}_{n,k}$ is given by $\frac{k-1}{2n} \cdot \frac{n!}{(k-2)!^{n/(k-1)}}$, Exercise 12.4.15, and for a set $S = \{e_1, e_2, \dots, e_s\} \subseteq X$ we have

$$\frac{|E(\mathscr{H}) \cap \langle S\rangle|}{|E(\mathscr{H})|} \leq \frac{(n-(k-1)s-1)!}{(k-2)!^{n/(k-1)-s}} \cdot \frac{2n(k-2)!^{n/(k-1)}}{(k-1)n!}$$

$$\leq \frac{2(k-2)!^s}{n^{(k-1)s}} \cdot \prod_{i=1}^{(k-1)s} \left(1-\frac{i}{n}\right)^{-1}$$

$$= \frac{2(k-2)!^s}{n^{(k-1)s}} \cdot \exp\left\{\sum_{i=1}^{(k-1)s} \frac{i}{n} + \frac{1}{2}\sum_{i=1}^{(k-1)s}\left(\frac{i}{n}\right)^2 + \cdots\right\} \leq \left(\frac{O(1)}{n^{k-1}}\right)^s. \quad (22.3)$$

Arguing as for the Shamir problem, we see that there exists $K > 0$ such if $m = Kn\log n$ then $\mathbb{H}_{n,m;k}$ has a loose hamilton cycle w.h.p. This being the result of [390], [305] and [307].

Powers of Hamilton cycles: The kth power of a Hamilton cycle in a graph $G = (V,E)$ is a permutation $x_1, x_2, \ldots, x_n$ of the vertices V such that $\{x_i, x_{i+j}\}$ is an edge of G for all $i \in [n], j \in [k]$. Kühn and Osthus [613] studied the existence of kth powers in $G_{n,p}$. They showed that for $k \geq 3$ one could use Riordan's Theorem [760] to show that if $np^k \to \infty$ then $G_{n,p}$ contains the kth power of a Hamilton cycle w.h.p. This is tight as the first moment method shows that if $np^k \to 0$ then w.h.p. there are no kth powers. The problem is more difficult for $k = 2$ and then after a series of papers, Nenadov and Škorić [712], Fischer, Škorić, Steger and Trujić [361], Montgomery [692] we have an upper bound of $p \gg \frac{\log^2 n}{n^{1/2}}$. Theorem 22.1 reduces the bound to $O(n^{3/2}\log n)$ in $G_{n,m}$, as we will now show.

We take $X = \binom{[n]}{2}$ and the edges of $\mathscr{H}$ correspond to the squares of Hamilton cycles of K_n. In which case we have for $|S| = s$, $r = n$ and

$$\frac{|E(\mathscr{H}) \cap \langle S\rangle|}{|E(\mathscr{H})|} \leq \frac{(n-2-\lfloor s/2\rfloor)!}{(n-1)!/2} \leq \left(\frac{e}{n-1}\right)^{\lfloor s/2\rfloor+1}. \quad (22.4)$$

We can therefore take $\kappa = e^{-1}n^{1/2}$, and then (22.2) yields the claimed upper bound of $O(n^{3/2}\log n)$ on the threshold for the existence of the square of a Hamilton cycle in $G_{n,m}$.

Bounded degree spanning trees: Let T_n be a sequence of spanning trees of K_n all of maximum degree $\Delta = O(1)$. We take $X = \binom{[n]}{2}$ and the edges of $\mathscr{H}$ correspond to the copies of T in K_n. We prove that (22.1) holds with $\kappa = n/\Delta$. If $S \subseteq X$ is not isomorphic to a subset of $E(T)$ then $E(\mathscr{H}) \cap \langle S\rangle = \emptyset$ and (22.1) holds. Suppose then that $S \subseteq X$ is isomorphic to a subset of $E(T)$. Then, where π is a random premutation of $[n]$ and

$$\pi(T) = ([n], \{\{\pi(v), \pi(w)\} : \{v,w\} \in E(T)\},$$

we have

$$\frac{|E(\mathscr{H}) \cap \langle S\rangle|}{|E(\mathscr{H})|} = \mathbb{P}(S \subseteq \pi(T)) \leq \kappa^{-|S|}. \quad (22.5)$$

We leave the verifcation of (22.5) as an exercise – Exercise 12.4.3. Consequently, we can apply Theorem 22.1 with $r=n-1$ that $O(n\log n)$ random edges are sufficient to contain a copy of T_n, w.h.p. Montgomery [690] gave the first proof of this result.

22.1 The Kahn-Kalai conjecture

Given a finite set X, write 2^X for the power set of X. For $p \in [0,1]$, let μ_p be the product measure on 2^X given by $\mu_p(A) = p^{|A|}(1-p)^{|X\setminus A|}$. Let $\langle \mathscr{H} \rangle = \bigcup_{S\in\mathscr{H}} \langle S \rangle$. The *threshold*, $p_c(\mathscr{H})$, is then the unique p for which

$$\mu_p(\langle \mathscr{H} \rangle) = \Pr(X_p \text{ contains an edge of } \mathscr{H}) = \frac{1}{2}.$$

We say $\mathscr{H}$ is *p-small* if there is $\mathscr{G} \subseteq 2^X$ such that

$$\mathscr{H} \subseteq \langle \mathscr{G} \rangle := \bigcup_{S\in\mathscr{G}} \{T : T \supseteq S\} \tag{22.6}$$

and

$$\sum_{S\in\mathscr{G}} p^{|S|} \leq 1/2. \tag{22.7}$$

Given that we are interested in the threshold for the existence of an edge of $\mathscr{H}$, we can assume that if $e,f \in \mathscr{H}$ then $e \not\subset f$. Note that we can always achieve this by replacing $\mathscr{H}$ by its minimal edges.

We say that $\mathscr{G}$ is a *cover* of $\mathscr{H}$ if (22.6) holds. The *expectation-threshold* of $\mathscr{H}$, $q(\mathscr{H})$, is defined to be the maximum p such that $\mathscr{H}$ is p-small. Observe that $q(\mathscr{H})$ is a trivial lower bound on $p_c(\mathscr{H})$, since

$$\mu_p(\mathscr{H}) \leq \mu_p(\langle \mathscr{G} \rangle) \leq \sum_{S\in\mathscr{G}} p^{|S|}. \tag{22.8}$$

Note that, with X_p the random variable whose distribution is μ_p, the right-hand side of (22.8) is $\mathbb{E}[|\{S \in \mathscr{G} : S \subseteq X_p\}|]$.

The following theorem resolves the expectation-threshold conjecture of Kahn and Kalai [537].

Theorem 22.2. *There is a universal constant K such that for every finite set X and hypergraph $\mathscr{H} \subseteq 2^X$,*

$$p_c(\mathscr{H}) \leq Kq(\mathscr{H}) \log r(\mathscr{H}).$$

The proof we give will be in terms of the following theorem: in the following $o_r(1)$ is a bound that tends to zero as $r \to \infty$.

Theorem 22.3. *There is a universal constant L such that for any r-bounded hypergraph $\mathscr{H}$ on X that is **not** p-small,*

$$\text{a uniformly random } ((Lp\log r)|X|)\text{-element subset of } X \text{ contains}$$
$$\text{an edge of } \mathscr{H} \text{ with probability } 1-o_r(1). \quad (22.9)$$

We leave it as an exercise to show that Theorem 22.3 implies Theorem 22.2. What we will show is:

Theorem 22.3 *implies Theorem* 22.1. The following argument was communicated to us by Tolson Bell [88]: it shows that if there exists a κ-spread measure μ on $\mathscr{H}$ then $\mathscr{H}$ is not κ^{-1}-small. Let $\mathscr{G}$ be such that $\mathscr{H} \subseteq \langle\mathscr{G}\rangle$. As μ is κ-spread we have that

$$\sum_{\substack{A\in\mathscr{H}\\ A\supseteq R}} \mu(A) \leq \frac{1}{\kappa^{|R|}} \text{ for all } R \in \mathscr{G}.$$

However, as every set in $\mathscr{H}$ contains some set in $\mathscr{G}$,

$$\sum_{R\in\mathscr{G}} \frac{1}{\kappa^{|R|}} \geq \sum_{R\in\mathscr{G}} \sum_{\substack{A\in\mathscr{H}\\ A\supseteq R}} \mu(A) \geq 1.$$

and so $\mathscr{G}$ is not κ^{-1}-small. □

Notations and Conventions. All logarithms are base 2 unless specified otherwise.

22.2 Proof of the Kahn-Kalai conjecture

We closely follow the argument of [730].

22.3 Constructing a cover

We use N for $|X|$ and let L be sufficiently large and let $\mathscr{H}$ be r-bounded. In the following S,S' etc. are edges of $\mathscr{H}$ and $W \in \binom{X}{w}$ is random w-subset of X where $w = LpN$.

Following [536], given S and W, we call a set of the form $S' \setminus W$ with S' contained in $S \cup W$ an (S,W)-*fragment*. Given S and W, define $T = T(S,W)$ to be a minimum size (S,W)-fragment. We use $t = t(S,W)$ for $|T(S,W)|$.

Given W, the *good set*, $\mathscr{G} = \mathscr{G}(W)$, is defined by

$$\mathscr{G}(W) := \{S \in \mathscr{H} : t(S,W) \geq .9r\}$$

$$= \left\{ S \in \mathscr{H} : |S' \setminus W| \geq .9r \text{ for all } S' \in \mathscr{H}, S' \subseteq S \cup W \right\}.$$

These are the edges $S \in \mathscr{H}$ such that $W \setminus S$ only contains a "small" piece of any edge of $\mathscr{H}$. W contains a significant portion of the remaining edges in $S \cup W$.

Given $\mathscr{G}$, we define $\mathscr{U}(W)$ as

$$\mathscr{U}(W) := \{T(S,W) : S \in \mathscr{G}(W)\}.$$

$\mathscr{U}(W)$ is a cover of $\mathscr{G}(W)$, since $T(S,W) \subseteq S$. We define

$$\mathscr{H}' = \mathscr{H}'(W) = \{T(S,W) : S \in \mathscr{H} \setminus \mathscr{G}(W)\}; \tag{22.10}$$

the hypergraph $\mathscr{H}'$, which is $.9r$-bounded, will be the host hypergraph in an iteration step (see (22.13)). Note that $\mathscr{H} \setminus \mathscr{G}(W) \subseteq \langle \mathscr{H}' \rangle$ so in particular, a cover of $\mathscr{H}'$ also covers $\mathscr{H} \setminus \mathscr{G}(W)$.

With these definitions, we can outline the strategy of the proof. Assume that $\mathscr{H}$ is not p-small. Starting with $W = W_1$ we produce $O(\log r)$ random disjoint sets $W_1, W_2, \ldots,$ of size $O(p|X|)$ and a sequence of hypergraphs $\mathscr{H}_i$ where $\mathscr{H}_i = \mathscr{H}'_{i-1}$, $i = 1, 2, \ldots$. At the same time we construct $\mathscr{G}_i$ and $\mathscr{U}_i = \mathscr{U}(W_i)$ and continue until either (i) $\mathscr{U} = \mathscr{U}_1 \cup \mathscr{U}_2 \cup \cdots$ covers $\mathscr{H}$ or (ii) $W_1 \cup W_2 \cup \cdots$ contains an edge of $\mathscr{H}$. We show that the former is unlikely by showing that w.h.p. $\mathscr{U}$ satisfies (22.7).

Note that if $S \in \mathscr{H}$ and $S_i = S \setminus \bigcup_{j=1}^{i} W_j \in \mathscr{H}_i$ then either S_i is put into $\mathscr{G}_{i+1}$ and gets covered or it shrinks by a factor 0.9.

22.4 Iteration

Recall that $N = |X|$, $r \to \infty$, and L is a large constant. Let $\gamma = \lfloor \log_{.9}(1/r) \rfloor + 1$. In the following definitions, $i = 1, 2, \ldots, \gamma$. Let $r_i = .9^i r$ and note that

$$0 < r_\gamma < 1. \tag{22.11}$$

Let $X_0 = X$ and W_i be uniform from $\binom{X_{i-1}}{w_i}$, where $X_i = X_{i-1} \setminus W_i$ and $w_i = L_i pn$ with

$$L_i = \begin{cases} L & \text{if} \quad i < \gamma - \sqrt{\log_{.9}(1/r)} \\ L\sqrt{\log r} & \text{if} \quad \gamma - \sqrt{\log_{.9}(1/r)} \leq i \leq \gamma. \end{cases}$$

At the end, $W := \bigcup_{i=1}^{\gamma} W_i$ is a uniformly random $(CLp\log r)N$-subset of X where $C \leq C'$ for some absolute constant $C' > 0$. Note that there is an absolute constant $c > 0$ for which

$$r_i > \exp(c\sqrt{\log r}), \quad \text{for all } i < \gamma - \sqrt{\log_{.9}(1/r)}. \tag{22.12}$$

We define a sequence $\{\mathscr{H}_i\}$ with $\mathscr{H}_0 = \mathscr{H}$ and

$$\mathscr{H}_i = \mathscr{H}'_{i-1} \tag{22.13}$$

(see (22.10) for the definition of $\mathscr{H}'$).

Note that each $\mathscr{H}_i$ is r_i-bounded, and associated to each set W_i in step i, we have a good set $\mathscr{G}_i = \mathscr{G}_i(W_i)$ and a cover $\mathscr{U}_i = \mathscr{U}_i(W_i)$ of $\mathscr{G}_i$.

We terminate the iteration process as soon as we have either

$$\exists S \in \mathscr{H} \text{ such that } S \subseteq \bigcup_{j=1}^{i} W_j \tag{22.14}$$

or

$$\mathscr{G}_i = \mathscr{H}_{i-1} \tag{22.15}$$

for some $i = i^*$. Note that $i^* \leq \gamma$ because of the upper bound on r_γ in (22.11) (so in particular, the iteration terminates with probability 1).

Proposition 22.4. *If $\mathscr{H}$ is not p-small then w.h.p., the process terminates in at most γ steps due to* (22.14)

Proof. $\mathscr{H}_i$ is r_i-bounded and $r_\gamma < 1$ and so the process cannot continue for more than γ steps. If the process stops because of (22.15) then $\mathscr{U}^* = \bigcup_{i \leq i^*} \mathscr{U}(W_i)$ covers $\mathscr{H}$, by construction. We show in (22.17) below that $\mathbb{E}\left(\sum_{S \in \mathscr{U}^*} p^{|S|}\right) = o(1)$. Thus, since $\mathscr{H}$ is not p-small,

$$\Pr(22.15) \leq \Pr\left(\sum_{S \in \mathscr{U}^*} p^{|S|} \geq \frac{1}{2}\right) = o(1),$$

from the Markov inequality. So if $\mathscr{H}$ is not p-small then the probability that the process stops due to (22.15) is $o(1)$.

Now consider $S \in \mathscr{H}$. Let $S_0 = S$ and $S_{i+1} = T(S_i, W_i)$ and note that $|S_{i+1}| \leq 0.9|S_i|$. Thus at some point either S_i gets put into $\mathscr{U}^*$, implying that S is covered by $\mathscr{U}^*$ or it becomes the emptyset, implying that (22.14) holds for this S. We have argued that w.h.p. the former cannot occur for every $S \in \mathscr{H}$ and so the latter must be true for some $S \in \mathscr{H}$. $\square$

Let $\mathscr{E}(W)$ be the event that $\mathscr{U}(W)$ covers $\mathscr{H}$. At this point we make the following claim about $\mathscr{U}(W)$ where $W = X_p$. Its proof is deferred to the end of the section.

Claim 9. *For W uniformly chosen from $\binom{X}{w}$, where $w = LpN$,*

$$\mathbb{E}\left(\sum_{U\in\mathcal{U}(W)} p^{|U|}\right) < L^{-.8r} \tag{22.16}$$

(where the expectation is over the choice of W).

Thus,

$$\begin{aligned}\mathbb{E}\left(\sum_{S\in\mathcal{U}^*} p^{|S|}\right) \overset{(22.16)}{<} \sum_{i\le\gamma} L_i^{-.8r_i} &= \sum_{i<\gamma-\sqrt{\log_{.9}(1/r)}} L_i^{-.8r_i} + \sum_{i=\gamma-\sqrt{\log_{.9}(1/r)}}^{\gamma} L_i^{-.8r_i} \\ &\overset{(22.11),(22.12)}{\le} 2L^{-.8\exp(c\sqrt{\log r})} + O((L\sqrt{\log r})^{-c'}) \\ &= (\log r)^{-c''}\end{aligned} \tag{22.17}$$

for some constant $c'' > 0$.

Proof of Claim 9 Observe that Claim 9 is equivalent to

$$\sum_{W\in\binom{X}{w}}\sum_{U\in\mathcal{U}(W)} p^{|U|} < \binom{N}{w} L^{-.8r}. \tag{22.18}$$

Proof of (22.18). Given W and $m \ge .9r$, let

$$\mathcal{G}_m(W) := \{S \in \mathcal{H} : t(S,W) = m\}$$

and

$$\mathcal{U}_m(W) := \{T(S,W) : S \in \mathcal{G}_m(W)\}.$$

Note that for any $U \in \mathcal{U}_m(W)$ we have $|U| = m$, so $\sum_{W\in\binom{X}{w}}\sum_{U\in\mathcal{U}_m(W)} p^{|U|}$ is equal to p^m multiplied by

$$\left|\left\{(W, T(S,W)) : W \in \binom{X}{w}, S \in \mathcal{H}, \text{ and } t(S,W) = m\right\}\right|. \tag{22.19}$$

We bound the number of choices of W and $T = T(S,W)$'s in the collection in (22.19) using the following specification steps.

Step 1. Pick $Z := W \cup T$. Since $|Z| = w + m$ (note W and T are always disjoint), the number of possibilities for Z is at most (recalling $w = LpN$)

$$\binom{N}{w+m} = \binom{N}{w}\cdot\prod_{j=0}^{m-1}\frac{N-w-j}{w+j+1} \le \binom{N}{w}(Lp)^{-m}.$$

Step 2. Pick an arbitrary $\hat{S} \in \mathscr{H}, \hat{S} \subseteq Z$, whose choice depends only on Z. Note that Z $(= W \cup T)$ must contain an edge of $\mathscr{H}$ by the definition of fragment. The choice of $\hat{S}$ is free given Z.

Here a crucial observation is that, since $T(S,W)$ is a minimum fragment,

$$T \subseteq \widehat{S}. \tag{22.20}$$

Indeed, since $\widehat{S}$ is contained in $T \cup W = S \cup W$, the failure of (22.20) implies that $\widehat{S} \setminus W$ is an (S,W)-fragment that is smaller than T, contradicting the minimality of T. Thus We see that $T \subseteq \widehat{S}$, and so the number of possibilities for T is at most 2^r.

Note that (W,T) is determined upon fixing a choice of Z and T. In sum, we have

$$\sum_{W \in \binom{X}{w}} \sum_{U \in \mathscr{U}_m(W)} p^{|U|} \leq p^m \binom{N}{w} (Lp)^{-m} 2^r = \binom{N}{w} L^{-m} 2^r,$$

and the left hand side of (22.18) is at most

$$\sum_{m \geq .9r} \binom{N}{w} L^{-m} 2^r \leq \binom{N}{w} L^{-.8r}$$

for L sufficiently large. □

22.5 Square of a Hamilton cycle and a little more

Kahn, Narayanan and Park [536] modified the proof of Theorem 22.1 and proved

Theorem 22.5. *If $m \geq Cn^{3/2}$ for sufficiently large C then w.h.p. $G_{n,m}$ contains a copy of H_2, the square of a Hamilton cycle.*

Spiro [814] introduced a refinement on the notion of spread. Let κ and $r_1 > \cdots > r_\ell$ be positive integers and let $r_{\ell+1} = 1$. We say that a hypergraph $\mathscr{H}$ is $(\kappa; r_1, \ldots, r_\ell)$-spread if $\mathscr{H}$ is non-empty and r_1-bounded and if for all $A \subseteq X = V(\mathscr{H})$ with $r_i \geq |A| \geq r_{i+1}$ for some $1 \leq i \leq \ell$ we have for all $j \geq r_{i+1}$ that,

$$M_t(A) = |\{S \in \mathscr{H} : |S \cap A| \geq t\}| \leq \kappa^{-j} |\mathscr{H}|. \tag{22.21}$$

Theorem 22.6. *Suppose that $\mathscr{H}$ is $(\kappa; r_1, \ldots, r_\ell)$-spread and r_1-uniform. Then, given $\varepsilon > 0$, there exists an absolute constant $C_\varepsilon > 0$ such that if $C \geq C_\varepsilon$ and W is a set of size $C\ell|X|/\kappa$ chosen uniformly randomly from X then*

$$\Pr(W \textit{ contains an edge of } \mathscr{H}) \geq 1 - \varepsilon$$

(This is a little weaker than what is proved in [814], but it suffices for the proof of Theorem 22.5.)

Proof. The proof described here is in large part due to the ideas of Jinyoung Park and Huy Pham.

Let $N=|X|$ and $m=\frac{CN}{\kappa}$. Let $p=\frac{2m}{N}$. Let $W_i, i=1,2,\ldots,\ell$ be independently distributed as X_m and let $W_{\leq i}=\bigcup_{j\leq i} W_j$ and $W=W_{\leq \ell}$. Also let $\eta_i(S)=S\setminus W_{\leq i}$ for $S\in\mathscr{H}$. We let $\mathscr{H}_1=\mathscr{H}$ and define $\mathscr{H}_i\subseteq\mathscr{H}, i\geq 2$ below. If $S\in\mathscr{H}_i$ then $T_i(S)$ minimises $|\eta_i(S')|$ over $S'\in\mathscr{H}_i$, $S'\subseteq S\cup W_i$.

We say that $(S,W_i), S\in\mathscr{H}_i$ is *i-bad*, $i=1,\ldots,\ell$, if $|T_i(S)|\geq r_{i+1}$. Otherwise (S,W_i) is *i-good*. Let $\mathscr{H}_{i+1}=\{S\in\mathscr{H}_i:\ (S,W_i) \text{ is } i\text{-good}\}$. Note that if (S,W_ℓ) is ℓ-good then $S\subseteq W_{\leq \ell}$.

We now estimate the number ν_{i-bad} of i-bad sets (S,W_i). We first specify $Z=W_i\cup T, T=T_i(S)$ in at most $\binom{N}{Np+t}, t\geq r_{i+1}$ ways. Then specify some arbitrary $S^*\in\mathscr{H}_i$, $\eta_i(S^*)\subseteq Z$. Note that for any $S\in\mathscr{H}_i$ such that $\eta_i(S)\cup W_i\subseteq Z$ we have $T\subseteq\eta_i(S)\cap\eta_i(S^*)$ and so $|\eta_i(S)\cap\eta_i(S^*)|\geq t$. When $|\eta_i(S)\cap\eta_i(S^*)|=t'\leq r_i$ we use (22.21) (with $A=\eta_i(S^*)$) and for $t'>r_i$ we use $2^{r_i}|\mathscr{H}_i|/\kappa^{t'}$ to bound the number of choices for S such that $\eta_i(S)$ contains T. Here 2^{r_i} bounds the number of choices for $\eta_i(S)\cap\eta_i(S^*)$ Finally, given S and S^* with $|\eta_i(S)\cap\eta_i(S^*)|=t'$, there are most $2^{t'}$ choices for $T\subseteq\eta_i(S)\cap\eta_i(S^*)$. Then, given Z and T, we also know $Z\setminus T$.

So,

$$\begin{aligned}
\nu_{i-bad} &\leq \sum_{t=r_{i+1}}^{r_i}\binom{N}{Np+t}\left(\sum_{t'=t}^{r_i}\frac{2^{t'}|\mathscr{H}_i|}{\kappa^{t'}}+\sum_{t'>r_i}\frac{2^{r_i+t'}|\mathscr{H}_i|}{\kappa^{t'}}|\right)\\
&\leq \binom{N}{Np}|\mathscr{H}_i|\sum_{t=r_{i+1}}^{r_i}\left(\frac{\kappa}{C}\right)^t\left(\sum_{t'=t}^{r_i}\frac{2^{t'}}{\kappa^{t'}}+\sum_{t'>r_i}\frac{2^{r_i+t'}}{\kappa^{t'}}\right)\\
&\leq 2\binom{N}{Np}\frac{2^{r_{i+1}}|\mathscr{H}_i|}{C^{r_{i+1}}}.
\end{aligned}$$

So, by the Markov inequality, we have, with $\alpha_i=2^{-r_{i+1}}$,

$$\Pr\left(|\mathscr{H}_{i+1}|\geq(1-\alpha_i)|\mathscr{H}_i|\right)\geq 1-\frac{2^{r_{i+1}+1}}{\alpha_i C^{r_{i+1}}},$$

and

$$\begin{aligned}
\Pr\left(|\mathscr{H}_\ell|\geq|\mathscr{H}|\prod_{i=1}^{\ell}(1-\alpha_i)\geq|\mathscr{H}|/2\right) &\geq \prod_{i=1}^{\ell}\left(1-\frac{2^{r_{i+1}+1}}{\alpha_i C^{r_{i+1}}}\right)\\
&\geq 1-\sum_{i=1}^{\ell}\frac{2^{r_{i+1}+1}}{\alpha_i C^{r_{i+1}}}\\
&\geq 1-\varepsilon,
\end{aligned}$$

if $C\geq C_\varepsilon=10\varepsilon^{-1}$. $\square$

We obtain Theorem 22.5 from the fact that $\mathscr{H}$ the hypergraph of squares of Hamilton cycles is $(Cn^{1/2};2n,Cn^{1/2},1)$-spread, for sufficiently large C, see Exercise 22.7.4.

22.6 Embedding a factor

In this section we discuss factors in random subgraphs of graphs with high minimum degree. Let F be a fixed graph. Let G be a graph on vertex set $[n]$ where $|V(F)| \mid n$. Let $\delta(G)$ denote the minimum degree in G. Let $\delta_F > 0$ be the minimum value such that if $\delta(G) \geq \delta_F n$ then G contains an F-factor i.e. $n/|V(F)|$ disjoint copies of F. We prove the following theorem which is a weakening of an example of one of the results from Kelly, Müyesser and Pokrovskiy [562]. For a graph G we let $d_1(G) = |E(G)|/(|V(G)|-1)$ and let $m_1(G) = \max_{G' \subseteq G: |V(G')|>1} d_1(G')$.

G_p is the random subgraph of G when each edge of G is independently included with probability p.

Theorem 22.7. *Let F be a fixed graph. Let $\alpha > 0$ be arbitrary and let G be a graph with vertex set $[n]$ and $\delta(G) \geq (\delta_F + \alpha)n$ where $|V(F)| \mid n$. Then w.h.p. G_p contains an F-factor for $p \geq Cn^{-1/m_1} \log^{(1+2/m_1)} n$, $m_1 = m_1(F)$.*

The proof of this requires a variation on spread and a lemma linking it to the original definition. We first define *vertex spread*.

Definition 22.8. Let X and Y be finite sets, and let μ be a probability distribution over injections $\varphi : X \to Y$. For $\kappa \geq 1$ we say that μ is κ-vertex-spread if for avery two sequences $x_1, x_2, \ldots, x_s \in X$ and $y_1, y_2, \ldots, y_s \in Y$,

$$\mu(\{\varphi : \varphi(x_i) = y_i, i = 1,2,\ldots,s\}) \leq \frac{1}{\kappa^s}.$$

A *graph embedding* $\varphi : G \hookrightarrow H$ of a graph G into graph H is an injective map $\varphi : V(G) \to V(H)$ that maps edges to edges. So there is an embedding of G into H if and only if H contains a subgraph isomorphic to G. The following lemma links the two notions of spread.

Lemma 22.9. *Fix $C, \Delta \geq 1$ and let $C_1 = (C^2\Delta)^{1/m_1(G)}$. Assume that n is sufficiently large n. Let G_1, G_2 be graphs on vertex set $[n]$. If there is an (n/C)-vertex-spread distribution on embeddings $G_1 \hookrightarrow G_2$ and $\Delta(G_1) \leq \Delta$, then there is an $\left(n^{1/m_1(G_1)}/C_1\right)$-spread distribution on copies of G_1 in G_2.*

Proof of Lemma 22.9*:* First we need an upper bound on the number of *partial embeddings* of G_1 into G_2. We prove that if $K \subseteq G_2$ has v vertices and c

components, then there are at most

$$n^c(e\Delta(G_1))^{v-c} \tag{22.22}$$

embeddings $\varphi : G_1[X] \hookrightarrow G_2[V(K)]$ where $X \in \binom{V(G_1)}{v}$ and $K \subseteq G_2[\varphi(X)]$.

Proof of (22.22)*:* Let $K_1,\dots,K_c$ be the components of K, and for each $i \in [c]$, let $V(K_i) = \{v_{i,j}, j = 1,2,\dots,|V(K_i)|\}$, where for every $j > 1$, there exists $j' < j$ such that $v_{i,j}$ and $v_{i,j'}$ are contained in a common edge of K_i. Each embedding $\varphi : G_1[X] \hookrightarrow G_2[V(K)]$ with $K \subseteq G_2[\varphi(X)]$ is then determined by

(i) the preimages $\varphi^{-1}(v_{1,1}),\dots,\varphi^{-1}(v_{c,1})$ of the "roots" and

(ii) for each $i \in [c]$, a sequence in $\Delta(G_1)^{|V(K_i)|-1}$, where the jth term in the sequence determines $\varphi^{-1}(v_{i,j+1})$ based on $\varphi^{-1}(v_{i,j'})$, where $j' \le j$ and $v_{i,j'}$ and $v_{i,j+1}$ are contained in a common edge of K_i.

Note that there are at most n^c choices for the preimages of the roots and at most $\Delta(G_1)^{v-c}$ choices for the sequences. Combining these choices yields the desired bound.

End of proof of (22.22).

We continue with the proof of Lemma 22.9. Let G_1, G_2 be n-vertex graphs, where $\Delta(G_1) \le \Delta$, and suppose there is an (n/C)-vertex-spread distribution μ on embeddings $G_1 \hookrightarrow G_2$. For every $F \subseteq G_2$ isomorphic to G_1, let

$$\mu'(F) = \mu\left(\{\varphi : \varphi(E(G_1)) = E(F)\}\right),$$

and note that μ' is a probability distribution on subgraphs of G_2 which are isomorphic to G_1. We prove that μ' is $\left(n^{1/m_1(G_1)}/C_1\right)$-spread. To that end, let $S \subseteq E(G_2)$, and let $T \subseteq G_2$ have edge set S and subject to that, the fewest number of vertices. We may assume T is isomorphic to a subgraph of G_1, or else $\mu'(\{F \subseteq G_2 : E(F) \supseteq S\}) = 0$. We may also assume $S \neq \emptyset$. Let v and c be the number of vertices and components of T, respectively. Note that $v \ge 2c$ since we can assume that T has no isolated vertices. By Lemma 22.9, the number of embeddings $\varphi : G_1[X] \hookrightarrow G_2[V(T)]$ where $X \in \binom{V(G_1)}{v}$ and $T \subseteq G_2[\varphi(X)]$ is at most $n^c\Delta^{v-c}$, so since μ is (C/n)-vertex-spread,

$$\begin{aligned}\mu'(\{F \subseteq G_2 : E(F) \supseteq S\}) &\le n^c\Delta^{v-c}\left(\frac{C}{n}\right)^v \le n^{v-c}(\Delta C^2)^{v-c}\\ &= \left(\frac{C_1^{m_1(G_1)}}{n}\right)^{v-c} = \left(\frac{C_1^{m_1(G_1)}}{n}\right)^{|S|(v-c)/|S|}.\end{aligned}$$

Let $T_1,\dots,T_c$ be the components of T. Since each T_i is isomorphic to a subgraph of G_1, for every $i \in [c]$ we have $|E(T_i)|/(|V(T_i)|-1) \le m_1(G_1)$. In particular,

$$|S| = \sum_{i=1}^{c} |E(T_i)| \le m_1(G_1) \sum_{i=1}^{c} (|V(T_i)|-1) = m_1(G_1)(v-c),$$

so $(v-c)/|S| \ge 1/m_1(G_1)$. Therefore

$$\mu'(\{F \subseteq G_2 : E(F) \supseteq S\}) \le \left(\frac{C_1^{m_1(G_1)}}{n}\right)^{|S|(v-c)/|S|} \le \left(\frac{C_1}{n^{1/m_1(G_1)}}\right)^{|S|},$$

as desired.
End of proof of Lemma 22.9.
Proof of Theorem 22.7 We randomly partition $V(G)$ into random *clusters* $U_1,\dots,U_m$ where $|V(F)| \mid |U_i|$ and $|U_i| \sim C\log n$ for $i = 1,2,\dots,m$. The degree of $v \in U_i$ in $G[U_i]$ dominates $Bin((\delta_F+\alpha)n, (|U_i|-1)/(n-1))$ and so if C is sufficiently large, then w.h.p. $\delta(U_i) \ge (\delta_F+\alpha/2)|U_i|$ for $i = 1,2,\dots,m$. So, w.h.p. each U_i contains an F-factor. A simple calculation shows that for every set of distinct vertices $x_1,\dots,x_s \in V(H)$ and every function $f : [s] \to [m]$,

$$\Pr\left(x_i \in U_{f(i)} \text{ for each } i \in [s]\right) \le \left(\frac{C\log n}{n}\right)^s. \tag{22.23}$$

Theorem 22.7 from Lemma 22.9 (with C replaced by $C\log n$) and Theorem 22.1. □

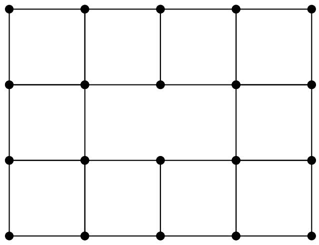
In the paper [562], the size of the U_i's are $O(1)$ and this makes the analysis much more difficult. The paper also deals with hypergraphs and Hamilton cycles.

22.7 Exercises

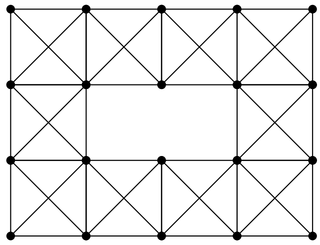
22.7.1 Let $U_1,U_2,\dots,U_k$ denote k disjoint sets of size n. Let $\mathscr{HP}_{n,m,k}$ denote the set of k-partite, k-uniform hypergraphs with vertex set $V = U_1 \cup U_2 \cup \cdots \cup U_k$ and m edges. Here each edge contains exactly one vertex from each $U_i, 1 \le i \le k$. The random hypergraph $HP_{n,m,k}$ is sampled uniformly from $\mathscr{HP}_{n,m,k}$. Prove the k-partite analogue of Shamir's problem viz. there exists a constant $K > 0$ such that if $m \ge Kn\log n$ then

$$\lim_{n\to\infty} \mathbb{P}(HP_{n,m,k} \text{ has a 1-factor}) = 1.$$

22.7.2 Find the threshold, up to a $\log n$ factor for the existence of the following structures in $G_{n,p}$: replace 4 by n in an appropriate way.



C_4-cycle, overlap 2



K_4-cycle, overlap 2

22.7.3 Verify (22.5).

22.7.4 Verify that $\mathscr{H}$, the hypergraph of squares of a Hamilton cycle is $(Cn^{1/2}; 2n, Cn^{1/2}, 1)$-spread for sufficiently large C.
(Hints: For $I \subseteq \binom{X}{2}$ let $\kappa(I)$ denote the number of components in the graph induced by I and the vertices incident with I.

(a) $|I| = t \leq n/3$ implies that $|\mathscr{H} \cap \langle I \rangle| \leq (16)^t \left(n - \left\lceil \frac{t+c}{2} \right\rceil - 1\right)!$.

(b) If $F \subseteq H_2$ and $|F| = h$ then the number of subgraphs of F with t edges and c components is at most $(8e)^t \binom{2h}{c}$.)

22.7.5 Show that if $\mathscr{H}$ is $(\kappa;\ r_1, \ldots, r_\ell, 1)$-spread, then it is κ-spread.

22.7.6 Show that if $\mathscr{H}$ is κ-spread and r_1-bounded, then it is $(\kappa/4;\ r_1, \ldots, r_\ell, 1)$-spread for any sequence r_i, satisfying $r_i > r_{i+1} \geq r_i/2$.

22.7.7 Produce a version of Theorem 22.7 for hypergraphs.

22.8 Notes

Perfect matchings in regular hypergraphs

The perfect matching problem turns out to be a much easier problem than that discussed in Section 22. Cooper, Frieze, Molloy and Reed [269] used small subgraph conditioning to prove that $H_{n,r;k}$ has a perfect matching w.h.p. iff $k > k_r$ where $k_r = \frac{\log r}{(r-1)\log\left(\binom{r}{r-1}\right)} + 1$.

Embedding into Dirac Graphs

The notion of vertex spread was introduced in Pham, Sah, Sawhney and Simkin [736]. Further results related to Section 22.6 can be found in Kang, Kelly, Kühn, Osthus and Pfenninger [542], Anastos, Chakraborti, Kang, Methuku and Pfenninger [43].

Chapter 23

Contiguity

Formally, let $\mathscr{G}_n$ be a family of graphs on vertex set $[n]$ and let $\mathscr{F}_n$ be the σ-field of subsets of $\mathscr{G}_n$. Denote $\mathbb{P}_n$ and $\mathbb{Q}_n$ as two different sequences of probability measures ruling the selection of a graph from the family $\mathscr{G}_n$. Then we say that $\mathbb{P}_n$ and $\mathbb{Q}_n$ are *contiguous* if for any sequence of events $A_n \in \mathscr{F}_n$,

$$\lim_{n\to\infty} \mathbb{P}_n(A_n \in \mathscr{F}_n) = 0 \Leftrightarrow \lim_{n\to\infty} \mathbb{Q}_n(A_n \in \mathscr{F}_n) = 0. \tag{23.1}$$

It means that we only demand that if a certain property holds w.h.p. for one of the probability measures (class of random graphs) it also holds w.h.p. for the other. Note that contiguity is a weaker property than *asymptotic equivalence* for which the following condition has to be satisfied:

$$\lim_{n\to\infty} (\mathbb{P}_n(A_n \in \mathscr{F}_n) - \mathbb{Q}_n(A_n \in \mathscr{F}_n)) = 0. \tag{23.2}$$

The following two propositions, formulated and proved by Janson in [499], characterize properties of the contiguity relation.

Proposition 23.1. *Suppose that $W_n = d\mathbb{Q}_n/d\mathbb{P}_n$, regarded as a random variable on $(\mathscr{G}_n, \mathscr{F}_n, \mathbb{P}_n)$, converges to a random variable W as $n \to \infty$. Then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous if and only if $W > 0$ a.s. and $\mathbb{E}W = 1$.*

$\square$

Proposition 23.2. *Suppose that $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous on $\mathscr{G}_n$. Then*

(i) *If A_n is any sequence of events such that $\liminf \mathbb{P}_n(A_n) > 0$ then the conditioned measures $\mathbb{P}_n(\cdot|A_n)$ and $\mathbb{Q}_n(\cdot|A_n)$ are contiguous.*

(ii) *If $f_n : \mathscr{G}_n \to \mathscr{G}_n'$ are measurable functions, then the induced measures $\mathbb{P}_n \circ f_n^{-1}$ and $\mathbb{Q}_n \circ f_n^{-1}$ on $\mathscr{G}_n'$ are contiguous.*

(iii) *If $\mathbb{P}_n'$ and $\mathbb{Q}_n'$ are contiguous probability measures on $\mathscr{G}_n'$, then the product measures $\mathbb{P}_n \times \mathbb{P}_n'$ and $\mathbb{Q}_n \times \mathbb{Q}_n'$ on $\mathscr{G}_n \times \mathscr{G}_n'$.*

□

23.1 Small subgraph conditioning for proving contiguity

Janson [499] (see also [686]) proved the following result, explaining and extending a general framework of the *method of small subgraph conditioning*, developed by Robinson and Wormald (see [767]).

Theorem 23.3. *Let $\lambda_i > 0$ and $\delta_i \geq -1,\ i = 1,2,\dots,$ be real numbers and suppose that for each n there are random variables $X_{i,n},\ i = 1,2,\dots,$ and Y_n defined on the same probability space $(\mathscr{G}_n, \mathbb{P}_n)$ such that the $X_{i,n}$ are non-negative integer valued, Y_n is non-negative and $\mathbb{E}\,Y_n > 0$.*

Suppose also that

(A1) for each $k \geq 1,\ X_{i,n} \xrightarrow{D} Z_i$ jointly for $i = 1,2,\dots,k$, where $Z_i \sim \mathrm{Po}(\lambda_i)$ are independent Poisson random variables, (i.e., $\mathbb{E}\,X_{i,n} \to \lambda_i$ as $n \to \infty$);

(A2)

$$\frac{\mathbb{E}(Y_n | X_{1,n} = j_1 \dots, X_{k,n} = j_k)}{\mathbb{E}\,Y_n} \to \prod_{i=1}^{k} ((1+\delta_i))^{j_i} e^{-\lambda_i \delta_i}, \ \ as \ \ n \to \infty,$$

for every finite sequence $j_1, j_2, \dots, j_k$ of non-negative integers;

(A3) $\sum_i \lambda_i \delta_i^2 < \infty$;

(A4) $\frac{\mathbb{E}\,Y_n^2}{(\mathbb{E}\,Y_n)^2} \to \exp\left(\sum_i \lambda_i \delta_i^2\right)$ as $n \to \infty$.

Then

$$\frac{Y_n}{\mathbb{E}\,Y_n} \xrightarrow{D} W = \prod_{i=1}^{\infty} (1+\delta_i)^{Z_i} e^{-\lambda_i \delta_i} \quad as \quad n \to \infty, \tag{23.3}$$

and the infinite product defining W converges a.s., with

$$\mathbb{E}\,W = 1 \ \ and \ \ \mathbb{E}\,W^2 = \exp\left(\sum_{i=1}^{\infty} \lambda_i \delta_i^2\right). \tag{23.4}$$

Furthermore, the event $W > 0$ equals, up to a set of probability zero, the event that $Z_i > 0$ for some i with $\delta_i = -1$. In particular, $W > 0$ a.s., if and only if every $\delta_i > -1$.

Proof. We start with proving (23.4). Observe first that, since $Z_i \sim \mathrm{Po}(\lambda_i)$,

$$\mathbb{E}(1+\delta_i)^{Z_i} = e^{\lambda_i(1+\delta_i)-\lambda_i} = e^{\lambda_i\delta_i}$$

and

$$\mathbb{E}\left((1+\delta_i)^{Z_i}\right)^2 = e^{\lambda_i(1+\delta_i)^2-\lambda_i} = e^{\lambda_i(2\delta_i+\delta_i^2)}.$$

Now, to prove that (23.4) holds, let

$$W^{(k)} = \prod_{i=1}^{k}(1+\delta_i)^{Z_i}e^{-\lambda_i\delta_i}. \tag{23.5}$$

Notice, that $W^{(k)}$ is a product of independent random variables with the expectation equal to 1, and with second moment

$$\mathbb{E}\left(W^{(k)}\right)^2 = \prod_{i=1}^{k} e^{\lambda_i\delta_i^2} = \exp\left(\sum_{i=1}^{k}\lambda_i\delta_i^2\right) \to \exp\left(\sum_{i=1}^{\infty}\lambda_i\delta_i^2\right) \quad \text{as } k\to\infty. \tag{23.6}$$

Hence the sequence $\left(W^{(k)}\right)_{k=1}^{\infty}$ is an L^2-bounded martingale and so $W = \lim_{k\to\infty} W^{(k)}$ exists a.s. and in L^2, and so (23.4) holds.

To prove the statement (23.3), i.e., that $Y_n/\mathbb{E}Y_n \to W$, let us introduce some simplifying assumptions, following Janson's reasoning (see [499]).
First, assume that $\mathbb{E}Y_n = 1$ and, second, that $X_{i,n} \to Z_i$ a.s. as $n\to\infty$ for each i. Next for fixed large integer k define the functions

$$f_n(j_1,\dots,j_k) = \mathbb{E}(Y_n|X_{1,n} = j_1\dots,X_{k,n} = j_k)$$
$$f_\infty(j_1,\dots,j_k) = \lim_{n\to\infty} f_n(j_1,\dots,j_k) = \prod_{i=1}^{k}((1+\delta_i))^{j_i}e^{-\lambda_i\delta_i},$$

(see assumption (A2)), and the random variable

$$Y_n^{(k)} = \mathbb{E}(Y_n|X_{1,n}\dots,X_{k,n}) = f_n(X_1,\dots,X_k).$$

Then

$$\mathbb{E}\left(Y_n^{(k)}\right)^2 = \sum_{j_1,\dots,j_k}\mathbb{P}(X_{1,n} = j_1\dots,X_{k,n} = j_k)f_n(j_1,\dots,j_k)^2$$

and thus , by Fatou's lemma and assumptions (A1) and (A2),

$$\begin{aligned}\liminf\mathbb{E}\left(Y_n^{(k)}\right)^2 &\geq \sum_{j_1,\dots,j_k}\lim_{n\to\infty}\mathbb{P}(X_{1,n} = j_1\dots,X_{k,n} = j_k)f_n(j_1,\dots,j_k)^2\\ &= \sum_{j_1,\dots,j_k}\mathbb{P}(Z_1 = j_1\dots,Z_k = j_k)f_\infty(j_1,\dots,j_k)^2\end{aligned}$$

$$= \mathbb{E} f_\infty(Z_1,\dots,Z_k)^2 = \mathbb{E}\left(W^{(k)}\right)^2 = \exp\left(\sum_{i=1}^k \lambda_i \delta_i^2\right).$$

On the other hand, since $Y_n^{(k)}$ is a conditional expectation of Y_n,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{E}(Y_n - Y_n^{(k)})^2 &= \limsup_{n\to\infty}(\mathbb{E} Y_n^2 - \mathbb{E}(Y_n^{(k)})^2 - 2\,\mathbb{E}(Y_n^{(k)}(Y_n - Y_n^{(k)})))\\ &= \limsup_{n\to\infty}(\mathbb{E} Y_n^2 - \mathbb{E}(Y_n^{(k)})^2)\\ &\le \exp\left(\sum_{i=1}^\infty \lambda_i \delta_i^2\right) - \exp\left(\sum_{i=1}^k \lambda_i \delta_i^2\right). \qquad (23.7)\end{aligned}$$

Furthermore, by the simplifying assumption that $X_{i,n} \to Z_i$ a.s. as $n \to \infty$ for each i, $X_{i,n} = Z_i,\ i \le k$, for large n, and thus

$$\lim_{n\to\infty} Y_n^{(k)} = \lim_{n\to\infty} f_n(Z_1,\dots,Z_k) = f_\infty(Z_1,\dots,Z_k) = W^{(k)} \ \text{ a.s.} \qquad (23.8)$$

Finally, applying Chebyshev's inequality (see Lemma 26.3) and (23.7) we get

$$\begin{aligned}&\limsup_{n\to\infty} \mathbb{P}(|Y_n - W| > 3\varepsilon) \le \limsup_{n\to\infty} \mathbb{P}(|Y_n - Y_n^{(k)}| > \varepsilon) +\\ &+ \limsup_{n\to\infty} \mathbb{P}(|Y_n^{(k)} - W^{(k)}| > \varepsilon) + \limsup_{n\to\infty} \mathbb{P}(|W^{(k)} - W| > \varepsilon)\\ &\le \varepsilon^{-2} \limsup_{n\to\infty} \mathbb{E}\,|Y_n - Y_n^{(k)}|^2 + 0 = \varepsilon^{-2}\,\mathbb{E}\,|W - W^{(k)}|^2.\end{aligned}$$

However notice, that

$$\varepsilon^{-2}\,\mathbb{E}\,|W - W^{(k)}|^2 = \mathbb{E} W^2 - \mathbb{E}(W^{(k)})^2 = \exp\left(\sum_{i=1}^\infty \lambda_i \delta_i^2\right) - \exp\left(\sum_{i=1}^k \lambda_i \delta_i^2\right),$$

and hence

$$\limsup_{n\to\infty} \mathbb{P}(|Y_n - W| > 3\varepsilon) \le 2\varepsilon^{-2}\left[\exp\left(\sum_{i=1}^\infty \lambda_i \delta_i^2\right) - \exp\left(\sum_{i=1}^k \lambda_i \delta_i^2\right)\right]. \qquad (23.9)$$

Now let $k \to \infty$ and keep ε fixed. Then the right hand side of the bound (23.9) tends to 0, so the left hand side,which does not depend on k, has to vanish for each $\varepsilon > 0$, which proves that Y_n tends to W in probability, which implies that (23.3) holds.

To prove the remaining part of the theorem, i.e., to show that $W > 0$ a.s., except when $Z_i > 0$ for some i with $\delta_i = -1$, let us break the product defining W in (23.3) into two parts:

$$W_1 = \prod_{I_1 = \{i:\delta_i < -1/2\}} (1+\delta_i)^{Z_i} e^{-\lambda_i \delta_i} \quad \text{and} \quad W_2 = \prod_{I_2 = \{i:\delta_i \ge -1/2\}} (1+\delta_i)^{Z_i} e^{-\lambda_i \delta_i}.$$

For W_1 observe that condition (A3) implies that

$$\mathbb{E}\sum_{I_1} Z_i = \sum_{I_1} \lambda_i \leq 4\sum_{I_1} \lambda_i \delta_i^2 < \infty \quad \text{and} \quad \sum_{I_1} |\lambda_i \delta_i| < \infty.$$

Hence, there are a.s. only finitely many non-zero $Z_i,\ i \in I_1$ and

$$W_1 = e^{-\sum_{I_1} \lambda_1 \delta_i} \prod_{I_1} (1+\delta_i)^{Z_i},$$

where $\prod_{I_1}(1+\delta_i)^{Z_i}$ is really a finite product which is positive unless some factor vanishes, i.e., unless $Z_i > 0$ for some i with $\delta_i = -1$.

For W_2 we define $\overline{\delta}_i = -\delta_i(1+\delta_i),\ \ i \in I_2$ and note that $\sum_{I_2} \lambda_i \overline{\delta}_i^2 < \infty$. So, the argument above shows that

$$\overline{W}_2 = \prod_{I_2=\{i:\delta_i \geq -1/2\}} (1+\overline{\delta}_i)^{Z_i} e^{-\lambda_i \overline{\delta}_i)}$$

converges a.s. with $\overline{W}_2 < \infty$. However since $(1+\delta_i)(1+\overline{\delta}_i) = 1$,

$$W_2\overline{W}_2 = \prod_{I_2} e^{-\lambda_i(\delta_i+\overline{\delta}_i)} = \exp\left(-\sum_{I_2} \lambda_i \delta_i^2/(1+\delta_i)\right) > 0,$$

so $W_2 > 0$ a.s. Therefore $W > 0$ a.s. except when some factor vanishes, which proves the second statement of the Theorem and completes the proof of the Theorem.

□

Consider the probability space $(\mathscr{G}_n, \mathbb{P}_n)$, where $\mathbb{P}_n$ is the uniform probability measure over $\mathscr{G}_n$. Let Y_n be a non-negative random variable defined on $\mathscr{G}_n$, with $\mathbb{E}\,Y_n > 0$. Let $\mathbb{Q}$ be a probability measure on $\mathscr{G}_n$ where for any $G \in \mathscr{G}_n$, let

$$\mathbb{Q}_n(G) = \frac{Y_n(G)}{\mathbb{E}(Y_n)|\mathscr{G}_n|}. \tag{23.10}$$

Theorem 23.4. *(see Molloy et. al. [686])*
Under the assumptions of Theorem 23.3*, probability measures* $\mathbb{P}_n$ *and* $\mathbb{Q}_n$*, defined above, conditioned on the event* $\bigcap_{\delta_i=-1}(X_i = 0)$*, are contiguous, provided* $\sum_{\delta_i=-1} \lambda_i < \infty$*.*

Proof. To see that measures $\mathbb{P}_n$ amd $\mathbb{Q}_n$ are contiguous recall that since $\mathbb{P}_n$ is a uniform probability measure over the set $\mathscr{G}_n$ of all graphs on vertex set $[n]$, so each element $G \in \mathscr{G}_n$ has weight $1/|\mathscr{G}_n|$, while in the probability space $(\mathscr{G}_n, \mathbb{Q}_n)$ each element G occurs with probability $Y_n(G)/\left(\mathbb{E}(Y_n)|\mathscr{G}_n|\right)$.

Suppose that an event A_n from $(\mathscr{G}_n, \mathbb{P}_n)$ occurs with probability $\mathbb{P}(A_n) = p(n) = o(1)$ Then, from the assumptions $(A3)$ and $(A4)$ (see Theorem 23.3), we get

$$\operatorname{Var} Y_n = \exp\left(\left(\sum_{i=1}^{\infty} \lambda_i \delta_i^2\right) - 1\right) (\mathbb{E}\, Y_n)^2 = O((\mathbb{E}\, Y_n)^2).$$

Therefore, by Chebyshev's bound, for each positive integer k the number of $G \in \mathscr{G}_n$ for which $Y_n(G) > k\mathbb{E}\, Y_n$ is at most $\frac{\operatorname{Var} Y_n}{k^2} |\mathscr{G}_n|$. Thus, taking

$$K = \left\{ k : \frac{\operatorname{Var} Y_n}{(k+1)^2} < p(n) \right\} = \left\{ k : k \geq \left(\frac{\operatorname{Var} Y_n}{p(n)} \right)^{1/2} \right\}$$

we get

$$\begin{aligned}
\mathbb{Q}(A_n) = & \sum_{\substack{G \in A_n \\ \frac{Y_n(G)}{\mathbb{E} Y_n} \leq \left(\frac{\operatorname{Var} Y_n}{p(n)}\right)^{1/2}}} \frac{Y_n(G)}{\mathbb{E}\, Y_n |\mathscr{G}_n|} + \sum_{\substack{G \in A_n \\ \frac{Y_n(G)}{\mathbb{E} Y_n} > \left(\frac{\operatorname{Var} Y_n}{p(n)}\right)^{1/2}}} \frac{Y_n(G)}{\mathbb{E}\, Y_n |\mathscr{G}_n|} \\
\leq & \sum_{\substack{G \in A_n \\ \frac{Y_n(G)}{\mathbb{E} Y_n} \leq \left(\frac{\operatorname{Var} Y_n}{p(n)}\right)^{1/2}}} \frac{1}{|\mathscr{G}_n|} \left(\frac{\operatorname{Var} Y_n}{p(n)} \right)^{1/2} + \sum_{k \in K} \sum_{G : \frac{Y_n(G)}{\mathbb{E} Y_n} \in (k, k+1]} \frac{k}{|\mathscr{G}_n|} \\
\leq & \; p(n) \left(\frac{\operatorname{Var} Y_n}{p(n)} \right)^{1/2} + \sum_{k \in K} k \left(\frac{\operatorname{Var} Y_n}{(k-1)^2} - \frac{\operatorname{Var} Y_n}{k^2} \right) \\
= & \; O\left(\sqrt{p(n)}\right) + \sum_{k \in K} O(k^{-2}) = o(1).
\end{aligned}$$

To show that also $\mathbb{P}_n(A_n) = o(1)$ when $\mathbb{Q}_n(A_n) = o(1)$ and thus to show that the other direction of (23.1) also holds, assume that event A_n has probability $q(n) = o(1)$ in $(\mathscr{G}_n, \mathbb{Q}_n)$.
Denote by

$$\mathscr{G}_n^* = \{ G \in \mathscr{G}_n : q(n)^{1/3} < Y_n(G)/\mathbb{E}\, Y_n < q(n)^{-1/3} \}.$$

Now,

$$\begin{aligned}
\mathbb{Q}(Y_n(G) < q(n)^{1/3}\, \mathbb{E}\, Y_n) &= \sum_{G : Y_n(G) < q(n)^{1/3}\, \mathbb{E}\, Y_n} \frac{Y_n(G)}{\mathbb{E}\, Y_n |\mathscr{G}_n|} \\
&\leq \frac{1}{|\mathscr{G}_n|} \sum_{G : Y_n(G) < q(n)^{1/3}\, \mathbb{E}\, Y_n} q(n)^{1/3}
\end{aligned}$$

$$\leq q(n)^{1/3} = o(1).$$

Furthermore, by Chebyshev's inequality,

$$\mathbb{Q}(Y_n(G) > q(n)^{-1/3}\,\mathbb{E}\,Y_n) \leq \frac{\operatorname{Var} Y_n}{((q(n)^{-1/3}-1)\,\mathbb{E}\,Y_n)^2} = o(1).$$

So, $\mathbb{Q}_n(\mathscr{G}_n^*) = 1 - o(1)$ and

$$\mathbb{Q}_n(A_n|\mathscr{G}_n^*) \leq \frac{q(n)}{1-o(1)}.$$

Furthermore, due to the choice of $\mathscr{G}_n^*$, we have

$$\mathbb{Q}_n(A_n|\mathscr{G}_n^*) \geq \sum_{G \in A_n : Y_n(G) \geq q(n)^{1/3}\,\mathbb{E}\,Y_n} \frac{Y_n(G)}{\mathbb{E}\,Y_n|\mathscr{G}_n|} \geq q(n)^{1/3}\,\mathbb{P}_n(A_n|\mathscr{G}_n^*).$$

Therefore,

$$\mathbb{P}_n(A_n|\mathscr{G}_n^*) \leq q(n)^{2/3}(1+o(1)) = o(1),$$

which completes the proof. □

23.2 Contiguity of random regular graphs and multigraphs

The notion of contiguity and Theorem 23.4 are particularly useful in studies of asymptotic properties of random r-regular graphs, i.e. graphs uniformly distributed over the family of all labeled graphs on n vertices, where each vertex has degree r. Since there is no simple exact formula for the number of such graphs (just an asymptotic approximation - see [91]) a useful way to bypass this main obstacle to study their properties is to generate a random r-regular multigraph with probability distribution close (contiguous) to the uniform distribution. One of the most effective ways to do so is to use the random configuration method presented in Section 9.1.

Denote by $\mathbb{G}_{n,r}$ a graph chosen at random from the family $\mathscr{G}_n$ of all r-regular graphs on vertex set $[n]$, and let $\mathbb{G}_{n,r}^{Y_n}$ stand for a graph sampled from $(\mathscr{G}_n, \mathbb{Q}_n)$, where $\mathbb{Q}_n$ and Y_n are defined as in (23.10).
The following contiguity relationship was a key element of Robinson and Wormald's [767] proof that a random 3-regular graph $\mathbb{G}_{n,3}$ is Hamiltonian w.h.p. For convenience, in what follows, we shall replace all statements that probability measures are contiguous by a statement that the respective random graphs are contiguous.

Theorem 23.5. *Let H_n be the number of Hamilton cycles in a random graph $\mathbb{G}_{n,3}$. Then random graphs $\mathbb{G}_{n,3}$ and $\mathbb{G}_{n,3}^{H_n}$ are contiguous.*

Instead of proving the above theorem we shall prove a simpler result, replacing (for $r=3$) a random r regular graph $\mathbb{G}_{n,r}$ by a random r-regular multigraph $\gamma(F)=\mathbb{M}_{n,r}$ on vertex set $[n]$, sampled uniformly from the family $\mathscr{M}_{n,r}$ of all r-regular multigraphs generated by configurations (pairings) F of $rn=2m$ points into m pairs (see Section 9.1 for details).

Theorem 23.6. *Let H_n be the number of Hamilton cycles in a random multigraph $\mathbb{M}_{n,3}$. Then random graphs $\mathbb{M}_{n,3}$ and $\mathbb{M}_{n,3}^{H_n}$ are contiguous.*

The proof of Theorem 23.6 reduces to checking if the conditions of Theorems 23.3 and 23.4 are satisfied. However in the proof we shall replace condition (A2) of Theorem 23.3 by a computationally more convenient condition (A2').

Lemma 23.7. *(Janson [499], Lemma 1) Suppose that condition (A1) of Theorem* 23.3 *holds, that $Y_n \geq 0$ and that*

$$(A2'): \qquad \frac{\mathbb{E}\left(Y_n(X_1)_{j_1}\cdots Y_n(X_k)_{j_k}\right)}{\mathbb{E}Y_n} \to \prod_{i=1}^{k}\mu_i^{j_i} \quad as \quad n\to\infty,$$

for some $\mu_i \geq 0$ and every finite sequence $j_1,\ldots,j_k$ of non-negative integers. Then condition (A2) holds with $\delta_i=\mu_i/\lambda_i-1$.

Proof. To show that condition (A2') implies condition (A2) of Theorem 23.3 observe that if the measure $\mathbb{Q}_n$ is defined as in formula (23.10), then, putting $\mu_i=\lambda_i(1+\delta_i)$, the condition (A2') can be written as

$$\mathbb{E}_{\mathbb{Q}_n}\left((X_1)_{j_1}\cdots(X_k)_{j_k}\right) \to \prod_{i=1}^{k}\mu_i^{j_i}, \text{ as } n\to\infty, \tag{23.11}$$

hence the method of moments (see Lemma 26.7) yields under $\mathbb{Q}_n$, $X_i \xrightarrow{D} \mathrm{Po}(\mu_i)$, jointly for all i with independent limits. This and condition (A1) gives, for any $j_1,j_2,\ldots,j_k$,

$$\begin{aligned}\frac{\mathbb{E}(Y_n|X_1=j_1\ldots,X_k=j_k)}{\mathbb{E}Y} &= \frac{\mathbb{E}(Y_nI(X_1=j_1\ldots,X_k=j_k)}{\mathbb{E}Y_n\mathbb{P}_n(X_1=j_1\ldots,X_k=j_k)}\\ &= \frac{\mathbb{Q}_n(X_1=j_1\ldots,X_k=j_k)}{\mathbb{P}_n(X_1=j_1\ldots,X_k=j_k)}\\ &\to \frac{\prod_{i=1}^k\mu_i^{j_i}e^{-\mu_i}/j_i!}{\prod_{i=1}^k\lambda_i^{j_i}e^{-\lambda_i}/j_i!} = \prod_{i=1}^{k}\left(\frac{\mu_i}{\lambda_i}\right)^{j_i}e^{\lambda_i-\mu_i}\end{aligned}$$

$$= \prod_{i=1}^{k} ((1+\delta_i))^{j_i} e^{-\lambda_i \delta_i}$$

which is the condition (A2) of Theorem 23.3. □

We shall also need two other Lemmas. The results in the first Lemma were derived for $r = 3$ in [767], stated in general without being used in [768], and then derived and applied by Frieze et al. [396].

Lemma 23.8. *For fixed $r \geq 3$, let H_n be the number of Hamilton cycles in a random r-regular multigraph $\mathbb{M}_{n,r}$. Then with rn restricted to even integers,*

$$\mathbb{E} H_n \sim \sqrt{\frac{\pi}{2n}} \left(\frac{(r-2)^{r-2}(r-1)^2}{r^{r-2}} \right)^{n/2}, \text{while } \frac{\mathbb{E} H_n^2}{(\mathbb{E} H_n)^2} \to \frac{r}{r-2}.$$

□

Finally, we shall apply the following asymptotic result due to Bollobás [146].

Lemma 23.9. *For r fixed, let $X_{i,n}, i \geq 1$ be the number of cycles of length i in a random regular multigraph $\mathbb{M}_{n,r}$. Then for each $k \geq 1$, $X_{i,n} \xrightarrow{D} Z_i$ jointly for $i = 1, 2, \ldots, k$, where $Z_i \sim \mathrm{Po}(\lambda_i)$ are asymptotically independent Poisson random variable with expectations $\lambda_i = (r-1)^i/(2i)$.*

□

Proof. (of Theorem 23.6 , see Wormald [859])
As we have already mentioned the proof reduces to verification that in this case the conditions (A1)-(A4) stated in Theorem 23.3 are satisfied.
For convenience, let H_n now denote the number of Hamilton cycles in a random 3-regular multigraph $\mathbb{M}_{n,3}$, while $X_i = X_{i,n}, i \geq 1$ is the number of cycles of length i in $\mathbb{M}_{n,3}$.

First notice that, by Lemma 23.8, $\mathbb{E} H_n^2/(\mathbb{E} H_n)^2 \to 3$, while Lemma 23.9 implies that the condition (A1) of Theorem 23.3 holds with $\lambda_i = 2^{i-1}/i$.

Instead of checking the second condition (A2) of 23.3 we use Lemma 23.7 and replace it by condition (A2') with $\mu_i = (1+\delta_i)\lambda_i$. As a first step to show that (A2') holds we shall prove that

$$\frac{\mathbb{E}(H_n X_{i,n})}{\mathbb{E} H_n} \to (1+\delta_i)\lambda_i \tag{23.12}$$

as $n \to \infty$, where

$$\delta_i = \frac{(-1)^i - 1}{2^i}.$$

We use the notation and construction of the configuration model introduced in Section 9.1 with each $d_i = 3$ and let D be a set of pairs of points corresponding to a Hamilton cycle in a partition F of a set of $3n = 2m$ points into m pairs (*a configuration* or *pairing*). By symmetry all copies of D are equivalent and so in $\mathbb{M}_{n,3} = \gamma(F)$

$$\frac{\mathbb{E}(H_n X_{i,n})}{\mathbb{E} H_n} = \mathbb{E}(X_{i,n} | D \subseteq F).$$

Let C be a set of pairs corresponding to a cycle of length i and consider paths in the graph $\gamma(D \cap C)$. Note that there must be at least one such path in $\gamma(D \cap C)$. All pairs in C cannot be in D if $n > i$. Give these paths and a consistent orientation along C (which multiplies the count by 2) and a distinguished pair as first (which multiplies the count by the number of paths and induces a linear ordering of paths around C). Thus

$$\frac{\mathbb{E}(H_n X_{i,n})}{\mathbb{E} H_n} = \sum_S \frac{1}{2|S|} \mathbb{E}(X_{i,n}(S) | D \subseteq \ F),$$

where S denotes the sequence of path lengths, $|S|$ is the number of paths in S and $X_{i,n}(S)$ is the number of cycles of length i in F consistent with S.
Fix S such that $|S| = k$. There are asymptotically $(2n)^k$ ways to choose the starting points of the paths on D together with their directions along D. Once starting points are chosen then pairs of points in C, which correspond to a cycle of length i yielding S, are determined. The probability that these pairs all occur in $\mathbb{M}_{n,3}$ conditional upon $D \subseteq F$ is asymptotically n^{-k}. Hence

$$\mathbb{E}(X_{i,n}(S) | D \subseteq F) \to 2^{|S|}$$

and so

$$\frac{\mathbb{E}(H_n X_{i,n})}{\mathbb{E} H_n} \to \sum_{k \geq 1} \frac{2^k}{2k} |\{S : |S| = k\}|.$$

The ordinary generating function for the number of sequences S with x marking the total number of vertices involved and y marking the number of paths is $g(x,y)/(1 - g(x,y))$, where $g(x,y) = yx^2/(1-x)$ is the generating function of one path. Thus, with square brackets denoting extraction of coefficients,

$$\begin{aligned}
\frac{\mathbb{E}(H_n X_{i,n})}{\mathbb{E} H_n} &\to \sum_{k \geq 1} \frac{2^k}{2k} [x^i y^k] \frac{yx^2}{1 - x - yx^2} \\
&= [x^i] \sum_{k \geq 1} \frac{1}{2k} [y^{k-1}] \frac{2x^2}{1 - x - yx^2} \\
&= [x^i] \frac{1}{2} \int_0^1 \frac{2x^2}{1 - x - yx^2} dy \\
&= -[x^i] \frac{1}{2} \log(1 - x - 2x^2) + \frac{1}{2} \log(1 - x),
\end{aligned}$$

and (23.12) follows. To prove that the condition (A2') (see Lemma 23.7) holds one has to extend (23.12) to higher moments. In fact one can show that the above argument works in such case as well.
Finally the condition (A3) of Theorem 23.3 is satisfied since

$$\sum_{i\geq 1} \lambda_i \delta_i^2 = \log 3,$$

which, combined with Lemma 23.8, shows that the condition $(A4)$ also holds.
To apply Theorem 23.4 notice that $\delta_i = -1$ only for $i = 1$ therefore in fact our conclusions are restricted to a random 3-regular multigraph $\widehat{\mathbb{M}}_{n,3}$ without loops, i.e., that $\widehat{\mathbb{M}}_{n,3}$ and $\widehat{\mathbb{M}}_{n,3}^{H_n}$ are contiguous. The theorem follows using the definition of contiguity by restricting to $X_2 > 0$, since in $\mathbb{M}_{n,3}$ probability of this event tends to zero (see Lemma 9.6). $\square$

23.3 Contiguity of superposition models

To prove that random r-regular graph is Hamiltonian for every $r \geq 3$ and so to strengthen their result for $r = 3$ Robinson and Wormald [768] found that their method of analysis of variance, based on conditioning on short cycles, and described in detail in the first part of this chapter, becomes technically too complicated. Instead they introduced a new method based on an idea of unions of regular graphs on the same vertex set. They create simple random graphs induced by multigraphs on n vertices using, for example, the superposition of r 1-regular graphs (perfect matchings, for n even) or $\lfloor r/2 \rfloor$ Hamiltonian cycles, or combination of random regular graphs of degree lower than r with random perfect matchings and/or random Hamiltonian cycles. Next they show that such random graphs are contiguous with $\mathbb{G}_{n,r}$, i.e., share asymptotic properties with them, including Hamiltonicty.

We shall introduce superposition of random regular graphs and multigraphs by defining a two related union operations $+$ and $\oplus$. Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two independent random graphs or multigraphs on the same set $[n]$ of vertices. Then a random (multi) graph $\mathbb{G}_1 + \mathbb{G}_2$ denotes its union (sum). Since, in general, $\mathbb{G}_1 + \mathbb{G}_2$ may not be a simple graph therefore we define a "simple sum" $\mathbb{G}_1 \oplus \mathbb{G}_2$ to be $\mathbb{G}_1 + \mathbb{G}_2$ conditioned on being simple. Notice that both operations $+$ and $\oplus$ are commutative and associative therefore we define

$$k\mathbb{G} = \mathbb{G} + \mathbb{G} + \cdots + \mathbb{G},$$

and

$$k\mathbb{G} = \mathbb{G} \oplus \mathbb{G} \oplus \cdots \oplus \mathbb{G},$$

with k terms on the right.

Finally, denote contiguity of two random multigraphs (graphs) $\mathbb{M}_n^{(1)}$ and $\mathbb{M}_n^{(2)}$ by $\mathbb{M}_n^{(1)} \approx \mathbb{M}_n^{(2)}$.

Lemma 23.10. *(Janson [499]) Suppose that $\mathbb{M}_n, \widehat{\mathbb{M}}_n, \mathbb{R}_n, \widehat{\mathbb{R}}_n$ are random multigraphs (graphs) on n vertices with some probability distributions such that $\mathbb{M}_n \approx \widehat{\mathbb{M}}_n$ and $\mathbb{R}_n \approx \widehat{\mathbb{R}}_n$. Then*

$$\mathbb{M}_n + \mathbb{R}_n \approx \widehat{\mathbb{M}}_n + \widehat{\mathbb{R}}_n.$$

Proof. It is a simple consequence of statements (ii) and (iii) of Proposition 23.2. Indeed, by the statement (iii) pairs $(\mathbb{M}_n, \mathbb{R}_n)$ and $(\widehat{\mathbb{M}}_n, \widehat{\mathbb{R}}_n)$ are contiguous, and the result follows by the statement (ii). □

The following extension of the above lemma is given by Wormald in [859].

Lemma 23.11. *Suppose that $\mathbb{M}_n, \widehat{\mathbb{M}}_n, \mathbb{R}_n, \widehat{\mathbb{R}}_n$ are random multigraphs (graphs) on n vertices with bounded degrees and with some label-independent probability distributions such that $\mathbb{M}_n \approx \widehat{\mathbb{M}}_n$ and $\mathbb{R}_n \approx \widehat{\mathbb{R}}_n$. Then*

$$\mathbb{M}_n \oplus \mathbb{R}_n \approx \widehat{\mathbb{M}}_n \oplus \widehat{\mathbb{R}}_n.$$

Proof. (sketch)

Let A_n be any sequence of events which is w.h.p. true in $M_n \oplus \mathbb{R}_n$. Due to definitions of operations $+$ and $\oplus$, sequence A_n holds w.h.p. in $\mathbb{M}_n + \mathbb{R}_n$ as well. Let B_n be the event that $\mathbb{G} \in \mathbb{M}_n + \mathbb{R}_n$ has a multiple edge. Then the union of events $A_n \cup B_n$ is w.h.p. true in $\mathbb{M}_n + \mathbb{R}_n$ and hence also, by Lemma 23.10, in $\widehat{\mathbb{M}}_n + \widehat{\mathbb{R}}_n$. To show that it also holds w.h.p. in $\widehat{\mathbb{M}}_n \oplus \widehat{\mathbb{R}}_n$ one has to show that the probability of the complement of B_n in $\widehat{\mathbb{M}}_n + \widehat{\mathbb{R}}_n$ is bounded below by a positive constant. For this it suffices to relabel separately each pair of graphs $\mathbb{G}_1, \mathbb{G}_2 \in \widehat{\mathbb{M}}_n$ and use the Method of Moments (see Chapter 26). One can verify that the expected number λ of edges in common between two such relabelings is exactly half of the product of the average degrees of vertices in $\mathbb{G}_1$ and $\mathbb{G}_2$ and that k-th factorial moments tend to λ^k. Hence, by Theorem 26.11 the probability of the complement of B_n is asymptotically $e^{-\lambda}$, which is bounded below in view of the assumption that vertex degrees are bounded.

The reverse argument, from $\widehat{\mathbb{M}}_n \oplus \widehat{\mathbb{R}}_n$ to $\mathbb{M}_n \oplus \mathbb{R}_n$ is identical.

□

Theorem 23.12. *The following pairs of random graphs are contiguous:*

(i) $\mathbb{G}_{n,r-1} \oplus \mathbb{G}_{n,1}$ *and* $\mathbb{G}_{n,r}$ *for* $r \geq 3$ *and n even,*

(ii) $\mathbb{G}_{n,r-2} \oplus \mathbb{G}_{n,2}$ *and* $\mathbb{G}_{n,r}$ *for* $r \geq 3$*,*

(iii) $3\mathbb{G}_{n,1}$ *and* $\mathbb{G}_{n,3}$ *, n even.*

Denote by $\mathbb{H}_n$ a random Hamilton cycle on vertex set $[n]$ and notice that $\mathbb{H}_n$ although is 2-regular, is clearly differs from a random graph $\mathbb{G}(n,2)$.

Theorem 23.13. *The following pairs of random graphs are contiguous:*

(i) $\mathbb{G}_{n,r-2} \oplus \mathbb{H}_n$ *and* $\mathbb{G}_{n,r}$ *for* $r \geq 3$,

(ii) $\mathbb{G}_{n,4}$ *and* $2\mathbb{H}_n$

Theorem 23.14. *Let* $r \geq 3$ *and suppose that* $r = 2j + \sum_{i=1}^{r-1} ik_i$, *with all terms non-negative. Then*

$$\mathbb{G}_{n,r} \approx j\mathbb{H}_n \oplus k_1\mathbb{G}_{n,1} \oplus k_2\mathbb{G}_{n,2} \oplus \cdots \oplus k_{r-1}\mathbb{G}_{n,r-1},$$

where n is restricted to even integers if $k_i \neq 0$ *for any odd i.*

Proof. (see [859])
Notice that using (ii) and (repeatedly) (iii) of Theorem 23.12, as well as Lemma 23.11, □

$$\mathbb{G}_{n,r} \approx r\mathbb{G}_{n,1}. \tag{23.13}$$

Moreover, applying (ii) of Theorem 23.12 and Theorem 23.13 we get

$$\mathbb{G}_{n,r} \approx j\mathbb{H}_n \oplus k\mathbb{G}_{n,2} \oplus \mathbb{G}_{n,r-2j-2k}, \tag{23.14}$$

for any $k \leq r/2 - 2j$. If $k_i = 0$ for all odd i, take $k = r/2 - 2j$ and combine the copies of $\mathbb{G}_{n,2}$ into the desired random graph $\mathbb{G}_{n,i}$ using the same result in reverse, with $j = 0$, for each of them. If not we can assume n is even, so by (23.14) with $k = k_2$ and by (23.13), we get

$$\mathbb{G}_{n,r} \approx j\mathbb{H}_n \oplus k_2\mathbb{G}_{n,2} \oplus (r - 2j - 2k_2)\mathbb{G}_{n,1}, \tag{23.15}$$

unless $r - 2j - 2k_2 = 2$. Next, recombine the copies of $\mathbb{G}_{n,1}$ into all the other terms required, using (23.13) in reverse for each term $\mathbb{G}_{n,i}$, $i \geq 3$. The required k_1 copies of $\mathbb{G}_{n,1}$ will be surplus. The only case left is $r - 2j - 2k_2 = 2$ and $k_1 = 2$, whence either $j > 0$ or $k_2 > 0$.
From above we have

$$\mathbb{G}_{n,r} \approx j\mathbb{H}_n \oplus (k_2 + 1)\mathbb{G}_{n,2},$$

and any two of these can be recombined to give $\mathbb{G}_{n,4}$ This can be split as desired since by (i) of Theorem 23.12 (repeated twice)

$$\mathbb{G}_{n,4} \approx \mathbb{G}_{n,3} \oplus \mathbb{G}_{n,1} \approx \mathbb{G}_{n,2} \oplus 2\mathbb{G}_{n,1}$$

while, by (i) of Theorem 23.13

$$\mathbb{G}_{n,3} \approx \mathbb{H}_n \oplus \mathbb{G}_{n,2}$$

□

23.4 Exercises

23.4.1 Prove Lemma 23.8

23.4.2 Prove Lemma 23.9

23.4.3 Show that (23.12) extends to higher moments.

23.4.4 Apply Theorem 23.12 to show that $\mathbb{G}_{n,9} \approx 5\mathbb{G}_{n,1} \oplus 2\mathbb{G}_{n,2}$ for n even. Explain why $\mathbb{G}_{n,1} \oplus \mathbb{G}_{n,1} \not\approx \mathbb{G}_{n,2}$.

23.5 Notes

Interest in the notion of contiguity of random graphs was stimulated by the results of Robinson and Wormald [767], [768] that random r-regular graphs, $r \geq 3, r = O(1)$ are Hamiltonian. As a result, we find that other non-uniform models of random regular graphs are contiguous to $\mathbb{G}_{n,r}$ e.g. the union $r\mathbb{G}_{n,1}$ of r random perfect matchings when n is even. (There is an implicit conditioning on $r\mathbb{G}_{n,1}$ being simple here). The most general result in this line is given by Wormald [859], improving on earlier results of Janson [499] and Molloy, Robalewska, Robinson and Wormald[686] and Kim and Wormald [567]. More precisely, those results are gathered in Theorems 23.12 and 23.13. The statement (i) of Theorem 23.12 was proved in [767] and [499], the statement (ii) was verified by Robalewska in [766], while the proof of (iii) is given in [499] and [686]. A proof of statement (i) of the Theorem 23.13 is due to Frieze, Jerrum, Molloy, Robinson and Wormald (see [396]), while the statement (ii) of the same theorem is given in [567]. In addition, Greenhill, Janson, Kim and Wormald considered in [451] the contiguity of random regular graphs and the unions of random permutations. They proved that if $\mathscr{P}_n$ denotes a graph formed from uniformly chosen permutation on n vertices without loops or multiple edges, then $\mathscr{P}_n \oplus \mathscr{P}_n \approx \mathbb{G}_{n,4}$ and $\mathbb{G}_{n,r-2} \oplus \mathscr{P}_n \approx \mathbb{G}_{n,r}$ for $r \geq 3$.

Janson in [504] discusses asymptotic equivalence and contiguity of sequences of homogeneous random graphs. In particular he studies the contiguity of generalized binomial random graphs on n vertices (see Section 8.1) with perturbed edge probabilities $p_{i,j}$. His results are extended by Kleijn and Rizzeli [571] who propose a more general form of asymptotic congruence called *remote contiguity* to show that connectivity properties are preserved in more heavily perturbed homogeneous random graphs.

Lefebvre [618] assumes that two models of random graphs are *first-order contiguous* if they satisfy w.h.p. the same sets of first-order formulas, and uses this notion to classify random structures from a logical point of view. In particular he studies

random graphs with a given degree sequence and characterize degree sequences that define contiguous random graph sequences.
For a comprehensive and detailed account of the contiguity of various models of random graphs we refer the reader to Wormald's review paper [859] and to Chapter 9 of Janson, Łuczak and Ruciński monograph [509].

Chapter 24

Random Walk on Random Graphs

In this chapter we are concerned with certain aspects of a random walk on a random graph. Given a connected graph or multi-graph G on vertex set $V=[n]$, a random walk $\mathscr{W}_v$ is a sequence of random variables $\mathscr{W}_v(t)\in[n], t\geq 0$, where $\mathscr{W}_v(0)$ is an arbitrarily chosen vertex v and $\mathscr{W}_v(t)$ is a uniformly random neighbor of $\mathscr{W}_v(t-1)$. (In the case of multi-graphs, we choose neighbors proportional to the multiplicity of the edge $\{\mathscr{W}_v(t-1),\mathscr{W}_v(t)\}$.) See Section 30 for a discussion of the results on random walks used here.

A random walk on a graph is a special case of a *Markov Chain*. This is a sequence of random variables $X(0),X(1),\ldots,X(t),\ldots$ taking values in a set Ω. The defining property is that

$$\Pr(X(t)=\omega_t \mid X(t-1)=\omega_{t-1},X(t-2)=\omega_{t-2},\ldots,X(0)=\omega_0)=$$
$$\Pr(X(t)=\omega_t \mid X_{t-1}=\omega_{t-1}),$$

i.e. the probability that $X(t)=\omega_t$ depends only on X_{t-1} and not on earlier *states*. Clearly $X(t)=\mathscr{W}_u(t)$ satifies these requirements.

We first focus on the *mixing time* of the walk. Let $P_v(w,t)$ denote the probability that $\mathscr{W}_v(t)=w$ and let $P_v(A,t)=\sum_{w\in A}P_v(w,t)$. It is well known that $\lim_{t\to\infty}P_v(w,t)=\pi(w)=\frac{\deg_G(w)}{2|E(G)|}$. The mixing time here will be

$$\tau(\varepsilon)=\min\{t:||\pi-P_v(\cdot,t)||_{TV}\leq\varepsilon\}. \tag{24.1}$$

Here the *total variation distance*

$$||\pi-P_v(\cdot,t)||_{TV}=\frac{1}{2}\sum_{w\in V}|P_v(w,t)-\pi(w)|=\max_{A\subseteq V}|P_v(A,t)-\pi(A)|.$$

So the mixing time is a statement about the rate at which the random walk converges to the *stationary distribution* π.

We also spend some time on the *cover time*. We let $C_G(v)$ denote the expected time for the random walk $\mathscr{W}_v$ to visit all the vertices of G. The cover time C_G of G is then $\max_v C_G(v)$.

As a final topic, we consider a random version of the so-called *Walker-Deletor* game. Here alternately, Walker takes a step of a random walk and Deletor deletes an edge that Walker has not yet traversed. The question to resolve is as to how many vertices Walker can visit in this process.

A random walk is of course an example of a Markov chain. A Markov chain is *lazy* if its transition matrix P satisfies $P(i,i) \geq 1/2$ for all $i \in [n]$. For technical reasons, we will assume that our walks are lazy. This can be achieved by adding a loop of probability 1/2 at each vertex. In terms of transition matrices, this replaces P by $(P+I)/2$. This at most doubles the mixing and cover times.

24.1 Mixing time

In this section, we discuss the mixing time of a random walk on $\mathbb{G}_{n,p}$ and on a random regular graph, $\mathbb{G}(n,r), r = O(1)$. We begin with $\mathbb{G}_{n,p}$ and estimate the *conductance* of the walk $\mathscr{W}$ using (30.7). (We omit the suffix u from $\mathscr{W}_u$ if the starting vertex is irrelevant.)

Mixing time for $\mathbb{G}_{n,p}$

In this section, we prove

Theorem 24.1. *If $p = \frac{c \log n}{n}$ where $c > 1$ is a constant then in $\mathbb{G}_{n,p}$, we have $\tau(\varepsilon) \leq 2\log(20n/\varepsilon)$.*

Proof. The conductance $\Phi = \Phi(\mathscr{M})$ of $\mathscr{M}$ is defined by

$$\Phi = \min\{\Phi_S : S \subseteq \Omega,\ 0 < \pi(S) \leq 1/2\}$$

where if $Q(i,j) = \pi_i P(i,j)$ and $\bar{S} = \Omega \setminus S$,

$$\Phi_S = \pi(S)^{-1} Q(S,\bar{S}).$$

Thus Φ_S is the steady state probability of moving from S to $\bar{S}$ in one step of the chain, conditional on being in S. Clearly $\Phi \leq 1/2$ if $\mathscr{M}$ is lazy. In terms of mixing time we have

Lemma 24.2. *If $\mathscr{M}$ is a lazy ergodic chain then*

$$\tau(\varepsilon) \leq \left\lceil \frac{2|\log \varepsilon \pi_{\min}|}{\Phi^2} \right\rceil.$$

For the proof of this see Section 30.

Lemma 24.3. *Suppose that $p = \frac{c\log n}{n}$ where $c > 1$ is a constant. Then w.h.p. in $\mathbb{G}_{n,p}$ the conductance $\Phi(\mathscr{W}) \geq 1/3$.*

Proof of Lemma 24.3
Let $N_\varepsilon = \{v : \deg(v) \notin (1 \pm \varepsilon)\log n\}$. The Chernoff bounds imply that $\mathbb{E}(|N_\varepsilon|) \leq n^{1-\varepsilon^2/2}$ and so by the Markov inequality we have that

$$|N_\varepsilon| \leq n^{1-\varepsilon^2/3} \ w.h.p. \tag{24.2}$$

We next argue that w.h.p.,

$$S \subseteq [n],\ |S| \leq n/\log^2 n \text{ implies that } e(S) \leq 3|S|. \tag{24.3}$$

Indeed,

$$\begin{aligned}
\Pr(\exists S, |S| \leq n/\log^2 n, e(S) \geq 3|S|) &\leq \sum_{s=7}^{n/\log^2 n} \binom{n}{s} \binom{\binom{s}{2}}{3s} \left(\frac{c\log n}{n}\right)^{3s} \\
&\leq \sum_{s=7}^{n/\log^2 n} \left(\frac{ne}{s}\right)^s \left(\frac{ces\log n}{6n}\right)^{3s} \\
&= \sum_{s=7}^{n/\log^2 n} \left(\left(\frac{s}{n}\right)^2 \cdot \frac{c^3 e^4 \log^3 n}{216}\right)^s = o(1).
\end{aligned}$$

Finally we have

$$|S| \in \left[\frac{n}{\log^2 n}, \frac{2n}{3}\right] \text{ implies that } e(S) \geq \frac{c|S|\log n}{2}. \tag{24.4}$$

Indeed,

$$\begin{aligned}
&\Pr(\exists S, |S| \geq n/\log^2 n, e(S) \leq cs\log n/2) \\
&\leq \sum_{s=n/\log^2 n}^{2n/3} \binom{n}{s} \Pr(Bin(s(n-s), p) \leq s(n-s)p/2) \\
&\leq \sum_{s=n/\log^2 n}^{2n/3} \left(\frac{ne}{s}\right)^s e^{-cs(n-s)\log n/8} \\
&\leq \sum_{s=n/\log^2 n}^{2n/3} \left(\frac{ne}{s} \cdot e^{-c\log n/25}\right)^s = o(1).
\end{aligned}$$

We can now estimate the conductance. We have that w.h.p.

$$\alpha_0 \log n \leq \deg(v) \leq \alpha_1 \log n \text{ for all } v \in [n],$$

for some constants $\alpha_i = \alpha_i(c), i = 0, 1$, see Exercises 3.3.7 and 3.3.8.
For a random walk, conductance reduces to the following. See (30.7) for the proof.

$$\Phi_S = \frac{e(S, \bar{S})}{\deg(S)} = \frac{\deg(S) - 2e(S)}{\deg(S)} \geq \frac{\deg(S) - 3|S|)}{\deg(S)} = 1 - o(1),$$

if $|S| \leq n/\log^2 n$.
Note next that if $|S| \geq 2n/3$ then w.h.p.

$$\pi(S) \geq \frac{(2n/3 - n^{1-\varepsilon^2/3})(1-\varepsilon)c\log n}{c\log n} > \frac{1}{2}.$$

And so if $n/\log^2 n \leq |S| \leq 2n/3$ then

$$\Phi_S = \frac{e(S, \bar{S})}{\deg(S)} \geq \frac{c|S|\log n/2}{(1+\varepsilon)c\log n + \alpha_1 n^{1-\varepsilon^2/3}\log n} \geq \frac{1}{3}.$$

End of proof of Lemma 24.3.
Theorem 24.1 follows from Lemma 24.3 and Corollary 30.6 and doubling through laziness. □

Mixing time for $\mathbb{G}(n,r)$

We now consider the random regular graph $\mathbb{G}(n,r)$. Friedman [380] proved that w.h.p., if $r \geq 3$,

$$l_1 \leq 2(r-1)^{1/2} + \eta$$

for any $\eta > 0$. Here l_1 is the second eigenvalue of the adjacency matrix of $\mathbb{G}(n,r)$.

It follows from this and Corollary 30.2 that after making the chain lazy, we have $l_{\max} \leq (1 + 2(r-1)^{1/2} + \eta)/2$. Hence,

Theorem 24.4. *If $r \geq 3$ then in $\mathbb{G}(n,r)$ we have $\tau(\varepsilon) \leq 2\log(2n/\varepsilon)$.*

24.2 Cover time

We let $C_G(u)$ denote the expected time for the random walk $\mathscr{W}_u$ to visit all the vertices of G. The cover time C_G of G is then $\max_u C_G(u)$.

The cover time of $\mathbb{G}_{n,p}$

Jonasson [528] initiated the study of the cover time of random graphs and showed that if $np \gg \log n$ then w.h.p. the cover time of $\mathbb{G}_{n,p}$ is aymptotic to $n\log n$, the time for the *coupon collector* to get all n coupons. We will prove

Theorem 24.5. *Suppose that $c > 1$ is constant and $p = \frac{c\log n}{n}$. Then w.h.p.,*

$$C_{\mathbb{G}_{n,p}} \sim \psi(c) n\log n, \quad \textit{where } \psi(c) = c\log\left(\frac{c}{c-1}\right).$$

This ws proved by (Cooper and Frieze [257] where $c = 1 + o(1)$ is allowed.

Proof.

First Visit Time Lemma Subsequent to [257], they tackled random regular graphs and in the process proved the *First Visit Time Lemma*: Let G denote a fixed connected graph, and let u be some arbitrary vertex from which a walk $\mathscr{W}_u$ is started. Let $\mathscr{W}_u(t)$ be the vertex reached at step t, let P be the matrix of transition probabilities of the walk, and let $P_u(v,t) = \Pr(\mathscr{W}_u(t) = v)$.

Let T be such that, for $t \geq T$

$$\max_{u,x\in V}\left|\frac{P_u(x,t) - \pi_x}{\pi_x}\right| \leq \frac{1}{\omega} \tag{24.5}$$

where $\omega = \omega(n) \to \infty$.

For a large constant $K > 0$, let

$$\eta = \frac{1}{KT}.$$

For $t \geq 0$, let $\mathscr{A}_t(v)$ be the event that $\mathscr{W}_u$ does not visit v in steps $T, T+1, \dots, t$. The vertex u will have to be implicit in this definition. We will use the following lemma. For the proof see Section 30.2. Let $r_t = \Pr(\mathscr{W}_v(t) = v)$ and let $R_v = \sum_{t=0}^T r_t$ denote the expected number of visits by $\mathscr{W}_v$ to vertex v in the time interval $[0,T]$.

Lemma 24.6. *[First Visit Time Lemma]*
Suppose that $T\pi_v = o(1)$ and that $\sum_{t=0}^T r_t(1+\eta)^t < 1$. Then for all $t \geq \max\{T, n\log n\}$,

$$\Pr(\mathscr{A}_t(v)) \sim \exp\left\{-\frac{t\pi_v}{R_v}(1 + O(T\pi_v))\right\}. \tag{24.6}$$

This is proved in Section 30.2.

Upper bound on cover time By putting $\varepsilon = 1/n$ in Theorem 24.1, we can take $T = 5\log n$ and $\omega = n$ in (24.5). Given this we can show

Lemma 24.7. *$R_v = 1 + o(1)$ for all $v \in [n]$.*

Proof. Observe next that w.h.p. a set of $s \leq \log\log n$ vertices contains at most s edges. Indeed,

$$\begin{aligned}\Pr(\exists |S| \leq \log\log n : e(S) \geq |S|+1) &\leq \sum_{s=4}^{\log\log n} \binom{n}{s}\binom{\binom{s}{2}}{s+1} p^{s+1} \\ &\leq ep\log\log n \sum_{s=4}^{\log\log n} \left(\frac{enp}{2}\right)^s = o(1).\end{aligned} \tag{24.7}$$

Let $r_t = P_v(v,t)$. Then $r_0 = 1$ and $r_1 = 0$ and $r_2 \leq 1/\alpha_0 \log n$. It follows from (24.7) that $r_3 = O(1/\log^2 n)$ and also, if $\mathscr{W}_v(t)$ is at distance 3 or more from v then $r_\tau = O((\sigma - t)/\log^3 n)$ for $\sigma > t$. Thus,

$$R_v - 1 \leq \sum_{t=1}^{T} r_t (1+\eta)^t = O\left(\sum_{t=1}^{T} r_t\right) = O\left(\frac{1}{\log n} + \frac{1}{\log^2 n} + \frac{T}{\log^3 n}\right) = o(1). \tag{24.8}$$

□

Note that (24.8) implies that Lemma 24.6 is applicable.

Let $T_G(u)$ be the time taken to visit every vertex of G by the random walk $\mathscr{W}_u$. Let U_s be the number of vertices of $G = \mathbb{G}_{n,p}$ which have not been visited by $\mathscr{W}_u$ at step s.

$$\begin{aligned} C_u = \mathbb{E}\, T_G(u) &= \sum_{s>0} \Pr(T_G(u) \geq s) = \sum_{s>0} \Pr(U_s > 0) \\ &\leq \sum_{s>0} \min\{1, \mathbb{E}\, U_s\} \leq t + \sum_{v\in V}\sum_{s>t} \Pr(\mathscr{A}_s(v)), \qquad (24.9) \\ &\leq t + (1+o(1)) \sum_{v\in V}\sum_{s>t} \exp\left\{-\frac{(1-o(1))s\deg(v)}{cn\log n}\right\} \quad \text{using (24.6)} \\ &= t + (1+o(1))n \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \sum_{s>t} \exp\left\{-\frac{(1-o(1))sk}{cn\log n}\right\} \\ &= t + (1+o(1))n^{1-c} \sum_{k=0}^{n} \frac{(c\log n)^k}{k!} \sum_{s>t} \exp\left\{-\frac{(1-o(1))sk}{cn\log n}\right\} \\ &= t + (1+o(1))n^{1-c} \sum_{k=0}^{n} \frac{(c\log n)^k}{k!} \frac{\exp\left\{-\frac{(1-o(1))tk}{cn\log n}\right\}}{1 - \exp\left\{-\frac{(1-o(1))k}{cn\log n}\right\}} \end{aligned}$$

$$\leq \tau n \log n + (1+o(1))cn^{2-c} \log n \sum_{k=0}^{n} \frac{(c\log n)^{k+1}}{k!k} \cdot e^{-\tau k/c}, \tag{24.10}$$

for $t = \tau n \log n$.

Putting $\tau_0 = \psi(c) = c\log\left(\frac{c}{c-1}\right)$, we see that

$$\begin{aligned}\sum_{k=0}^{n}\left(\frac{(c\log n)^k}{k!k}\cdot e^{-\tau_0 k/c}\right) &\leq \frac{2}{c\log n}\sum_{k=0}^{n}\left(\frac{(c\log n)^{k+1}}{(k+1)!}\cdot e^{-\tau_0 k/c}\right)\\ &\leq \frac{2}{c\log n}\exp\left\{c\log n\cdot\frac{c-1}{c}\right\}\\ &= \frac{2n^{c-1}}{c\log n}.\end{aligned} \tag{24.11}$$

It follows from (24.10), (24.11) that if $t_0 = \tau_0 n \log n$ then

$$C_u \leq (1+o(1))t_0 + O(n) \lesssim \psi(c) n \log n. \tag{24.12}$$

Lower bound on cover time For any vertex u, we can find a set of vertices S such that at time $t_1 = t_0(1-\varepsilon_1)$, $\varepsilon_1 = 1/\log\log n$, the probability the set S is covered by the walk $\mathscr{W}_u$ tends to zero. Hence $T_G(u) > t_1$ w.h.p. which implies that $C_G \geq (1-o(1))t_0$.

We construct S as follows. Let

$$k^* = \lceil (c-1)\log n \rceil \text{ and } V^* = \{v : \deg(v) = k^*\}.$$

Then, w.h.p. see Exercise ??,

$$|V^*| \geq \frac{n^{\theta(c)}}{\log n} \text{ where } \theta(c) = (c-1)\log\left(\frac{c}{c-1}\right). \tag{24.13}$$

If $v \in V^*$ then

$$\begin{aligned}\Pr(\mathscr{A}_{t_1}(v)) &\sim \exp\{-(1+O(1/\log n)+T\pi_v)(1-\varepsilon_1)\theta(c)\log n\}\\ &\geq n^{-(1-\varepsilon_1/2)\theta(c)}.\end{aligned} \tag{24.14}$$

Because the maximum degree in $\mathbb{G}_{n,p}$ is $O(\log n)$, we can choose a subset $S_1 \subseteq V^*$ if size $\Omega(|V^*|/\log^4 n)$ such that the distance between $v,w \in S_1$ is at least 4. Now if S is the set of vertices not visited in the interval $[T, t_1]$ then

$$\mathbb{E}(|S|) \geq \frac{\Omega(n^{\varepsilon_1\theta(c)/2})}{\log^5 n} - T \to \infty.$$

We have subtracted the upper bound T for the set of vertices in S that have been visited up to time T.

We now use the second moment method to prove that $S \neq \emptyset$ w.h.p. This implies that $C_G \geq t_1$. So, let $v, w \in S$. Let Γ be obtained from G by merging v, w into a single node z. This node has degree $2k^*$.

There is a natural measure preserving mapping from the set of walks in G which start at u and do not visit v or w, to the corresponding set of walks in Γ which do not visit z. Thus the probability that $\mathscr{W}_u$ does not visit v or w in the first t steps is equal to the probability that a random walk $\widehat{\mathscr{W}_u}$ in Γ which also starts at u does not visit z in the first t steps. We apply Lemma 24.6 to Γ.

We have that $\pi_z = \frac{2k^*}{cn\log n}$ and the argument for (24.8) implies that $R_z = 1 + o(1)$. Merging v, w does not decrease conductance and so we can use $T = 5\log n$ as a mixing time. Thus

$$\Pr(\mathscr{A}_{t_1}(Z)) = \exp\left\{-\frac{2t_1k^*}{cn\log n}\left(1 + O\left(\frac{\log n}{n}\right)\right)\right\} \sim \Pr(\mathscr{A}_{t_1}(v)\Pr(\mathscr{A}_{t_1}(w)).$$

It follows that $\mathbb{E}|S|^2 \sim (\mathbb{E}|S|)^2$ and the second moment method implies that $|S| \neq \emptyset$ w.h.p. This in conjunction with (24.11), completes the proof of Theorem 24.5. $\square$

The cover time of $\mathbb{G}(n,r)$

We will prove

Theorem 24.8. *Suppose that $r \geq 3$ is constant. Then w.h.p.*

$$C_{\mathbb{G}(n,r)} \sim \frac{r-1}{r-2} n\log n.$$

Proof. By putting $\varepsilon = 1/n$ in Theorem 24.4, we can take $T = 5\log n$ and $\omega = n$ in (24.5).

Let

$$\sigma = \lfloor \log\log\log n \rfloor .$$

For $v \in V$ and $k \geq 0$, let $N_k(v) = \{w : dist(v,w) = k\}$ be the set of vertices at distance k from v. Let $M_l(v) = \cup_{j=0}^{l} N_j(v)$, and let $G_l(v)$ be the subgraph of $G = \mathbb{G}(n,r)$ induced by $M_l(v)$.

We say that v is *locally tree-like* if $G_\sigma(v)$ is a tree. A cycle C is *small* if $|C| \leq \sigma$.

Lemma 24.9.

(a) There are at most $r^{2\sigma}$ small cycles.

(b) There do not exist 2 small cycles within distance σ of each other.

(c) There are at most $r^{3\sigma}$ non tree-like vertices.

(d) A non tree-like vertex v is within distance 2σ of a tree-like vertex.

Proof. (a) We use the configuration model of Chapter ??. F denotes a random configuration.

The expected number of small cycles in $\gamma(F)$ is bounded by

$$\sum_{k=3}^{\sigma}\binom{n}{k}\frac{(k-1)!}{2}\frac{(r(r-1))^k\varphi(rn-2k)}{\varphi(rn)} \leq \sum_{k=3}^{\sigma}\binom{n}{k}\frac{(k-1)!}{2}\left(\frac{r}{n}\right)^k \leq r^{\sigma},$$

and so (a) follows from the Markov inequality.

The expected number of pairs of small cycles within σ of each other is bounded by

$$\sum_{a=3}^{\sigma}\sum_{b=3}^{\sigma}\binom{n}{a}\binom{n}{b-1}\frac{(a-1)!}{2}\frac{(b-1)!}{2}\left(\frac{r}{n}\right)^{a+b}+$$
$$\sum_{a=3}^{\sigma}\sum_{b=3}^{\sigma}\sum_{c=1}^{\sigma}\binom{n}{a}\binom{n}{b}\binom{n}{c}\frac{(a-1)!}{2}\frac{(b-1)!}{2}ab\left(\frac{r}{n}\right)^{a+b+c+1}=o(1).$$

This verifies (b).

Part (c) follows from (a) and (b) and the fact that a non tree-like vertex is within σ of a small cycle. Part (d) follows from (b). □

We now estimate R_v for a locally tree-like vertex. (R_v is defined just prior to Lemma 24.6.)

Lemma 24.10. *If v is locally tree-like then*

(a) $R_v = \frac{r-1}{r-2} + o(\sigma^{-1})$.

(b) The conditions of Lemma 24.6 are satisfied.

Proof. Let T_r be the infinite r-regular tree, rooted at v. Let $\mathscr{X}$ be a random walk on T_r starting at v. Let ρ_i be the probability that $\mathscr{X}$ is at v at step i. Now we can project the walk $\mathscr{X}$ onto a walk $\mathscr{Y}$ on $\{0,1,2,\dots,\}$ where the particle moves right with probability $q = \frac{r-1}{r}$ and left with probability $p = \frac{1}{r}$, except of course at the origin, where it must move right. Let E_i be the expected number of visits to 0 for $\mathscr{Y}$ starting at i. Then

$$E_0 = 1 + E_1 = 1 + E_0 p/q.$$

This is because E_1 is E_0 times the expected number of visits to 0 between right moves from 1. Solving gives

$$\sum_{t=0}^{\infty} \rho_t = E_0 = \frac{r-1}{r-2}. \tag{24.15}$$

Note next that for $t \geq 0$ we have $\rho_{2t+1} = 0$ and we will argue that

$$\rho_{2t} \leq \binom{2t}{t} \frac{(r-1)^t}{r^{2t}} \leq \left(\frac{2(r-1)}{r^2}\right)^t \leq \left(\frac{4}{9}\right)^t. \tag{24.16}$$

Now,

$$r_t = \rho_t \text{ for } t \leq \sigma, \tag{24.17}$$

and from (30.1), for $\sigma \leq t \leq T$,

$$r_t \leq \pi_v + l_{\max}^t \leq \pi_v + \left(\frac{2\sqrt{2}+o(1)}{3}\right)^t. \tag{24.18}$$

So,

$$\sum_{t=1}^{T} r_t(1+l)^t \leq$$

$$\sum_{t=1}^{\sigma} \left(\frac{4(1+l)^2}{9}\right)^t + \sum_{t=\sigma+1}^{T} \left(\pi_v + \left(\frac{2\sqrt{2}+o(1)}{3}\right)^t\right)(1+l)^{2t} < 1, \tag{24.19}$$

if K is sufficiently large.

Equations (24.17), (24.18) and (24.19) complete the proof of the lemma, modulo the proof of (24.16).

Proof of (24.16)**:** First observe that the RHS of (24.16) is the probability that a walk $\mathscr{Y}_1$ is at the origin after $2i$ steps. Here $\mathscr{Y}_1$ is the walk on $\{0, \pm 1, \pm 2, \ldots, \}$ where the particle moves right with probability $q = \frac{r-1}{r}$ and left with probability $p = \frac{1}{r}$ i.e. there is no barrier at the origin. We can couple $\mathscr{Y}, \mathscr{Y}_1$ so that $\mathscr{Y}(t) \geq |\mathscr{Y}_1(t)|$. When $\mathscr{Y}_1(t) > 0$ we can move them in the same direction and when $\mathscr{Y}_1 < 0$ then we can move $\mathscr{Y}$ further from the origin whenever $\mathscr{Y}_1$ moves further from the origin. Finally, $\Pr(\mathscr{Y}_1(2t) = 0) = \Pr(Bin(2t, p) = t)$. □

Upper bound on cover time Let $t_0 = \frac{r-1}{r-2} n \log n$. We prove that for any vertex $u \in V$,

$$C_u \leq t_0 + o(t_0). \tag{24.20}$$

As in (24.9) we have

$$C_u \leq t + \sum_{s \geq t} \mathbb{E} U_s = t + \sum_{v \in V} \sum_{s > t} \Pr(A_s(v)). \tag{24.21}$$

Let V_1 be the set of locally tree-like vertices and let $V_2 = V - V_1$. For $v \in V_1$ we have

$$\begin{aligned}
\sum_{s \geq t_0} \Pr(A_s(v)) &\leq \sum_{t \geq t_0} \exp\left\{-\frac{(r-2)t}{(r-1)n}\left(1 + O\left(\frac{\log n}{n}\right)\right)\right\} \\
&\leq 2\exp\left\{-\frac{(r-2)t_0}{(r-1)n}\right\} \cdot \frac{1}{1 - \exp\left\{-\frac{(r-2)}{(r-1)n}\right\}} \\
&\leq 3\frac{r-1}{r-2}.
\end{aligned}$$

Furthermore, we see that in particular,

$$\Pr(A_{5n}(v)) \leq 2e^{-1}. \tag{24.22}$$

Suppose next that $v \in V_2$. Lemma 24.9(d) shows we can find $w \in V_1$ such that $\mathrm{dist}(\mathrm{v}, \mathrm{w}) \leq 2\sigma$. So from (24.22), with $\nu = 5n + 2\sigma$, we have

$$\Pr(A_\nu(v)) \leq 1 - (1 - 2e^{-1})r^{-2\sigma}$$

since if our walk visits w, it will with probability at least $r^{-2\sigma}$ visit v within the next 2σ steps. Thus if $\xi = (1 - 2e^{-1})r^{-2\sigma}$,

$$\sum_{s \geq t_0} \Pr(A_s(v)) \leq \sum_{s \geq t_0} (1 - \xi)^{\lfloor s/\nu \rfloor} \leq \sum_{s \geq t_0} (1 - \xi)^{s/(2\nu)} = \frac{(1 - \xi)^{t_0/(2\nu)}}{1 - (1 - \xi)^{1/(2\nu)}} \leq 3\nu\xi^{-1}.$$

Thus, for all $u \in V$,

$$C_u \leq t_0 + 3\frac{r-1}{r-2}|V_1| + 3|V_2|\nu\xi^{-1} = t_0 + O(r^{5\sigma}n) = t_0 + o(t_0),$$

as $\sigma = \lfloor \log\log\log n \rfloor$.

Lower bound on cover time For any vertex u, we can find a set of vertices S such that at time $t_1 = t_0(1 - \varepsilon)$, $\varepsilon \to 0$, the probability the set S is covered by the walk $\mathcal{W}_u$ tends to zero. Hence $T_G(u) > t_1$ w.h.p. which implies that $C_G \geq t_0 - o(t_0)$.

We construct S as follows. Let $S \subseteq V_1$ be some maximal set of locally tree-like vertices all of which are at least distance $2\sigma + 1$ apart. Thus $|S| \geq (n - r^{3\sigma})r^{-(2\sigma+1)}$.

Let $S(t)$ denote the subset of S which has not been visited by $\mathscr{W}_u$ after step t. Now, provided $t \geq T$

$$\mathbb{E}|S(t)| \geq (1-o(1)) \sum_{v \in S} \exp\left\{-\frac{t}{nR_v}\right\}.$$

Let u be a fixed vertex of S. Setting $t = t_1 = (1-\varepsilon)t_0$ where $\varepsilon = 2\sigma^{-1}$, we have

$$\mathbb{E}|S(t_1)| = (1+o(1))|S|e^{-(1-\varepsilon)t_0/(R_v n)} \geq n^{1/2\sigma}. \tag{24.23}$$

To finish we can follow the same second moment argument as for the lower bound in $\mathbb{G}_{n,p}$ to show that t_1 is a lower bound on C_u. □

24.3 Walker-Deletor

This is a game played by Walker and Deletor on a graph G with vertex set $[n]$ and minimum degree αn. In a move, Walker makes one step of a simple random walk $\mathscr{W}$ and Deletor deletes an edge of G that has not been visited by Walker. The game ends when it is not possible for Walker to move to an unvisited vertex. Let V^* denote the set of vertices visited by Walker before the end of the game. We prove the following theorem from Espig, Frieze and Pegden [341]:

Theorem 24.11.

(a) *If G has minimum co-degree at least αn for some absolute constant $\alpha > 0$ then under optimum play (by Deletor), Walker visits at least $n - c\log n$ vertices of G w.h.p., for a constant c depending on α. (Here the co-degree of a pair of vertices v, w is the size of their common neighborhood.)*

(b) *If G has minimum degree at least αn for some absolute constant $\alpha > 0$ then under optimum play (by Deletor), Walker visits at most $n - c\log n$ vertices w.h.p., for any constant $c < \alpha$.*

Proof.

Lower bound on $|V^*|$ Let $\beta = \alpha^3/36$ and consider the first $t_0 = 4\beta^{-1} n \log n$ moves. We will show that Walker will w.h.p. visit the required number of vertices within this time. Let G_t be the subgraph of G induced by the edges not acquired by Deletor after t moves. Let L_t be the set of vertices incident with more than $\alpha n/3$ deleted edges after the completion of t moves by Deletor. Clearly $|L_t| \leq C_0 \log n$, where $C_0 = 12(\alpha\beta)^{-1}$.

Let v_t denote the current vertex being visited by Walker. We let U_t denote the set of vertices that are not in L_t and are currently unvisited. Then

(a) If $v_t \notin L_t$ then the probability that Walker visits U_t within two steps is at least $\beta|U_t|/n$. To see this let $Z = |N(v_{t+1}) \cap U_t|$. Then $\mathbb{E}(Z) \geq \alpha|U_t|/3$. This is because v_t and any $w \in U_t$ have at least $\alpha n - 2\alpha n/3 = \alpha n/3$ common neighbors in G_t. Thus, if $\bar{Z} = |U_t| - Z$ then $\mathbb{E}(\bar{Z}) \leq (1-\alpha/3)|U_t|$. It follows from the Markov inequality that $\Pr(\bar{Z} \geq (1-\alpha^2/9)|U_t|) \leq \frac{1}{1+\alpha/3}$ and so $\Pr(Z \geq \alpha^2|U_t|/9) \geq \frac{\alpha}{3+\alpha}$. Finally observe that $\Pr(v_{t+2} \in U_t \mid Z) \geq (Z-1)/n$, where we have subtracted 1 to account for Deletor's next move.

(b) We divide our moves up into periods $A_1, B_1, A_2, B_2, \ldots$, where A_j is a sequence of moves taking place entirely outside L_t and B_j is a sequence of moves entirely within L_t. During a time period A_j, the probability this period ends is at most $\frac{3C_0 \log n}{2\alpha n}$. So the number of time periods is dominated by the binomial $Bin(4\beta^{-1} n \log n, 3C_0 \log n/(2\alpha n))$ and so with probability $1 - o(n^{-3})$ the number of periods is less than $10C_0(\alpha\beta)^{-1} \log^2 n$.

(c) We argue next that with probability $1 - o(n^{-3})$

$$\text{each } B_j \text{ takes up at most } O(\log^6 n) \text{ moves.} \tag{24.24}$$

Suppose that B_j begins with a move from $v \notin L_t$ to $w \in L_t$. Let $L^* = L_t \cup \{v\}$ and let H^* denote the subgraph induced by the edges contained in L^* that have not been acquired by Deletor. Walker's moves in period B_j constitute a random walk on (part) of the graph H^*. This is not quite a simple random walk, since H^* changes due to the fact that Deletor can delete some of the edges available to Walker. Nevertheless, Walker will always be in a component of H^* containing v_t. This is because Walker has arrived at the current vertex via a walk from v_t. Now consider running this walk for $C_1 \log^5 n$ steps, where C_1 is some sufficiently large constant. Observe that Deletor can claim at most $C_0^2 \log^2 n$ edges inside this component of H^*. Hence there will be an interval of length $C_2 \log^3 n$, $C_2 = C_1/C_0^2$ where Deletor does not claim any edge inside H^*. This means that in this interval we perform a simple random walk on a connected graph with at most $1 + C_0 \log n$ vertices. If we start this interval at a certain vertex x, then we are done if the random walk visits v. It follows from Theorem 30.9 that the expected time for the walk to visit v can be bounded by $C_0^3 \log^3 n$. So, if $C_2 > 2C_0^3$ then v will be visited with probability at least 1/2.

Suppose that time has increased from the time t when B_j began to t' when v is first re-visited. If $v \notin L_{t'}$ then B_j is complete. If however $v \in L_{t'}$ then we know that v is incident with at most $\alpha n/3 + C_1 \log^5 n$ Deletor edges. So the probability that Walker leaves $L_{t'}$ in her next step is at least

$$\frac{d_G(v) - (\alpha n/3 + C_1 \log^5 n)}{d_G(v)} \geq \frac{1}{2}. \tag{24.25}$$

So the probability that B_j ends after $C_1 \log^5 n$ steps is at least 1/4. Suppose on the other hand that B_j does not end and that we return to v for kth time where $k \leq 20 \log n$. The effect of this is to replace $C_1 \log^5 n$ in (24.25) by $kC_1 \log^5 n$. This does not however affect the final inequality. So if C_1 is sufficiently large, the probability that B_j does not end after $20C_1 \log^6 n$ steps is at most $(3/4)^{20 \log n} = o(n^{-4})$. Estimate (24.24) follows immediately.

(d) Combining the discussion in (b), (c) we see that w.h.p. $\left|\bigcup_j B_j\right| = O(\log^8 n)$, which is negligible compared with t_0; i.e., Walker spends almost all of her time outside L_{t_0}. Let $X_i, 1 \leq i \leq k = n - C_0 \log n$, be the time needed to add the ith vertex to the list of vertices visited by Walker. (Here we exclude any time spent in $\bigcup_j B_j$.) It follows from (a) that $X_i/2$ is dominated by a geometric random variable with probability of success $\frac{(n-i-C_0 \log n)\beta}{n}$. This is true regardless of $X_1, X_2, \ldots, X_{i-1}$. So $(X_1 + \cdots + X_k)/2$ is dominated by the sum of independent geometric random variables. Furthermore, $\mathbb{E}(X_1 + \cdots + X_k) \leq \frac{2}{\beta} n \log n$ and it is not difficult to show that $X_1 + \cdots + X_k \leq t_0$ w.h.p. Indeed, the variance of a geometric random variable with probability of success p is given by $1/p^2 - 1/p \leq 1/p^2$. So, by the Chebyshev inequality,

$$\Pr(X_1 + \cdots + X_k > t_0) \leq \frac{\beta^2}{4n^2 \log^2 n} \sum_{i=1}^{k} \frac{4n^2}{\beta^2 (n - i - C_0 \log n)^2} = O\left(\frac{1}{\log^2 n}\right).$$

This completes the proof of Theorem 24.11(a).

Upper bound on $|V^*|$ We assume that Walker chooses a vertex a_0 to start at and then Deletor chooses an edge to acquire.

Deletor's strategy will be to choose an arbitrary unvisited vertex v_1 and protect it by always on his turn taking the edge (v_1, w) where w is the current vertex being visited by Walker, if $(v_1, w) \in E(G)$. If Deletor has already acquired (v_1, w) or $(v_1, w) \notin E(G)$ then he will choose an unacquired edge incident with v_1. This continues until Deletor has acquired all of the edges incident with v_1. He then chooses v_2 and protects it. This continues until there are no unvisited vertices to protect.

After Deletor has protected $v_1, v_2, \ldots, v_{k-1}$ and while he is protecting v_k, Walker finds herself doing a random walk on a dense graph with $n - k$ vertices. Let the moves spent protecting v_k be denoted by *round* k. (More precisely, round k consists of the moves, after round $k-1$, until Deletor has acquired all edges incident with v_k.)

Fix $k = O(\log n)$ and let Z_k be the number of unvisited, unprotected vertices when Deletor begins protecting v_k. Because Deletor has taken the edges incident with $v_1, v_2, \ldots, v_{k-1}$, it will take at most $n-k$ more moves to protect v_k. If w is an unvisited, unprotected vertex at the start of the round, then it remains unvisited with probability at least $\left(1 - \frac{1}{\alpha n - k}\right)^{n-k-1} = e^{-1/\alpha} + O(1/(n-k))$. It follows that $\mathbb{E}(Z_{k+1}) \sim Z_k/e^{1/\alpha}$. To show that it is close to this w.h.p. we proceed as follows: Suppose that we throw $n-k$ balls randomly into $\alpha n - k$ boxes, of which Z_k are special. Then Z_{k+1} dominates the number of empty special boxes. We can use Talagrand's inequality, Theorem 20.3 to see that if $Z_k \gg \log n$ then Z_{k+1} will be concentrated around its mean.

It follows that w.h.p. $Z_k \sim ne^{-k/\alpha}$ for $k \leq (1-\varepsilon)\alpha \log n$ where $0 < \varepsilon < 1$ is a positive constant. Thus Deletor will w.h.p. be able to protect $(1-\varepsilon)\alpha \log n$ vertices and we can choose any $c < \alpha$ in Theorem 24.11(b).

□

24.4 Exercises

24.4.1 Verify equation (24.13).

24.4.2 The vacant set V_t of a random walk on a graph G at time t is the set of vertices so far unvisited. We let $\mathbb{G}_t$ denote the subgraph induced by V_t.Suppose that $G = \mathbb{G}_{n,p}$ where $np = c\log n$ where $c > 1$ is constant. Let $t_\theta = n(\log\log n + (1+\theta)c)$. Show that if $t \leq t_{-\varepsilon}$ then w.h.p. the size of the largest component in $\mathbb{G}_t$ is of order $|V_t|$ and that if $t \geq t^{\varepsilon}$ then this drops to $O(\log n)$ w.h.p.

24.4.3 The trace of a random walk at time t is the graph/digraph induced by the the t edges crossed. Show that if $t \geq Kn\log n$ then w.h.p. the trace of a random walk on K_n is Hamiltonian.

24.5 Notes

Mixing time Fountoulakis and Reed [375] and Benjamini, Kozma and Wormald [95] proved that the mixing time of a random walk on the giant component $\mathbb{G}_{n,m}, p =$

$c/n, c > 1$ is $O(\log^2 n)$ w.h.p. Whereas Nachmias and Peres [704] show that in the case $c \sim 1$, the mixing time is in $[A^{-1}n, An]$ with probability at least $1-\varepsilon$, for $A = A(\varepsilon)$. Ding, Lubetzky and Peres [297] show that mixing time for the emerging giant at $p = (1+\varepsilon)/n$ where $\lambda = \varepsilon^3 n \to \infty$ is of order $(n/\lambda)(\log \lambda)^2$.

For random regular graphs, we have shown that the mixing time is $O(\log n)$. Lubetzky and Sly [630] proved that the mixing time exhibits a *cut-off* phenomenon i.e. the variation distance goes from near one to near zero very rapidly.

Boyd, Ghosh, Prabhakar and Shah [191] considered random walk on a random geometric graph. They show that w.h.p. the mixing time is $\Theta(r^{-2})$ where r is maximum distance between adjacent points.

Cover time Cooper and Frieze used the First Visit Time Lemma to establish asymptotically correct estimates for the cover time of the giant component of $\mathbb{G}_{n,p}, p = c/n, c > 1$ [259], $C_G \sim \frac{cx(2-x)}{4(cx-\log c)} n \log^2 n$ where $xe^{-x} = ce^{-c}$; the preferential attachment graph [258], $C_G \sim \frac{2m}{m-1} n \log n$; the random geometric graph in $d \geq 3$ dimensions [260], $C_G \sim \psi(c) n \log n$ assuming that the distance r between adjacent vertices is $(c \log n)/(\Psi_d n)^{1/2}$ where Ψ_d is the volume of the ball of radius 1 in $\mathbb{R}^d$ (the case of $d = 2$ is less well understood, although Avin and Ercal [56] prove that w.h.p. it is $\Theta(n \log n)$).; the random digraph $D_{n,p}, p = c \log n/n$ [261], $C_D \sim \psi(c) n \log n$; random graphs with a fixed degree sequence [263]. Cooper, Frieze and Radzik considered random walks on uniform hypergraphs. Frieze, Pegden and Tkocz [412] analysed the cover time of the emerging giant in $\mathbb{G}_{n,p}, p = (1+\varepsilon)/n, \varepsilon^3 n \to \infty$. They show that w.h.p. the cover time $\sim n \log^2(\varepsilon^3 n)$. This improved the results of Barlow, Ding, Nachmias and Peres [80]. Martinez and Mitsche [654] prove a lower bound of $\Omega(n \log^2 n)$ for the cover time of the largest component when r is below the threshold for connectivity.

Trace The trace of a random walk at time t is the graph/digraph induced by the t edges crossed. Frieze, Krivelevich, Michaeli and Peled [400] considered the trace of random walks on $G_{n,p}$ and K_n. For $Gn, p, p = C \log n/n, C = C(\varepsilon)$ the trace at time $(1+\varepsilon) n \log n$ is Hamiltonian w.h.p. and $\beta \log n$ connected for $\beta(\varepsilon) > 0$. In the case of K_n they prove hitting time results for Hamiltonicity and connectivity.

Krivelevich and Michaeli [599] discuss the threshold for the appearance of a fixed small subgraph H in the trace. In the case of K_n the results are analogous to those for $\mathbb{G}_{n,m}$, see Chapter 3. They also consider $\mathbb{G}_{n,p}$.

Vacant set The vacant set V_t of a random walk at time t is the set of vertices so far unvisited. We let $\mathbb{G}_t$ denote the subgraph induced by V_t. Černy, Teixeira and Windisch [210] studied the vacant set of a random walk on $\mathbb{G}(n,r)$. Let

$t^* = \frac{r(r-1)\log(r-1)}{(r-2)^2}n$. They showed that w.h.p. if $t \geq (1+\varepsilon)t^*$ then the maximum component size in $\mathbb{G}_t$ is $O(\log n)$ and conjectured that there is a phase transition at t^*. This was confirmed by Cooper and Frieze [262] who gave an asymptotic formula for the size of the largest component. Černy and Teixeira [210] show that in a critical window around t^* the size of the largest component is of order $n^{2/3}$. Cooper and Frieze also considered the vacant set for random walks on $\mathbb{G}_{n,p}$, see Exercise 24.4.1.

Chapter 25

Brief notes on uncovered topics

There are several topics that we have not been able to cover and that might be of interest to the reader. For these topics, we provide some short synopses and some references that the reader may find useful.

Edge Colored Random Graphs

Suppose that we color the edges of a graph G. A set of edges S is said to be *rainbow* colored if each edge of S has a different color. Consider first the existence of a rainbow spanning tree. We consider the graph process where the edges are randomly colored using $q \geq n-1$ colors. Let τ_a be the hitting time for $n-1$ colors to appear in the process and let τ_c be the hitting time for connectivity and let $\tau^* = \max\{\tau_a, \tau_c\}$. Frieze and McKay [407] showed that w.h.p. G_{τ^*} contains a rainbow spanning tree. This is clearly best possible. Bal, Bennett, Frieze and Pralat [61] consider the case where each edge has a choice of k random colors. This reduces τ_a, but the result still holds.

The existence of rainbow Hamilton cycles is different. The existence of a rainbow spanning tree can be checked in polynomial time and this leads to a simple criterion for non-existence. This is clearly not likely for Hamilton cycles. Cooper and Frieze [253] proved that if $m \geq Kn\log n$ and $q \geq Kn$ then w.h.p. $\mathbb{G}_{n,m}$ contains a rainbow Hamilton cycle. This was improved to $m \geq \frac{1+o(1)}{2} n \log n$ and $q \geq (1+o(1))n$ by Frieze and Loh [402]. Bal and Frieze [63] show that if $m \geq Kn\log n$ and $q = n$ and n is even there is a rainbow Hamilton cycle w.h.p. Ferber [352] removed the requirement that n be even. Bal and Frieze also considered rainbow perfect matchings in k-uniform hypergraphs. Janson and Wormald [512] considered random coloring's of r-regular graphs. They proved that if $r \geq 4, r = O(1)$ and the edges of $G_{n,2r}$ are randomly colored so that each color is used r times, then w.h.p. there is a rainbow Hamilton cycle. Ferber, Kronenberg, Mousset and Shikhelman [355] give results on packing rainbow structures such as Hamilton cycles. Ferber, Nenadov and Peter prove that if $p \gg n^{-1/d}(\log n)^{1/d}$ and H is a fixed graph of density at most d then w.h.p. $\mathbb{G}_{n,p}$ contains a rainbow copy of H if it is randomly colored with $(1+\varepsilon)|E(H)|$ colors, for any fixed $\varepsilon > 0$.

Cooper and Frieze [250] found the threshold for the following property: If $k = O(1)$ and $\mathbb{G}_{n,m}$ is arbitrarily edge colored so that no color is used more than k times, then $\mathbb{G}_{n,m}$ contains a rainbow Hamilton cycle.

Games

Positional games can be considered to be a generalisation of the game of "Noughts and Crosses" or "Tic-Tac-Toe". There are two players A (Maker) and B (Breaker) and in the context for this section, the board will be a graph G. Each player in turn chooses an edge and at the end of the game, the winner is determined by the partition of the edges claimed by the players. As a typical example, in the connectivity game, player A is trying to ensure that the edges she collects contain a spanning tree of G and player B is trying to prevent this. See Chvátal and Erdős [230] for one of the earliest papers on the subject and books by Beck [83] and Hefetz, Krivelevich, Stojaković and Szabó [476]. Most of the analyses have considered $G = K_n$ and to make the problem interesting [230] introduced the notion of *bias*. Thus in the connectivity game, player B is allowed to collect b edges for each edge of A. Now the question becomes what is the largest value of b for which A has a winning strategy. There is a striking though somewhat mysterious connection between the optimal values of b for various games and thresholds for associated properties in random graphs. For example in the connectivity game, the threshold bias $b \approx \frac{n}{\log n}$ i.e. player A collects about $\frac{1}{2}n \log n$ edges, see Gebauer and Szabó [427]. Another example is the biased H-game where Maker wins if she can create a copy of some fixed graph H with at least two adjacent edges. The optimal threshold bias b for this game is of order $\Theta\left(n^{1/m_2(H)}\right)$, Bednarska and Łuczak [85]. For sufficiently small constant $c > 0$, if $b \leq cn^{1/m_2(H)}$, then Maker can create $\Theta(EX_H)$ copies of H in K_n, where X_H is the number of copies of H in $G_{n,1/b}$. Furthermore, if Maker plays randomly, she achieves this goal w.h.p.

Recently Stojaković and Szabó [820] began research on random boards i.e. where G is a random graph. Ben-Shimon, Ferber, Hefetz and Krivelevich [99] prove a hitting time result for the $b = 1$ Hamilton cycle game on the graph process. Assuming that player A wants to build a Hamilton cycle and player B starts first, player A will have a winning strategy in G_m iff $m \geq m_4^*$. This is best possible. Biased Hamiltonicity games on $\mathbb{G}_{n,p}$ were considered in Ferber, Glebov, Krivelevich and Naor [353] where it was shown that for $p \gg \frac{\log n}{n}$, the threshold bias b_{HAM} satisfies $b_{HAM} \approx \frac{np}{\log n}$ w.h.p. The H-game where A wins if she can create a copy of some fixed graph H was first studied by Stojakovic and Szabo [820] in the case of H is a clique on k vertices. This was strengthened by Müller and Stojaković [701]. They show that if $p \leq cn^{-2/(k+1)}$, then w.h.p. B can win this game. For $p \geq Cn^{-2/(k+1)}$ one can use the results of [772] to argue that A wins w.h.p. This result was generalised to arbitrary graphs H (satisfying certain mild conditions) by Nenadov, Steger and Stojaković [713] where they showed that the threshold is where one would expect it to be - at the 2-density of H. As we have seen there are other models of random graphs and Beveridge, Dudek, Frieze, Müller and Stojaković [106] studied these games on random geometric graphs.

The *game chromatic number* $\chi_g(G)$ of a graph G can be defined as follows. Once again there are two players A,B and they take it in turns to properly color vertices of G with one of q colors. Thus if $\{u,v\}$ is an edge and u is colored with color c and v is uncolored at the start of any turn, then v may not be colored with c by either player. The goal of A is to ensure that the game ends with every vertex colored and the goal of B is to prevent this by using all q colors in the neighborhood of some uncolored vertex. The game chromatic number is the minimum q for which A can win. For a survey on results on this parameter see Bartnicki, Grytczuk, Kierstead and and Zhu [81]. Bohman, Frieze and Sudakov [142] studied χ_g for dense random graphs and proved that for such graphs, χ_g is within a constant factor of the chromatic number. Keusch and Steger [564] proved that this factor is asymptotically equal to two. Frieze, Haber and Lavrov [395] extended the results of [142] to sparse random graphs.

Graph Searching

Cops and Robbers

A collection of cops are placed on the vertices of a graph by player C and then a robber is placed on a vertex by player R. The players take turns. C can move all cops to a neighboring vertex and R can move the robber. The cop number of a graph is them minimum number of cops needed so that C can win. The basic rule being that if there is a cop occupying the same vertex as the robber, then C wins. Łuczak and Pralat [643] proved a remarkable “zigzag” theorem giving the cop number of a random graph. This number being n^{α} where $\alpha = \alpha(p)$ follows a saw-toothed curve. Pralat and Wormald [748] proved that the cop number of the random regular graph $\mathbb{G}_{n,r}$ is $O(n^{1/2})$. It is worth noting that Meyniel has conjectured $O(n^{1/2})$ as a bound on the cop number of any connected n-vertex graph. There are many variations on this game and the reader is referred to the monograph by Bonato and Pralat [179].

Graph Cleaning

Initially, every edge and vertex of a graph G is dirty, and a fixed number of brushes start on a set of vertices. At each time-step, a vertex v and all its incident edges that are dirty may be cleaned if there are at least as many brushes on v as there are incident dirty edges. When a vertex is cleaned, every incident dirty edge is traversed (that is, cleaned) by one and only one brush, and brushes cannot traverse a clean edge. The *brush number* $b(G)$ is the minimum number of brushes needed to clean G. Pralat [749], [750] proved that w.h.p. $b(\mathbb{G}_{n,p}) \approx \frac{1-e^{-2d}}{4}n$ for $p = \frac{d}{n}$ where $d < 1$ and w.h.p. $b(\mathbb{G}_{n,p}) \leq (1+o(1))\left(d+1-\frac{1-e^{-2d}}{2d}\right)\frac{n}{4}$ for $d > 1$. For the random d-regular graph $\mathbb{G}_{n,d}$, Alon, Pralat and Wormald [30] proved that w.h.p. $b(\mathbb{G}_{n,d}) \geq \frac{dn}{4}\left(1-\frac{2^{3/2}}{d^{1/2}}\right)$.

Acquaintance Time

Let $G = (V,E)$ be a finite connected graph. We start the process by placing one agent on each vertex of G. Every pair of agents sharing an edge are declared to be acquainted, and remain so throughout the process. In each round of the process, we choose some matching M in G. The matching M need not be maximal; perhaps it is a single edge. For each edge of M, we swap the agents occupying its endpoints, which may cause more agents to become acquainted. We may view the process as a graph searching game with one player, where the player's strategy consists of a sequence of matchings which allow all agents to become acquainted. Some strategies may be better than others, which leads to a graph optimisation parameter. The acquaintance time of G, denoted by $\mathscr{A}(G)$, is the minimum number of rounds required for all agents to become acquainted with one another. The parameter $\mathscr{A}(G)$ was introduced by Benjamini, Shinkar and Tsur [90], who showed that $\mathscr{A}(G) = O\left(\frac{n^2\log\log n}{\log n}\right)$ for an n vertex graph. The $\log\log n$ factor was removed by Kinnersley, Mitsche and Pralat [569]. The paper [569] also showed that w.h.p. $\mathscr{A}(\mathbb{G}_{n,p}) = O\left(\frac{\log n}{p}\right)$ for $\frac{(1+\varepsilon)\log n}{n} \leq p \leq 1-\varepsilon$. The lower bound here was relaxed to $np - \log n \to \infty$ in Dudek and Pralat [313]. A lower bound, $\Omega\left(\frac{\log n}{p}\right)$ for $\mathbb{G}_{n,p}$ and $p \geq n^{-1/2+\varepsilon}$ was proved in [569].

H-free process

In an early attempt to estimate the Ramsey number $R(3,t)$, Erdős, Suen and Winkler [339] considered the following process for generating a triangle free graph. Let $e_1, e_2, \ldots, e_N, N = \binom{n}{2}$ be a random ordering of the complete graph K_n. Let

$\mathscr{P}$ be a graph property e.g. being triangle free. We generate a sequence of random graphs $\Gamma_0, \Gamma_1, \ldots, \Gamma_N$ where $\Gamma_{i+1} = \Gamma_i + e_{i+1}$ if adding e_{i+1} does not destroy $\mathscr{P}$, otherwise $\Gamma_{i+1} = \Gamma_i$. In this way we can generate a random graph that is guaranteed to have property $\mathscr{P}$.

For $\mathscr{P}$ is "bipartite" they show in [339] that Γ_N has expected size greater than $(n^2 - n)/4$. When $\mathscr{P}$ is "triangle free" they show that w.h.p. that Γ_N has size $\Omega(n^{3/2})$ w.h.p. Bollobás and Riordan [175] studied the general H-free process. More recently, Bohman [132] showed in the case of the triangle free process, that w.h.p. Γ_N has size $\Theta(n^{3/2}(\log n)^{1/2})$. This provides an alternative proof to that of Kim [565] that $R(3,t) = \Omega\left(\frac{t^2}{\log t}\right)$. He made use of a careful use of the differential equations method, see Chapter 28. Bohman and Keevash [143] and Fiz Pontiveros, Griffiths and Morris [362] have improved this result and shown that w.h.p. Γ_N has size asymptotically equal to $\frac{1}{2\sqrt{2}} n^{3/2} (\log n)^{1/2}$. They also show that the independence number of Γ_N is bounded by $(1 + o(1))(2n \log n)^{1/2}$. This shows that $R(3,t) > \left(\frac{1}{4} - o(1)\right) t^2 / \log t$.

Bohman, Mubayi and Picolleli [145] considered an r-uniform hypergraph version. In particular they studied the $T^{(r)}$-free process, where $T^{(r)}$ generalises a triangle in a graph. It consists of $S \cup \{a_i\}, i = 1, 2, \ldots, r$ where $|S| = r - 1$ and a further edge $\{a_1, a_2, \ldots, a_r\}$. Here hyperedges are randomly added one by one until one is forced to create a copy of T^r. They show that w.h.p. the final hypergraph produced has independence number $O((n \log n)^{1/r})$. This proves a lower bound of $\Omega\left(\frac{s^r}{\log s}\right)$ for the Ramsey number $R(T^{(r)}, K_s^{(r)})$. The analysis is based on a paper on the random greedy hypergraph independent set process by Bennett and Bohman [96].

There has also been work on the related triangle removal process. Here we start with K_n and repeatedly remove a random triangle until the graph is triangle free. The main question is as to how many edges are there in the final triangle free graph. A proof of a bound of $O(n^{7/4+o(1)})$ was outlined by Grable [448]. A simple proof of $O(n^{7/4+o(1)})$ was proved in Bohman, Frieze and Lubetzky [139]. Furthermore, Bohman, Frieze and Lubetzky [140] have proved a tight result of $n^{3/2+o(1)}$ for the number of edges left. This is close to the $\Theta(n^{3/2})$ bound conjectured by Bollobás and Erdős in 1990.

An earlier paper by Ruciński and Wormald [779] consider the d-process. Edges were now rejected if they raised the degree of some vertex above d. Answering a question of Erdős, they proved that the resulting graph was w.h.p. d-regular.

Logic and Random Graphs

The *first order theory of graphs* is a language in which one can describe some, but certainly not all, properties of graphs. It can describe G has a triangle, but not G is connected. Fagin [346] and Glebskii, Kogan, Liagonkii and Talanov [437] proved that for any property $\mathscr{A}$ that can be described by a first order sentence, $\lim_{n\to\infty}\mathbb{P}(\mathbb{G}_{n,1/2}\in\mathscr{A})\in\{0,1\}$. We say that $p=1/2$ obeys a 0-1 law. One does not need to restrict oneself to $\mathbb{G}_{n,1/2}$. Shelah and Spencer [802] proved that if α is irrational then $p=n^{-\alpha}$ also obeys a 0-1 law, while if α is rational, then there are first order sentences such that $\lim_{n\to\infty}\mathbb{P}(\mathbb{G}_{n,1/2}\in\mathscr{A})$ does not exist. See the book by Spencer [812] for much more on this subject.

Planarity

We have said very little about random planar graphs. This is partially because there is no simple way of generating a random planar graph. The study begins with the seminal work of Tutte [836], [837] on counting planar maps. The number of rooted maps on surfaces was found by Bender and Canfield [93]. The size of the largest components were studied by Banderier, Flajolet, Schaeffer and Soria [71].

When it comes to random labeled planar graphs, McDiarmid, Steger and Welsh [668] showed that if $pl(n)$ denotes the number of labeled planar graphs with n vertices, then $(pl(n)/n!)^{1/n}$ tends to a limit γ as $n\to\infty$. Osthus, Prömel and Taraz [720] found an upper bound for γ, Bender, Gao and Wormald [94] found a lower bound for γ. Finally, Giménez and Noy [434] proved that $pl(n)\approx cn^{-7/2}\gamma^n n!$ for explicit values of c,γ.

Next let $pl(n,m)$ denote the number of labelled planar graphs with n vertices and m edges. Gerke, Schlatter, Steger and Taraz [430] proved that if $0\leq a\leq 3$ then $(pl(n,an)/n!)^{1/n}$ tends to a limit γ_a as $n\to\infty$. Giménez and Noy [434] showed that if $1<a<3$ then $pl(n,an)\approx c_a n^{-4}\gamma_a^n n!$. Kang and Łuczak [544] proved the existence of two critical ranges for the sizes of complex components.

Planted Cliques, Cuts and Hamilton cycles

The question here is the following: Suppose that we plant an unusual object into a random graph. Can someone else find it? One motivation being that if finding the planted object is hard for someone who does not know where it is planted, then this modified graph can be used as a signature. To make this more precise, consider starting with $\mathbb{G}_{n,1/2}$, choosing an s-subset S of $[n]$ and then making S into a clique. Let the modified graph be denoted by Γ. Here we assume that $s \gg \log n$ so that S should stand out. Can we find S, if we are given Γ, but we are not told S. Kucera [612] proved that if $s \geq C(n \log n)^{1/2}$ for a sufficiently large C then w.h.p. one can find S by looking at vertex degrees. Alon, Krivelevich and Sudakov [34] improved this to $s = \Omega(n^{1/2})$. They show that the second eigenvector of the adjacency matrix of Γ contains enough information so that w.h.p. S can be found. Frieze and Kannan [398] related this to a problem involving optimisation of a tensor product. Recently, Feldman, Grigorescu, Reyzin, Vempala and Xiao [349] showed that a large class of algorithms will fail w.h.p. if $s \leq n^{1/2-\delta}$ for some positive constant δ.

There has also been a considerable amount of research on planted cuts. Beginning with the paper of Bui, Chaudhuri, Leighton and Sipser [202] there have been many papers that deal with the problem of finding a cut in a random graph of unusual size. By this we mean that starting with $\mathbb{G}_{n,p}$, someone selects a partition of the vertex set into $k \geq 2$ sets of large size and then alters the edges between the subsets of the partition so that it is larger or smaller than can be usually found in $\mathbb{G}_{n,p}$. See Coja–Oghlan [231] for a recent paper with many pertinent references.

As a final note on this subject of planted objects. Suppose that we start with a Hamilton cycle C and then add a copy of $\mathbb{G}_{n,p}$ where $p = \frac{c}{n}$ to create Γ. Broder, Frieze and Shamir [198] showed that if c is sufficiently large then w.h.p. one can in polynomial time find a Hamilton cycle H in Γ. While H may not necessarily be C, this rules out a simple use of Hamilton cycles for a signature scheme.

Random Lifts

For a graph K, an *n-lift* G of K has vertex set $V(K) \times [n]$ where for each vertex $v \in V(K)$, $\{v\} \times [n]$ is called the *fiber* above v and will be denoted by Π_v. The edge set of a an n-lift G consists of a perfect matching between fibers Π_u and Π_w for each edge $\{u,w\} \in E(K)$. The set of n-lifts will be denoted $\Lambda_n(K)$. In a random n-lift, the matchings between fibers are chosen independently and uniformly at random.

Lifts of graphs were introduced by Amit and Linial in [39] where they proved that if K is a connected, simple graph with minimum degree $\delta \geq 3$, and G is a random n-lift of K then G is $\delta(G)$-connected w.h.p., where the asymptotics are for $n \to \infty$. They continued the study of random lifts in [40] where they proved expansion properties of lifts. Together with Matoušek, they gave bounds on the independence number and chromatic number of random lifts in [41]. Linial and Rozenman [627] give a tight analysis for when a random n-lift has a perfect matching. Greenhill, Janson and Ruciński [450] consider the number of perfect matchings in a random lift.

Łuczak, Witkowski and Witkowski [646] proved that a random lift of H is Hamiltonian w.h.p. if H has minimum degree at least 5 and contains two disjoint Hamiltonian cycles whose union is not a bipartite graph. Chebolu and Frieze [214] considered a directed version of lifts and showed that a random lift of the complete digraph $\vec{K}_h$ is Hamiltonian w.h.p. provided h is sufficiently large.

Random Simplicial Complexes

Linial and Meshulam [626] pioneered the extension of the analysis of $\mathbb{G}_{n,p}$ to higher dimensional complexes. We are at the beginning of research in this area and can look forward to exciting connections with Algebraic Topology. For more details see the survey of Kahle [539].

Stable Matching

In the stable matching problem we have a complete bipartite graph on vertex sets A,B where $A = \{a_1,a_2,\ldots,a_n\}, B = \{b_1,b_2,\ldots,b_n\}$. If we think of A as a set of women and B as a set of men, then we refer to this as the *stable marriage* problem. Each $a \in A$ has a total ordering p_a of B and each $b \in B$ has a total ordering p_b of B. The problem is to find a perfect matching $(a_i,b_i), i = 1,2,\ldots,n$ such that there *does not exist* a pair i,j such that $b_j > b_i$ in the order p_{a_i} and $a_i > b_j$ in the order p_{b_j}. The existence of i,j leads to an unstable matching. Gale and Shapley [422] proved that there is always a stable matching and gave an algorithm for finding one. We focus on the case where p_a, p_b are uniformly random for all $a \in A, b \in B$. Wilson [855] showed that the expected number of proposals in a sequential version of the Gale-Shapley algorithm is asymptotically equal to $n\log n$. Knuth, Motwani and Pittel [573] studied the likely number of stable husbands for an element of $A \cup B$. I.e. they show that w.h.p. there are constants $c < C$ such that for a fixed $a \in A$ there are between $c\log n$ and $C\log n$ choices $b \in B$ such that a and b are matched together in some stable matching. The question of how many distinct stable matchings there are likely to be was raised in Pittel [739] who showed that w.h.p. there are at least $n^{1/2-o(1)}$. More recently, Lennon and Pittel [620] show that there are at least $n\log n$ with probability at least 0.45. Thus the precise growth rate of the number of stable matchings is not clear at the moment. Pittel, Shepp and Veklerov [743] considered the number $Z_{n,m}$ of $a \in A$ that have exactly m choices of stable husband. They show that $\lim_{n\to\infty} \frac{\mathbb{E}(Z_{n,m})}{(\log n)^{m+1}} = \frac{1}{(m-1)!}$.

Universal graphs

A graph G is *universal* for a class of graphs $\mathscr{H}$ if G contains a copy of every $H \in \mathscr{H}$. In particular, let $\mathscr{H}(n,d)$ denote the set of graphs with vertex set $[n]$ and maximum degree at most d. One question that has concerned researchers, is to find the threshold for $\mathbb{G}_{n,p}$ being universal for $\mathscr{H}(n,d)$. A counting argument shows that any $\mathscr{H}(n,d)$ universal graph has $\Omega(n^{2-2/d})$ edges. For random graphs this can be improved to $\Omega(n^{2-2/(d+1)}(\log n)^{O(1)})$. This is because to contain the union of $\left\lfloor \frac{n}{d+1} \right\rfloor$ disjoint copies of K_{d+1}, all but at most d vertices must lie in a copy of K_{d+1}. This problem was first considered in Alon, Capalbo, Kohayakawa, Rödl, Ruciński and Szemerédi [29]. Currently the best upper bound on the value of p needed to make $\mathbb{G}_{n,m}$ $\mathscr{H}(n,d)$ universal is $O(n^{2-1/d}(\log n)^{1/d})$ in Dellamonica, Kohayakawa, Rödl, and Ruciński [281]. Ferber, Nenadov and Peter [356] prove that if $p \gg \Delta^8 n^{-1/2} \log n$ then $\mathbb{G}_{n,p}$ is universal for the set of trees with maximum degree Δ.

Part V

Tools and Methods

Chapter 26

Moments

26.1 First and Second Moment Method

Lemma 26.1 (The Markov Inequality)**.** *Let X be a non-negative random variable. Then, for all $t > 0$,*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}.$$

Proof. Let

$$I_A = \begin{cases} 1 & \text{if event } A \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Notice that

$$X = XI_{\{X \geq t\}} + XI_{\{X < t\}} \geq XI_{\{X \geq t\}} \geq tI_{\{X \geq t\}}.$$

Hence,

$$\mathbb{E}X \geq t\mathbb{E}I_{\{X \geq t\}} = t\,\mathbb{P}(X \geq t).$$

□

As an immediate corollary, we obtain

Lemma 26.2 (First Moment Method)**.** *Let X be a non-negative integer valued random variable. Then*

$$\mathbb{P}(X > 0) \leq \mathbb{E}X.$$

Proof. Put $t = 1$ in the Markov inequality. □

The following inequality is a simple consequence of Lemma 26.1.

Lemma 26.3 (Chebyshev Inequality)**.** *If X is a random variable with a finite mean and variance, then, for $t > 0$,*

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \frac{\operatorname{Var} X}{t^2}.$$

Proof.

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) = \mathbb{P}((X - \mathbb{E}X)^2 \geq t^2) \leq \frac{\mathbb{E}(X - \mathbb{E}X)^2}{t^2} = \frac{\operatorname{Var} X}{t^2}.$$

□

Throughout the book the following consequence of the Chebyshev inequality plays a particularly important role.

Lemma 26.4 (Second Moment Method)**.** *If X is a non-negative integer valued random variable then*

$$\mathbb{P}(X = 0) \leq \frac{\operatorname{Var} X}{(\mathbb{E}X)^2} = \frac{\mathbb{E}X^2}{(\mathbb{E}X)^2} - 1$$

Proof. Set $t = \mathbb{E}X$ in the Chebyshev inequality. Then

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}X| \geq \mathbb{E}X) \leq \frac{\operatorname{Var} X}{(\mathbb{E}X)^2}$$

□

Lemma 26.5 ((Strong) Second Moment Method)**.** *If X is a non-negative integer valued random variable then*

$$\mathbb{P}(X = 0) \leq \frac{\operatorname{Var} X}{\mathbb{E}X^2} = 1 - \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}.$$

Proof. Notice that

$$X = X \cdot I_{\{X \geq 1\}}.$$

Then, by the Cauchy-Schwarz inequality,

$$(\mathbb{E}X)^2 = \left(\mathbb{E}(X \cdot I_{\{X \geq 1\}})\right)^2 \leq \mathbb{E}I_{\{X \geq 1\}}^2 \, \mathbb{E}X^2 = \mathbb{P}(X \geq 1)\, \mathbb{E}X^2.$$

□

The bound in Lemma 26.5 is stronger than the bound in Lemma 26.4, since $\mathbb{E}X^2 \geq (\mathbb{E}X)^2$. However, for many applications, these bounds are equally useful since the Second Moment Method can be applied if

$$\frac{\operatorname{Var}X}{(\mathbb{E}X)^2} \to 0, \tag{26.1}$$

or, equivalently,

$$\frac{\mathbb{E}X^2}{(\mathbb{E}X)^2} \to 1, \tag{26.2}$$

as $n \to \infty$. In fact if (26.1) holds, then much more than $\mathbb{P}(X > 0) \to 1$ is true. Note that

$$\begin{aligned}
\frac{\operatorname{Var}X}{(\mathbb{E}X)^2} &= \operatorname{Var}\left(\frac{X}{\mathbb{E}X}\right) = \mathbb{E}\left(\frac{X}{\mathbb{E}X}\right)^2 - \left(\mathbb{E}\left(\frac{X}{\mathbb{E}X}\right)\right)^2 \\
&= \mathbb{E}\left(\frac{X}{\mathbb{E}X} - 1\right)^2
\end{aligned}$$

Hence

$$\mathbb{E}\left(\frac{X}{\mathbb{E}X} - 1\right)^2 \to 0 \text{ if } \frac{\operatorname{Var}X}{(\mathbb{E}X)^2} \to 0.$$

It simply means that

$$\frac{X}{\mathbb{E}X} \xrightarrow{L^2} 1. \tag{26.3}$$

In particular, it implies (as does the Chebyshev inequality) that

$$\frac{X}{\mathbb{E}X} \xrightarrow{\mathbb{P}} 1, \tag{26.4}$$

i.e., for every $\varepsilon > 0$,

$$\mathbb{P}((1-\varepsilon)\mathbb{E}X < X < (1+\varepsilon)\mathbb{E}X) \to 1. \tag{26.5}$$

So, we can only apply the Second Moment Method, if the random variable X has its distribution asymptotically concentrated at a single value (X can be approximated by the non-random value $\mathbb{E}X$, as stated at (26.3), (26.4) and (26.5)).

We complete this section with another lower bound on the probability $\mathbb{P}(X_n \geq 1)$, when X_n is a sum of (asymptotically) negatively correlated indicators. Notice that in this case we do not need to compute the second moment of X_n.

Lemma 26.6. *Let $X_n = I_1 + I_2 + \cdots + I_n$, where $\{I_i\}_{i=1}^n$ be a collection of $0-1$ random variables, such that*

$$\mathbb{P}(I_i = I_j = 1) \leq (1+\varepsilon_n)\,\mathbb{P}(I_i = 1)\,\mathbb{P}(I_j = 1)$$

for $i \neq j = 1, 2, \ldots, n$. Here $\varepsilon_n \to 0$ as $n \to \infty$. Then

$$\mathbb{P}(X_n \geq 1) \geq \frac{1}{1+\varepsilon_n + 1/\mathbb{E}X_n}.$$

Proof. By the (strong) second moment method (see Lemma 26.5)

$$\mathbb{P}(X_n \geq 1) \geq \frac{(\mathbb{E}X_n)^2}{\mathbb{E}X_n^2}.$$

Now

$$\begin{aligned}
\mathbb{E}X_n^2 &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(I_i I_j) \\
&\leq \mathbb{E}X_n + (1+\varepsilon_n) \sum_{i \neq j} \mathbb{E}I_i \mathbb{E}I_j \\
&= \mathbb{E}X_n + (1+\varepsilon_n) \left(\left(\sum_{i=1}^n \mathbb{E}I_i \right)^2 - \sum_{i=1}^n (\mathbb{E}I_i)^2 \right) \\
&\leq \mathbb{E}X_n + (1+\varepsilon_n)(\mathbb{E}X_n)^2.
\end{aligned}$$

□

26.2 Convergence of Moments

Let X be a random variable such that $\mathbb{E}|X|^k < \infty$, $k \geq 1$, i.e., all *k-th moments* $\mathbb{E}X^k$ exist and are finite. Let the distribution of X be completely determined by its moments. It means that all random variables with the same moments as X have the same distribution as X. In particular, this is true when X has the Normal or the Poisson distribution.

The method of moments provides a tool to prove the convergence in distribution of a sequence of random variables with finite moments (see Durrett [318] for details).

Lemma 26.7 (Method of Moments)**.** *Let X be a random variable with probability distribution completely determined by its moments. If $X_1, X_2, \ldots X_n, \ldots$ are random variables with finite moments such that $\mathbb{E}X_n^k \to \mathbb{E}X^k$ as $n \to \infty$, for every integer $k \geq 1$, then the sequence of random variables $\{X_n\}$ converges in distribution to random variable X, denoted as $X_n \xrightarrow{D} X$.*

The next result, which can be deduced from Theorem 26.7, provides a tool to prove asymptotic Normality.

Corollary 26.8. *Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables with finite moments and let $a_1, a_2, \ldots, a_n, \ldots$ be a sequence of positive numbers, such that*

$$\mathbb{E}(X_n - \mathbb{E}X_n)^k = \begin{cases} \frac{(2m)!}{2^m m!} a_n^k + o(a_n^k), & \text{when } k = 2m,\ m \geq 1, \\ o(a_n^k), & \text{when } k = 2m-1,\ m \geq 2, \end{cases}$$

as $n \to \infty$. Then

$$\frac{X_n - \mathbb{E}X_n}{a_n} \xrightarrow{D} Z, \quad \text{and} \quad \tilde{X}_n = \frac{X_n - \mathbb{E}X_n}{\sqrt{\operatorname{Var} X_n}} \xrightarrow{D} Z,$$

where Z is a random variable with the standard Normal distribution $N(0,1)$.

A similar result for convergence to the Poisson distribution can also be deduced from Theorem 26.7. Instead, we will show how to derive it directly from the *Inclusion-Exclusion Principle*.

The following lemma sometimes simplifies the proof of some probabilistic inequalities: A boolean function f of events $A_1, A_2, \ldots, A_n \subseteq \Omega$ is a random variable where $f(A_1, A_2, \ldots, A_n) = \bigcup_{S \in \mathscr{S}} \left((\bigcap_{i \in S} A_i) \cap \left(\bigcap_{i \notin S} A_i^c \right) \right)$ for some collection $\mathscr{S}_i$ of subsest of $[r] = \{1, 2, \ldots, r\}$.

Lemma 26.9 (Rényi's Lemma)**.** *Suppose that $A_1, A_2, \ldots, A_r$ are events in some probability space Ω, $f_1, f_2, \ldots, f_s$ are boolean functions of $A_1, A_2, \ldots, A_r$, and $\alpha_1, \alpha_2, \ldots, \alpha_s$ are reals. Then, if*

$$\sum_{i=1}^{s} \alpha_i \mathbb{P}(f_i(A_1, A_2, \ldots, A_r)) \geq 0, \tag{26.6}$$

whenever $\mathbb{P}(A_i) = 0$ or 1, then (26.6) holds in general.

Proof. Write

$$f_i = \bigcup_{S \in \mathscr{S}_i} \left(\left(\bigcap_{i \in S} A_i \right) \cap \left(\bigcap_{i \notin S} A_i^c \right) \right),$$

for some collection $\mathscr{S}_i$ of subsets of $[r] = \{1, 2, \ldots, r\}$.

Then,

$$\mathbb{P}(f_i) = \sum_{S \in \mathscr{S}_i} \mathbb{P} \left(\left(\bigcap_{i \in S} A_i \right) \cap \left(\bigcap_{i \notin S} A_i^c \right) \right),$$

and then the left hand side of (26.6) becomes

$$\sum_{S\subseteq[r]} \beta_S \mathbb{P}\left(\left(\bigcap_{i\in S} A_i\right) \cap \left(\bigcap_{i\notin S} A_i^c\right)\right),$$

for some real β_S. If (26.6) holds, then $\beta_S \geq 0$ for every S, since we can choose $A_i = \Omega$ if $i \in S$, and $A_i = \emptyset$ for $i \notin S$. □

For $J \subseteq [r]$ let $A_J = \bigcap_{i\in J} A_i$, and let $S = |\{j : A_j \text{ occurs}\}|$ denote the number of events that occur. Then let

$$B_k = \sum_{J:|J|=k} \mathbb{P}(A_J) = \mathbb{E}\binom{S}{k}.$$

Let $\mathscr{E}_j = \bigcup_{|S|=j} (\bigcap_{i\in S} A_i \cap \bigcap_{i\notin S} A_i^c)$ be the event that exactly j among the events $A_1, A_2, \ldots, A_r$ occur.

The expression in Lemma 26.10 can be motivated as follows. For $\omega \in \Omega$, we let $\theta_{\omega,i} \in \{0,1\}$ satisfy $\theta_{\omega,i} = 1$ if and only if $\omega \in A_i$. Then

$$\begin{aligned}
\Pr(\mathscr{E}_j) &= \sum_{\omega\in\Omega} \Pr(\omega) \sum_{|I|=j} \prod_{i\in I} \theta_{\omega,i} \cdot \prod_{i\notin I} (1-\theta_{\omega,i}), \\
&= \sum_{\omega\in\Omega} \Pr(\omega) \sum_{|I|=j} \sum_{J\supseteq I} (-1)^{|J\setminus I|} \theta_{\omega,J}, \qquad \text{where } \theta_{\omega,J} = \prod_{i\in J} \theta_{\omega,i}, \\
&= \sum_{|J|\geq j} \sum_{\substack{I\subseteq J \\ |I|=j}} (-1)^{J\setminus I} \Pr(A_J), \qquad \text{where } A_J = \bigcap_{i\in J} A_i \\
&= \sum_{k=j}^{r} (-1)^{k-j} \binom{k}{j} \sum_{|J|=k} \Pr(A_J) \\
&= \sum_{k=j}^{r} (-1)^{k-j} \binom{k}{j} B_k.
\end{aligned}$$

More precisely,

Lemma 26.10.

$$\mathbb{P}(\mathscr{E}_j) \begin{cases} \leq \sum_{k=j}^{s} (-1)^{k-j} \binom{k}{j} B_k & s-j \textit{ even}. \\ \geq \sum_{k=j}^{s} (-1)^{k-j} \binom{k}{j} B_k & s-j \textit{ odd} \\ = \sum_{k=j}^{s} (-1)^{k-j} \binom{k}{j} B_k & s = r. \end{cases}$$

Proof. It follows from Lemma 26.9 that we only need to check the truth of the statement for

$$\begin{aligned}\mathbb{P}(A_i) &= 1 \quad 1 \le i \le \ell,\\ \mathbb{P}(A_i) &= 0 \quad \ell < i \le r.\end{aligned}$$

where $0 \le \ell \le r$ is arbitrary.

Now

$$\mathbb{P}(S = j) = \begin{cases} 1 & \text{if } j = \ell,\\ 0 & \text{if } j \ne \ell,\end{cases}$$

and

$$B_k = \binom{\ell}{k}.$$

So,

$$\begin{aligned}\sum_{k=j}^{s} (-1)^{k-j} \binom{k}{j} B_k &= \sum_{k=j}^{s} (-1)^{k-j} \binom{k}{j} \binom{\ell}{k}\\ &= \binom{\ell}{j} \sum_{k=j}^{s} (-1)^{k-j} \binom{\ell - j}{k - j}. \qquad (26.7)\end{aligned}$$

If $\ell < j$ then $\mathbb{P}(\mathscr{E}_j) = 0$ and the sum in (26.7) reduces to zero. If $\ell = j$ then $\mathbb{P}(\mathscr{E}_j) = 1$ and the sum in (26.7) reduces to one. Thus in this case, the sum is exact for all s. Assume then that $r \ge \ell > j$. Then $\mathbb{P}(\mathscr{E}_j) = 0$ and

$$\sum_{k=j}^{s} (-1)^{k-j} \binom{\ell - j}{k - j} = \sum_{t=0}^{s-j} (-1)^t \binom{\ell - j}{t} = (-1)^{s-j} \binom{\ell - j - 1}{s - j}.$$

This explains the alternating signs of the theorem. Finally, observe that $\binom{\ell-j-1}{r-j} = 0$, as required. □

Now we are ready to state the main tool for proving convergence to the Poisson distribution.

Theorem 26.11. *Let $S_n = \sum_{i \ge 1} I_i$ be a sequence of random variables, $n \ge 1$ and let $B_k^{(n)} = \mathbb{E}\binom{S_n}{k}$. Suppose that there exists $\lambda \ge 0$, such that for every fixed $k \ge 1$,*

$$\lim_{n\to\infty} B_k^{(n)} = \frac{\lambda^k}{k!}.$$

Then, for every $j \ge 0$,

$$\lim_{n\to\infty} \mathbb{P}(S_n = j) = e^{-\lambda} \frac{\lambda^j}{j!},$$

i.e., S_n converges in distribution to the Poisson distributed random variable with expectation λ ($S_n \xrightarrow{D} \mathrm{Po}(\lambda)$).

Proof. By Lemma 26.10, for $l \geq 0$,

$$\sum_{k=j}^{j+2l+1} (-1)^{k-j} \binom{k}{j} B_k^{(n)} \leq \mathbb{P}(S_n = j) \leq \sum_{k=j}^{j+2l} (-1)^{k-j} \binom{k}{j} B_k^{(n)}.$$

So, as n grows to ∞,

$$\sum_{k=j}^{j+2l+1} (-1)^{k-j} \binom{k}{j} B_k^{(n)} \leq \liminf_{n\to\infty} \mathbb{P}(S_n = j)$$
$$\leq \limsup_{n\to\infty} \mathbb{P}(S_n = j) \leq \sum_{k=j}^{j+2l} (-1)^{k-j} \binom{k}{j} B_k^{(n)}.$$

But,

$$\sum_{k=j}^{j+m} (-1)^{k-j} \binom{k}{j} \frac{\lambda^k}{k!} = \frac{\lambda^j}{j!} \sum_{t=0}^{m} (-1)^t \frac{\lambda^t}{t!} \to \frac{\lambda^j}{j!} e^{-\lambda},$$

as $m \to \infty$. □

Notice that the falling factorial

$$(S_n)_k = S_n(S_n - 1) \cdots (S_n - k + 1)$$

counts number of ordered k-tuples of events with $I_i = 1$. Hence the binomial moments of S_n can be replaced in Theorem 26.11 by the factorial moments, defined as

$$\mathbb{E}(S_n)_k = \mathbb{E}[S_n(S_n - 1) \cdots (S_n - k + 1)],$$

and one has to check whether, for every $k \geq 1$,

$$\lim_{n\to\infty} \mathbb{E}(S_n)_k = \lambda^k.$$

26.3 Stein–Chen Method

Stein in [817] introduced a powerful technique for obtaining estimates of the rate of convergence to the standard normal distribution. His approach was subsequently extended to cover convergence to the Poisson distribution by Chen [216], while Barbour [73] ingeniously adapted both methods to random graphs.

The Stein–Chen approach has some advantages over the method of moments. The principal advantage is that a rate of convergence is automatically obtained.

Also the computations are often easier and fewer moment assumptions are required. Moreover, it frequently leads to conditions for convergence weaker than those obtainable by the method of moments.

Consider a sequence of random variables $(X_n)_{n=1}^{\infty}$ and let $(\lambda_n)_{n=1}^{\infty}$ be a sequence of positive integers, and let $\mathrm{Po}(\lambda)$ denote, as before, the Poisson distribution with expectation λ. We say that X_n is *Poisson convergent* if the *total variation distance* between the distribution $\mathscr{L}(X_n)$ of X_n and $\mathrm{Po}(\lambda_n)$, $\lambda_n = \mathbb{E}X_n$, distribution, tends to zero as n tends to infinity. So, we ask for

$$d_{TV}(\mathscr{L}(X_n), \mathrm{Po}(\lambda_n)) = \sup_{A \subseteq \mathbb{Z}^+} \left| \mathbb{P}(X_n \in A) - \sum_{k \in A} \frac{\lambda_n^k}{k!} e^{-\lambda_n} \right| \to 0, \tag{26.8}$$

as $n \to \infty$, where $\mathbb{Z}^+ = \{0, 1, \ldots\}$.

Notice, that if X_n is Poisson convergent and $\lambda_n \to \lambda$, then X_n converges in distribution to the $\mathrm{Po}(\lambda)$ distributed random variable. Furthermore, if $\lambda_n \to 0$, then X_n converges to a random variable with distribution degenerated at 0. More importantly, if $\lambda_n \to \infty$, then the central limit theorem for Poisson distributed random variables implies, that $\tilde{X}_n = (X_n - \lambda_n)/\sqrt{\lambda_n}$ converges in distribution to a random variable with the standard normal random distribution $N(0,1)$.

The basic feature and advantage of the Stein–Chen approach is that it gives computationally tractable bounds for the distance d_{TV}, when the random variables in question are sums of indicators with a fairly general dependence structure.

Let $\{I_a\}_{a \in \Gamma}$, be a family of indicator random variables, where Γ is some index set. To describe the relationship between these random variables we define a *dependency graph* $L = (V(L), E(L))$, where $V(L) = \Gamma$. Graph L has the property that whenever there are no edges between A and B, $A, B \subseteq \Gamma$, then $\{I_a\}_{a \in A}$ and $\{I_b\}_{b \in B}$ are mutually independent families of random variables. The following general bound on the total variation distance was proved by Barbour, Holst and Janson [74] via the Stein–Chen method.

Theorem 26.12. *Let $X = \sum_{a \in \Gamma} I_a$ where the I_a are indicator random variables with a dependency graph L. Then, with $\pi_a = \mathbb{E} I_a$ and $\lambda = \mathbb{E}X = \sum_{a \in \Gamma} \pi_a$,*

$$d_{TV}(\mathscr{L}(X), \mathrm{Po}(\lambda)) \leq \min(\lambda^{-1}, 1) \left(\sum_{a \in V(L)} \pi_a^2 + \sum_{ab \in E(L)} \{\mathbb{E}(I_a I_b) + \pi_a \pi_b\} \right),$$

where $\sum_{ab \in E(L)}$ means summing over all ordered pairs (a,b), such that $\{a,b\} \in E(L)$.

Finally, let us briefly mention, that the original Stein method investigates the convergence to the normal distribution in the following metric

$$d_S(\mathscr{L}(X_n), N(0,1)) = \sup_h ||h||^{-1} \left| \int h(x) dF_n(x) - \int h(x) d\Phi(x) \right|, \tag{26.9}$$

where the supremum is taken over all bounded test functions h with bounded derivative, $||h|| = \sup|h(x)| + \sup|h'(x)|$.

Here F_n is the distribution function of X_n, while Φ denotes the distribution function of the standard normal distribution. So, if $d_S(\mathscr{L}(X_n), N(0,1)) \to 0$ as $n \to \infty$, then $\tilde{X}_n$ converges in distribution to $N(0,1)$ distributed random variable.

Barbour, Karoński and Ruciński [76] obtained an effective upper bound on $d_S(\mathscr{L}(X_n), N(0,1))$ if S belongs to a general class of *decomposable* random variables. This bound involves the first three moments of S only.

For a detailed and comprehensive account of the Stein–Chen method the reader is referred to the book by Barbour, Holst and Janson [74], or to Chapter 6 of the book by Janson, Łuczak and Ruciński [509], where other interesting approaches to study asymptotic distributions of random graph characteristics are also discussed. For some applications of the Stein–Chen method in random graphs, one can look at a survey by Karoński [548].

Chapter 27

Inequalities

27.1 Binomial Coefficient Approximation

We state some important inequalities. The proofs of all but (g) are left as exercises:

Lemma 27.1. *(a)*

$$1 + x \le e^x, \qquad \forall x.$$

(b)

$$1 - x \ge e^{-x/(1-x)}, \qquad 0 \le x < 1.$$

(c)

$$\binom{n}{k} \le \left(\frac{ne}{k}\right)^k, \qquad \forall n, k.$$

(d)

$$\sum_{i=0}^{k} \binom{n}{i} \le \left(\frac{ne}{k}\right)^k, \qquad \forall n, k.$$

(e)

$$\binom{n}{k} \le \frac{n^k}{k!}\left(1 - \frac{k}{2n}\right)^{k-1}, \qquad \forall n, k.$$

(f)

$$\frac{n^k}{k!}\left(1 - \frac{k(k-1)}{2n}\right) \le \binom{n}{k} \le \frac{n^k}{k!} e^{-k(k-1)/(2n)}, \qquad \forall n, k.$$

(g)

$$\binom{n}{k} \approx \frac{n^k}{k!}, \qquad if\ k^2 = o(n).$$

(h) If $a \geq b$ then

$$\left(\frac{t-b}{n-b}\right)^b \left(\frac{n-t-a+b}{n-a}\right)^{a-b} \leq \frac{\binom{n-a}{t-b}}{\binom{n}{t}} \leq \left(\frac{t}{n}\right)^b \left(\frac{n-t}{n-b}\right)^{a-b}.$$

Proof.
(d) We write

$$\sum_{i=0}^{k} \binom{n}{i} \left(\frac{k}{n}\right)^k \leq \sum_{i=0}^{k} \frac{n^i}{i!} \cdot \frac{k^k}{n^k} = \sum_{i=0}^{k} \frac{k^i}{i!} \cdot \left(\frac{k}{n}\right)^{k-i} \leq e^k.$$

(g)

$$\begin{aligned}
&\frac{\binom{n-a}{t-b}}{\binom{n}{t}} = \frac{(n-a)!t!(n-t)!}{n!(t-b)!(n-t-a+b)!} \\
&= \frac{t(t-1)\cdots(t-b+1)}{n(n-1)\cdots(n-b+1)} \times \frac{(n-t)(n-t-1)\cdots(n-t-a+b+1)}{(n-b)(n-b-1)\cdots(n-a+1)} \\
&\leq \left(\frac{t}{n}\right)^b \times \left(\frac{n-t}{n-b}\right)^{a-b}.
\end{aligned}$$

The lower bound follows similarly. □

We will need also the following estimate for binomial coefficients. It is a little more precise than those given in Lemma 27.1.

Lemma 27.2. *Let $k = o(n^{3/4})$. Then*

$$\binom{n}{k} \approx \frac{n^k}{k!} \exp\left\{-\frac{k^2}{2n} - \frac{k^3}{6n^2}\right\}.$$

Proof.

$$\begin{aligned}
\binom{n}{k} &= \frac{n^k}{k!} \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\
&= \frac{n^k}{k!} \exp\left\{\sum_{i=0}^{k-1} \log\left(1 - \frac{i}{n}\right)\right\} \\
&= \frac{n^k}{k!} \exp\left\{-\sum_{i=0}^{k-1} \left(\frac{i}{n} + \frac{i^2}{2n^2}\right) + O\left(\frac{k^4}{n^3}\right)\right\} \\
&= (1+o(1)) \frac{n^k}{k!} \exp\left\{-\frac{k^2}{2n} - \frac{k^3}{6n^2}\right\}.
\end{aligned}$$

□

27.2 Balls in Boxes

Suppose that we have M boxes and we independently place N distinguishable balls into them. Let us assume that a ball goes into box i with probability p_i where $p_1 + \cdots + p_M = 1$. Let W_i denote the number of balls that are placed in box i and for $S \subseteq [M]$, let $W_S = \sum_{i \in S} W_i$. The following looks obvious and is extremely useful.

Theorem 27.3. *Let S, T be disjoint subsets of $[M]$ and let s, t be non-negative integers. Then*

$$\mathbb{P}(W_S \leq s \mid W_T \leq t) \leq \mathbb{P}(W_S \leq s). \tag{27.1}$$
$$\mathbb{P}(W_S \geq s \mid W_T \leq t) \geq \mathbb{P}(W_S \geq s). \tag{27.2}$$
$$\mathbb{P}(W_S \geq s \mid W_T \geq t) \leq \mathbb{P}(W_S \geq s). \tag{27.3}$$
$$\mathbb{P}(W_S \leq s \mid W_T \geq t) \geq \mathbb{P}(W_S \leq s). \tag{27.4}$$

Proof. Equation (27.2) follows immediately from (27.1). Also, equation (27.4) follows immediately from (27.3). The proof of (27.3) is very similar to that of (27.1) and so we will only prove (27.1).

Let
$$\pi_i = \mathbb{P}(W_S \leq s \mid W_T = i).$$
Given $W_T = i$, we are looking at throwing $N - i$ balls into $M - 1$ boxes. It is clear therefore that π_i is monotone increasing in i. Now, let $q_i = \mathbb{P}(W_T = i)$. Then,

$$\mathbb{P}(W_S \leq s) = \sum_{i=0}^{N} \pi_i q_i.$$
$$\mathbb{P}(W_S \leq s \mid W_T \leq t) = \sum_{i=0}^{t} \pi_i \frac{q_i}{q_0 + \cdots + q_t}.$$

So, (27.1) reduces to

$$(q_0 + \cdots + q_N) \sum_{i=0}^{t} \pi_i q_i \leq (q_0 + \cdots + q_t) \sum_{i=0}^{N} \pi_i q_i,$$

or

$$(q_{t+1} + \cdots + q_N) \sum_{i=0}^{t} \pi_i q_i \leq (q_0 + \cdots + q_t) \sum_{i=t+1}^{N} \pi_i q_i,$$

or

$$\sum_{j=t+1}^{N} \sum_{i=0}^{t} q_i q_j \pi_i \leq \sum_{j=0}^{t} \sum_{i=t+1}^{N} q_i q_j \pi_i.$$

The result now follows from the monotonicity of π_i. $\square$

The following is an immediate corollary:

Corollary 27.4. *Let $S_1,S_2,\dots,S_k$ be disjoint subsets of $[M]$ and let $s_1,s_2,\dots,s_k$ be non-negative integers. Then*

$$\mathbb{P}\left(\bigcap_{i=1}^{k}\{W_{S_i}\le s_i\}\right)\le\prod_{i=1}^{k}\mathbb{P}(\{W_{S_i}\le s_i\}).$$
$$\mathbb{P}\left(\bigcap_{i=1}^{k}\{W_{S_i}\ge s_i\}\right)\le\prod_{i=1}^{k}\mathbb{P}(\{W_{S_i}\ge s_i\}).$$

27.3 FKG Inequality

A function $f : C_N = \{0,1\}^{[N]} \to \mathbb{R}$ is said to be *monotone increasing* if whenever $x = (x_1,x_2,\dots,x_N), y = (y_1,y_2,\dots,y_N) \in C_N$ and $x \le y \in C_N$ (i.e. $x_j \le y_j, j = 1,2,\dots,N$) then $f(x) \le f(y)$. Similarly, f is said to be *monotone decreasing* if $-f$ is monotone increasing.

An important example for us is the case where f is the indicator function of some subset $\mathscr{A}$ of $2^{[N]}$. Then

$$f(x)=\begin{cases}1 & x\in\mathscr{A}\\ emptyset & x\notin\mathscr{A}\end{cases}.$$

A typical example for us would be $N=\binom{n}{2}$ and then each $G\in 2^{[N]}$ corresponds to a graph with vertex set $[n]$. Then $\mathscr{A}$ will be a set of graphs i.e. a graph property. Suppose that f is the indicator function for $\mathscr{A}$. Then f is monotone increasing, if whenever $G \in \mathscr{A}$ and $e \notin E(G)$ we have $G+e \in \mathscr{A}$ i.e. adding an edge does not destroy the property. We will say that the set/property is monotone increasing. For example if $\mathscr{H}$ is the set of Hamiltonian graphs then $\mathscr{H}$ is monotone increasing. If $\mathscr{P}$ is the set of planar graphs then $\mathscr{P}$ is monotone decreasing. In other words a property is monotone increasing iff its indicator function is monotone increasing.

Suppose next that we turn C_N into a probability space by choosing some $p_1,p_2,\dots,p_N \in [0,1]$ and then for $\mathbf{x} = (x_1,x_2,\dots,x_N) \in C_N$ letting

$$\mathbb{P}(\mathbf{x})=\prod_{j:x_j=1}p_j\prod_{j:x_j=0}(1-p_j). \tag{27.5}$$

If $N=\binom{n}{2}$ and $p_j = p, j = 1,2,\dots,N$ then this model corresponds to $\mathbb{G}_{n,p}$.

The following is a special case of the FKG inequality, Harris [469] and Fortuin, Kasteleyn and Ginibre [370]:

Theorem 27.5. *If f,g are monotone increasing functions on C_N then $\mathbb{E}(fg) \ge \mathbb{E}(f)\,\mathbb{E}(g)$.*

Proof. We will prove this by induction on N. If $N = 0$ then $\mathbb{E}(f) = a, \mathbb{E}(g) = b$ and $\mathbb{E}(fg) = ab$ for some constants a, b.

So assume the truth for $N-1$. Suppose that $\mathbb{E}(f \mid x_N = 0) = a$ and $\mathbb{E}(g \mid x_N = 0) = b$ then

$$\mathbb{E}((f-a)(g-b)) - \mathbb{E}(f-a)\,\mathbb{E}(g-b) = \mathbb{E}(fg) - \mathbb{E}(f)\,\mathbb{E}(g).$$

By replacing f by $f-a$ and g by $g-b$ we may therefore assume that $\mathbb{E}(f \mid x_N = 0) = \mathbb{E}(g \mid x_N = 0) = 0$. By monotonicity, we see that $\mathbb{E}(f \mid x_N = 1), \mathbb{E}(g \mid x_N = 1) \geq 0$.

We observe that by the induction hypothesis that

$$\begin{aligned}\mathbb{E}(fg \mid x_N = 0) &\geq \mathbb{E}(f \mid x_N = 0)\,\mathbb{E}(g \mid x_N = 0) = 0\\ \mathbb{E}(fg \mid x_N = 1) &\geq \mathbb{E}(f \mid x_N = 1)\,\mathbb{E}(g \mid x_N = 1) \geq 0\end{aligned}$$

Now, by the above inequalities,

$$\begin{aligned}\mathbb{E}(fg) &= \mathbb{E}(fg \mid x_N = 0)(1-p_N) + \mathbb{E}(fg \mid x_N = 1)p_N\\ &\geq \mathbb{E}(f \mid x_N = 1)\,\mathbb{E}(g \mid x_N = 1)p_N. \end{aligned} \tag{27.6}$$

Furthermore,

$$\begin{aligned}\mathbb{E}(f)\,\mathbb{E}(g) = &\\ &(\mathbb{E}(f \mid x_N = 0)(1-p_N) + \mathbb{E}(f \mid x_N = 1)p_N)\times\\ &(\mathbb{E}(g \mid x_N = 0)(1-p_N) + \mathbb{E}(g \mid x_N = 1)p_N)\\ &= \mathbb{E}(f \mid x_N = 1)\,\mathbb{E}(g \mid x_N = 1)p_N^2. \end{aligned} \tag{27.7}$$

The result follows by comparing (27.6) and (27.7) and using the fact that $\mathbb{E}(f \mid x_N = 1), \mathbb{E}(g \mid x_N = 1) \geq 0$ and $0 \leq p_N \leq 1$. □

In terms of monotone increasing sets $\mathscr{A}, \mathscr{B}$ and the same probability (27.5) we can express the FKG inequality as

$$\mathbb{P}(\mathscr{A} \mid \mathscr{B}) \geq \mathbb{P}(\mathscr{A}). \tag{27.8}$$

27.4 Sums of Independent Bounded Random Variables

Suppose that S is a random variable and $t > 0$ is a real number. We will be concerned here with bounds on the *upper and lower tail* of the distribution of S, i.e., on $\mathbb{P}(S \geq \mu + t)$ and $\mathbb{P}(S \leq \mu - t)$, respectively, where $\mu = \mathbb{E}S$.

The basic observation which leads to the construction of such bounds is due to Bernstein [104]. Let $\lambda \geq 0$, then

$$\mathbb{P}(S \geq \mu + t) = \mathbb{P}(e^{\lambda S} \geq e^{\lambda(\mu+t)}) \leq e^{-\lambda(\mu+t)}\mathbb{E}(e^{\lambda S}), \tag{27.9}$$

by the Markov inequality (see Lemma 26.1). Similarly for $\lambda \leq 0$,

$$\mathbb{P}(S \leq \mu - t) \leq e^{-\lambda(\mu-t)}\mathbb{E}(e^{\lambda S}). \tag{27.10}$$

Combining (27.9) and (27.10) one can obtain a bound for $\mathbb{P}(|S-\mu| \geq t)$.

Now let $S_n = X_1 + X_2 + \cdots + X_n$ where $X_i, i = 1, \ldots, n$ are independent random variables. Assume that $0 \leq X_i \leq 1$ and $\mathbb{E}X_i = \mu_i$ for $i = 1, 2, \ldots, n$. Let $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$. Then for $\lambda \geq 0$

$$\mathbb{P}(S_n \geq \mu + t) \leq e^{-\lambda(\mu+t)} \prod_{i=1}^{n} \mathbb{E}(e^{\lambda X_i}) \tag{27.11}$$

and for $\lambda \leq 0$

$$\mathbb{P}(S_n \leq \mu - t) \leq e^{-\lambda(\mu-t)} \prod_{i=1}^{n} \mathbb{E}(e^{\lambda X_i}). \tag{27.12}$$

Note that $\mathbb{E}(e^{\lambda X_i})$ in (27.11) and (27.12), likewise $\mathbb{E}(e^{\lambda S})$ in (27.9) and (27.10) are the moment generating functions of the X_i's and S, respectively. So finding bounds boils down to the estimation of these functions.

Now the convexity of e^x and $0 \leq X_i \leq 1$ implies that

$$e^{\lambda X_i} \leq 1 - X_i + X_i e^{\lambda}.$$

Taking expectations we get

$$\mathbb{E}(e^{\lambda X_i}) \leq 1 - \mu_i + \mu_i e^{\lambda}.$$

Equation (27.11) becomes, for $\lambda \geq 0$,

$$\begin{aligned} \mathbb{P}(S_n \geq \mu + t) &\leq e^{-\lambda(\mu+t)} \prod_{i=1}^{n} (1 - \mu_i + \mu_i e^{\lambda}) \\ &\leq e^{-\lambda(\mu+t)} \left(\frac{n - \mu + \mu e^{\lambda}}{n} \right)^n. \end{aligned} \tag{27.13}$$

The second inequality follows from the fact that the geometric mean is at most the arithmetic mean i.e. $(x_1 x_2 \cdots x_n)^{1/n} \leq (x_1 + x_2 + \cdots + x_n)/n$ for non-negative $x_1, x_2, \ldots, x_n$. This in turn follows from Jensen's inequlaity and the concavity of $\log x$.

The right hand side of (27.13) attains its minimum, as a function of λ, at

$$e^{\lambda} = \frac{(\mu+t)(n-\mu)}{(n-\mu-t)\mu}. \tag{27.14}$$

Hence, by (27.13) and(27.14), assuming that $\mu+t<n$,

$$\mathbb{P}(S_n \geq \mu+t) \leq \left(\frac{\mu}{\mu+t}\right)^{\mu+t}\left(\frac{n-\mu}{n-\mu-t}\right)^{n-\mu-t}, \tag{27.15}$$

while for $t > n-\mu$ this probability is zero.

Now let

$$\varphi(x) = (1+x)\log(1+x) - x, \quad x \geq -1,$$

and let $\varphi(x) = \infty$ for $x < -1$. Now, for $0 \leq t < n-\mu$, we can rewrite the bound (27.15) as

$$\mathbb{P}(S_n \geq \mu+t) \leq \exp\left\{-\mu\varphi\left(\frac{t}{\mu}\right) - (n-\mu)\varphi\left(\frac{-t}{n-\mu}\right)\right\}. \tag{27.16}$$

Since $\varphi(x) \geq 0$ for every x, we get

$$\mathbb{P}(S_n \geq \mu+t) \leq e^{-\mu\varphi(t/\mu)}. \tag{27.17}$$

Similarly, putting $n-S_n$ for S_n, or by an analogous argument, using (27.12), we get for $0 \leq t \leq \mu$,

$$\mathbb{P}(S_n \leq \mu-t) \leq \exp\left\{-\mu\varphi\left(\frac{-t}{\mu}\right) - (n-\mu)\varphi\left(\frac{t}{n-\mu}\right)\right\}. \tag{27.18}$$

Hence,

$$\mathbb{P}(S_n \leq \mu-t) \leq e^{-\mu\varphi(-t/\mu)}. \tag{27.19}$$

We can simplify the expressions (27.17) and (27.19) by observing that

$$\varphi(x) \geq \frac{x^2}{2(1+x/3)}. \tag{27.20}$$

To see this observe that for $|x| \leq 1$ we have

$$\varphi(x) - \frac{x^2}{2(1+x/3)} = \sum_{k=2}^{\infty}(-1)^k\left(\frac{1}{k(k-1)} - \frac{1}{2\cdot 3^{k-2}}\right)x^k.$$

Equation (27.20) for $|x| \leq 1$ follows from $\frac{1}{k(k-1)} - \frac{1}{2\cdot 3^{k-2}} \geq 0$ for $k \geq 2$. We leave it as an exercise to check that (27.20) remains true for $x > 1$.

Taking this into account we arrive at the following theorem, see Hoeffding [480].

Theorem 27.6 (Chernoff/Hoeffding inequality)**.** *Suppose that* $S_n = X_1 + X_2 + \cdots + X_n$ *where (i)* $0 \le X_i \le 1$ *and* $\mathbb{E}X_i = \mu_i$ *for* $i = 1,2,\ldots,n$*, (ii)* $X_1, X_2, \ldots, X_n$ *are independent. Let* $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$*. Then for* $t \ge 0$,

$$\mathbb{P}(S_n \ge \mu + t) \le \exp\left\{-\frac{t^2}{2(\mu + t/3)}\right\} \tag{27.21}$$

and for $t \le \mu$,

$$\mathbb{P}(S_n \le \mu - t) \le \exp\left\{-\frac{t^2}{2(\mu - t/3)}\right\}. \tag{27.22}$$

Putting $t = \varepsilon\mu$, for $0 < \varepsilon < 1$, one can immediately obtain the following bounds.

Corollary 27.7. *Let* $0 < \varepsilon < 1$*, then*

$$\mathbb{P}(S_n \ge (1+\varepsilon)\mu) \le \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu} \le \exp\left\{-\frac{\mu\varepsilon^2}{3}\right\}, \tag{27.23}$$

while

$$\mathbb{P}(S_n \le (1-\varepsilon)\mu) \le \exp\left\{-\frac{\mu\varepsilon^2}{2}\right\} \tag{27.24}$$

Proof. The formula (27.24) follows directly from (27.22) and (27.23) follows from (27.16). □

One can "tailor" Chernoff bounds with respect to specific needs. For example, for small ratios t/μ, the exponent in (27.21) is close to $t^2/2\mu$, and the following bound holds.

Corollary 27.8.

$$\mathbb{P}(S_n \ge \mu + t) \le \exp\left\{-\frac{t^2}{2\mu} + \frac{t^3}{6\mu^2}\right\} \tag{27.25}$$

$$\le \exp\left\{-\frac{t^2}{3\mu}\right\} \qquad \textit{for } t \le \mu. \tag{27.26}$$

Proof. Use (27.21) and note that

$$(\mu + t/3)^{-1} \ge (\mu - t/3)/\mu^2.$$

□

For large deviations we have the following result.

Corollary 27.9. *If $c > 1$ then*

$$\mathbb{P}(S_n \geq c\mu) \leq \left\{ \frac{e}{ce^{1/c}} \right\}^{c\mu}. \tag{27.27}$$

Proof. Put $t = (c-1)\mu$ into (27.17). $\square$

We also have the simple easiest to use version:

Corollary 27.10. *Suppose that $X_1, X_2, \ldots, X_n$ are independent random variable and that $a_i \leq X_i \leq b_i$ for $i = 1, 2, \ldots, n$. Let $S_n = X_1 + X_2 + \cdots + X_n$ and $\mu_i = \mathbb{E}(X_i)$, $i = 1, 2, \ldots, n$ and $\mu = \mathbb{E}(S_n)$. Then for $t > 0$ and $c_i = b_i - a_i$, $i = 1, 2, \ldots, n$, we have*

$$\mathbb{P}(S_n \geq \mu + t) \leq \exp\left\{ -\frac{2t^2}{c_1^2 + c_2^2 + \cdots + c_n^2} \right\}. \tag{27.28}$$

$$\mathbb{P}(S_n \leq \mu - t) \leq \exp\left\{ -\frac{2t^2}{c_1^2 + c_2^2 + \cdots + c_n^2} \right\}. \tag{27.29}$$

Proof. We can assume without loss of generality that $a_i = 0, i = 1, 2, \ldots, n$. We just subtract $A = \sum_{i=1}^n a_i$ from S_n. We proceed as before.

$$\mathbb{P}(S_n \geq \mu + t) = \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda(\mu+t)}) \leq e^{-\lambda(\mu+t)} \mathbb{E}(e^{\lambda S_n}) = e^{-\lambda t} \prod_{i=1}^n \mathbb{E}(e^{\lambda(X_i - \mu_i)}).$$

Note that $e^{\lambda x}$ is a convex function of x, and since $0 \leq X_i \leq c_i$, we have

$$e^{\lambda(X_i - \mu_i)} \leq e^{-\lambda\mu_i} \left(1 - \frac{X_i}{c_i} + \frac{X_i}{c_i} e^{\lambda c_i} \right)$$

and so

$$\begin{aligned} \mathbb{E}(e^{\lambda X_i}) &\leq e^{-\lambda\mu_i} \left(1 - \frac{\mu_i}{c_i} + \frac{\mu_i}{c_i} e^{\lambda c_i} \right) \\ &= e^{-\theta_i p_i} \left(1 - p_i + p_i e^{\theta_i} \right), \end{aligned} \tag{27.30}$$

where $\theta_i = \lambda c_i$ and $p_i = \mu_i / c_i$.

Then, taking the logarithm of the RHS of (27.30), we have

$$\begin{aligned} f(\theta_i) &= -\theta_i p_i + \log\left(1 - p_i + p_i e^{\theta_i} \right). \\ f'(\theta_i) &= -p_i + \frac{p_i}{1 - p_i + p_i e^{\theta_i}}. \end{aligned}$$

$$f''(\theta_i) = \frac{p_i(1-p_i)e^{-\theta_i}}{((1-p_i)e^{-\theta_i}+p_i)^2}.$$

Now $\frac{\alpha\beta}{(\alpha+\beta)^2} \leq 1/4$ and so $f''(\theta_i) \leq 1/4$ anf therefore

$$f(\theta_i) \leq f(0) + f'(0)\theta_i + \frac{1}{8}\theta_i^2 = \frac{\lambda^2 c_i^2}{8}.$$

It follows then that

$$\mathbb{P}(S_n \geq \mu + t) \leq e^{-\lambda t} \exp\left\{\sum_{i=1}^n \frac{\lambda^2 c_i^2}{8}\right\}.$$

We obtain (27.28) by puttng $\lambda = \frac{4}{\sum_{i=1}^n c_i^2}$ and (27.29) is proved in a similar manner. $\square$

Our next bound incorporates the variance of the X_i's.

Theorem 27.11 (Bernstein's Theorem)**.** *Suppose that* $S_n = X_1 + X_2 + \cdots + X_n$ *where (i)* $|X_i| \leq 1$ *and* $\mathbb{E}X_i = 0$ *and* $\operatorname{Var} X_i = \sigma_i^2$ *for* $i = 1, 2, \ldots, n$*, (ii)* $X_1, X_2, \ldots, X_n$ *are independent. Let* $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$*. Then for* $t \geq 0$*,*

$$\mathbb{P}(S_n \geq t) \leq \exp\left\{-\frac{t^2}{2(\sigma^2 + t/3)}\right\} \tag{27.31}$$

and

$$\mathbb{P}(S_n \leq -t) \leq \exp\left\{-\frac{t^2}{2(\sigma^2 + t/3)}\right\}. \tag{27.32}$$

Proof. The strategy is once again to bound the moment generating function. Let

$$F_i = \sum_{r=2}^{\infty} \frac{\lambda^{r-2}\mathbb{E}X_i^r}{r!\sigma_i^2} \leq \sum_{r=2}^{\infty} \frac{\lambda^{r-2}\sigma_i^2}{r!\sigma_i^2} = \frac{e^\lambda - 1 - \lambda}{\lambda^2}.$$

Here $\mathbb{E}X_i^r \leq \sigma_i^2$, since $|X_i| \leq 1$.

We then observe that

$$\begin{aligned}
\mathbb{E}(e^{\lambda X_i}) &= 1 + \sum_{r=2}^{\infty} \frac{\lambda^r \mathbb{E}X_i^r}{r!} \\
&= 1 + \lambda^2 \sigma_i^2 F_i \\
&\leq e^{\lambda^2 \sigma_i^2 F_i} \\
&\leq \exp\left\{(e^\lambda - \lambda - 1)\sigma_i^2\right\}.
\end{aligned}$$

So,

$$\begin{aligned}\mathbb{P}(S_n \geq t) &\leq e^{-\lambda t} \prod_{i=1}^{n} \exp\left\{(e^{\lambda} - \lambda - 1)\sigma_i^2\right\} \\ &= e^{\sigma^2(e^{\lambda}-\lambda-1)-\lambda t} \\ &= \exp\left\{-\sigma^2 \varphi\left(\frac{t}{\sigma^2}\right)\right\}.\end{aligned}$$

after assigning

$$\lambda = \log\left(1 + \frac{t}{\sigma^2}\right).$$

To obtain (27.31) we use (27.20). To obtain (27.32) we apply (27.31) to $Y_i = -X_i, i = 1, 2, \ldots, n$. □

27.5 Sampling Without Replacement

Let a multi-set $A = \{a_1, a_2, \ldots, a_N\} \subseteq \mathbf{R}$ be given. We consider two random variables. For the first let $X = a_i$ where i is chosen uniformly at random from $[N]$. Let

$$\mu = \mathbb{E}X = \frac{1}{N}\sum_{i=1}^{N} a_i \text{ and } \sigma^2 = \mathrm{Var}X = \frac{1}{N}\sum_{i=1}^{N}(a_i - \mu)^2.$$

Now let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of n independent copies of X. Next let $W_n = \sum_{i \in \mathbf{X}} a_i$ where $\mathbf{X}$ is a uniformly random n-subset of $[N]$. We have $\mathbb{E}S_n = \mathbb{E}W_n = n\mu$ but as shown in Hoeffding [480], W_n is more tightly concentrated around its mean than S_n. This will follow from the following:

Lemma 27.12. *Let $f : \mathbf{R} \to \mathbf{R}$ be continuous and convex. Then*

$$\mathbb{E}f(W_n) \leq \mathbb{E}f(S_n).$$

Proof. We write, where $(A)_n$ denotes the set of sequences of n distinct members of A and $(N)_n = N(N-1)\cdots(N-n+1) = |(A)_n|$,

$$\mathbb{E}f(S_n) = \frac{1}{N^n}\sum_{\mathbf{y} \in A^n} f(y_1 + \cdots + y_n) = \\ \frac{1}{(N)_n}\sum_{\mathbf{x} \in (A)_n} g(x_1, x_2, \ldots, x_n) = \mathbb{E}g(\mathbf{X}), \quad (27.33)$$

where g is a symmetric function of $\mathbf{x}$ and

$$g(x_1, x_2, \ldots, x_n) = \sum_{k, \mathbf{i}, \mathbf{r}} \psi(k, \mathbf{i}, \mathbf{r}) f(r_{i_1} x_{i_1} + \cdots + r_{i_k} x_{i_k}).$$

Here $\mathbf{i}$ ranges over sequences of k distinct values $i_1, i_2, \dots, i_k \in [n]$ and $r_{i_1} + \cdots + r_{i_k} = n$. The factors $\psi(k, \mathbf{i}, \mathbf{r})$ are independent of the function f.

Putting $f = 1$ we see that $\sum_{k,\mathbf{i},\mathbf{r}} \psi(k, \mathbf{i}, \mathbf{r}) = 1$. Putting $f(x) = x$ we see that g is a linear symmetric function and so

$$\sum_{k,\mathbf{i},\mathbf{r}} \psi(k, \mathbf{i}, \mathbf{r})(r_{i_1} x_{i_1} + \cdots + r_{i_k} x_{i_k}) = K(x_1 + \cdots + x_n),$$

for some K. Equation (27.33) implies that $K = 1$.

Applying Jensen's inequality we see that

$$g(\mathbf{x}) \geq f(x_1 + \cdots + x_n).$$

It follows that

$$\mathbb{E}\, g(\mathbf{X}) \geq \mathbb{E}\, f(W_n)$$

and the Lemma follows from (27.33). $\square$

As a consequence we have that (i) $\operatorname{Var} W_n \leq \operatorname{Var} S_n$ and (ii) $\mathbb{E}\, e^{\lambda W_n} \leq \mathbb{E}\, e^{\lambda S_n}$ for any $\lambda \in \mathbf{R}$.

Thus all the inequalities developed in Section 27.4 can a fortiori be applied to W_n in place of S_n. Of particular importance in this context, is the *hypergeometric distribution*: Here we are given a set of $S \subseteq [N], |S| = m$ and we choose a random set X of size k from $[N]$. Let $Z = |X \cap S|$. Then

$$\mathbb{P}(Z = t) = \frac{\binom{m}{t}\binom{N-m}{k-t}}{\binom{N}{k}}, \quad \text{for } 0 \leq t \leq k.$$

27.6 Janson's Inequality

In Section 27.4 we found bounds for the upper and lower tails of the distribution of a random variable S_n composed of *n independent* summands. In the previous section we allowed some dependence between the summands. We consider another case where the random variables in question are not necessarily independent. In this section we prove an inequality of Janson [495]. This generalised an earlier inequality of Janson, Łuczak and Ruciński [508], see Corollary 27.14.

Fix a family of n subsets $D_i, i \in [n]$. Let R be a random subset of $[N]$ such that for $s \in [N]$ we have $0 < \mathbb{P}(s \in R) = q_s < 1$. The elements of R are chosen independently of each other and the sets $D_i, i = 1, 2, \dots, n$. Let $\mathscr{A}_i$ be the event that D_i is a subset of R. Moreover, let I_i be the indicator of the event $\mathscr{A}_i$. Note that, I_i and I_j are independent iff $D_i \cap D_j = \emptyset$. One can easily see that the I_i's are increasing.

We let

$$S_n = I_1 + I_2 + \cdots + I_n,$$

and

$$\mu = \mathbb{E}S_n = \sum_{i=1}^{n} \mathbb{E}(I_i).$$

We write $i \sim j$ if $D_i \cap D_j \neq \emptyset$. Then, let

$$\overline{\Delta} = \sum_{\{i,j\}:i\sim j} \mathbb{E}(I_i I_j) = \mu + \Delta \tag{27.34}$$

where

$$\Delta = \sum_{\substack{\{i,j\}:i\sim j \\ i\neq j}} \mathbb{E}(I_i I_j). \tag{27.35}$$

As before, let $\varphi(x) = (1+x)\log(1+x) - x$. Now, with $S_n, \overline{\Delta}, \varphi$ given above one can establish the following upper bound on the lower tail of the distribution of S_n.

Theorem 27.13 (Janson's Inequality)**.** *For any real t, $0 \leq t \leq \mu$,*

$$\mathbb{P}(S_n \leq \mu - t) \leq \exp\left\{-\frac{\varphi(-t/\mu)\mu^2}{\overline{\Delta}}\right\} \leq \exp\left\{-\frac{t^2}{2\overline{\Delta}}\right\}. \tag{27.36}$$

Proof. We begin as we did in Section 27.4. Put $\psi(\lambda) = \mathbb{E}(e^{-\lambda S_n}), \lambda \geq 0$. By the Markov inequality we have

$$\mathbb{P}(S_n \leq \mu - t) \leq e^{\lambda(\mu - t)} \mathbb{E}e^{-\lambda S_n}.$$

Therefore,

$$\log \mathbb{P}(S_n \leq \mu - t) \leq \log \psi(\lambda) + \lambda(\mu - t). \tag{27.37}$$

Now let us estimate $\log\psi(\lambda)$ and minimise the right-hand-side of (27.37) with respect to λ.

Note that

$$-\psi'(\lambda) = \mathbb{E}(S_n e^{-\lambda S_n}) = \sum_{i=1}^{n} \mathbb{E}(I_i e^{-\lambda S_n}). \tag{27.38}$$

Now for every $i \in [n]$, split S_n into Y_i and Z_i, where

$$Y_i = \sum_{j:j\sim i} I_j, \quad Z_i = \sum_{j:j\not\sim i} I_j, \quad S_n = Y_i + Z_i.$$

Then by the FKG inequality (applied to the random set R and conditioned on $I_i = 1$) we get, setting $p_i = \mathbb{E}(I_i) = \prod_{s\in D_i} q_s$,

$$\mathbb{E}(I_i e^{-\lambda S_n}) = p_i \mathbb{E}(e^{-\lambda Y_i} e^{-\lambda Z_i} \mid I_i = 1) \geq p_i \mathbb{E}(e^{-\lambda Y_i} \mid I_i = 1) \mathbb{E}(e^{-\lambda Z_i} \mid I_i = 1).$$

Since Z_i and I_i are independent we get

$$\mathbb{E}(I_i e^{-\lambda S_n}) \geq p_i \mathbb{E}(e^{-\lambda Y_i} \mid I_i = 1)\mathbb{E}(e^{-\lambda Z_i}) \geq p_i \mathbb{E}(e^{-\lambda Y_i} \mid I_i = 1)\psi(\lambda). \quad (27.39)$$

From (27.38) and (27.39), applying Jensen's inequality to get (27.40) and remembering that $\mu = \mathbb{E}S_n = \sum_{i=1}^n p_i$, we get

$$\begin{aligned} -(\log \psi(\lambda))' &= -\frac{\psi'(\lambda)}{\psi(\lambda)} \\ &\geq \sum_{i=1}^n p_i \mathbb{E}(e^{-\lambda Y_i} \mid I_i = 1) \\ &\geq \mu \sum_{i=1}^n \frac{p_i}{\mu} \exp\left\{-\mathbb{E}(\lambda Y_i \mid I_i = 1)\right\} \\ &\geq \mu \exp\left\{-\frac{1}{\mu}\sum_{i=1}^n p_i \mathbb{E}(\lambda Y_i \mid I_i = 1)\right\} \qquad (27.40) \\ &= \mu \exp\left\{-\frac{\lambda}{\mu}\sum_{i=1}^n \mathbb{E}(Y_i I_i)\right\} \\ &= \mu e^{-\lambda\overline{\Delta}/\mu}. \end{aligned}$$

So

$$-(\log \psi(\lambda))' \geq \mu e^{-\lambda\overline{\Delta}/\mu} \quad (27.41)$$

which implies that

$$-\log \psi(\lambda) \geq \int_0^\lambda \mu e^{-z\overline{\Delta}/\mu} dz = \frac{\mu^2}{\overline{\Delta}}(1 - e^{-\lambda\overline{\Delta}/\mu}). \quad (27.42)$$

Hence by (27.42) and (27.37)

$$\log \mathbb{P}(S_n \leq \mu - t) \leq -\frac{\mu^2}{\overline{\Delta}}(1 - e^{-\lambda\overline{\Delta}/\mu}) + \lambda(\mu - t), \quad (27.43)$$

which is minimized by choosing $\lambda = -\log(1 - t/\mu)\mu/\overline{\Delta}$. It yields the first bound in (27.36), while the final bound in (27.36) follows from the fact that $\varphi(x) \geq x^2/2$ for $x \leq 0$. □

The following Corollary is very useful:

Corollary 27.14 (Janson, Łuczak, Ruciński Inequality)**.**

$$\mathbb{P}(S_n = 0) \leq e^{-\mu+\Delta}.$$

Proof. We put $t = \mu$ into (27.36) giving $\mathbb{P}(S_n = 0) \leq \exp\left\{-\frac{\varphi(-1)\mu^2}{\overline{\Delta}}\right\}$. Now note that $\varphi(-1) = 1$ and $\frac{\mu^2}{\overline{\Delta}} \geq \frac{\mu^2}{\mu+\Delta} \geq \mu - \Delta$. □

27.7 Martingales. Azuma-Hoeffding Bounds

Before we present the basic results of this chapter we have to briefly introduce *martingales* and concentration inequalities for martingales. Historically, martingales were applied to random graphs for the first time in the context of the chromatic number of $\mathbb{G}_{n,p}$.
Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. If the sample space Ω is finite, then $\mathscr{F}$ is the algebra of all subsets of Ω. For simplicity, let us assume that we deal with this case.

Recall that if $\mathscr{D} = \{D_1, D_2, \ldots, D_m\}$ is a *partition* of Ω, i.e., $\bigcup_{i=1}^m D_i = \Omega$ and $D_i \cap D_j = \emptyset$ if $i \neq j$, then it generates an algebra of subsets $\mathscr{A}(\mathscr{D})$ of Ω. The algebra generated by the partition $\mathscr{D}$ and denoted by $\mathscr{A}(\mathscr{D})$ is the family of all unions of the events (sets) from $\mathscr{D}$, with $\emptyset$ obtained by taking an empty union.

Let $\mathscr{D} = \{D_1, D_2, \ldots, D_m\}$ be a partition of Ω and A be any event, $A \subset \Omega$ and let $\mathbb{P}(A|\mathscr{D})$ be the random variable defined by

$$\begin{aligned}\mathbb{P}(A|\mathscr{D})(\omega) &= \sum_{i=1}^m \mathbb{P}(A|D_i) I_{D_i}(\omega)\\ &= \mathbb{P}(A|D_{i(\omega)}) \text{ where } \omega \in D_{i(\omega)}.\end{aligned}$$

Note that if $\mathscr{D}$ a trivial partition, i.e., $\mathscr{D} = \mathscr{D}_0 = \{\Omega\}$ then $\mathbb{P}(A|\mathscr{D}_0) = \mathbb{P}(A)$, while, in general,

$$\mathbb{P}(A) = \mathbb{E}\,\mathbb{P}(A|\mathscr{D}). \tag{27.44}$$

Suppose that X is a discrete random variable taking values $\{x_1, x_2, \ldots, x_l\}$ and write X as

$$X = \sum_{j=1}^l x_j I_{A_j}, \tag{27.45}$$

where $A_j = \{\omega : X(\omega) = x_j\}$. Notice that the random variable X generates a partition $\mathscr{D}_X = \{A_1, A_2, \ldots, A_l\}$.

Now the conditional expectation of X with respect to a partition $\mathscr{D}$ of Ω is given as

$$\mathbb{E}(X|\mathscr{D}) = \sum_{j=1}^l x_j \mathbb{P}(A_j|\mathscr{D}). \tag{27.46}$$

Hence, $\mathbb{E}(X|\mathscr{D})(\omega_1)$ is the expected value of X conditional on the event $\{\omega \in D_{i(\omega_1)}\}$.

Suppose that $\mathscr{D}$ and $\mathscr{D}'$ are two partitions of Ω. We say that $\mathscr{D}'$ is *finer* than $\mathscr{D}$ if $\mathscr{A}(\mathscr{D}) \subseteq \mathscr{A}(\mathscr{D}')$ and denote this as $\mathscr{D} \prec \mathscr{D}'$.

If $\mathscr{D}$ is a partition of Ω and Y is a discrete random variable defined on Ω, then Y is $\mathscr{D}$-measurable if $\mathscr{D}_Y \prec \mathscr{D}$, i.e., if the partition $\mathscr{D}$ is *finer* than the partition induced by Y. It simply means that Y takes constant values y_i on the atoms D_i of $\mathscr{D}$, so Y can be written as $Y = \sum_{i=1}^{m} y_i I_{D_i}$, where some y_i may be equal. Note that a random variable Y is $\mathscr{D}_0$-measurable if Y has a degenerate distribution, i.e., it takes a constant value on all $\omega \in \Omega$. Also, trivially, the random variable Y is $\mathscr{D}_Y$-measurable.

Note that if $\mathscr{D}'$ is finer than $\mathscr{D}$ then

$$\mathbb{E}(\mathbb{E}(X \mid \mathscr{D}') \mid \mathscr{D}) = \mathbb{E}(X \mid \mathscr{D}). \tag{27.47}$$

Indeed, if $\omega \in \Omega$ then

$$\begin{aligned}
&\mathbb{E}(\mathbb{E}(X \mid \mathscr{D}') \mid \mathscr{D})(\omega) = \\
&= \sum_{\omega' \in D_{i(\omega)}} \left(\sum_{\omega'' \in D'_{i(\omega')}} X(\omega'') \frac{\mathbb{P}(\omega'')}{\mathbb{P}(D'_{i(\omega')})} \right) \frac{\mathbb{P}(\omega')}{\mathbb{P}(D_{i(\omega)})} \\
&= \sum_{\omega'' \in D_{i(w)}} X(\omega'') \mathbb{P}(\omega'') \sum_{\omega' \in D'_{i(\omega'')}} \frac{\mathbb{P}(\omega')}{\mathbb{P}(D'_{i(\omega')}) \mathbb{P}(D_{i(\omega)})} \\
&= \sum_{\omega'' \in D_{i(w)}} X(\omega'') \mathbb{P}(\omega'') \sum_{\omega' \in D'_{i(\omega'')}} \frac{\mathbb{P}(\omega')}{\mathbb{P}(D'_{i(\omega'')}) \mathbb{P}(D_{i(\omega)})} \\
&= \sum_{\omega'' \in D_{i(w)}} X(\omega'') \frac{\mathbb{P}(\omega'')}{\mathbb{P}(D_{i(\omega)})} \\
&= \mathbb{E}(X \mid \mathscr{D})(\omega).
\end{aligned}$$

Note that despite all the algebra, (27.47) just boils down to saying that the properly weighted average of averages is just the average.

Finally, suppose a partition $\mathscr{D}$ of Ω is induced by a sequence of random variables $\{Y_1, Y_2, \ldots, Y_n\}$. We denote such partition as $\mathscr{D}_{Y_1, Y_2, \ldots, Y_n}$. Then the atoms of this partition are defined as

$$D_{y_1, y_2, \ldots, y_n} = \{\omega : Y_1(\omega) = y_1, Y_2(\omega) = y_2, \ldots, Y_n(\omega) = y_n\},$$

where the y_i range over all possible values of the Y_i's. $\mathscr{D}_{Y_1, Y_2, \ldots, Y_n}$ is then the *coarsest* partition such that $Y_1, Y_2, \ldots, Y_n$ are all constant over the atoms of the partition. For convenience, we simply write
$\mathbb{E}(X | Y_1, Y_2, \ldots, Y_n)$, instead of $\mathbb{E}(X | \mathscr{D}_{Y_1, Y_2, \ldots, Y_n})$.

Now we are ready to introduce an important class of *dependent* random variables called *martingales*.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a finite probability space and $\mathscr{D}_0 \prec \mathscr{D}_1 \prec \mathscr{D}_2 \prec \ldots \prec \mathscr{D}_n = \mathscr{D}^*$ be a nested sequence of partitions of Ω (a *filtration* of Ω), where $\mathscr{D}_0$ is a trivial partition, while $\mathscr{D}^*$ stands for the discrete partition (i.e., $\mathscr{A}(\mathscr{D}_0) = \{\emptyset, \Omega\}$, while $\mathscr{A}(\mathscr{D}^*) = 2^\Omega = \mathscr{F}$).

A sequence of random variables $X_0, X_1, \ldots, X_n$ is called (a) a *martingale*, (b) a *super-martingale*, (c) a *sub-martingale*, with respect to the partition $\mathscr{D}_0 \prec \mathscr{D}_1 \prec \mathscr{D}_2 \prec \ldots \prec \mathscr{D}_n = \mathscr{D}^*$ if

$$X_k \text{ is } \mathscr{D}_k\text{-measurable}$$

and

(a) $\mathbb{E}(X_{k+1} \mid \mathscr{D}_k) = X_k \quad k = 0, 1, \ldots, n-1.$

(b) $\mathbb{E}(X_{k+1} \mid \mathscr{D}_k) \leq X_k \quad k = 0, 1, \ldots, n-1.$

(c) $\mathbb{E}(X_{k+1} \mid \mathscr{D}_k) \geq X_k \quad k = 0, 1, \ldots, n-1.$

If the partition $\mathscr{D}$ of Ω is generated by a sequence of random variables $Y_1, \ldots, Y_n$ then the sequence $X_1, \ldots, X_n$ is called a martingale with respect to the sequence $Y_1, \ldots, Y_n$. In particular, when $Y_1 = X_1, \ldots, Y_n = X_n$, i.e., when $\mathscr{D}_k = \mathscr{D}_{X_1, \ldots, X_k}$, then we simply say that X is a martingale with respect to itself. Observe also that $\mathbb{E} X_k = \mathbb{E} X_1 = X_0$, for every k. Analogous statements hold for super- and sub-martingales.

Martingales are ubiquitous, we can obtain a martingale from essentially any random variable. Let $Z = Z(Y_1, Y_2, \ldots, Y_n)$ be a random variable defined on the random variables $Y_1, Y_2, \ldots, Y_n$. The sequence of random variables

$$X_k = \mathbb{E}(Z \mid Y_1, Y_2, \ldots, Y_k), \quad k = 0, 1, \ldots, n$$

is called the *Doob Martingale* of Z.

Theorem 27.15. *We have (i) $X_0 = \mathbb{E} Z$, (ii) $X_n = Z$ and (iii) the sequence $X_0, X_1, \ldots, X_n$ is a martingale with respect to (the partition defined by) $Y_1, Y_2, \ldots, Y_n$.*

Proof. Only (iii) needs to be explicitly checked.

$$\begin{aligned}
\mathbb{E}(X_k \mid Y_1, \ldots, Y_{k-1}) &= \mathbb{E}(\mathbb{E}(Z \mid Y_1, \ldots, Y_k) \mid Y_1, \ldots, Y_{k-1}) \\
&= \mathbb{E}(Z \mid Y_1, \ldots, Y_{k-1}) \\
&= X_{k-1}.
\end{aligned}$$

Here the second equality comes from (27.47). □

We next show how one can define the so called, *vertex and edge exposure martingales*, on the space of random graphs. Consider the binomial random graph

$\mathbb{G}_{n,p}$. Let us view $\mathbb{G}_{n,p}$ as a vector of random variables $(I_1, I_2, \ldots, I_{\binom{n}{2}})$, where I_i is the indicator of the event that the ith edge is present, with $\mathbb{P}(I_i = 1) = p$ and $\mathbb{P}(I_i = 0) = 1 - p$ for $i = 1, 2, \ldots, \binom{n}{2}$. These random variables are independent of each other. Hence, in this case, Ω consists of all $(0,1)$-sequences of length $\binom{n}{2}$.

Now given any graph invariant (a random variable) $X : \Omega \to \mathbb{R}$, (for example, the chromatic number, the number of vertices of given degree, the size of the largest clique, etc.), we will define a martingale generated by X and certain sequences of partitions of Ω.

Let the random variables $I_1, I_2, \ldots, I_{\binom{n}{2}}$ be listed in a lexicographic order. Define $\mathscr{D}_0 \prec \mathscr{D}_1 \prec \mathscr{D}_2 \prec \ldots \prec \mathscr{D}_n = \mathscr{D}^*$ in the following way: $\mathscr{D}_k$ is the partition of Ω induced by the sequence of random variables $I_1, \ldots, I_{\binom{k}{2}}$, and $\mathscr{D}_0$ is the trivial partition. Finally, for $k = 1, \ldots, n$,

$$X_k = \mathbb{E}(X \mid \mathscr{D}_k) = \mathbb{E}(X \mid \mathscr{D}_{I_1, I_2, \ldots, I_{\binom{k}{2}}}).$$

Hence, X_k is the conditional expectation of X, given that we "uncovered" the set of edges induced by the first k vertices of our random graph $\mathbb{G}_{n,p}$. A martingale determined through such a sequence of nested partitions is called a *vertex exposure martingale*.

An *edge exposure martingale* is defined in a similar way. The martingale sequence is defined as follows

$$X_k = \mathbb{E}(X \mid \mathscr{D}_k) = \mathbb{E}(X \mid \mathscr{D}_{I_1, I_2, \ldots, I_k}),$$

where $k = 1, 2, \ldots, \binom{n}{2}$, i.e., we uncover the edges of $\mathbb{G}_{n,p}$ one by one.

We next give upper bounds for both the lower and upper tails of the probability distributions of certain classes of martingales.

Theorem 27.16 (Azuma-Hoeffding bound). *Let $\{X_k\}_0^n$ be a sequence of random variables such that $|X_k - X_{k-1}| \le c_k$, $k = 1, \ldots, n$ and X_0 is constant.*

(a) If $\{X_k\}_0^n$ is a super-martingale then for all $t > 0$ we have

$$\mathbb{P}(X_n \ge X_0 + t) \le \exp\left\{ -\frac{t^2}{2\sum_{i=1}^n c_i^2} \right\}.$$

(b) If $\{X_k\}_0^n$ is a sub-martingale then for all $t > 0$ we have

$$\mathbb{P}(X_n \le X_0 - t) \le \exp\left\{ -\frac{t^2}{2\sum_{i=1}^n c_i^2} \right\}.$$

(c) If $\{X_k\}_0^n$ is a martingale then for all $t > 0$ we have

$$\mathbb{P}(|X_n - X_0| \geq t) \leq 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right\}.$$

Proof. We only need to prove (a), since (b), (c) will then follow easily, since $\{X_k\}_0^n$ is a sub-martingale iff $-\{X_k\}_0^n$ is a super-martingale and $\{X_k\}_0^n$ is a martingale iff it is a super-martingale and a sub-martingale.

Define the *martingale difference sequence* by $Y_1 = 0$ and

$$Y_k = X_k - X_{k-1} \quad , \quad k = 1, \ldots, n.$$

Then

$$\sum_{k=1}^n Y_k = X_n - X_0,$$

and

$$\mathbb{E}(Y_{k+1} \mid Y_0, Y_1, \ldots, Y_k) \leq 0. \tag{27.48}$$

Let $\lambda > 0$. Then

$$\begin{aligned}\mathbb{P}(X_n - X_0 \geq t) &= \mathbb{P}\left(\exp\left\{\lambda\sum_{i=1}^n Y_i\right\} \geq e^{\lambda t}\right) \\ &\leq e^{-\lambda t}\,\mathbb{E}\left(\exp\left\{\lambda\sum_{i=1}^n Y_i\right\}\right),\end{aligned}$$

by the Markov inequality.

Note that $e^{\lambda x}$ is a convex function of x, and since $-c_i \leq Y_i \leq c_i$, we have

$$\begin{aligned}e^{\lambda Y_i} &\leq \frac{1 - Y_i/c_i}{2}e^{-\lambda c_i} + \frac{1 + Y_i/c_i}{2}e^{\lambda c_i} \\ &= \cosh(\lambda c_i) + \frac{Y_i}{c_i}\sinh(\lambda c_i).\end{aligned}$$

It follows from (27.48) that

$$\mathbb{E}(e^{\lambda Y_n} \mid Y_0, Y_1, \ldots, Y_{n-1}) \leq \cosh(\lambda c_n). \tag{27.49}$$

We then see that

$$\mathbb{E}\left(\exp\left\{\lambda\sum_{i=1}^n Y_i\right\}\right)$$

$$= \mathbb{E}\left(\mathbb{E}(e^{\lambda Y_n} \mid Y_0, Y_1, \ldots, Y_{n-1}) \times \exp\left\{\lambda \sum_{i=1}^{n-1} Y_i\right\}\right)$$
$$\leq \cosh(\lambda c_n)\,\mathbb{E}\left(\exp\left\{\lambda \sum_{i=1}^{n-1} Y_i\right\}\right) \leq \prod_{i=1}^{n} \cosh(\lambda c_i).$$

The expectation in the middle term is over $Y_0, Y_1, \ldots, Y_{n-1}$ and the last inequality follows by induction on n.

By the above equality and the Taylor expansion, we get

$$\begin{aligned} e^{\lambda t}\,\mathbb{P}(X_n - X_0 \geq t) &\leq \prod_{i=1}^{n} \cosh(\lambda c_i) = \prod_{i=1}^{n} \sum_{m=0}^{\infty} \frac{(\lambda c_i)^{2m}}{(2m)!} \\ &\leq \prod_{i=1}^{n} \sum_{m=0}^{\infty} \frac{(\lambda c_i)^{2m}}{2^m m!} = \exp\left\{\frac{1}{2}\lambda^2 \sum_{i=1}^{n} c_i^2\right\}. \end{aligned}$$

Putting $\lambda = t/\sum_{i=1}^n c_i^2$ we arrive at the theorem. □

We end by describing a simple situation where we can apply these inequalities.

Lemma 27.17 (McDiarmid's Inequality)**.** *Let $Z = Z(W_1, W_2, \ldots, W_n)$ be a random variable that depends on n independent random variables $W_1, W_2, \ldots, W_n$. Suppose that*

$$|Z(W_1, \ldots, W_i, \ldots, W_n) - Z(W_1, \ldots, W_i', \ldots, W_n)| \leq c_i$$

for all $i = 1, 2, \ldots, n$ and $W_1, W_2, \ldots, W_n, W_i'$. Then for all $t > 0$ we have

$$\mathbb{P}(Z \geq \mathbb{E}Z + t) \leq \exp\left\{-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right\},$$

and

$$\mathbb{P}(Z \leq \mathbb{E}Z - t) \leq \exp\left\{-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right\}.$$

Proof. We consider the martingale

$$X_k = X_k(W_1, W_2, \ldots, W_k) = \mathbb{E}(Z \mid W_1, W_2, \ldots, W_k).$$

Then

$$X_0 = \mathbb{E}Z \text{ and } X_n = Z.$$

We only have to show that the martingale differences $Y_k = X_k - X_{k-1}$ are bounded. But,

$$|X_k(W_1, \ldots, W_k) - X_{k-1}(W_1, \ldots, W_{k-1})|$$

$$\leq \sum_{W_k', W_{k+1}, \ldots, W_n} |Z(W_1, \ldots, W_k, \ldots, W_n) - Z(W_1, \ldots, W_k', \ldots, W_n))|$$
$$\times \mathbb{P}(W_k') \prod_{i=k+1}^{n} \mathbb{P}(W_i)$$
$$\leq \sum_{W_k', W_{k+1}, \ldots, W_n} c_k \mathbb{P}(W_k') \prod_{i=k+1}^{n} \mathbb{P}(W_i)$$
$$= c_k.$$

□

27.8 Talagrand's Inequality

In this section we describe a concentration inequality that is due to Talagrand [828] that has proved to be very useful. It can often overcome the following problem with using Theorems 27.16, 27.17: If $\mathbb{E} X_n = O(n^{1/2})$ then the bounds they give are weak. Our treatment is a re-arrangement of the treatment in Alon and Spencer [36].

Let $\Omega = \prod_{i=1}^n \Omega_i$, where each Ω_i is a probability space and Ω has the product measure. Let $A \subseteq \Omega$ and let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \Omega$.

For $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ we let

$$d_\alpha(A, \mathbf{x}) = \inf_{\mathbf{y} \in A} \sum_{i: y_i \neq x_i} \alpha_i.$$

Then we define

$$\rho(A, \mathbf{x}) = \sup_{|\alpha|=1} d_\alpha(A, \mathbf{x}),$$

where $|\alpha|$ denotes the Euclidean norm, $(\alpha_1^2 + \cdots + \alpha_n^2)^{1/2}$.

We then define, for $t \geq 0$,

$$A_t = \{\mathbf{x} \in \Omega : \rho(A, \mathbf{x}) \leq t\}.$$

The following theorem is due to Talagrand [828]:

Theorem 27.18.

$$\mathbb{P}(A_t)(1 - \mathbb{P}(A_t)) \leq e^{-t^2/4}.$$

Theorem 27.18 follows from

Lemma 27.19.

$$\int_\Omega \exp\left\{\frac{1}{4}\rho^2(A, \mathbf{x})\right\} d\mathbf{x} \leq \frac{1}{\mathbb{P}(A)}.$$

Proof. Indeed, fix A and consider $X = \rho(A, \mathbf{x})$. Then,

$$1 - \mathbb{P}(A_t) = \mathbb{P}(X > t) = \mathbb{P}(e^{X^2/4} > e^{t^2/4}) \leq \mathbb{E}(e^{X^2/4})e^{-t^2/4}.$$

The lemma states that $\mathbb{E}(e^{X^2/4}) \leq \frac{1}{\mathbb{P}(A)}$. $\square$

The following alternative description of ρ is important. Let

$$U(A, \mathbf{x}) = \{\mathbf{s} \in \{0,1\}^n : \exists \mathbf{y} \in A \ s.t. \ s_i = 0 \text{ implies } x_i = y_i\}$$

and let $V(A, \mathbf{x})$ be the convex hull of $U(A, \mathbf{x})$. Then

Lemma 27.20.

$$\rho(A, \mathbf{x}) = \min_{\mathbf{v} \in V(A, \mathbf{x})} |\mathbf{v}|.$$

Here $|v|$ denotes the Euclidean norm of v. We leave the proof of this lemma as a simple exercise in convex analysis.

We now give the proof of Lemma 27.19.

Proof. We use induction on the dimension n. For $n = 1, \rho(A, \mathbf{x}) = 1_{\mathbf{x} \notin A}$ so that

$$\int_\Omega \exp\left\{\frac{1}{4}\rho^2(A, \mathbf{x})\right\} = \mathbb{P}(A) + (1 - \mathbb{P}(A))e^{1/4} \leq \frac{1}{\mathbb{P}(A)}$$

which follows from $u + (1-u)e^{1/4} \leq u^{-1}$ for $0 < u \leq 1$.

Assume the result for n. Write $\Psi = \prod_{i=1}^n \Omega_i$ so that $\Omega = \Psi \times \Omega_{n+1}$. Any $\mathbf{z} \in \Omega$ can be written uniquely as $\mathbf{z} = (\mathbf{x}, \omega)$ where $\mathbf{x} \in \Psi$ and $\omega \in \Omega_{n+1}$. Set

$$B = \{\mathbf{x} \in \Psi : (\mathbf{x}, \omega) \in A \text{ for some } \omega \in \Omega_{n+1}\}$$

and for $\omega \in \Omega_{n+1}$ set

$$A_\omega = \{\mathbf{x} \in \Psi : (\mathbf{x}, \omega) \in A\}.$$

Then

$$\mathbf{s} \in U(B, \mathbf{x}) \Longrightarrow (\mathbf{s}, 1) \in U(A, (\mathbf{x}, \omega)).$$
$$\mathbf{t} \in U(A_\omega, \mathbf{x}) \Longrightarrow (\mathbf{t}, 0) \in U(A, (\mathbf{x}, \omega)).$$

If $\mathbf{s} \in V(B, \mathbf{x})$ and $\mathbf{t} \in V(A_\omega, \mathbf{x})$ then $(\mathbf{s}, 1)$ and $(\mathbf{t}, 0)$ are both in $V(A, (\mathbf{x}, \omega))$ and hence for any $\lambda \in [0,1]$,

$$((1-\lambda)\mathbf{s} + \lambda\mathbf{t}, 1-\lambda) \in V(A, (\mathbf{x}, \omega)).$$

Then,

$$\rho^2(A,(\mathbf{x},\omega)) \le (1-\lambda)^2 + |(1-\lambda)\mathbf{s}+\lambda\mathbf{t}|^2 \le (1-\lambda)^2 + (1-\lambda)|\mathbf{s}|^2 + \lambda|\mathbf{t}|^2,$$

where the second inequality uses the convexity of $|\cdot|^2$.

Selecting, $\mathbf{s},\mathbf{t}$ with minimal norms yields the critical inequality,

$$\rho^2(A,(\mathbf{x},\omega)) \le (1-\lambda)^2 + \lambda\rho^2(A_\omega,\mathbf{x}) + (1-\lambda)\rho^2(B,\mathbf{x}).$$

Now fix ω and bound,

$$\int_{\mathbf{x}\in\Psi} \exp\left\{\frac{1}{4}\rho^2(A,(\mathbf{x},\omega))\right\} \le$$
$$e^{(1-\lambda)^2/4}\int_{\mathbf{x}\in\Psi} \exp\left\{\frac{1}{4}\rho^2(A_\omega,\mathbf{x})\right\}^{\lambda} \exp\left\{\frac{1}{4}\rho^2(B,\mathbf{x})\right\}^{1-\lambda}.$$

By Hölder's inequality this is at most

$$e^{(1-\lambda)^2/4}\left(\int_{\mathbf{x}\in\Psi} \exp\left\{\frac{1}{4}\rho^2(A_\omega,\mathbf{x})\right\}\right)^{\lambda} \left(\int_{\mathbf{x}\in\Psi} \exp\left\{\frac{1}{4}\rho^2(B,\mathbf{x})\right\}\right)^{1-\lambda},$$

which by induction is at most

$$e^{(1-\lambda)^2/4}\frac{1}{\mathbb{P}(A_\omega)^\lambda}\cdot\frac{1}{\mathbb{P}(B)^{1-\lambda}} = \frac{1}{\mathbb{P}(B)}e^{(1-\lambda)^2/4}r^{-\lambda}$$

where $r = \mathbb{P}(A_\omega)/\mathbb{P}(B) \le 1$.

Using calculus, we minimise $e^{(1-\lambda)^2/4}r^{-\lambda}$ by choosing $\lambda = 1+2\log r$ for $e^{-1/2} \le r \le 1$, $\lambda = 0$ otherwise. Further calculation shows that $e^{(1-\lambda)^2/4}r^{-\lambda} \le 2-r$ for this value of λ. Thus,

$$\int_{\mathbf{x}\in\Psi} \exp\left\{\frac{1}{4}\rho^2(A,(\mathbf{x},\omega))\right\} \le \frac{1}{\mathbb{P}(B)}\left(2-\frac{\mathbb{P}(A_\omega)}{\mathbb{P}(B)}\right).$$

We integrate over ω to give

$$\int_{\omega\in\Omega_{n+1}}\int_{\mathbf{x}\in\Psi} \exp\left\{\frac{1}{4}\rho^2(A,(\mathbf{x},\omega))\right\} \le$$
$$\frac{1}{\mathbb{P}(B)}\left(2-\frac{\mathbb{P}(A)}{\mathbb{P}(B)}\right) = \frac{1}{\mathbb{P}(A)}x(2-x),$$

where $x = \mathbb{P}(A)/\mathbb{P}(B) \le 1$. But $x(2-x) \le 1$, completing the induction and hence the theorem. □

We call $h : \Omega \to \mathbb{R}$ *Lipschitz* if $|h(\mathbf{x}) - h(\mathbf{y})| \le 1$ whenever $\mathbf{x},\mathbf{y}$ differ in at most one coordinate.

Definition 27.21. Let $f : \mathbb{N} \to \mathbb{N}$. h is *f-certifiable* if whenever $h(\mathbf{x}) \geq s$ then there exists $I \subseteq [n]$ with $|I| \leq f(s)$ so that if $\mathbf{y} \in \Omega$ agrees with $\mathbf{x}$ on coordinates I then $h(\mathbf{y}) \geq s$.

Theorem 27.22. *Suppose that h is Lipschitz and f-certifiable. Then if $X = h(\mathbf{x})$ for $\mathbf{x} \in \Omega$, then for all b and for all $t \geq 0$,*

$$\mathbb{P}(X \leq b - t\sqrt{f(b)})\,\mathbb{P}(X \geq b) \leq e^{-t^2/4}.$$

Proof. Set $A = \left\{\mathbf{x} : h(\mathbf{x}) < b - t\sqrt{f(b)}\right\}$. Now suppose that $h(\mathbf{y}) \geq b$. We claim that $\mathbf{y} \notin A_t$. Let I be a set of indices of size at most $f(b)$ that certifies $h(y) \geq b$ as given above. Define $\alpha_i = 0$ when $i \notin I$ and $\alpha_i = |I|^{-1/2}$ when $i \in I$. Using Lemma 27.20 we see that if $\mathbf{y} \in A_t$ then there exists a $\mathbf{z} \in A$ that differs from $\mathbf{y}$ in at most $t|I|^{1/2} \leq t\sqrt{f(b)}$ coordinates of I, though at arbitrary coordinates outside I. Let $\mathbf{y}'$ agree with $\mathbf{y}$ on I and agree with $\mathbf{z}$ outside I. By the certification $h(\mathbf{y}') \geq b$. Now $\mathbf{y}', \mathbf{z}$ differ in at most $t\sqrt{f(b)}$ coordinates and so, by Lipschitz,

$$h(\mathbf{z}) \geq h(\mathbf{y}') - t\sqrt{f(b)} \geq b - t\sqrt{f(b)},$$

but then $\mathbf{z} \notin A$, a contradiction. So, $\mathbb{P}(X \geq b) \leq 1 - \mathbb{P}(A_t)$ and from Theorem 27.18,

$$\mathbb{P}(X < b - t\sqrt{f(b)})\,\mathbb{P}(X \geq b) \leq e^{-t^2/4}.$$

As the RHS is continuous in t, we may replace "$<$" by "$\leq$" giving Theorem 27.22. □

Next let m denote the median of X so that $\mathbb{P}(X \geq m) \geq 1/2$ and $\mathbb{P}(X \leq m) \geq 1/2$.

Corollary 27.23.

(a) $\mathbb{P}(X \leq m - t\sqrt{f(m)}) \leq 2e^{-t^2/4}$.

(b) Suppose that $b - t\sqrt{f(b)} \geq m$, then $\mathbb{P}(X \geq b) \leq 2e^{-t^2/4}$.

27.9 Dominance

We say that a random variable X *stochastically dominates* a random variable Y if

$$\mathbb{P}(X \geq t) \geq \mathbb{P}(Y \geq t) \quad \text{for all real } t.$$

There are many cases when we want to use our inequalities to bound the upper tail of some random variable Y and (i) Y does not satisfy the necessary conditions to apply the relevant inequality, but (ii) Y is dominated by some random variable X that does. Clearly, we can use X as a surrogate for Y.

The following case arises quite often. Suppose that $Y = Y_1 + Y_2 + \cdots + Y_n$ where $Y_1, Y_2, \ldots, Y_n$ are not independent, but instead we have that for all t in the range $[A_i, B_i]$ of Y_i,

$$\mathbb{P}(Y_i \geq t \mid Y_1, Y_2, \ldots, Y_{i-1}) \leq \Phi(t)$$

where $\Phi(t)$ decreases monotonically from 1 to 0 in $[A_i, B_i]$.

Let X_i be a random variable taking values in the same range as Y_i and such that $\mathbb{P}(X_i \geq t) = \Phi(t)$. Let $X = X_1 + \cdots + X_n$ where $X_1, X_2, \ldots, X_n$ are independent of each other and $Y_1, Y_2, \ldots, Y_n$. Then we have

Lemma 27.24. *X stochastically dominates Y.*

Proof. Let $X^{(i)} = X_1 + \cdots + X_i$ and $Y^{(i)} = Y_1 + \cdots + Y_i$ for $i = 1, 2, \ldots, n$. We will show by induction that $X^{(i)}$ dominates $Y^{(i)}$ for $i = 1, 2, \ldots, n$. This is trivially true for $i = 1$ and for $i > 1$ we have

$$\begin{aligned}\mathbb{P}(Y^{(i)} \geq t \mid Y_1 \ldots, Y_{i-1}) &= \mathbb{P}(Y_i \geq t - (Y_1 + \cdots + Y_{i-1}) \mid Y_1 \ldots, Y_{i-1}) \\ &\leq \mathbb{P}(X_i \geq t - (Y_1 + \cdots + Y_{i-1}) \mid Y_1 \ldots, Y_{i-1}).\end{aligned}$$

Removing the conditioning we have

$$\mathbb{P}(Y^{(i)} \geq t) \leq \mathbb{P}(Y^{(i-1)} \geq t - X_i) \leq \mathbb{P}(X^{(i-1)} \geq t - X_i) = \mathbb{P}(X^{(i)} \geq t),$$

where the seond inequality follows by induction. □

Chapter 28

Differential Equations Method

Let $D \subseteq \mathbb{R}^2$ be open and bounded and connected. Consider a general random process

$$X(0), X(1), \ldots, X(t), \ldots, X(n) \in \mathbb{Z}.$$

where $X(0)$ is fixed and $\left(0, \frac{X(0)}{n}\right) \in D$.

Let H_t denote the history $X(0), X(1), \ldots, X(t)$ of the process to time t. Let T_D be the stopping time which is the minimum t such that $(t/n, X(t)/n) \notin D$. We further assume

(P1) $|X(t)| \leq C_0 n, \ \forall t < T_D$, where C_0 is a constant.

(P2) $|X(t+1) - X(t)| \leq \beta = \beta(n) \geq 1, \ \forall t < T_D$.

(P3) $|\mathbb{E}(X(t+1) - X(t)|H_t, \mathscr{E}) - f(t/n, X(t)/n)| \leq \lambda, \forall t < T_D$.
Here $\mathscr{E}$ is some likely event that holds with probability at least $1 - \gamma$.

(P4) $f(t,x)$ is continuous and satisfies a Lipschitz condition
$|f(t,x) - f(t',x')| \leq L\|(t,x) - (t',x)'\|_\infty$
for $(t,x), (t',x') \in D \cap \{(t,x) : t \geq 0\}$

Theorem 28.1. *Suppose that*

$$\lambda = o(1) \text{ and } \alpha = \frac{n\lambda^3}{\beta^3} \gg 1.$$

$$\sigma = \inf\{\tau : (\tau, z(\tau)) \notin D_0 = \{(t,z) \in D : l^\infty \text{ distance of } (t,z) \text{ from the boundary of } D \geq 2\lambda\}\}$$

Let $z(\tau)$, $0 \le \tau \le \sigma$ be the unique solution to the differential equation

$$z'(\tau) = f(\tau, z(\tau)) \tag{28.1}$$

$$z(0) = \frac{X(0)}{n} \tag{28.2}$$

Then,

$$X(t) = nz(t/n) + O(\lambda n), \tag{28.3}$$

uniformly in $0 \le t \le \sigma n$, with probability $1 - O(\gamma + \beta e^{-\alpha}/\lambda)$.

Proof. The γ in the probability of success will be handled by conditioning on $\mathscr{E}$. Now let

$$\omega = \left\lceil \frac{n\lambda}{\beta} \right\rceil.$$

We study the difference $X(t+\omega) - X(t)$. Assume that $(t/n, X(t)/n) \in D_0$. For $0 \le k \le \omega$ we have from (P2) that

$$\left| \frac{X(t+k)}{n} - \frac{X(t)}{n} \right| \le \frac{k\beta}{n} \le \frac{\omega\beta}{n},$$

so

$$\left\| \left(\frac{t+k}{n}, \frac{X(t+k)}{n} \right) - \left(\frac{t}{n}, \frac{X(t)}{n} \right) \right\|_\infty \le 2\lambda,$$

and so $\left(\frac{t+k}{n}, \frac{X(t+k)}{n} \right)$ is in D.

Therefore, using (P3),

$$\begin{aligned}
\mathbb{E}(X(t+k+1) - X(t+k) | H_{t+k}, \mathscr{E}) = \\
f\left(\frac{t+k}{n}, \frac{X(t+k)}{n} \right) + \theta_k = \\
f\left(\frac{t}{n}, \frac{X(t)}{n} \right) + \theta_k + \psi_k = \\
f\left(\frac{t}{n}, \frac{X(t)}{n} \right) + \rho,
\end{aligned}$$

where $|\rho| \le (2L+1)\lambda$, since $|\theta_k| \le \lambda$ (by (P3)) and $|\psi_k| \le \frac{L\beta k}{n}$ (by (P4)). Now, given H_t, let

$$Z_k = \begin{cases} X(t+k) - X(t) - kf\left(\frac{t}{n}, \frac{X(t)}{n} \right) - (2L+1)k\lambda & \mathscr{E} \\ \emptyset & \neg\mathscr{E} \end{cases}.$$

Then

$$\mathbb{E}(Z_{k+1} - Z_k | Z_0, Z_1, \ldots, Z_k) \leq 0,$$

i.e., $Z_0, Z_1, \ldots, Z_\omega$ is a super-martingale.
Also

$$|Z_{k+1} - Z_k| \leq \beta + \left| f\left(\frac{t}{n}, \frac{X(t)}{n}\right) \right| + (2L+1)\lambda \leq K_0 \beta,$$

where $K_0 = O(1)$, since $f\left(\frac{t}{n}, \frac{X(t)}{n}\right) = O(1)$ by continuity and boundedness of D.
So, using Theorem 27.16 we see that conditional on $H_t, \mathscr{E}$,

$$\mathbb{P}\left(X(t+\omega) - X(t) - \omega f(t/n, X(t)/n) \geq (2L+1)\omega\lambda + K_0 \beta \sqrt{2\alpha\omega}\right) \\ \leq \exp\left\{-\frac{2K_0^2 \beta^2 \alpha\omega}{2\omega K_0^2 \beta^2}\right\} = e^{-\alpha}. \quad (28.4)$$

Similarly,

$$\mathbb{P}\left(X(t+\omega) - X(t) - \omega f(t/n, X(t)/n) \leq -(2L+1)\omega\lambda - K_0 \beta \sqrt{2\alpha\omega}\right) \\ \leq e^{-\alpha}. \quad (28.5)$$

Thus

$$\mathbb{P}\left(|X(t+\omega) - X(t) - \omega f(t/n, X(t)/n)| \geq (2L+1)\omega\lambda + K_0 \beta \sqrt{2\alpha\omega}\right) \\ \leq 2e^{-\alpha}.$$

We have that $\omega\lambda$ and $\beta\sqrt{2\alpha\omega}$ are both $\Theta(n\lambda^2/\beta)$ giving

$$(2L+1)\omega\lambda + K_0 \beta \sqrt{2\alpha\omega} \leq K_1 \frac{n\lambda^2}{\beta}.$$

Now let $k_i = i\omega$ for $i = 0, 1, \ldots, i_0 = \lfloor \sigma n/\omega \rfloor$. We will show by induction that

$$\mathbb{P}\left(\exists j \leq i : |X(k_j) - z(k_j/n)n| \geq B_j\right) \leq 2ie^{-\alpha}, \quad (28.6)$$

where

$$B_j = B\left(\left(1 + \frac{L\omega}{n}\right)^{j+1} - 1\right) \frac{n\lambda}{L} \quad (28.7)$$

and where B is another constant.
The induction begins with $z(0) = \frac{X(0)}{n}$ and $B_0 = 0$. Note that

$$B_{i_0} \leq \frac{Be^{\sigma L}\lambda}{L} n = O(\lambda n). \quad (28.8)$$

Now write

$$
\begin{aligned}
&|X(k_{i+1}) - z(k_{i+1}/n)n| = |A_1 + A_2 + A_3 + A_4|,\\
&A_1 = X(k_i) - z(k_i/n)n,\\
&A_2 = X(k_{i+1}) - X(k_i) - \omega f(k_i/n, X(k_i)/n),\\
&A_3 = \omega z'(k_i/n) + z(k_i/n)n - z(k_{i+1}/n)n,\\
&A_4 = \omega f(k_i/n, X(k_i)/n) - \omega z'(k_i/n).
\end{aligned}
$$

We now bound each of these terms individually.

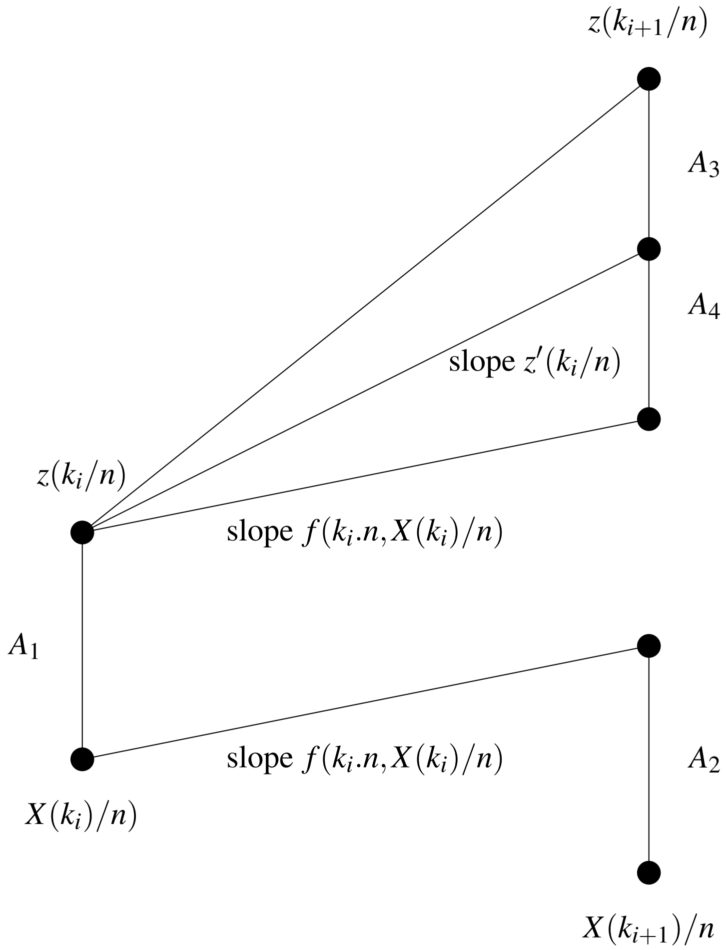


Figure 28.1: Error terms

Our induction gives that with probability at most $2ie^{-\alpha}$,

$$|A_1| \le B_i.$$

Equations (28.4) and (28.5) give

$$|A_2| \le K_1 \frac{n\lambda^2}{\beta},$$

with probability $1-2e^{-\alpha}$.

$$A_3 = \omega z'(k_i/n) + z(k_i/n)n - z(k_{i+1}/n)n$$

Now

$$z(k_{i+1}/n) - z(k_i/n) = \frac{\omega}{n} z'(k_i/n + \hat{\omega}/n)$$

for some $0 \le \hat{\omega} \le \omega$ and so (P4) implies that

$$|A_3| = \omega |z'(k_i/n + \omega/n) - z'(k_i/n + \hat{\omega}/n)| \le L\frac{\omega^2}{n} \le 2L\frac{n\lambda^2}{\beta^2}.$$

Finally, (P4) gives

$$|A_4| \le \frac{\omega L |A_1|}{n} \le \frac{\omega L}{n} B_i.$$

Thus for some $B > 0$,

$$\begin{aligned} B_{i+1} &\le |A_1| + |A_2| + |A_3| + |A_4| \\ &\le \left(1 + \frac{\omega L}{n}\right) B_i + Bn\frac{\lambda^2}{\beta}. \end{aligned}$$

A little bit of algebra verifies (28.6) and (28.7).

Finally consider $k_i \le t < k_{i+1}$. From "time" k_i to t the change in X and nz is at most $\omega\beta = O(n\lambda)$. □

Remark 28.2. The above proof generalises easily to the case where
$X(t)$ is replaced by $X_1(t), X_2(t), \ldots, X_a(t)$ where $a = O(1)$.

The earliest mention of differential equations with respect to random graphs was in the paper by Karp and Sipser [558]. The paper by Ruciński and Wormald [779] was also influential. See Wormald [860] for an extensive survey on the differential equations method.

Chapter 29

Branching Processes

In the Galton-Watson branching process, we start with a single particle comprising generation 0. In general, the nth generation consists of Z_n particles and each member x of this generation independently gives rise to a random number X of descendants in generation $n+1$. In the book we need the following theorem about the probability that the process continues indefinitely: Let

$$p_k = \mathbb{P}(X = k), \qquad k = 0, 1, 2, \dots.$$

Let

$$G(z) = \sum_{k=0}^{\infty} p_k z^k$$

be the probability generating function (p.g..f.) of X. Let $\mu = \mathbb{E}X$. Let

$$\eta = \mathbb{P}\left(\bigcup_{n\geq 0} \{Z_n = 0\}\right) \tag{29.1}$$

be the probability of ultimate extinction of the process.

Theorem 29.1. *η is the smallest non-negative root to the equation $G(s) = s$. Here $\eta = 1$ if $\mu < 1$.*

Proof. If $G_n(z)$ is the p.g.f. of Z_n, then $G_n(z) = G(G_{n-1}(z))$. This follows from the fact that Z_n is the sum of Z_{n-1} independent copies of G. Let $\eta_n = \mathbb{P}(Z_n = 0)$. Then

$$\eta_n = G_n(0) = G(G_{n-1}(0)) = G(\eta_{n-1}).$$

It follows from (29.1) that $\eta_n \nearrow \eta$. Let ψ be any other non-negative solution to $G(s) = s$. We have

$$\eta_1 = G(0) \leq G(\psi) = \psi.$$

Now assume inductively that $\eta_n \leq \psi$ for some $n \geq 1$. Then

$$\eta_{n+1} = G(\eta_n) \leq G(\psi) = \psi.$$

$\square$

Chapter 30

Random Walk

30.1 Mixing time

Spectral gap

We let $\Omega = \{0, 1, \ldots, N-1\}$ denote the state space of a Markov chain $\mathscr{M} = (X_0, X_1, \ldots)$. Let P be the transition matrix of $\mathscr{M}$ i.e. $P(i,j) = \Pr(X_{t+1} = j \mid X_t = i)$. We assume that $\mathscr{M}$ is *ergodic*, i.e that each state is reachable from every other state. In which case there will be a unique *steady state distribution* satisfying $\pi P = \pi$ and $\lim_{t\to\infty} P^t(i,j) = \pi_j$ for all i,j.

We assume that $\mathscr{M}$ is reversible, i.e. $\pi_i P(i,j) = \pi_j P(j,i)$ for all i,j. Note that a random walk on a graph G is reversible as can be seen by puting $\pi_i = \deg(i)/|E(G)|$ and $P(i,j) = 1/\deg(i)$.

Let the eigenvalues of P be $1 = l_0 > l_1 \geq l_2 \geq \cdots \geq l_{N-1}$. They are all real valued. Let $l_{\max} = \max\{|l_i| : i > 0\}$. Let $D^{1/2}$ be the diagonal $\Omega \times \Omega$ matrix with diagonal entries $\sqrt{\pi(\omega)}$, $\omega \in \Omega$ and let $D^{-1/2}$ be its inverse. Then the reversibility of of the chain implies that the matrix $S = D^{1/2} P D^{-1/2}$ is symmetric. It has the same eigenvalues as P and its symmetry means that these are all real.

The fact that $l_{\max} < 1$ is a classical result of the theory of non-negative matrices. The *spectral gap* $1 - l_{\max}$ determines the mixing rate of the chain i.e. how fast $P^t(i,j)$ tends to π_j. The larger it is, the more rapidly does the chain mix. For $U \subseteq \Omega$ let

$$\Delta_U(t) = \max_{i,j\in U} \left\{ \frac{|P^t(i,j) - \pi(j)|}{\pi(j)} \right\}.$$

Theorem 30.1.

$$\Delta_U \leq \frac{l_{\max}^t}{\min\limits_{i\in U} \pi(i)}.$$

Proof. We can select an orthonormal basis of column vectors $\mathbf{e}i, i \in \Omega$ for $\mathbb{R}^\Omega$ consisting of left eigenvectors of S where $\mathbf{e}i$ has associated eigenvalue l_i and $\mathbf{e}0 = \pi^T D^{-1/2}$. S has the *spectral decomposition*

$$S = \sum_{i=0}^{N-1} l_i \mathbf{e}i(\mathbf{e}i)^T = \sum_{i=0}^{N-1} l_i E^{(i)},$$

where $E^{(i)} = \mathbf{e}i(\mathbf{e}i)^T$. Note that $E^{(i)}E^{(j)} = 0$ for $i \neq j$ and ${E^{(i)}}^2 = E^{(i)}$. It follows that for any $t = 0, 1, 2, \ldots$, $S^t = \sum_{i=0}^{N-1} l_i^t E^{(i)}$. Hence

$$\begin{aligned} P^t = D^{-1/2} S^t D^{1/2} &= \sum_{i=0}^{N-1} l_i^t (D^{-1/2}\mathbf{e}i))((\mathbf{e}i)^T D^{1/2}) \\ &= \mathbf{1}_N \pi^T + \sum_{i=1}^{N-1} l_i^t (D^{-1/2}\mathbf{e}i)((\mathbf{e}i)^T D^{1/2}), \end{aligned}$$

where $\mathbf{1}_N$ is the N-vector all of whose components are 1. In component form, we get with the help of the Cauchy-Schwartz inequality:

$$\begin{aligned} |P^t(j,k) - \pi_k| &= \left| \sqrt{\frac{\pi_k}{\pi_j}} \sum_{i=1}^{N-1} l_i^t \mathbf{e}i_j \mathbf{e}i_k \right| \\ &\leq \sqrt{\frac{\pi_k}{\pi_j}} l_{\max}^t \left(\sum_{i=0}^{N-1} {\mathbf{e}i_j}^2 \right)^{1/2} \left(\sum_{i=0}^{N-1} {\mathbf{e}i_k}^2 \right)^{1/2} \\ &= \sqrt{\frac{\pi_k}{\pi_j}} l_{\max}^t. \end{aligned} \tag{30.1}$$

The theorem follows by substitution of the above inequality in the definition of Δ_U. □

In terms of mixing time (see (24.1)) we have

Corollary 30.2.

$$\tau(\varepsilon) \leq \left\lceil \frac{\log \varepsilon \pi_{\min}}{\log l_{\max}} \right\rceil.$$

Proof. For $A \subseteq \Omega$ we have

$$p_t(A) - \pi(A) \leq \frac{l_{\max}^t}{\pi_{\min}} \pi(A) \leq \frac{l_{\max}^t}{\pi_{\min}}. \tag{30.2}$$

□

Conductance

The conductance $\Phi = \Phi(\mathscr{M})$ of $\mathscr{M}$ is defined by

$$\Phi = \min\{\Phi_S : S \subseteq \Omega,\ 0 < \pi(S) \leq 1/2\}$$

where if $Q(i,j) = \pi_i P(i,j)$ and $\bar{S} = \Omega \setminus S$,

$$\Phi_S = \pi(S)^{-1} Q(S,\bar{S}).$$

Thus Φ_S is the steady state probability of moving from S to $\bar{S}$ in one step of the chain, conditional on being in S. We say that $\mathscr{M}$ is *lazy* if $P(i,i) \geq 1/2$ for $i \in \Omega$. Clearly $\Phi \leq 1/2$ if $\mathscr{M}$ is lazy. Note that

$$\Phi_S \pi(S) = Q(S,\bar{S}) = Q(\bar{S},S) = \Phi_{\bar{S}} \pi(\bar{S}). \tag{30.3}$$

Indeed,

$$Q(S,\bar{S}) = Q(\Omega,\bar{S}) - Q(\bar{S},\bar{S}) = \pi(\bar{S}) - Q(\bar{S},\bar{S}) = Q(\bar{S},S).$$

Let $\pi_{\min} = \min\{\pi(\omega) : \omega \in \Omega > 0\}$ and $\pi_{\max} = \max\{\pi(\omega) : \omega \in \Omega\}$.

We now show how conductance gives us an estimate of the spectral gap of a reversible chain. To make the chain lazy, we replace the transition matrix P by $Q = (P+I)/2$. (This introduces a loop of probability 1/2 at each state.)

Lemma 30.3. *If $\mathscr{M}$ is ergodic then replacing P by Q makes all the eigenvalues positive.*

Proof. $Q \geq 0$ is stochastic and has eigenvalues $\mu_i = (1+l_i)/2, i = 0,1,\ldots N-1$. The result follows from $l_i > -1, i = 0,1,\ldots N-1$, by the Perron-Frobenius theorem. □

For $y \in \mathbb{R}^N$ let

$$\mathscr{E}(y,y) = \sum_{i<j} \pi_i P_{i,j} (y_i - y_j)^2.$$

Lemma 30.4. *If $\mathscr{M}$ is reversible then*

$$1 - l_1 = \min_{\pi^T y = 0} \frac{\mathscr{E}(y,y)}{\sum_i \pi_i y_i^2}.$$

Proof. Let $D, S, \mathbf{e}0$ be as above. Then by the Rayleigh principle,

$$l_1 = \max_{\mathbf{x}:\pi^T D^{-1/2}\mathbf{x}=0} \frac{\mathbf{x}^T D^{1/2} P D^{-1/2} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}.$$

Thus

$$\begin{aligned} 1-l_1 &= \min_{\pi^T D^{-1/2}x=0} \frac{x^T D^{1/2}(I-P)D^{-1/2}x}{x^T x} \\ &= \min_{\pi^T y=0} \frac{y^T D(I-P)y}{y^T Dy}. \end{aligned} \tag{30.4}$$

Now

$$\begin{aligned} y^T D(I-P)y &= -\sum_{i\neq j} y_i y_j \pi_i P_{i,j} + \sum_i \pi_i (1-P_{i,i}) y_i^2 \\ &= -\sum_{i\neq j} y_i y_j \pi_i P_{i,j} + \sum_{i\neq j} \pi_i P_{i,j} \frac{y_i^2+y_j^2}{2} = \sum_{i<j} \pi_i P_{i,j} (y_i - y_j)^2 \\ &= \mathscr{E}(y,y), \end{aligned}$$

and the lemma follows from (30.4). □

Theorem 30.5. *If $\mathscr{M}$ is a reversible chain then*

$$1-l_1 \geq \frac{\Phi^2}{2}.$$

Proof. Assume now that $\pi^T y = 0$, $y_1 \geq y_2 \geq \cdots \geq y_N$ and that

$$\pi_1 + \pi_2 + \cdots + \pi_{r-1} \leq \frac{1}{2} < \pi_1 + \pi_2 + \cdots + \pi_r.$$

Let $z_i = y_i - y_r$ for $i = 1,2,\ldots,n$. Then $z_1 \geq z_2 \geq \cdots \geq z_r = 0 \geq z_{r+1} \geq \cdots \geq z_N$, and

$$\begin{aligned} \frac{\mathscr{E}(y,y)}{\sum_i \pi_i y_i^2} &= \frac{\mathscr{E}(z,z)}{-y_r^2 + \sum_i \pi_i z_i^2} \geq \frac{\mathscr{E}(z,z)}{\sum_i \pi_i z_i^2}. \\ &= \frac{\left(\sum_{i<j} \pi_i P_{i,j} (z_i - z_j)^2\right)\left(\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2\right)}{\left(\sum_i \pi_i z_i^2\right)\left(\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2\right)} = \frac{A}{B}, \qquad \text{say.} \end{aligned}$$

Now,

$$\begin{aligned} A &\geq \left(\sum_{i<j} \pi_i P_{i,j} |z_i - z_j| (|z_i| + |z_j|)\right)^2 \qquad \text{by Cauchy-Schwartz} \\ &\geq \left(\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|\right)^2. \end{aligned} \tag{30.5}$$

To prove (30.5), we show that if $i<j$ then

$$|z_i - z_j|(|z_i| + |z_j|) \geq \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|. \tag{30.6}$$

If $r \notin \{i, i+1, \ldots, j\}$ i.e., if z_i, z_j have the same sign then LHS(30.6)=RHS(30.6)=$|z_i^2 - z_j^2|$. Otherwise LHS(30.6)=$(|z_i| + |z_j|)^2$ and RHS(30.6)=$z_i^2 + z_j^2$, proving (30.5).

Now,

$$\sum_{i<j} \pi_i P_{i,j}(|z_i| + |z_j|)^2 \leq 2\sum_{i<j} \pi_i P_{i,j}(z_i^2 + z_j^2) \leq 2\sum_i \pi_i z_i^2.$$

So,

$$\frac{\mathscr{E}(y,y)}{\sum_i \pi_i y_i^2} \geq \frac{A}{B} \geq \frac{\left(\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|\right)^2}{2\left(\sum_i \pi_i z_i^2\right)^2.}$$

Now let $S_k = \{1, 2, \ldots, k\}$ and $C_k = \{(i,j) : i \leq k < j\}$. Then

$$\begin{aligned}
\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2| &= \sum_{k=1}^{N-1} |z_{k+1}^2 - z_k^2| \sum_{(i,j)\in C_k} \pi_i P_{i,j} \\
&\geq \Phi\left(\sum_{k=1}^{r-1} (z_k^2 - z_{k+1}^2)\pi(S_k) + \sum_{k=r}^{N-1} (z_{k+1}^2 - z_k^2)(1 - \pi(S_k))\right) \\
&= \Phi\left(\sum_{k=1}^{N-1} (z_k^2 - z_{k+1}^2)\pi(S_k) + (z_N^2 - z_r^2)\right) \\
&= \Phi\left(\sum_{k=1}^{N} \pi_k z_k^2\right)
\end{aligned}$$

since $z_r = 0$.

Thus if $\pi^T y = 0$ then

$$\frac{\mathscr{E}(y,y)}{\sum_i \pi_i y_i^2} \geq \frac{\Phi^2}{2}$$

and Theorem 30.5 follows. □

In terms of mixing time we obtain from Corollary 30.2,

Corollary 30.6. *If $\mathscr{M}$ is a lazy ergodic chain then*

$$\tau(\varepsilon) \leq \left\lceil \frac{2|\log \varepsilon \pi_{\min}|}{\Phi^2} \right\rceil.$$

Proof. Lemma 30.3 implies that $l_1 = l_{\max}$ and then

$$\frac{1}{\log l_{\max}^{-1}} \le \frac{1}{\log(1-\Phi^2/2)^{-1}} \le \frac{2}{\Phi^2}.$$

□

Now consider the conductance of a random walk on a graph $G=(V,E)$. Then, by definition,

$$\Phi_S = \frac{\displaystyle\sum_{(v,w)\in E(S,\bar{S})} \frac{\deg(v)}{2|E|}\frac{1}{d_v}}{\displaystyle\sum_{v\in S}\frac{\deg(v)}{2|E|}} = \frac{e(S:\bar{S})}{\displaystyle\sum_{v\in S}\deg(v)}. \tag{30.7}$$

In particular when G is an r-regular graph

$$\Phi = r^{-1} \min_{|S|\le|V|/2} \frac{e(S:\bar{S})}{|S|}. \tag{30.8}$$

The minimand above is referred to as the *expansion* of G. This graphs with good expansion (*expander graphs*) have large conductance and random walks on them mix rapidly.

We finish this section by proving a sort of converse to Theorem 30.5.

Theorem 30.7. *If $\mathscr{M}$ is a reversible chain then*

$$1 - l_1 \le 2\Phi$$

Proof. We use Lemma 30.4. Let S be a set of states which minimises Φ_S and define y by $y_j = \frac{1}{\pi(S)}$ if $j \in S$ and $y_j = -\frac{1}{\pi(\bar{S})}$ if $j \in \bar{S}$. It is easy to check that $\pi^T y = 0$. Then

$$\mathscr{E}(y,y) = \left(\frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}\right)^2 Q(S,\bar{S}) \text{ and } \sum \pi_i y_i^2 = \frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}.$$

Thus

$$1 - l_{\max} \le \Phi_S \pi(S)\left(\frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}\right) \le 2\Phi_S = 2\Phi.$$

□

30.2 First Visit Time Lemma

Let G denote a fixed connected graph, and let u be some arbitrary vertex from which a walk $\mathscr{W}_u$ is started. Let $\mathscr{W}_u(t)$ be the vertex reached at step t, let P be the matrix of transition probabilities of the walk, and let $P_u^{(t)}(v) = \Pr(\mathscr{W}_u(t) = v)$. Let π be the steady state distribution of the random walk $\mathscr{W}_u$. Let π_v denote the stationary distribution of the vertex v. For an unbiased ergodic random walk on a graph G with $m = m(G)$ edges, $\pi_v = \frac{\deg(v)}{2m}$, where $\deg(v)$ denotes the degree of v in G.

Let $d(t) = \max_{u,x\in V} |P_u^{(t)}(x) - \pi_x|$, and let T be such that, for $t \geq T$

$$\max_{u,x\in V} \left| \frac{P_u(x,t) - \pi_x}{\pi_x} \right| \leq \frac{1}{\omega} \tag{30.9}$$

where $\omega = \omega(n) \to \infty$.

It follows from e.g. Aldous and Fill [21] that $d(s+t) \leq 2d(s)d(t)$ and so for $t \geq T$ and $k = \lfloor t/T \rfloor$,

$$\max_{u,x\in V} \left| \frac{P_u(x,t) - \pi_x}{\pi_x} \right| \leq \frac{2^{k-1}}{\omega^k}. \tag{30.10}$$

Fix two vertices u,v. Let $h_t = \Pr(\mathscr{W}_u(t) = v)$ be the probability that the walk $\mathscr{W}_u$ visits v at step t. Let

$$H(z) = \sum_{t=T}^{\infty} h_t z^t \tag{30.11}$$

generate h_t for $t \geq T$. Next, considering the walk $\mathscr{W}_v$, starting at v, let $r_t = \Pr(\mathscr{W}_v(t) = v)$ be the probability that this walk returns to v at step $t = 0, 1, \ldots$. Let

$$R(z) = \sum_{t=0}^{\infty} r_t z^t$$

generate r_t. Our definition of return involves $r_0 = 1$.

For $t \geq T$ let $f_t = f_t(u,v)$ be the probability that the first visit of the walk $\mathscr{W}_u$ to v in the period $[T, T+1, \ldots]$ occurs at step t. Let

$$F(z) = \sum_{t=T}^{\infty} f_t z^t$$

generate f_t. Then we have

$$H(z) = F(z)R(z). \tag{30.12}$$

Finally, for $R(z)$ let

$$R_T(z) = \sum_{j=0}^{T-1} r_j z^j. \tag{30.13}$$

For a large constant $K > 0$, let

$$\eta = \frac{1}{KT}.$$

For $t \geq 0$, let $\mathscr{A}_t(v)$ be the event that $\mathscr{W}_u$ does not visit v in steps $T, T+1, \ldots, t$. The vertex u will have to be implicit in this definition.

Lemma 30.8. *[First Visit Time Lemma]*
Suppose that

(a) *For some constant $\theta > 0$, we have*

$$\min_{|z| \leq 1+l} |R_T(z)| \geq \theta. \tag{30.14}$$

(b) $T\pi_v = o(1)$.

Let

$$p_v = \frac{\pi_v}{R_T(1)(1 + O(T\pi_v))},$$

*where $R_v = R_T(1)$ is from (*30.13*).*

Then for all $t \geq T$,

$$\begin{aligned}
\Pr(\mathscr{A}_t(v)) &= \frac{(1 + O(T\pi_v))}{(1 + p_v)^t} + O(T^2\pi_v e^{-lt/2}) \\
&= \exp\left\{-\frac{t\pi_v}{R_v}(1 + O(T\pi_v))\right\} + O(T^2\pi_v e^{-lt/2}).
\end{aligned} \tag{30.15}$$

(The term $O(T^2\pi_v e^{-lt/2})$ is negligible and can be dropped for $t \geq n\log n$, as it is in (24.6).)

Proof. Write

$$R(z) = R_T(z) + \hat{R}_T(z) + \frac{\pi_v z^T}{1 - z}, \tag{30.16}$$

where $R_T(z)$ is given by (30.13) and

$$\hat{R}_T(z) = \sum_{t \geq T} (r_t - \pi_v) z^t$$

generates the error in using the stationary distribution π_v for r_t when $t \geq T$. Similarly,

$$H(z) = \widehat{H}_T(z) + \frac{\pi_v z^T}{1 - z}. \tag{30.17}$$

Equation (30.10) implies that the radii of convergence of both $\hat{R}_T$ and $\widehat{H}_T$ exceed $1+2l$. Moreover, for $Z = H, R$ and $|z| \le 1+l$, we see from (30.10) that

$$|\widehat{Z}_T(z)| \le \frac{\pi_v}{2} \sum_{t \ge T} \left(\frac{2(1+\eta)}{\omega} \right)^{\lfloor t/T \rfloor} \le \frac{2(1+\eta)T\pi_v}{\omega} = O(\omega^{-2}). \qquad (30.18)$$

Using (30.16), (30.17) we rewrite $F(z) = H(z)/R(z)$ from (30.12) as $F(z) = B(z)/A(z)$ where

$$\begin{aligned} A(z) &= \pi_v z^T + (1-z)(R_T(z) + \hat{R}_T(z)), && (30.19) \\ B(z) &= \pi_v z^T + (1-z)\widehat{H}_T(z). && (30.20) \end{aligned}$$

For real $z \ge 1$ and $Z = H, R$, we have

$$Z_T(1) \le Z_T(z) \le Z_T(1)z^T.$$

Let $z = 1 + \beta\pi_v$, where $0 \le \beta \le 1$. Since $T\pi_v \le \omega^{-1}$ we have

$$Z_T(z) = Z_T(1)(1+\xi_1) \;\text{ where } |\xi_1| \le (1+\beta\pi_v)^T - 1 \le \frac{2\beta}{\omega}.$$

$T\pi_v \le \omega^{-1}$ and $R_v \ge 1$ implies that

$$A(z) = \pi_v(1 - \beta R_v(1+\xi_1)) \text{ where } |\xi_1| = O(\omega^{-1}).$$

It follows that $A(z)$ has a real zero at z_0, where

$$z_0 = 1 + \frac{\pi_v}{R_v(1+\xi_1)} = 1 + p_v. \qquad (30.21)$$

We also see that since $|z_0^T| \le 1 + 2\omega^{-1}$,

$$\begin{aligned} A'(z_0) &= T\pi_v z_0^{T-1} - (R_T(z_0) + \hat{R}_T(z_0)) - p_v(R'_T(z_0) + \hat{R}'_T(z_0)) \\ &= O(\omega^{-1}) - \left(R_v + O(\omega^{-1}) + o(\omega^{-1})\right) - o(\pi_v) \\ &= -R_v + O(\omega^{-1}) \\ &\ne 0. \end{aligned}$$

and thus z_0 is a simple zero (see e.g. [201] p193). The value of $B(z)$ at z_0 is

$$B(z_0) = \pi_v\left(1 + O(\omega^{-1}) + o(\omega^{-1})\right) = \pi_v\left(1 + O(\omega^{-1})\right) \ne 0. \qquad (30.22)$$

Thus,

$$\frac{B(z_0)}{A'(z_0)} = -(1+\xi_2)\,p_v \text{ where } |\xi_2| = O(\omega^{-1}). \qquad (30.23)$$

Thus (see e.g. [201] p195) the principal part of the Laurent expansion of $F(z)$ at z_0 is

$$f(z) = \frac{B(z_0)/A'(z_0)}{z - z_0}. \tag{30.24}$$

To approximate the coefficients of the generating function $F(z)$, we now use a standard technique for the asymptotic expansion of power series (see e.g.[854] Theorem 5.2.1).

We prove below that $F(z) = f(z) + g(z)$, where $g(z)$ is analytic in $C_l = \{|z| \leq 1 + l\}$ and that

$$M = \max_{z \in C_l} |g(z)| = O(\omega^{-1}).$$

Let $a_t = [z^t]g(z)$, then (see e.g.[201] p143), $a_t = g^{(t)}(0)/t!$. By the Cauchy Inequality (see e.g. [201] p130) we see that $|g^{(t)}(0)| \leq Mt!/(1+\eta)^t$ and thus

$$|a_t| \leq \frac{M}{(1+\eta)^t} \leq Me^{-t\eta/2}.$$

As $[z^t]F(z) = [z^t]f(z) + [z^t]g(z)$ and $[z^t]1/(z - z_0) = -1/z_0^{t+1}$ we have

$$[z^t]F(z) = \frac{-B(z_0)/A'(z_0)}{z_0^{t+1}} + \eta_1(t) \text{ where } |\eta_1(t)| \leq Me^{-t\eta/2}. \tag{30.25}$$

Thus, we obtain

$$[z^t]F(z) = \frac{(1+\xi_2)p_v}{(1+p_v)^{t+1}} + \eta_1(t).$$

Now

$$\Pr(\mathbf{A}_t(v)) = \sum_{\tau > t} f_\tau(u \to v) = \sum_{\tau > t} \left(\frac{(1+\xi_2)p_v}{(1+p_v)^{\tau+1}} + \eta_1(\tau) \right) = \frac{1+\xi_2}{(1+p_v)^{t+1}} + \eta_2(t),$$

where

$$\eta_2(t) = \sum_{\tau > t} \eta_1(t) \leq \frac{Me^{-lt/2}}{1 - e^{-l/2}} = o(Te^{-l/2}).$$

This completes the proof of (30.15).

Now $M = \max_{z \in C_l} |g(z)| \leq \max |f(z)| + \max |F(z)| = O(T\pi_v) + \max |F(z)| = O(\omega^{-1}) + \max |F(z)|$, where $F(z) = B(z)/A(z)$. On C_l we have, using (30.18)-(30.20),

$$|F(z)| \leq \frac{\pi_v z^T + o(\pi_v)}{\pi_v z^T + l(|R_T(z)| - O(\omega^{-2}))} = O\left(\frac{\pi_v z^T}{T^{-1}R_v} \right) = O(\omega^{-1}).$$

We now prove that z_0 is the only zero of $A(z)$ inside the circle C_l and this implies that $F(z) - f(z)$ is analytic inside C_l. We use Rouché's Theorem (see e.g.

[201]), the statement of which is as follows: *Let two functions $\varphi(z)$ and $\gamma(z)$ be analytic inside and on a simple closed contour C. Suppose that $|\varphi(z)| > |\gamma(z)|$ at each point of C, then $\varphi(z)$ and $\varphi(z)+\gamma(z)$ have the same number of zeroes, counting multiplicities, inside C.*

Let the functions $\varphi(z), \gamma(z)$ be given by $\varphi(z) = (1-z)R_T(z)$ and $\gamma(z) = \pi_v z^T + (1-z)\hat{R}_T(z)$.

$$\frac{|\gamma(z)|}{|\varphi(z)|} \leq \frac{\pi_v(1+l)^T}{\eta\theta} + \frac{|\hat{R}_T(z)|}{\theta} = o(1).$$

As $\varphi(z)+\gamma(z) = A(z)$ we conclude that $A(z)$ has only one zero inside the circle C_l. This is the simple zero at z_0. □

In our use of Lemma 24.6 in Chapter 24.2, we never mentioned the condition (30.16). If $r_t \leq \rho^t + o(1/T)$ then

$$\begin{aligned} R_T(z) &\geq 1 - \left(\sum_{t=1}^{T/2}\left((\rho(1+\eta))^t + o\left(\frac{1}{T}\right)\right)\right) \\ &\geq \frac{\rho(1+\eta)}{1-\rho(1+\eta)} + o(1) > 0, \end{aligned}$$

for K sufficiently large.

This completes dealing with condition (30.16) in our context. We should note that Manzo, Quattropani, Scoppola [652] gave a completely different proof of Lemma 30.8 that removes condition (30.16) but does not explicitly give exponential tail bounds.

An early theorem We end with the classic theorem of Aleliunas, Karp, Lipton., Lovász and Rackoff [24].

Theorem 30.9. *If $G = (V,E)$ is a connected graph then*

$$C_G \leq 2|E|(|V|-1).$$

Proof. We present a proof due to Palacios [721]. We let $T_{u,v}$ denote the expected time for $\mathcal{W}_u$ to first reach v and $M = |E|, N = |V|$.

We double the edges of G and let $v_0, v_1, \ldots, v_{2M}$ be a walk through G that traverses each edge twice, once in each direction. Now

$$C_{v_0} \leq T_{v_0,v_1} + T_{v_1,v_2} + \cdots + T_{2M-1,2M}. \tag{30.26}$$

It is well-known, see for example [21], that $T_{u,u} = 2M/\deg(u)$, i.e. the inverse of the steady state probability of being at u. Therefore,

$$\frac{2M}{\deg(u)} = \sum_{v:\{u,v\}\in E} \frac{1}{\deg(u)}(1+T_{v,u}) = 1 + \frac{1}{\deg(u)} \sum_{v:\{u,v\}\in E} T_{v,u}$$

so that

$$2M = \deg(u) + \sum_{v:\{u,v\}\in E} T_{v,u}.$$

Summing over u, gives

$$2MN = 2M + \sum_{u,v:\{u,v\}\in E} T_{v,u}.$$

The theorem now follows from (30.26). $\square$

Chapter 31

Entropy

31.1 Basic Notions

Entropy is a useful tool in many areas. The entropy we talk about here was introduced by Shannon in [800]. We need some results on entropy in Chapter 12. We collect them here for convenience. For more on the subject we refer the reader to Cover and Thomas [275], or Gray [449] or Martin and England [674].

Let X be a random variable taking values in a finite set R_X. Let $p(x) = \mathbb{P}(X = x)$ for $x \in R_X$. Then the *entropy* of X is given by

$$h(X) = -\sum_{x \in R_X} p(x) \log p(x).$$

We have a choice for the base of the logarithm here. We use the natural logarithm, for use in Chapter 12.

Note that if X is chosen uniformly from R_X, i.e. $\mathbb{P}(X = x) = 1/|R_X|$ for all $x \in R_X$ then then

$$h(X) = \sum_{x \in R_X} \frac{\log |R_X|}{|R_X|} = \log |R_X|.$$

We will see later that the uniform distribution maximises entropy.

If Y is another random variable with a finite range then we define the *conditional entropy*

$$h(X \mid Y) = \sum_{y \in R_Y} p(y) h(X_y) = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)}, \tag{31.1}$$

where X_y is the random variable with $\mathbb{P}(X_y = x) = \mathbb{P}(X = x \mid Y = y)$. Here $p(y) = \mathbb{P}(Y = y)$. The summation is over y such that $p(y) > 0$. We will use notation like this from now on, without comment.

Chain Rule:

Lemma 31.1.

$$h(X_1,X_2,\dots,X_m) = \sum_{i=1}^{m} h(X_i \mid X_1,X_2,\dots,X_{i-1}). \tag{31.2}$$

Proof. This follows by induction on m, once we have verified it for $m=2$. For then

$$h(X_1,X_2,\dots,X_m) = h(X_1,X_2,\dots,X_{m-1}) + h(X_m \mid X_1,X_2,\dots,X_{m-1}).$$

Now,

$$\begin{aligned}
h(X_2 \mid X_1) &= -\sum_{x_1,x_2} p(x_1,x_2)\log\frac{p(x_1,x_2)}{p(x_1)} \\
&= -\sum_{x_1,x_2} p(x_1,x_2)\log p(x_1,x_2) + \sum_{x_1,x_2} p(x_1,x_2)\log p(x_1) \\
&= h(X_1,X_2) + \sum_{x_1} p(x_1)\log p(x_1) \\
&= h(X_1,X_2) - h(X_1).
\end{aligned}$$

□

Inequalities:

Entropy is a measure of uncertainty and so we should not be surprised to learn that $h(X \mid Y) \le h(X)$ for all random variables X,Y – here conditioning on Y represents providing information. Our goal is to prove this and a little more.

Let p,q be probability measures on the finite set X. We define the *Kullback-Liebler* distance

$$D(p||q) = \sum_{x\in A} p(x)\log\frac{p(x)}{q(x)}$$

where $A = \{x : p(x) > 0\}$.

Lemma 31.2.

$$D(p||q) \ge 0$$

with equality iff $p = q$.

Proof. Let

$$\begin{aligned}
-D(p||q) &= \sum_{x\in A} p(x)\log\frac{q(x)}{p(x)} \\
&\le \log\sum_{x\in A} p(x)\frac{q(x)}{p(x)} \qquad (31.3)
\end{aligned}$$

$$
\begin{aligned}
&= \log 1 \\
&= 0.
\end{aligned}
$$

Inequality (31.3) follows from Jensen's inequality and the fact that log is a concave function. Because log is strictly concave, will have equality in (31.3) iff $p = q$. $\square$

It follows from this that

$$h(X) \leq \log |R_X|. \tag{31.4}$$

Indeed, let u denote the uniform distribution over R_X i.e. $u(x) = 1/|R_X|$. Then

$$0 \leq D(p||u) = \sum_x p(x)(\log p(x) + \log |R_X|) = -h(X) + \log |R_X|.$$

We can now show that conditioning does not increase entropy.

Lemma 31.3. *For random variables X, Y, Z,*

$$h(X \mid Y, Z) \leq h(X \mid Y).$$

Taking Y to be a constant e.g. $Y = 1$ with probability one, we see

$$h(X \mid Z) \leq h(X).$$

Proof.

$$
\begin{aligned}
&h(X \mid Y) - h(X \mid Y, Z) \\
&= -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)} + \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)}{p(y,z)} \\
&= -\sum_{x,y,z} p(x,y,z) \log \frac{p(x,y)}{p(y)} + \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)}{p(y,z)} \\
&= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)p(y)}{p(x,y)p(y,z)} \\
&= D(p_{x,y,z}||p(x,y)p(y,z)/p(y)) \\
&\geq 0.
\end{aligned}
$$

Note that $\sum_{x,y,z} p(x,y)p(y,z)/p(y) = 1$. $\square$

Working through the above proof we see that $h(X) = h(X \mid Z)$ iff $p(x,z) = p(x)p(z)$ for all x, z, i.e. iff X, Z are independent.

31.2 Shearer's Lemma

The original proof is from Chung, Frankl, Graham and Shearer [219]. The following proof is from Radakrishnan [755].

Lemma 31.4. *Let $X = (X_1, X_2, \ldots, X_N)$ be a (vector) random variable and $\mathscr{A} = \{A_i :\ i \in I\}$ be a collection of subsets of $[N]$ such that each $j \in [N]$ appears in at least k members of $\mathscr{A}$. For $A \subseteq [N]$, let $X_A = (X_j : j \in A)$. Then,*

$$h(X) \leq \frac{1}{k} \sum_{i \in I} h(X_{A_i}).$$

Proof. We have, from Lemma 31.1 that

$$h(X) = \sum_{j \in [N]} h(X_j \mid X_1, X_2, \ldots, X_{j-1}) \tag{31.5}$$

and

$$h(X_{A_i}) = \sum_{j \in A_i} h(X_j \mid X_\ell,\ \ell \in A_i, \ell < j). \tag{31.6}$$

We sum (31.6) for all $i \in I$. Then

$$\begin{aligned}
\sum_{i \in I} h(X_{A_i}) &= \sum_{i \in I} \sum_{j \in A_i} h(X_j \mid X_\ell,\ \ell \in A_i, \ell < j) \\
&= \sum_{j \in [N]} \sum_{A_i \ni j} h(X_j \mid X_\ell,\ \ell \in A_i, \ell < j) && (31.7) \\
&\geq \sum_{j \in [N]} \sum_{A_i \ni j} h(X_j \mid X_1, X_2, \ldots, X_{j-1}) && (31.8) \\
&\geq k \sum_{j \in [N]} h(X_j \mid X_1, X_2, \ldots, X_{j-1}) && (31.9) \\
&= kh(X). && (31.10)
\end{aligned}$$

Here we obtain (31.8) from (31.7) by applying Lemma 31.3. We obtain (31.9) from (31.8) and the fact that each $j \in [N]$ appears in at least k A_i's. We then obtain (31.10) by using (31.5). □

Bibliography

[1] M.Abdullah and N. Fountoulakis, A phase transition on the evolution of bootstrap percolation processes on preferential attachment graphs, *Random Structures and Algorithms* 52 (2018) 379-418.

[2] M. Abdullah, C.Cooper and A.M. Frieze, Cover time of a random graph with given degree sequence, *Discrete Mathematics* 312 (2012) 3146-3163.

[3] D. Achlioptas, J-H. Kim, M. Krivelevich and P. Tetali, Two-Coloring Random Hypergraphs, *Random Structures and Algorithms* 20 (2002) 249-259.

[4] D. Achlioptas and C. Moore, Almost all Graphs with Average Degree 4 are 3-colorable, *Journal of Computer and System Sciences* 67 (2003) 441-471.

[5] D. Achlioptas and C. Moore, The Chromatic Number of Random Regular Graphs, *Proceedings of RANDOM 2004*.

[6] D. Achlioptas and A. Naor, The two possible values of the chromatic number of a random graph, *Annals of Mathematics* 162 (2005) 1335-1351.

[7] R. Adamczak and P. Wolff, Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order, *Probability Theory and Related Fields* 161 (2014) 1-56.

[8] W. Aiello, A. Bonato, C. Cooper, J. Janssen and P. Pralat, A spatial web graph model with local influence regions, *Internet Mathematics* 5 (2009) 175-196.

[9] E. Aigner-Horev, D. Hefetz, and M. Krivelevich, Minors, connectivity, and diameter in randomly perturbed sparse graphs. arXiv:2212.07192

[10] M. Ajtai, J. Komlós and E. Szemerédi, Largest random component of a k-cube, *Combinatorica* 2 (1982) 1-7.

[11] E. Aigner-Horev, D. Hefetz, and Krivelevich, Cycle lengths in randomly perturbed graphs. *Random Structures and Algorithms* 63 (2023) 867-884.

[12] E. Aigner-Horev, D. Hefetz, and M. Schacht, Ramsey properties of randomly perturbed hypergraphs. *APPROX/RANDOM 2024*, 59:1-59:18.

[13] M. Ajtai, J. Komlós and E. Szemerédi, Topological complete subgraphs in random graphs, *Studia Sci. Math. Hungar.* 14 (1979) 293-297.

[14] M. Ajtai, J. Komlós and E. Szemerédi, The longest path in a random graph, *Combinatorica* 1 (1981) 1-12.

[15] M. Ajtai, J. Komlós and E. Szemerédi, The first occurrence of Hamilton cycles in random graphs, *Annals of Discrete Mathematics* 27 (1985) 173-178.

[16] D. Aldous, Exchangeability and related topics,*Lecture Notes in Mathematics*, 1117, Springer Verlag, New York (1985)

[17] D. Aldous, Asymptotics in the random assignment problem, *Probability Theory and Related Fields* 93 (1992) 507-534.

[18] D. Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* 1 (1991) 228-266.

[19] D. Aldous, Brownian excursions, critical random graphs and the multiplicative coalescent, *Annals of Probability* 25 (1997) 812-854.

[20] D. Aldous, The $\zeta(2)$ limit in the random assignment problem, *Random Structures and Algorithms* 4 (2001) 381-418.

[21] D. Aldous and J. Fill, Reversible Markov Chains and Random Walks on Graphs, see https://www.stat.berkeley.edu/users/aldous/RWG/book.html

[22] D. Aldous, G. Miermont and J. Pitman, Brownian bridge asymptotics for random p-mappings, *Electronic J. Probability* 9 (2004) 37-56.

[23] D. Aldous, G. Miermont and J. Pitman, Weak convergence of random p-mappings and the exploration process of inhomogeneous continuum random trees, *Probab. Th. Related Fields* 133 (2005) 1-17.

[24] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff, Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems. *Proceedings of the 20th Annual IEEE Symposium on Foundations of Computer Science* (1979) 218-223.

[25] P. Allen, J. Böttcher, Y. Kohayakawa and Y. Person, Tight Hamilton cycles in random hypergraphs, see arxiv.org.

[26] N. Alon, Choice numbers of graphs; a probabilistic approach, *Combinatorics, Probability and Computing* 1 (1992) 107-114.

[27] N. Alon, Restricted colorings of graphs, in *Surveys of Combinatorics 1993, London Mathematical Society Lecture Notes Series 187 (K. Walker, ed.)*, Cambridge University Press 1993, 1-33.

[28] N.Alon, A note on network reliability, In Discrete Probability and Algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., 72, Springer, New York (1995) 11-14.

[29] N. Alon, M. Capalbo, Y. Kohayakawa, V. Rödl, A. Ruciński and E. Szemerédi, Universality and tolerance (extended abstract), In 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), IEEE Computer Society Press Los Alamitos, CA, 2000, pp. 14-21.

[30] N. Alon, P. Pralat, N. Wormald, Cleaning regular graphs with brushes, *SIAM Journal on Discrete Mathematics* 23 (2008) 233-250.

[31] N. Alon and Z. Füredi, Spanning subgraphs of random graphs, *Graphs and Combinatorics* 8 (1992) 91-94.

[32] N. Alon and M. Krivelevich, The concentration of the chromatic number of a random graph, *Combinatorica* 17 (1997) 303-313.

[33] N. Alon, M. Krivelevich and B. Sudakov, List coloring of random and pseudo random-graphs, *Combinatorica* 19 (1999) 453-472.

[34] N. Alon, M. Krivelevich and B. Sudakov, Finding a large hidden clique in a random graph, *Random Structures and Algorithm* 13 (1998) 457-466.

[35] N. Alon, M. Krivelevich and V. H. Vu, On the concentration of eigenvalues of random symmetric matrices, *Israel Journal of Mathematics* 131 (2002) 259-267.

[36] N. Alon and J. Spencer, The probabilistic method, Third Edition, John Wiley and Sons (2008).

[37] N. Alon and B. Sudakov, Increasing the chromatic number of a random graph, *Journal of Combinatorics* 1 (2010) 345–356.

[38] N. Alon and R. Yuster, Threshold functions for H-factors, *Combinatorics, Probability and Computing* 2 (1993) 137-144.

[39] A. Amit and N. Linial, Random Graph Coverings I: General Theory and Graph Connectivity, *Combinatorica* 22 (2002), 1-18.

[40] A. Amit and N. Linial, Random Lifts of Graphs II: Edge Expansion, *Combinatorics Probability and Computing* 15 (2006) 317-332.

[41] A. Amit, N. Linial and J. Matoušek, Random Lifts of Graphs III: Independence and Chromatic Number, *Random Structures and Algorithms* 20 (2002) 1-22.

[42] M. Anastos, S. Diskin, D. Elboim and M. Krivelevich, Climbing up a random subgraph of the hypercube, see arxiv.org.

[43] M. Anastos, D. Chakraborti, D. Kang, A. methuku and V. Pfenninger, Robust Hamiltonicity of Dirac hypergraph, see arxiv.org.

[44] O. Angel, A. Flaxman and D. Wilson, A sharp threshold for minimum bounded-depth and bounded-diameter spanning trees and steiner trees in random networks, *Combinatorica* (2012) 1-33.

[45] S. Anoulova, J. Bennies, J. Lenhard, D. Metzler, Y. Sung and A. Weber (1999) *The American Mathematical Monthly* 106 (199) 345-351.

[46] S. Antoniuk, A. Dudek, C. Reiher, A. Ruciński, and M. Schacht, High powers of Hamiltonian cycles in randomly augmented graphs. *J. Graph Theory* 98 (2021) 255–284.

[47] S. Antoniuk, A. Dudek, and A. Ruciński, Powers of Hamilton cycles in randomly augmentet Dirac graphs - the complete collection. *J. of Graph Theory* 104 (2023) 811-835.23.

[48] U. Ambroggio, Small maximal clusters are very unlikely in critical random graphs, see arxiv.org.

[49] U. Ambroggio and M. Roberts, Unusually large components in near-critical Erdős–Rényi graphs via ballot theorems, *Combinatorics, Probability and Computing* (2022) 31 (2022) 840-869.

[50] U. Ambroggio and M. Roberts, The probability of unusually large components for critical percolation on random d-regular graphs, *Electronic Journal of Probability* 28 (2023) 1-55.

[51] T. Apostol, Mathematical Analysis, Second Edition, Addison-Wesley, 1973.

[52] J. Aronson, A.M. Frieze and B. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser revisited, *Random Structures and Algorithms* 12 (1998) 111-178.

[53] I. Araujo, J. Balogh, R.A. Krueger, S. Piga, and A. Treglown, On oriented cycles in randomly perturbed digraphs. *Combinatorics, Probability and Computing* 33 (2024) 157-178.

[54] R. Arratia, A.D. Barbour and S. Tavaré, A tale of three couplings: Poisson-Dirichlet and GEM approximations for random permutations, *Combinatorics, Probability and Computing* 15 (2006) 31-62.

[55] R. Arratia and S. Tavaré, The cycle structure of random permutations, *The Annals of Probability* 20 (1992) 1567-1591.

[56] C. Avin and G. Ercal, On the Cover Time of Random Geometric Graphs, *Automata, Languages and Programming, Lecture Notes in Computer Science* 3580 (2005) 677-689.

[57] L.Babai, P. Erdős, S.M. Selkow, Random graph isomorphism, *SIAM Journal on Computing* 9 (1980) 628-635.

[58] L. Babai, M. Simonovits and J. Spencer, Extremal Subgraphs of Random Graphs, *Journal of Graph Theory* 14 (1990) 599-622.

[59] A. Backhaus and T.F. Móri, Local degree distribution in scale free random graphs, *Electronic Journal of Probability* 16 (2011) 1465-1488.

[60] D. Bal, P. Bennett, A. Dudek and A.M. Frieze, The t-tone chromatic number of random graphs, see arxiv.org.

[61] D. Bal, P. Bennett, A.M. Frieze and P. Pralat, Power of k choices and rainbow spanning trees in random graphs, see arxiv.org.

[62] J. Balogh, T. Bohman and D. Mubayi, Erdős-Ko-Rado in Random Hypergraphs, *Combinatorics, Probability and Computing* 18 (2009) 629-646.

[63] D. Bal and A.M. Frieze, Rainbow Matchings and Hamilton Cycles in Random Graphs, see arxiv.org.

[64] J. Balogh, B. Bollobás, M. Krivelevich, T. Müeller and M. Walters, Hamilton cycles in random geometric graphs, *Annals of Applied Probability* 21 (2011) 1053-1072.

[65] P. Balister and B. Bollobás, Percolation in the k-nearest neighbour graph.

[66] P. Balister, B. Bollobás, A. Sarkar and M. Walters, Connectivity of random k-nearest neighbour graphs, *Advances in Applied Probability* 37 (2005) 1-24.

[67] P. Balister, B. Bollobás, A. Sarkar and M. Walters, A critical constant for the k-nearest neighbour model, *Advances in Applied Probability* 41 (2009) 1-12.

[68] F. Ball, D. Mollison and G. Scalia-Tomba, Epidemics with two levels of mixing, *Ann. Appl. Prob.* 7 (1997) 46-89.

[69] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, see arxiv.org.

[70] J.Balogh, A.Treglown, and A.Z. Wagner, Tilings in randomly perturbed dense graphs. *Combinatorics Probability and Computing* 28(2019) 159–176.

[71] C. Banderier, P. Flajolet, G. Schaeffer, and M. Soria, Random maps, coalescing saddles, singularity analysis, and Airy phenomena, *Random Structures and Algorithms* 19 (2001) 194-246.

[72] L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* 286 (1999) 509-512.

[73] A.D Barbour, Poisson convergence and random graph, *Math. Proc. Camb. Phil. Soc.* 92 (1982) 349-359.

[74] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, 1992, Oxford University Press, Oxford.

[75] A.D. Barbour, S. Janson, M. Karoński and A. Ruciński, Small cliques in random graphs, *Random Structures and Algorithms* 1 (1990) 403-434.

[76] A.D. Barbour, M. Karoński and A. Ruciński, A central limit theorem for decomposable random variables with applications to random graphs, *J. Combin. Theory B* 47 (1989) 125-145.

[77] A.D. Barbour and G. Reinert, The shortest distance in random multi-type intersection graphs, *Random Structures Algorithms* 39 (2011) 179-209.

[78] A.D. Barbour and G. Reinert, Small worlds, *Random Structures and Algorithms* 19(2001) 54–74.

[79] A.D. Barbour and G. Reinert, Discrete smallworld networks, *Electronic Journal on Probability* 11 (2006) 1234–1283.

[80] M. Barlow, J. Ding, A. Nachmias and Y. Peres, The evolution of the cover time, The evolution of the cover time, *Combinatorics, Probability and Computing* 20 (2011) 331-345.

[81] T. Bartnicki, J. Grytczuk, H.A. Kierstead and X. Zhu, The Map-Coloring Game, *American Mathematical Monthly*, 114 (2007) 793-803.

[82] M. Bayati, D. Gamarnik and P. Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs, *Annals of Probability* 41 (2013) 4080-4115.

[83] J. Beck, Combinatorial Games: Tic-Tac-Toe theory, Encyclopedia of Mathematics and its Applications 114, Cambridge University Press 2008.

[84] W. Bedenknecht, J. Han, Y. Kohayakawa, and G. Oliveira Mota, Powers of tight Hamilton cycles in randomly perturbed hypergraphs. *Random Structures and Algorithms* 55 (2019) 795-807.

[85] M. Bednarska and T. Łuczak, *Biased positional games for which random strategies are nearly optimal*, Combinatorica 20 (2000) 477-488.

[86] M. Behrisch, Component evolution in random intersection graphs, *The Electronic Journal of Combinatorics* 14 (2007) #R17

[87] M. Behrisch, A. Taraz and M. Ueckerdt, Coloring random intersection graphs and complex networks, *SIAM J. Discrete Math.* 23 (2009) 288-299.

[88] T. Bell, The Park-Pham Theorem with Optimal Convergence Rate, see arxiv.org.

[89] I. Benjamini, G. Kozma and N. Wormald, The mixing time of the giant component of a random graph, *Random Structures and Algorithms* 45 (2014) 383-407.

[90] I. Benjamini, I. Shinkar, G. Tsur, Acquaintance time of a graph, see arxiv.org.

[91] E.A. Bender and E.R. Canfield, The asymptotic number of labelled graphs with given degree sequences, *Journal of Combinatorial Theory A* 24 (1978) 296-307.

[92] E.A. Bender, E.R. Canfield and B.D. McKay, The asymptotic number of labelled connected graphs, *Random Structures and Algorithms* 1 (1990) 127-169.

[93] E.A. Bender and E.R. Canfield, The number of rooted maps on an orientable surface. *Journal of Combinatorial Theory* B 53 (1991) 293-299.

[94] A. Bender, Z. Gao, and N. C. Wormald, The number of labeled 2-connected planar graphs, *Electronic Journal of Combinatorics* 9 (2002).

[95] I. Benjamini, G. Kozma and N. Wormald, The Mixing Time of the Giant Component of a Random Graph, *random Structures and Algorithms* 45 (2014) 383-407.

[96] P. Bennett and T. Bohman, A note on the random greedy independent set algorithm, see arxiv.org.

[97] S. Ben-Shimon, M. Krivelevich and B. Sudakov, Local resilience and Hamiltonicity Maker-Breaker games in random regular graph, *Combinatorics, Probability and Computing* 20 (2011) 173-211.

[98] S. Ben-Shimon, M. Krivelevich and B. Sudakov, On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs, *SIAM Journal of Discrete Mathematics* 25 (2011) 1176-1193.

[99] S. Ben-Shimon, A. Ferber, D. Hefetz and M. Krivelevich, Hitting time results for Maker-Breaker games, *Random Structures and Algorithms* 41 (2012) 23-46.

[100] S. Berg, Random contact processes, snowball sampling and and factorial series distributions, *Journal of Applied Probability* 20 (1983) 31-46.

[101] C. Berge, Graphs and Hypergraphs, North-Holland publishing company,Amsterdam 1973.

[102] F. Bergeron, P. Flajolet and B. Salvy, Varieties of increasing trees, *Lecture Notes in Computer Science* 581 (1992) 24-48.

[103] J.Bertoin, G. Uribe Bravo, Super-critical percolation on large scale-free random trees, *The Annals of Applied Probability* 25 (2015) 81-103.

[104] S.N. Bernstein, Theory of Probability, (Russian), Moscow, 1927.

[105] J. Bertoin, Sizes of the largest clusters for supercritical percolation on random recursive trees, *Random Structures Algorithms* 44 (2014) 29-44.

[106] A. Beveridge, A. Dudek, A.M. Frieze, T.Müller and M. Stojaković, Maker-Breaker games on random geometric graphs, see arxiv.org in *Random Structures and Algorithms*.

[107] A.Beveridge, A.M. Frieze and C.McDiarmid, Random minimum length spanning trees in regular graphs, *Combinatorica* 18 (1998) 311-333.

[108] S. Bhamidi and R. van der Hofstad, Weak disorder asymptotics in the stochastic mean-field model of distance, *Annals of Applied Probability* 22 (2012) 29-69.

[109] S. Bhamidi, R. van der Hofstad and G. Hooghiemstra, Universality for first passage percolation on sparse uniform and rank-1 random graphs, see arxiv.org.

[110] S. Bhamidi, S. Sen and X. Wang, Continuum limit of inhomogeneous random graphs, see arxiv.org.

[111] Ph. Biane, J. Pitman and M. Yor, Probability laws related to the Jacobi theta and Riemann zeta functions, and brownian excursions, *Bulletin of the AMS* 38 (2001) 435-465.

[112] S.R. Blackbourn and S. Gerke, Connectivity of the uniform random intersection graph, *Discrete Mathematics* 309 (2009) 5130-5140.

[113] P. Blackwell, M. Edmonson-Jones and J. Jordan, Spectra of adjacency matrices of random geometric graphs, see arxiv.org.

[114] M. Bloznelis, Degree distribution of a typical vertex in a general random intersection graph, *Lithuanian Mathematical Journal* 48 (2008) 38-45.

[115] M. Bloznelis, Component evolution in general random intersection graphs, *SIAM J. Discrete Mathematics* 24 (2010) 639-654.

[116] M. Bloznelis, A random intersection digraph: indegree and outdegree distributions, *Discrete Mathematics* 310 (2010) 2560-2566.

[117] M. Bloznelis, The largest component in an inhomogeneous random intersection graph with clustering, *The Electronic Journal of Combinatorics* 17 (2010) #R110.

[118] M. Bloznelis and J. Damarackas, Degree distribution of an inhomogeneous random intersection graph, *The Electronic Journal of Combinatorics* 20(3) (2013) #P3.

[119] M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas, and K. Rybarczyk, Recent progress in complex network analysis - models of random intersection graphs, In:*B. Lausen, S. Krolak-Schwerdt, M. Bhmer, Eds., Data Science, Learning by Latent Structures, and Knowledge Discovery, Springer* (2015) 59-68.

[120] M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas, and K. Rybarczyk, Recent progress in complex network analysis - properties of random intersection graphs, In:*B. Lausen, S. Krolak-Schwerdt, M. Bhmer, Eds., Data Science, Learning by Latent Structures, and Knowledge Discovery, Springer* (2015) 69-78.

[121] M. Bloznelis and F. Goetze, Preferred attachment model of affiliation network, *Journal of Statistical Physics* 156 (2014) 800-821.

[122] M. Bloznelis, F. Göetze and J. Jaworski, Birth of a strongly connected giant in an inhomogeneous random digraph, *Journal of Applied Probability* 49 (2012), 601-611.

[123] M. Bloznelis, J. Jaworski and K. Rybarczyk, Component evolution in a secure wireless sensor network, *Networks* 53 (2009) 19-26.

[124] M. Bloznelis and M. Karoński, Random intersection graph process, to appear in Internet Mathematics.

[125] M. Bloznelis and T. Łuczak, Perfect matchings in random intersection graphs, *Acta Mathematica Hungarica*, 138 (2013) 15-33.

[126] M. Bloznelis and I. Radavicius, A note on Hamiltonicity of uniform random intersection graphs, *Lithuanian Mathematical Journal*, 51 (2011) 155-161.

[127] M. Bloznelis and K. Rybarczyk, k-connectivity of uniform s-intersection graphs, *Discrete Mathematics*, 333 (2014) 94-100.

[128] M. Bode, N. Fountoulakis and T. Müller, The probability that the hyperbolic random graph is connected, see arxiv.org.

[129] M. Bode, M. Fountoulakis, T. Müller, On the largest component of a hyperbolic model of complex networks, see arxiv.org.

[130] E. Bodine-Baron, B. Hassibi and A. Wierman, Generalizing Kronecker graphs in order to model searchable networks, in *47th Annual Allerton Conference on Communication, Control, and Computing* (2009) 194-201.

[131] E. Bodine-Baron, B. Hassibi and A. Wierman, Distance-dependent Kronecker graphs for modeling social networks, *IEEE J. of Selected Topics in Signal Processing* 4 (2010) 718-731.

[132] T. Bohman, The Triangle-Free Process, *Advances in Mathematics* 221 (2009) 1653-1677.

[133] T. Bohman, C. Cooper, A.M. Frieze, R. Martin and M. Ruszinko, On Randomly Generated Intersecting Hypergraphs, *Electronic Journal on Combinatorics* 2003) R29.

[134] T. Bohman, A. M. Frieze, and R. Martin, How many random edges make a dense graph Hamiltonian? *Random Structures and Algorithms*, 22 (2003) 33-42.

[135] T. Bohman, A.M. Frieze, M. Krivelevich, and R. Martin, Adding random edges to dense graphs. *Random Structures and Algorithms*, 24 (2004) 105-117.

[136] T. Bohman, A.M. Frieze, R. Martin and M. Ruszinko and C. Smyth, On Randomly Generated Intersecting Hypergraphs II, *Random Structures and Algorithms* 30 (2007) 17-34.

[137] T. Bohman and A.M. Frieze, Avoiding a giant component, *Random Structures and Algorithms* (2001) 75-85.

[138] T. Bohman and A.M. Frieze, Hamilton cycles in 3-out, *Random Structures and Algorithms* 35 (2009) 393-417.

[139] T. Bohman, A.M. Frieze and E. Lubetzky, A note on the random greedy triangle-packing algorithm, *Journal of Combinatorics* 1 (2010) 477-488.

[140] T. Bohman, A.M. Frieze and E. Lubetzky, Random triangle removal, see arxiv.org.

[141] T. Bohman, A.M. Frieze, O. Pikhurko and C. Smyth, Anti-Ramsey Properties of Random Graphs, *Journal of Combinatorial Theory* B 100 (2010) 299-312.

[142] T. Bohman, A.M. Frieze and B.Sudakov, The game chromatic number of random graphs, *Random Structures and Algorithms* 32 (2008) 223-235.

[143] T. Bohman and P. Keevash, Dynamic concentration of the triangle-free process, see arxiv.org.

[144] T. Bohman and D. Kravitz, Creating a giant component, *Combinatorics, Probability and Computing* 15 (2006) 489-511.

[145] T. Bohman, D. Mubayi and M. Picolleil, The independent neighborhoods process, see arxiv.org.

[146] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal on Combinatorics* 1 (1980) 311-316.

[147] B. Bollobás, Random Graphs, In *Combinatorics, Proceedings, Swansea* (1981) 80-102.

[148] B. Bollobás, The diameter of random graphs, *Tranasactions of the American Mathematical Society* 267 (1981) 41-52.

[149] B. Bollobás, Degree sequence of random graphs, *Discrete Mathematics* 13 (1981) 1-19.

[150] B. Bollobás, Long paths in sparse random graphs, *Combinatorica* 2 (1982) 223-228.

[151] B. Bollobás, Vertices of given degree in a random graph, *Journal of Graph Theory* 6 (1982) 147-155.

[152] B. Bollobás, Distinguishing vertices of random graphs, *Annals of Discrete Mathematics* 13 (1982) 33-50.

[153] B. Bollobás, The evolution of random graphs, *Transactions of the Americam Mathematical Society* 286 (1984) 257-274.

[154] B. Bollobás, The evolution of sparse graphs, *Proceedings of a Cambridge Combinatorial Conference in honour of Paul Erős* (1984) 35-57.

[155] B. Bollobás, Random Graphs, First Edition, Academic Press, London 1985, Second Edition, Cambridge University Press, 2001.

[156] B. Bollobás, The chromatic number of random graphs, *Combinatorica* 8 (1988) 49-56.

[157] B. Bollobás, Modern Graph Theory, (1998) Springer, New York.

[158] B. Bollobás, Complete matchings in random subgraphs of the cube, *Random Structures and Algorithms* 1 (1990) 95-104.

[159] B. Bollobás, P. Catlin and P. Erdős, Hadwiger's conjecture is true for almost every graph, *European Journal of Combinatorics* 1 (1980) 195-199.

[160] B. Bollobás and F. Chung, The diameter of a cycle plus a random matching, *SIAM Journal on Discrete Mathematics* 1 (1988) 328-333.

[161] B. Bollobás, C. Cooper, T.I. Fenner and A.M. Frieze, On Hamilton cycles in sparse random graphs with minimum degree at least k, *Journal of Graph Theory* 34 (2000) 42-59.

[162] B. Bollobás and P. Erdős, Cliques in random graphs, *Mathematical Proceedings of the Cambridge Philosophical Society* 80 (1976) 419-427.

[163] B. Bollobás, T.I. Fenner and A.M. Frieze, An algorithm for finding hamilton paths and cycles in random graphs, *Combinatorica* 7 (1987) 327-341.

[164] B. Bollobás, T.I. Fenner and A.M. Frieze, Hamilton cycles in random graphs with minimal degree at least k, in *A tribute to Paul Erdos, edited by A.Baker, B.Bollobás and A.Hajnal* (1990) 59-96.

[165] B. Bollobás, T.I. Fenner and A.M. Frieze, Long cycles in sparse random graphs, In *Graph Theory and Combinatorics, Proceedings, Cambridge Combinatorial Conference in honour of Paul Erdős (B. Bollobás Ed.)* (1984) 59-64.

[166] B. Bollobás and A. Frieze, On matchings and hamiltonian cycles in random graphs, *Annals of Discrete Mathematics* 28 (1985) 23-46.

[167] B. Bollobás and A. Frieze, Spanning Maximal Planar Subgraphs of Random Graphs, *Random Structures and Algorithms* 2 (1991) 225-231.

[168] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures and Algorithms* 31 (2007) 3-122.

[169] B. Bollobás, D. Gamarnik, O. Riordan and B. Sudakov, On the Value of a Random Minimum Length Steiner Tree, *Combinatorica* 24 (2004) 187-207.

[170] B. Bollobás, Y. Kohayakawa and T. Łuczak, The evolution of random subgraphs of the cube, *Random Structures and Algorithms* 3 (1992) 55-90.

[171] B. Bollobás, Y. Kohayakava and T. Łuczak, On the diameter and radius of random subgraphs of the cube, *Random Structures and Algorithms*, 5 (1994) 627–648.

[172] B. Bollobás, Y. Kohayakawa and T. Łuczak, Connectivity properties of random subgraphs of the cube, *Random Structures and Algorithms* 6 (1995) 221-230.

[173] B. Bollobás, R. Kozma and D. Miklós, Handbook of Large-Scale Random Networks, Springer and János Bolyai Mathematical Society, Budapest.

[174] B. Bollobás and I. Leader, Exact face-isoperimetric inequalities, *Europ. J. Combinatorics*, 11 (1990) 335-340.

[175] B. Bollobás and O. Riordan, Constrained graph processes, *Electronic Journal of Combinatorics* 7 (2000) R18.

[176] B. Bollobás and O. Riordan, The diameter of a scale free random graph, *Combinatorica* 24 (2004) 5-34.

[177] B. Bollobás, O. Riordan, J. Spencer and G. Tusnády, The degree sequence of a scale-free random graph process, *Random Structures and Algorithms* 18 (2001) 279-290.

[178] B. Bollobás and A. Thomason, Threshold functions, *Combinatorica* 7 (1987) 35-38.

[179] A. Bonato and P. Pralat, Probabilistic Graph Searching, see arxiv.org.

[180] C. Borgs, M. Brautbar, J. Chayes, S. Khanna and B. Lucier, The Power of Local Information in Social Networks, see arxiv.org.

[181] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: I. The scaling window under the triangle condition, *Random Struct. Alg.* 27 (2005)137-184.

[182] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer, Random subgraphs of finite graphs: II. The lace expansion and the triangle condition, *Ann. Probab.* 33 (2005) 1886-1944.

[183] C. Borgs, J.T. Chayes, R. van der Hofstad, G. Slade and J. Spencer. Random subgraphs of finite graphs: III. The phase transition for the n-cube, *Combinatorica* 26 (2006) 395-410.

[184] J. Bourgain and G. Kalai, Influences of variables and threshold intervals under group symmetries, *Geometry and Functional Analysis* 7 (1997) 438-461.

[185] J. Bourgain, Appendix to [379].

[186] J. Böttcher, J. Han, Y. Kohayakawa, R. Montgomery, O. Parczyk, and Y. Person, Universality for bounded degree spanning trees in randomly perturbed graphs. *Random Structures and Algorithms* 55 (2019) 854-864.

[187] J. Böttcher, R. Montgomery, O. Parczyk, and Y. Person, Embedding spanning bounded degree graphs in randomly perturbed graphs. *Mathematika* 66 (2020) 422-447.

[188] J. Böttcher, O. Parczyk, A. Sgueglia, and J.Skokan,Cycle factors in randomly perturbed graphs. *Procedia Computer Science* 195 (2021) 404–411

[189] J. Böttcher, O. Parczyk, A. Sgueglia, and J.Skokan, triangles in randomly perturbed graphs. *Combinatorics, Probability and Computing* 32 (2023)91–121.

[190] J. Böttcher, O. Parczyk, A. Sgueglia, and J.Skokan, The square of a Hamilton cycle in randomly perturbed graphs. *Random Structures and Algorithms* 65 (2024) 342-386.

[191] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, Mixing Times for Random Walks on Geometric Random Graphs, *Proceedings of Analytic Algorithms and Combinatorics (ALENEX/ANALCO)* (2005) 240-249.

[192] M. Bradonjić, R. Elsässer, T. Friedrich, T. Sauerwald and A. Stauffer, Efficient broadcast on random geometric graphs, *21st Annual ACM-SIAM Symposium on Discrete Algorithms* (2010) 1412-1421.

[193] M. Bradonjić, A. Hagberg, N.W. Hengartner and A.G. Percus, Component evolution in general random intersection graphs, In *Algorithms and Models for the Web-Graph, LNCS 6516, Springer* (2010) 36-43.

[194] G.R. Brightwell and M.J. Luczak Vertices of high degree in the preferential attachment tree, *Elect. J. Probab.* 17 (2012) 43pp.

[195] G.R. Brightwell, K. Panagiotou and A. Steger, Extremal subgraphs of random graphs, *Random Structures and Algorithms* 41 (2012) 147-178.

[196] T. Britton, M. Deijfen, A. Lageras and M. Lindholm, Epidemics on random graphs with tunable clustering *J. Applied Probability*, 45 (2008) 743-756.

[197] T. Britton, M. Deijfen, and A. Martin-Löf, Generating simple random graphs with prescribed degree distribution, *J. Stat. Phys.* 124 (2006) 1377-1397.

[198] A. Broder, A.M. Frieze and E. Shamir, Finding hidden Hamilton cycles, *Random Structures and Algorithms* 5 (1994) 395-410.

[199] A. Broder, A.M. Frieze, S. Suen and E. Upfal, An efficient algorithm for the vertex-disjoint paths problem in random graphs, *Proceedings of SODA '96*, 261-268.

[200] A. Broder, A.M. Frieze, S. Suen and E. Upfal, Optimal construction of edge-disjoint paths in random graphs, *SIAM Journal on Computing* 28 (1999) 541-574.

[201] J. Brown and R. Churchill, *Complex Variables and Applications,* (Sixth Edition) McGraw-Hill (1996).

[202] T. Bui, S. Chaudhuri, T. Leightonand M. Sipser, Graph bisection algorithms with good average case behavior, *Combinatorica* 7 (1987) 171-191.

[203] J. D. Burtin, Extremal metric characteristics of a random graph *I, Teor. Veroyatnost. i Primenen* 19 (1974) 740-754.

[204] J. D. Burtin, Extremal metric characteristics of a random graph II, *Teor. Veroyatnost. i Primenen* 20 (1975) 82-99.

[205] Y.D. Burtin, On the probability of connectedness of a random subgraph of the n-cube, (in Russian) *Problemy Peredaci Informacii* 13 (1977) 90-95.

[206] Y.D. Burtin, On a simple formula for a random mappings and its applications, *J. Applied Probability* 17 (1980) 403-414.

[207] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman and S.H. Strogatz, Are randomly grown graphs really random? *Phys.Rev. E* 64 (2001) 041902.

[208] E. Candellero and N. Fountoulakis, Clustering and the hyperbolic geometry of complex networks, see arxiv.org.

[209] Y. Caro, A. Lev, Y. Roditty, Z. Tuza, and R. Yuster, *On rainbow connection*, *Electronic Journal of Combinatorics* 15 (2008).

[210] J. Cerny and A. Teixeira, Critical window for the vacant set left by random walk on random regular graphs, see arxiv.org.

[211] D. Chakrabarti, Y. Zhan and C. Faloutsos, R-MAT: A recursive model for graph mining, In: *Proceedings of 4th International Conference on Data Mining* (2004) 442-446.

[212] Y. Chang, J. Han, and L. Thoma, On powers of tight Hamilton cycles in randomly perturbed hypergraphs. *Random Structures and Algorithms* 63 (2023) 591-609.

[213] S. Chatterjee and A. Dembo, Nonlinear large deviations, see arxiv.org.

[214] P. Chebolu and A.M. Frieze, Hamilton cycles in random lifts of complete directed graphs, *SIAM Journal on Discrete Mathematics* 22 (2008) 520-540.

[215] P.Chebolu, A.M. Frieze and P.Melsted, Finding a Maximum Matching in a Sparse Random Graph in $O(n)$ Expected Time, *Journal of the ACM* 57 (2010) 1-27.

[216] L.H.Y. Chen, Poisson approximation for dependent trials, *Annals of Probability* 3 (1975) 534-545.

[217] W-C. Chen and W-C. Ni, Heap-ordered trees, 2-partitions and continued fractions, *European Journal of Combinatorics* 15 (1994) 513-517.

[218] W. Chen, W. Fang, G. Hu and M. Mahoney, On the hyperbolicity of small-world and tree-like random graphs, ISAAC 2012.

[219] F. Chung, F. Frank, R. Graham and J. Shearer, Some intersection theorems for ordered sets and graphs, *Journal of Combinatorial Theory A* 43 (86) 23-37.

[220] F. Chung and L. Lu, Connected components in random graphs with given expected degree sequence *Annals of Combinatorics*, 6 (2002) 125-145.

[221] F. Chung and L. Lu. The average distances in random graphs with given expected degree, *Internet Math.* 1 (2003) 91-113.

[222] F. Chung and L. Lu, The volume of the giant component of a random graph with given expected degrees, *SIAM Journal on Discrete Mathematics* 20 (2006) 395-411.

[223] F. Chung and L. Lu, Coupling on-line and off-line analyses for random power law graphs, *Internet Mathematics* 1 (2004) 409-461.

[224] F. Chung and L. Lu. *Complex Graphs and Networks*, American Mathematical Society, Providence, RI, 2006.

[225] F. Chung, L. Lu and V. Vu, Eigenvalues of random power law graphs, *Annals of Combinatorics* 7 (2003) 21-33.

[226] F. Chung, L. Lu and V. Vu, The spectra of random graphs with expected degrees, *Proceedings of the National Academy of Sciences* 100 (2003) 6313-6318.

[227] V. Chvátal and P. Erdős, A note on Hamiltonian circuits, *Discrete Math* 2 (1972) 111-113.

[228] V. Chvátal, DETERMINING THE STABILITY NUMBER OF A GRAPH, *SIAM Journal on Computing* 6 (1977) 643-662.

[229] V. Chvátal, Almost all graphs with 1.44n edges are 3-colorable, *Random Structures and Algorithms* 2 (1991) 11-28.

[230] V. Chvátal and P. Erdős, Biased positional games, *Annals of Discrete Mathematics* 2 (1978) 221-228.

[231] A. Coja–Oghlan, Graph Partitioning via Adaptive Spectral Techniques, *Combinatorics, Probability and Computing* 19 (2010) 227-284.

[232] A. Coja–Oghlan, On the Laplacian eigenvalues of $G(n,p)$, *Combinatorics, Probability and Computing* 16 (2007) 923-946.

[233] A. Coja–Oghlan, Upper bounding the k-colorability threshold by counting covers, *Electronic Journal of Combinatorics* 20 (2013) P32.

[234] A. Coja–Oghlan, K. Panagiotou and A. Steger, On the chromatic number of random graphs, *Journal of Combinatorial Theory, B* 98 (2008) 980-993.

[235] A. Coja–Oghlan and D. Vilenchik, Chasing the k-colorability threshold, see arxiv.org.

[236] P. Condon, A. Espuny Díaz, A. Girão, D. Kühn and D. Osthus, Hamiltonicity of random subgraphs of the hypercube, see arxiv.com

[237] N. A. Cook and A. Dembo, Large deviations of subgraph counts for sparse Erdő s–Rényi graphs, see arxiv.org.

[238] A. Collevecchio, A. Mehrabian and N. Wormald, Longest paths in random Apollonian networks and largest r-ary subtrees of random d-ary recursive trees, see arxiv.org.

[239] P. Condon, A. Espuny Dáz, J. Kim, D. Kühn and D. Osthus, Resilient degree sequences with respect to Hamilton cycles and matchings in random graphs, see arxiv.org.

[240] O. Cooley, M. Kang and C. Koch, The size of the giant component in random hypergraphs, see arxiv.org.

[241] D. Conlon, T. Gowers, Combinatorial theorems in sparse random sets, see arxiv.org.

[242] C. Cooper, Pancyclic Hamilton cycles in random graphs, *Discrete Mathematics* 91 (1991) 141-148.

[243] C. Cooper, On the thickness of sparse random graphs, *Combinatorics Probability and Computing* 1 (1992) 303-309.

[244] C. Cooper, 1-pancyclic Hamilton cycles in random graphs, *Random Structures and Algorithms* 3 (1992) 277-287.

[245] C. Cooper, The size of the cores of a random graph with a given degree sequence, *Random Structures and Algorithms* (2004) 353-375.

[246] C. Cooper, The age specific degree distribution of web-graphs, *Combinatorics Probability and Computing* 15 (2006) 637 - 661.

[247] C. Cooper and A.M. Frieze, On the number of hamilton cycles in a random graph, *Journal of Graph Theory* 13 (1989) 719-735.

[248] C. Cooper and A.M. Frieze, Pancyclic random graphs, in *Random Graphs, (Ed: Karonski M, Javorski J, and Rucinski A.)* Wiley (1990).

[249] C. Cooper and A.M. Frieze, The limiting probability that a-in, b-out is strongly connected, *Journal of Combinatorial Theory* B 48 (1990) 117-134.

[250] C. Cooper and A.M. Frieze, Multicoloured Hamilton cycles in random graphs: an anti-Ramsey threshold, *Electronic Journal of Combinatorics* 2, R19.

[251] C. Cooper and A.M. Frieze, On the connectivity of random k-th nearest neighbour graphs, *Combinatorics, Probability and Computing* 4 (196) 343-362.

[252] C. Cooper and A.M. Frieze, Hamilton cycles in random graphs and directed graphs, *Random Structures and Algorithms* 16 (2000) 369-401.

[253] C. Cooper and A.M. Frieze, Multi-coloured Hamilton cycles in randomly coloured random graphs, *Combinatorics, Probability and Computing* 11 (2002) 129-134.

[254] C. Cooper and A.M. Frieze, On a general model of web graphs *Random Structures and Algorithms* 22 (2003) 311-335.

[255] C. Cooper and A.M. Frieze, The Size of the Largest Strongly Connected Component of a Random Digraph with a Given Degree Sequence, *Combinatorics, Probability and Computing* 13 (2004) 319-337.

[256] C. Cooper and A.M. Frieze, The cover time of random regular graphs, *SIAM Journal on Discrete Mathematics* 18 (2005) 728-740.

[257] C. Cooper and A.M. Frieze,The cover time of sparse random graphs, *Random Structures and Algorithms* 30 (2007) 1-16.

[258] C. Cooper and A.M. Frieze, The cover time of the preferential attachment graph, *Journal of Combinatorial Theory* B 97 (2007) 269-290.

[259] C. Cooper and A.M. Frieze, The cover time of the giant component of a random graph, *Random Structures and Algorithms* 32 (2008) 401-439.

[260] C. Cooper and A.M. Frieze, The cover time of random geometric graphs, *Random Structures and Algorithms* 38 (2011) 324-349.

[261] C. Cooper and A.M. Frieze, Stationary distribution and cover time of random walks on random digraphs, *Journal of Combinatorial Theory* B 102 (2012) 329-362.

[262] C. Cooper and A.M. Frieze, Component structure induced by a random walk on a random graph, *Random structures and Algorithms* 42 (2013) 135-158.

[263] C. Cooper, A.M. Frieze and E. Lubetzky, Cover time of a random graph with given degree sequence II: Allowing vertices of degree two, *Random Structures and Algorithms* 45 (2014) 627-674.

[264] C. Cooper and A.M. Frieze, Long paths in random Apollonian networks, see arxiv.org.

[265] C. Cooper, A.M. Frieze, N. Ince, S. Janson and J. Spencer, On the length of a random minimum spanning tree, see arxiv.org.

[266] C.Cooper, A.M. Frieze and M.Krivelevich, Hamilton cycles in random graphs with a fixed degree sequence, *SIAM Journal on Discrete Mathematics* 24 (2010) 558-569.

[267] C.Cooper, A.M. Frieze and E. Lubetzky, Cover time of a random graph with given degree sequence II: Allowing vertices of degree two, *Random Structures and Algorithms* 45 (2014) 627-674.

[268] C.Cooper, A.M. Frieze and M. Molloy, Hamilton cycles in random regular digraphs, *Combinatorics, Probability and Computing* 3 (1994) 39-50.

[269] C.Cooper, A.M. Frieze, M. Molloy and B. Reed, Perfect matchings in random r-regular, s-uniform hypergraphs, *Combinatorics, Probability and Computing* 5 (1996) 1-15.

[270] C. Cooper, A,M, Frieze and P. Pralat, Some typical properties of the Spatial Preferred Attachment model, *Internet Mathematics* 10 (2014), 27-47.

[271] C.Cooper, A.M. Frieze and T. Radzik, The cover time of random walks on random uniform hypergraphs, Proceeedings of SIROCCO 2011, 210-221.

[272] C. Cooper, A.M. Frieze and B. Reed, Random regular graphs of non-constant degree: connectivity and Hamilton cycles, *Combinatorics, Probability and Computing* 11 (2002) 249-262.

[273] C. Cooper, A.M. Frieze and J. Vera, Random deletions in a scale free random graph process, *Internet Mathematics* 1 (2004) 463-483.

[274] D. Coppersmith and G. Sorkin, Constructive bounds and exact expectations for the random assignment problem, *Random Structures and Algorithms* 15 (1999) 133-144.

[275] T. Cover and J. Thomas, Elements of information theory, 2nd Edition, Wiley-Interscience, 2006.

[276] V. Dani and C. Moore, Independent sets in random graphs from the weighted second moment method, *Proceedings of RANDOM 2011* (2011) 472-482.

[277] A. Darrasse, H-K. Hwang and M. Soria, Shape measures of random increasing k-trees, manuscript (2013).

[278] S. Das, P. Morris, and A. Treglown, Vertex Ramsey properties of randomly perturbed graphs. *Random Structures and Algorithms* 57 (2020) 983-1006.

[279] S. Das and A. Treglown, Ramsey properties of randomly perturbed graphs: cliques and cycles. *Combinatorics, Probability and Computing* 29 (2020) 830-867.

[280] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, *Probab. Engrg. Inform. Sci.* 23 (2009) 661-674.

[281] D. Dellamonica, Jr., Y. Kohayakawa, V. Rödl, and A. Ruciński, An improved upper bound on the density of universal random graphs, *Random Structures and Algorithms*, see arxiv.org.

[282] B. DeMarco and J. Kahn, Tight upper tail bounds for cliques, *Random Structures and Algorithms* 41 (2012)

[283] B. DeMarco and J. Kahn, Mantel's theorem for random graphs, see arxiv.org.

[284] B. DeMarco and J. Kahn, Turán's theorem for random graphs, see arxiv.org.

[285] L. Devroye, Branching processes in the analysis of the heights of trees, *Acta Informatica* 24 (1987) 277-298.

[286] L. Devroye, O. Fawzi and N. Fraiman, Depth properties of scaled attachment random recursive tree, *Random Structures and Algorithms* 41 (2012) 66-98.

[287] L. Devroye and N. Fraiman, Connectivity in inhomogeneous random graphs, *Random Structures and Algorithms* 45 (2013) 408-420.

[288] L. Devroye and H-K. Hwang, Width and mode of the profile for some random trees of logarithmic height, *The Annals of Applied Probability* 16 (2006) 886-918.

[289] L. Devroye and S. Janson, Protected nodes and fringe subtrees in some random trees, *Electronic Communications in Probability* 19 (2014) 1-10.

[290] L. Devroy and J. Lu, The strong convergence of maximal degrees in uniform random recursive trees and dags, *Random Structure and Algorithms* 7 (1995) 1-14.

[291] J. Díaz, D. Mitsche and X. Pérez-Giménez, Sharp threshold for hamiltonicity of random geometric graphs, *SIAM Journal on Discrete Mathematics* 21 (2007) 57-65.

[292] J. Díaz, X. Pérez-Giménez, J. Serna and N. Wormald, Walkers on the Cycle and the Grid, *SIAM Journal on Discrete Mathematics* 22 (2008) 747-775.

[293] J. Díaz, F. Grandoni and A. Marchetti-Spaccamela, Balanced Cut Approximation in Random Geometric Graphs, *Theoretical Computer Science* 410 (2009) 527-536.

[294] J. Díaz, M.D. Penrose, J. Petit and M.J. Serna, Approximating layout problems on random geometric graphs, *Journal of Algorithms* 39 (2001) 78-116.

[295] R. Diestel, *Graph Theory*, Springer-Verlag, 1997.

[296] J. Ding, J.H. Kim, E. Lubetzky and Y. Peres, Anatomy of a young giant component in the random graph, *Random Structures and Algorithms* 39 (2011) 139-178.

[297] J. Ding, E. Lubetzky and Y. Peres, Mixing time of near-critical random graphs, *Annals of Probability* 40 3 (2012) 979-1008.

[298] R. DiPietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, Redoubtable sensor networks, *ACM Trans. Inform. Systems Security* 11 (2008) 1-22.

[299] S. Diskin and A. Geisler, Perfect matching in product graphs and in their random subgraphs, see arxiv.org.

[300] S. Diskin and M. Krivelevich, Hitting time of connectedness in the random hypercube process, (2024), see arxiv.org.

[301] R.P. Dobrow and R.T. Smythe, Poisson approximations for functionals of random trees, *Random Structures and Algorithms* 9 (1996) 79-92.

[302] M. Dondajewski and J. Szymański, On a generalized random recursive tree, unpublished manuscript.

[303] M. Drmota, *Random Trees*, Springer, Vienna, 2009.

[304] M. Drmota and H-K. Hwang, Profiles of random trees: Correlation and width of random recursive trees and binary search trees, *Adv. in Appl. Probab.* 37 (2005) 321-341.

[305] A. Dudek and A.M. Frieze, Loose Hamilton Cycles in Random k-Uniform Hypergraphs, *Electronic Journal of Combinatorics* (2011) P48.

[306] A. Dudek and A.M. Frieze, Tight Hamilton Cycles in Random Uniform Hypergraphs, *Random structures and Algorithms* 42 (2012) 374-385.

[307] A. Dudek, A.M. Frieze, P. Loh and S. Speiss Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs, *Electronic Journal of Combinatorics* 19 (2012).

[308] A. Dudek, A.M. Frieze, A. Ruciński and M. Šileikis, Loose Hamilton Cycles in Regular Hypergraphs, *Combinatorics, Probability and Computing* 24 (2015) 179-194.

[309] A. Dudek, A.M. Frieze, A. Ruciński and M. Šilekis, Embedding the Erdős-Rényi Hypergraph into the Random Regular Hypergraph and Hamiltonicity, see arxiv.org.

[310] A. Dudek, A.M. Frieze and C. Tsourakakis, Rainbow connection of random regular graphs, see arxiv.org.

[311] A. Dudek, D. Mitsche and P. Pralat, The set chromatic number of random graphs, see arxiv.org.

[312] A. Dudek and P. Pralat, An alternative proof of the linearity of the size-Ramsey number of paths, to appear in *Combinatorics, Probability and Computing*.

[313] A. Dudek, P. Pralat, Acquaintance time of random graphs near connectivity threshold, see arxiv.org.

[314] A. Dudek, C. Reiher, A. Ruciński, and M. Schacht, Powers of Hamiltonian cycles in randomly augmented graphs. *Random Structures Algorithms* 56 (2020) 122–141.

[315] R. Durrett, Rigoous result for CHKNS random graph model, In: *Proc. Discrete Random Walks (Paris 2003), Discrete Mathematics and Theoretical Computer Science AC* (2003) 95-104.

[316] R. Durrett, Random Graph Dynamics, *Cambridge University Press*, 2007.

[317] R. Durrett, *Probability: theory and examples*, Cambridge University Press, 2019.

[318] R. Durrett, Probability: Theory and Examples, 4th Edition, *Cambridge University Press* 2010.

[319] M. Dwass, The total progeny in a branching process, *Journal of Applied Probability* 6 (1969) 682-686.

[320] M.E.Dyer, A.M.Frieze and L.R. Foulds, On the strength of connectivity of random subgraphs of the n-cube, in *Random Graphs '85, Annals of Discrete Mathematics 33 (Karoński, Palka Eds.), North-Holland* (1987) 17-40.

[321] M. Dyer, A. Flaxman, A. M. Frieze and E. Vigoda, Random colouring sparse random graphs with fewer colours than the maximum degree, *Random Structures and Algorithms* 29 (2006) 450-465.

[322] M.E.Dyer, A.M.Frieze and C. Greenhill, On the chromatic number of a random hypergraph, see arxiv.org.

[323] M.E.Dyer, A.M.Frieze and C.J.H.McDiarmid, Linear programs with random costs, *Mathematical Programming* 35 (1986) 3-16.

[324] M.E.Dyer, A.M.Frieze and B. Pittel, The average performance of the greedy matching algorithm, *The Annals of Applied Probability* 3 (1993) 526-552.

[325] E. Ebrahimzadeh, L. Farczadi, P. Gao, A. Mehrabian, C. Sato, N. Wormald and J. Zung, On the Longest Paths and the Diameter in Random Apollonian Networks.

[326] J. Edmonds, Paths, Trees and Flowers, *Canadian Journal of Mathematics* 17 (1965) 449-467.

[327] C. Efthymiou, MCMC sampling colourings and independent sets of $G(n,d/n)$ near uniqueness threshold, *Proceedings of SODA 2014* 305-316.

[328] G.P. Egorychev, A solution of the Van der Waerden's permanent problem, *Preprint IFSO-L3 M Academy of Sciences SSSR, Krasnoyarks* (1980).

[329] J. Erde , M. Kang and M. Krivelevich, Expansion in supercritical random subgraphs of the hypercube and its consequences. see arxiv.com.

[330] P. Erdős, Graph Theory and Probability, *Canadian Journal of Mathematics* 11 (1959) 34-38.

[331] P. Erdős and A. Rényi, On random graphs I, *Publ. Math. Debrecen* 6 (1959) 290-297.

[332] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61.

[333] P. Erdős and A. Rényi, On the strength of connectedness of a random graph, *Acta. Math. Acad. Sci. Hungar.*, 8 (1961) 261-267.

[334] P. Erdős and A. Rényi, On random matrices, *Publ. Math. Inst. Hungar. Acad. Sci.* 8 (1964) 455-461.

[335] P. Erdős and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta. Math. Acad. Sci. Hungar. 17 (1966) 359-368.

[336] P. Erdős, M. Simonovits and V. T. Sós, Anti-Ramsey Theorems, *Colloquia Mathematica Societatis János Bolya* 10, Infinite and Finite Sets, Keszethely, 1973.

[337] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* 3 (1983) 181-192.

[338] P. Erdős and J. Spencer, Evolution of the n-cube, *Computers & Mathematics with Applications* 5 (1979) 33-39.

[339] P. Erdős, S. Suen and P. Winkler, On the size of a random maximal graph, *Random Structures and Algorithms* 6 (1995) 309-318.

[340] L. Eschenauer and V.D. Gligor, A key managment scheme for distributed sensor networks, In: *Proceedings of the* 9*th ACM Conference on Computer and Communication Security* (2002) 41-47.

[341] L. Espig, M. Krivelevich and W. Pegden, Walker-Breaker games *SIAM Journal on Discrete Mathematics* 29 (2015) 1476-1485.

[342] A. Espuny Díaz, Hamiltonicity of graphs perturbed by a random geometric graph. *J. Graph Theory* 103 (2023) 12-22.

[343] A. Espuny Díaz and A. Girão, Hamiltonicity of graphs perturbed by a random regular graph. *Random Structures Algorithms* 62 (2023) 857 - 886.

[344] A. Espuny Díaz, and J. Hyde, Powers of Hamilton cycles in dense graphs perturbed by a random geometric graph. *European Journal of Combinatorics* 121 C (2024) 14 pages.

[345] H. van den Esker, A geometric preferential attachment model with fitness, see arxiv.org.

[346] R. Fagin, Probabilities in Finite Models, *Journal of Symbolic Logic* 41 (1976) 50-58.

[347] D.I. Falikman, The proof of the Van der Waerden's conjecture regarding to doubly stochastic matrices, *Mat. Zametki* 29 (1981).

[348] M. Faloutsos, P. Faloutsos and C. Faloutsos, On Power-Law Relationships of the Internet Topology, *ACM SIGCOMM, Boston* (1999).

[349] V. Feldman, E. Grigorescu, L. Reyzin, S. Vempala and Y. Xiao, Statistical Algorithms and a lower bound for detecting planted cliques, STOC 2013.

[350] W. Feller, An Introduction to Probability Theory and its Applications, 3rd Ed., Wiley, New York (1968).

[351] Q. Feng, H.M. Mahmoud and A. Panholzer, Phase changes in subtree varieties in random recursive and binary search trees, *SIAM Journal Discrete Math.* 22 (2008) 160-184.

[352] A. Ferber, Closing gaps in problems related to Hamilton cycles in random graphs and hypergraphs, see arxiv.org.

[353] A. Ferber, R. Glebov, M. Krivelevich and A. Naor, Biased games on random boards, to appear in *Random Structures and Algorithms*.

[354] A. Ferber, G. Kronenberg and E. Long, Packing, Covering and Counting Hamilton Cycles in Random Directed Graphs, see arxiv.org.

[355] A. Ferber, G. Kronenberg, F. Mousset and C. Shikhelman, Packing a randomly edge-colored random graph with rainbow k-outs, see arxiv.org.

[356] A. Ferber, R. Nenadov and U. Peter, Universality of random graphs and rainbow embedding, see arxiv.org.

[357] A. Ferber, R. Nenadov, A. Noever, U. Peter and N. Škorić, Robust hamiltonicity of random directed graphs, see arxiv.org.

[358] T.I. Fenner and A.M. Frieze, On the connectivity of random m-orientable graphs and digraphs, *Combinatorica* 2 (1982) 347-359.

[359] D. Fernholz ande V. Ramachandran, The diameter of sparse random graphs, *Random Structures and Algorithms* 31 (2007) 482-516.

[360] J.A. Fill, E.R. Scheinerman and K.B. Singer-Cohen, Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution the evolution of the $G(n,m,p)$ and $G(n,p)$ models, *Random Structures and Algorithms* 16 (2000) 156-176.

[361] M. Fischer, N. Škorić, A. Steger and M. Trujić, Triangle resilience of the square of a Hamilton cycle in random graphs, see arxiv.org.

[362] G. Fiz Pontiveros, S. Griffiths and R. Morris, The triangle-free process and $R(3,k)$, see arxiv.org.

[363] A. Flaxman, The Lower Tail of the Random Minimum Spanning Tree, *The Electronic Journal of Combinatorics* 14 (2007) N3.

[364] A. Flaxman, A.M. Frieze and T. Fenner, High degree vertices and eigenvalues in the preferential attachment graph, *Internet Mathematics* 2 (2005) 1-20.

[365] A.Flaxman, A.M. Frieze and J. Vera, A Geometric Preferential Attachment Model of Networks, *Internet Mathematics* 3 (2007) 187-205.

[366] A.Flaxman, A.M. Frieze and J.Vera, A Geometric Preferential Attachment Model of Networks, *Internet Mathematics* 4 (2007) 87-112.

[367] A. Flaxman, A.M. Frieze and J. Vera, Adversarial deletions in a scale free random graph process, *Combinatorics, Probability and Computing* 16 (2007) 261-270.

[368] A. Flaxman, D. Gamarnik and G.B. Sorkin, Embracing the giant component, *Random Structures and Algorithms* 27 (2005) 277-289.

[369] J.E. Folkert, The distribution of the number of components of a random mapping function, PhD Dissertation, Michigan State University, (1955).

[370] C.M. Fortuin, P.W. Kasteleyn and J. Ginibre, Correlation inequalities on some partially ordered sets, Communications in Mathematical Physics 22 (1971) 89-103.

[371] N. Fountoulakis, On the evolution of random graphs on spaces of negative curvature, see arxiv.org.

[372] N. Fountalakis and K. Panagiotou, Sharp load thresholds for cuckoo hashing, *Random Structures and Algorithms* 41 (2012) 306-333.

[373] N. Fountalakis, M. Khosla and K. Panagiotou, The Multiple-orientability Thresholds for Random Hypergraphs, see arxiv.org.

[374] N. Fountoulakis, D. Kühn and D. Osthus, The order of the largest complete minor in a random graph, *Random Structures and Algorithms* 33 (2008) 127-141.

[375] N. Fountoulakis and B. A. Reed, The evolution of the mixing rate of a simple random walk on the giant component of a random graph, *Random Structures Algorithms* 33 (2008) 68-86.

[376] K. Frankston, J. Kahn, B. Narayanan and J. Park, Thresholds versus fractional expectation thresholds, *Annals of Mathematics* 194 (2021) 475-495.

[377] E. Friedgut, Hunting for Sharp Thresholds, *Random Structures Algorithms* 26 (2005) 37-51.

[378] E. Friedgut and G. Kalai, Every Monotone Graph Property Has a Sharp Threshold, *Proceedings of the American Mathematical Society* 124 (1996) 2993-3002.

[379] E. Friedgut, Sharp Thresholds of Graph Proprties, and the k-sat Problem, *Journal of the American Mathematical Society* 12 (1999) 1017-1054.

[380] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, Memoirs of the American Mathematical Society, 2008.

[381] T. Friedrich and A. Krohmer, Cliques in hyperbolic random graphs, IEEE INFOCOM (2015).

[382] T. Friedrich, T. Sauerwald and A. Stauffer, Diameter and Broadcast Time of Random Geometric Graphs in Arbitrary Dimensions, *Algorithmica* 67 (2013) 65-88.

[383] A.M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* 10 (1985) 47-56.

[384] A.M. Frieze, On large matchings and cycles in sparse random graphs, Discrete Mathematics 59 (1986) 243-256.

[385] A.M. Frieze, Maximum matchings in a class of random graphs, *Journal of Combinatorial Theory B* 40 (1986) 196-212.

[386] A.M. Frieze, An algorithm for finding hamilton cycles in random digraphs, *Journal of Algorithms* 9 (1988) 181-204.

[387] A.M. Frieze, On random regular graphs with non-constant degree, http://www.math.cmu.edu/ af1p/Texfiles/REGNONCONSTANT.pdf.

[388] A.M. Frieze, On the independence number of random graphs, *Discrete Mathematics* 81 (1990) 171-176.

[389] A.M. Frieze, Perfect matchings in random bipartite graphs with minimal degree at least 2, *Random Structures and Algorithms* 26 (2005) 319-358.

[390] A.M. Frieze, Loose Hamilton Cycles in Random 3-Uniform Hypergraphs, Electronic Journal of Combinatorics 17 (2010) N28.

[391] A.M. Frieze, On a Greedy 2-Matching Algorithm and Hamilton Cycles in Random Graphs with Minimum Degree at Least Three, *Random Structures and Algorithms* 45 (2014) 443-497.

[392] A.M. Frieze, N. Goyal, L. Rademacher and S. Vempala, Expanders via Random Spanning Trees, *SIAM Journal on Computing* 43 (2014) 497-513.

[393] A.M. Frieze and S. Haber, An almost linear time algorithm for finding Hamilton cycles in sparse random graphs with minimum degree at least three, to appear in *Random Structures and Algorithms*.

[394] A.M. Frieze, S. Haber and M. Lavrov, On the game chromatic number of sparse random graphs, *SIAM Journal of Discrete Mathematics* 27 (2013) 768-790.

[395] A.M. Frieze, S. Haber and M. M. Lavrov, On the game chromatic number of sparse random graphs, *SIAM Journal of Discrete Mathematics* 27 (2013) 768-790.

[396] A.M. Frieze, M.R. Jerrum, M.Molloy, R.Robinson and N.C.Wormald, Generating and counting Hamilton cycles in random regular graphs, *Journal of Algorithms* 21 (1996) 176-198.

[397] A.M. Frieze and T. Johansson, On random k-out sub-graphs of large graphs, see arxiv.org.

[398] A.M. Frieze and R. Kannan, A new approach to the planted clique problem, *Proceedings of Foundations of Software Technology and Theoretical Computer Science*, Bangalore, India (2008).

[399] A.M. Frieze and M. Krivelevich, On two Hamilton cycle problems in random graphs, *Israel Journal of Mathematics* 166 (2008) 221-234.

[400] A.M. Frieze, M. Krivelevich, P. Michaeli and R. Peled, On the trace of random walks on random graphs, *Proceedings of the London Mathematical Society* 16 (2018) 847-877.

[401] A.M. Frieze, M. Krivelevich and C. Smyth, On the chromatic number of random graphs with a fixed degree sequence, *Combinatorics, Probability and Computing* 16 (2007) 733-746.

[402] A.M. Frieze and P. Loh, Rainbow Hamilton cycles in random graphs, *Random Structures and Algorithms* 44 (2014) 328-354.

[403] A.M. Frieze and T. Łuczak, Hamiltonian cycles in a class of random graphs: one step further, in em Proceedings of Random Graphs '87, Edited by M.Karonski, J.Jaworski and A.Rucinski, John Wiley and Sons (1990) 53-59.

[404] A.M. Frieze and T. Łuczak, Edge disjoint trees in random graphs, *Periodica Mathematica Hungarica* 21 (1990) 28-30.

[405] A.M. Frieze and T. Łuczak, On the independence and chromatic numbers of random regular graphs, *Journal of Combinatorial Theory* 54 (1992) 123-132.

[406] A.M. Frieze and C.J.H. McDiarmid, On random minimum length spanning trees, *Combinatorica* 9 (1989) 363 - 374.

[407] A.M. Frieze and B.D. McKay, Multicoloured trees in random graphs, *Random Structures and Algorithms* 5 (1994) 45-56.

[408] A.M. Frieze and P. Melsted, Maximum Matchings in Random Bipartite Graphs and the Space Utilization of Cuckoo Hashtables, *Random Structures and Algorithms* 41 (2012) 334-364.

[409] A.M. Frieze, D. Mitsche, X. Pérez-Giménez and P. Pralat, On-line list colouring of random graphs, see arxiv.org.

[410] A.M. Frieze and W. Pegden, Between 2- and 3-colorability, see arxiv.org.

[411] A.M. Frieze and W. Pegden, Looking for vertex number one, see arxiv.org.

[412] A.M. Frieze, W. Pegden and T. Tkocz, On the cover time of the emerging giant, *SIAM Journal on Discrete Mathematic*s 36 (2022) 441 - 468.

[413] A.M. Frieze and B. Pittel, Perfect matchings in random graphs with prescribed minimal degree, *Trends in Mathematics, Birkhauser Verlag, Basel* (2004) 95-132.

[414] A.M. Frieze, M.Ruszinko and L.Thoma, A note on random minimum length spanning trees, *Electronic Journal of Combinatorics* 7 (2000) R41.

[415] A.M. Frieze and G. Sorkin, Efficient algorithms for three-dimensional axial and planar random assignment problems, *Random Structures and Algorithms* 46 (2015) 160-196.

[416] A.M. Frieze and C. Tsourakakis, High Degree Vertices, Eigenvalues and Diameter of Random Apollonian Networks, *WAW2012* (2012).

[417] A.M. Frieze and C. E. Tsourakakis, Rainbow connectivity of sparse random graphs, *Electronic Journal of Combinatorics* 19 (2012).

[418] A.M. Frieze and L. Zhao, Optimal Construction of Edge-Disjoint Paths in Random Regular Graphs, *Combinatorics, Probability and Computing* 9 (2000) 241-264.

[419] M. Fuchs, The subtree size profile of plane-oriented recursive trees, manuscript

[420] M. Fuchs, H-K. Hwang, R. Neininger, Profiles of random trees: Limit theorems for random recursive trees and binary search trees,*Algorithmica* 46 (2006) 367-407.

[421] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrics, *Combinatorica* 1 (1981) 233-241.

[422] D. Gale and L. S. Shapley, College Admissions and the Stability of Marriage, *American Mathematical Monthly* 69 (1962) 9-14.

[423] P. Gao and M. Molloy, The stripping process can be slow, see arxiv.org.

[424] P. Gao, X. Pérez-Giménez and C. M. Sato, Arboricity and spanning-tree packing in random graphs with an application to load balancing, see arxiv.org.

[425] P. Gao and N. Wormald, Orientability thresholds for random hypergraphs, see arxiv.org.

[426] J. Gastwirth, A probability model of a pyramid scheme, *American Statistician* 31 (1977) 79-82.

[427] H. Gebauer and T. Szabó, Asymptotic random graph intuition for the biased connectivity game, *Random Structures and Algorithms* 35 (2009) 431-443.

[428] S. Gerke, H-J. Prömel, T. Schickinger, A. Steger and A. Taraz, K_4-free subgraphs of random graphs revisited, *Combinatorica* 27 (2007) 329-365.

[429] S. Gerke, H-J. Prömel, T. Schickinger, K_5-free subgfraphs of random graphs, *Random Structures and Algorithms* 24 (2004) 194-232.

[430] S. Gerke, D. Schlatter, A. Steger and A. Taraz, The random planar graphs process, *Random Structures and Algorithms* 32 (2008) 236-261.

[431] I.B. Gertsbakh, Epidemic processes on random graphs: Some preliminary results, *J. Appl. Prob.* 14 (1977) 427-438.

[432] I. Gessel and R.P. Stanley, Stirling polynomials, *Journal of Combinatorial Theory A* 24 (1978) 24-33.

[433] E.N. Gilbert, Random graphs, *Annals of Mathematical Statistics* 30 (1959) 1141-1144.

[434] O. Giménez and M. Noy, Asymptotic enumeration and limit laws of planar graphs, *Journal of the American Mathematical Society* 22 (2009) 309-329.

[435] R. Glebov and M. Krivelevich, On the number of Hamilton cycles in sparse random graphs, *SIAM Journal on Discrete Mathematics* 27 (2013) 27-42.

[436] R. Glebov, H. Naves and B. Sudakov, The threshold probability for cycles, see arxiv.org.

[437] Y.V. Glebskii, D.I. Kogan, M.I. Liagonkii and V.A. Talanov, Range and degree of realizability of formulas in the restricted predicate calculus, *Cybernetics 5* (1969) 142-154.

[438] E. Godehardt and J.Jaworski, On the connectivity of a random interval graph, *Random Structures and Algorithms* 9 (1996) 137-161.

[439] E. Godehardt and J. Jaworski, Two models of random intersection graphs for classification, in: *Data Analysis and Knowledge Organization, Springer, Berlin*, 22 (2003) 67-82.

[440] E. Godehardt, J. Jaworski and K. Rybarczyk, On the connectivity of random intersection graphs, In: *Studies in Classification, Data Analysis and Knowledge Organization 33, Lenz H. J., Decker R., (Eds), Springer Verlag,Heidelberg - Berlin* (2007) 67-74.

[441] A. Goel, S. Rai and B. Krishnamachari, Monotone properties of Random geometric graphs have sharp thresholds, *Annals of Applied Probability* 15 (2005) 2535-2552.

[442] W. Goh, E. Schmutz, Limit distribution for the maximum degree of a random recursive tree, *Journal of Computing and Applied Mathematics* 142 (2002) 61-82.

[443] Ch. Goldschmidt and J.B. Martin, Random recursive trees and the Bolthausen-Sznitman coalescent, *Electronic Journal of Probability* 10 (2005) 718-745.

[444] S.W. Golomb, Random permutations, *Bulletin of the American Mathematical Society* 70 (1964) 747.

[445] E. Gomez-Leos and R.R. Martin, Tiling randomly perturbed bipartite graphs, arXiv.2411.12738.

[446] V.L. Goncharov, Sur la distribution des cycles dans les permutations *C.R. Acad. Sci USSR (N.S.)* 35, (1942) 267-269.

[447] V.L. Goncharov, Some facts in combinatorics, *Izv. Acad. Nauk SSSR, Ser. Mat.* 8 (1944) 3-48 (in Russian). [See also: On the field of combinatory analysis, *Trans. Amer. Math. Soc.* 19 (1962) 1-46.]

[448] D. A. Grable, On random greedy triangle packing, *Electronic Journal on Combinatorics* 4 (1997) R11.

[449] R. Gray, Entropy and Information Theory, Springer, 2011.

[450] C. Greenhill, S. Janson and A. Ruciński, On the number of perfect matchings in random lifts, *Combinatorics, Probability and Computing* 19 (2010) 791-817.

[451] C. Greenhill, S. Janson, J.H. Kim, N.C. Wormald, Permutation pseudographs and contiguity, *Combinatorics, Probability and Computing* 11 (2002) 273-298.

[452] G.R. Grimmett, Random labelled trees and their branching networks, *Journal of the Australian Mathematical Society* 30 (1980) 229-237.

[453] G. Grimmett and C. McDiarmid, On colouring random graphs, *Mathematical Proceedings of the Cambridge Philosophical Society* 77 (1975) 313-324.

[454] C. Groër, B.D. Sullivan and S. Poole, A mathematical analysis of the R-MAT random graph generator, *Networks*,58 (2011) 159-170.

[455] L. Gugelmann, K. Panagiotou and U. Peter, Random hyperbolic graphs: Degree sequence and clustering, *Proceedings of the 39th International Colloquium Conference on Automata, Languages, and Programming (ICALP)* (2012) 573-585.

[456] P. Gupta and P.R. Kumar, Critical Power for Asymptotic Connectivity in Wireless Networks, in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming, W. M. McEneany, G. Yin, and Q. Zhang, Eds. Boston, MA: Birkhauser*, (1998) 547-566.

[457] Y. Gurevich and S. Shelah, Expected computation time for Hamiltonian path problem, *SIAM Journal on Computing* 16 (1987) 486-502.

[458] E. Győri, B. Rorthchild and A. Ruciński, Every graph is contained in a sparest possible balanced graph, *Mathematical Proceedings of the Cambridge Philosophical Society* 98 (1985) 397-401.

[459] S. Haber and M. Krivelevich, The logic of random regular graphs, *Journal of Combinatorics* 1 (2010) 389-440.

[460] M. Hahn-Klimroth, G. Maesaka, Y. Mogge, S. Mohr, and O. Parczyk, Random perturbation of sparse graphs. *The Electronic Journal of Combinatorics* 28 (2021) P2.26.

[461] On resilience of connectivity in the evolution of random graphs, see arxiv.org.

[462] A. Hamm and J. Kahn, On Erdős-Ko-Rado for random hypergraphs I, see arxiv.org.

[463] A. Hamm and J. Kahn, On Erdős-Ko-Rado for random hypergraphs II, see arxiv.org.

[464] J. Han and Y. Zhao, Hamiltonicity in randomly perturbed hypergraphs. *J. Combin. Theory Ser. B* 144 (2020) 14-31.

[465] J. Han, P. Morris, and A. Treglown, Tilings in randomly perturbed graphs: bridging the gap between Hajnal–Szemerédi and Johansson–Kahn–Vu. *Random Structures and Algorithms* 58 (2021) 480-516.

[466] J.C. Hansen and J. Jaworski, Random mappings with exchangeable in-degrees, *Random Structures and Algorithms* 33 (2008) 105-126.

[467] J.C. Hansen and J. Jaworski, Predecessors and successors in random mappings with exchangeable in-degrees, *J. Applied Probability* 50 (2013) 721-740.

[468] H.L. Harper, Optimal assignments of number of vertice, *SIAM J. Appl. Math.*, 12 (1964) 131-135.

[469] T. Harris, A lower bound for the critical probability in a certain percolation, *Proceedings of the Cambridge Philosophical Society* 56 (1960) 13-20.

[470] H. Hatami, Random cubic graphs are not homomorphic to the cycle of size 7, *Journal of Combinatorial Theory B* 93 (2005) 319-325.

[471] H. Hatami and M. Molloy, The scaling window for a random graph with a given degree sequence, *Random Structures and Algorithms* 41 (2012) 99-123.

[472] P. Haxell, Y. Kohayakawa and T. Łuczak, Turáns extremal problem in random graphs: forbidding even cycles, *Journal of Combinatorial Theory B* 64 (1995) 273-287.

[473] P. Haxell, Y. Kohayakawa and T. Łuczak, Turáns extremal problem in random graphs: forbidding odd cycles, *Combinatorica* 16 (1995) 107-122.

[474] J. He and H. Liang, On rainbow-k-connectivity of random graphs, *Information Processing Letters* 112 (2012) 406-410.

[475] A. Heckel and O. Riordan, The hitting time of rainbow connection number two, *Electronic Journal on Combinatorics* 19 (2012).

[476] D. Hefetz, M. Krivelevich, M. Stojaković and T. Szabó, Positional Games (Oberwolfach Seminars), Birkhauser 2014.

[477] D. Hefetz, M. Krivelevich and T. Szabo, Sharp threshold for the appearance of certain spanning trees in random graphs, *Random Structures and Algorithms* 41 (2012) 391-412.

[478] D. Hefetz, A. Steger and B. Sudakov, Random directed graphs are robustly Hamiltonian, *Random Structures and Algorithms* 49 (2016) 345-362.

[479] D. Hefetz, A. Steger and B. Sudakov, Random directed graphs are robustly hamiltonian, see arxiv.org.

[480] W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58 (1963) 13-30.

[481] R. van der Hofstad, Random graphs and complex networks: Volume 1, see arxiv.org.

[482] R. van der Hofstad, Critical behaviour in inhomogeneous random graphs, *Random Structures and Algorithms* 42 (2013) 480-508.

[483] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, On the covariance of the level sizes in random recursive trees, *Random Structures and Algorithms* 20 (2002) 519-539.

[484] R. van der Hofstad, S. Kliem and J. van Leeuwaarden, Cluster tails for critical power-law inhomogeneous random graphs, see arxiv.org.

[485] R. van der Hofstad and G. Slade. Asymptotic expansion in n^{-1} for percolation critical values on the n-cube and $\mathbb{Z}^n$, *Random Struct. Alg.*, 27 (2005) 331-357.

[486] R. van der Hofstad and G. Slade, Expansion in n^{-1} for percolation critical values on the n-cube and $\mathbb{Z}^n$: the first three terms, *Combinatorics, Probability and Computing* 15 (2006) 695-713.

[487] R. van der Hofstad and A. Nachmias, Hypercube percolation, *J. Eur. Math. Soc.*, 19 (2017) 725–814.

[488] C. Holmgren and S. Janson, Limit laws for functions of fringe trees for binary search trees and recursive trees, see arxiv.org

[489] P. Horn and M. Radcliffe, Giant components in Kronecker graphs, *Random Structures and Algorithms* 40 (2012) 385-397.

[490] H-K. Hwang, Profiles of random trees: plane-oriented recursive trees, *Random Structures and Algorithms* 30 (2007) 380-413.

[491] G.I. Ivchenko, On the asymptotic behaviour of the degrees of vertices in a random graph, *Theory of Probability and Applications* 18 (1973) 188-195.

[492] S. Janson, Poisson convergence and Poisson processes with applications to random graphs, *Stochastic Process. Appl.* 26 (1987) 1-30.

[493] S. Janson, Normal convergence by higher semi-invariants with applications to random graphs, *Ann. Probab.* 16 (1988) 305-312.

[494] S. Janson, A functional limit theorem for random graphs with applications to subgraph count statistics *Random Structures and Algorithms* 1 (1990) 15-37.

[495] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms* 1 (1990) 221-230.

[496] S. Janson, Multicyclic components in a random graph process, *Random Structures and Algorithms* 4 (1993) 71-84.

[497] S. Janson, The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph, *Combin. Probab. Comput.* 3 (1994) 97-126.

[498] S. Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph, *Random Structures Algorithms* 7 (1995) 337-355.

[499] S. Janson, Random regular graphs: Asymptotic distributions and contiguity, *Combinatorics, Probability and Computing* 4 (1995) 369-405.

[500] S. Janson, One, two and three times $\log n/n$ for paths in a complete graph with random weights, *Combinatorics, Probability and Computing* 8 (1999) 347-361.

[501] S. Janson, Asymptotic degree distribution in random recursive trees, *Random Structures and Algorithms*, 26 (2005) 69-83.

[502] S. Janson, Monotonicity, asymptotic normality and vertex degrees in random graphs, *Bernoulli* 13 (2007) 952-965.

[503] S. Janson, Plane recursive trees, Stirling permutations and an urn model,*Fifth Colloquium on Mathematics and Computer Science, Discrete Math. Theor. Comput. Sci. Proc., AI, Assoc. Discrete Math. Theor. Comput. Sci., Nancy*, (2008) 541-547.

[504] S. Janson, Asymptotic equivalence and contiguity of some random graphs, *Random StructuresAlgorithms* 36 (2010) 26-45.

[505] S. Janson, M. Kuba and A. Panholzer, Generalized Stirling permutations, families of increasing trees and urn models, *Journal of Combinatorial Theory A* 118 (2010) 94-114.

[506] S. Janson and M.J. Luczak, A simple solution to the k-core problem, *Random Structures Algorithms* 30 (2007) 50-62.

[507] S. Janson, K. Oleszkiewicz and A. Ruciński, Upper tails for subgraph counts in random graphs, *Israel Journal of Mathematics* 142 (2004) 61-92.

[508] S. Janson, T. Łuczak, A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, *M. Karoński (ed.) J. Jaworski (ed.) A. Ruciński (ed.) , Random Graphs '87 , Wiley* (1990) 73-87.

[509] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, John Wiley and Sons, New York, 2000.

[510] S. Janson and K. Nowicki, The asymptotic distributions ofgeneralized U-statistics with applications to random graphs, *Probability Theory and Related Fields* 90 (1991) 341-375.

[511] S. Janson and O. Riordan, Susceptibility in inhomogeneous random graphs, *The Electr. J. Comb.* 19 (2012) #P31

[512] S. Janson and N. Wormald, Rainbow Hamilton cycles in random regular graphs, *Random Structures Algorithms* 30 (2007) 35-49.

[513] S. Janson, D.E. Knuth, T. Łuczak and B.G. Pittel, The birth of the giant component, *Random Structures and Algorithms* 4 (1993) 233-358.

[514] S. Janson and L. Warnke, Preferential attachment without vertex growth: Emergence of the giant component, *Annals of Applied Probability* 31 (2021) 1523-1547.

[515] J. Janssen, P. Pralat, and R. Wilson, Geometric Graph Properties of the Spatial Preferred Attachment model, *Advances in Applied Mathematics* 50 (2013) 243-267.

[516] J. Janssen, P. Pralat, and R. Wilson, Asymmetric Distribution of Nodes in the Spatial Preferred Attachment Model, *Proceedings of the 10th Workshop on Algorithms and Models for the Web Graph (WAW 2013), Lecture Notes in Computer Science 8305, Springer* (2013) 1-13.

[517] J. Jaworski, On a random mapping (T, P_j), *J. Applied Probability* 21 (1984) 186-191.

[518] J. Jaworski, Random mappings with independent choices of images, In: *Random Graphs'87, Wiley, New York* (1990) 89-101.

[519] J. Jaworski, Predecessors in a random mapping, *Random Structures and Algorithms* 13 (1998) 501-519.

[520] J. Jaworski, Epidemic processes on digraphs of random mappings, *J. Applied Probabability* 36 (1999) 780-798.

[521] J. Jaworski and T. Łuczak, Cycles in a uniform graph process, *Combinatorics, Probability and Computing* 1 (1992) 223-239.

[522] J. Jaworski and Z. Palka, Remarks on a general model of a random digraph,*Ars Combinatoria* 65 (2002) 135-144.

[523] J. Jaworski and I.H. Smit, On a random digraph, *Annals of Discrete Math.* 33 (1987) 111-127.

[524] J. Jaworski, M. Karoński and D. Stark, The degree of a typical vertex in generalized random intersection graph model, *Discrete Mathematics* 30 (2006) 2152-2165.

[525] J. Jaworski and D. Stark, The vertex degree distribution of passive random intersection graph models, *Combin. Prob Comput.* 17 (2008) 549-558.

[526] M.R. Jerrum, Large Cliques Elude the Metropolis Process, *Random Structures and Algorithms* 3 (1992) 347-359.

[527] J. Jiang, M. Mitzenmacher and J. Thaler, Parallel Peeling Algorithms, see arxiv.org.

[528] J. Jonasson, On the cover time of random walks on random graphs, *Combinatorics, Probability and Computing* 7 (1998), 265-279.

[529] A. Johansson, J. Kahn and V. Vu, Factors in Random Graphs, *Random Structures and Algorithms* 33 (2008) 1-28.

[530] F. Joos and J. Kim, Spanning trees in randomly perturbed graphs. *Random Structures Algorithms* 56 (2018) 169-219.

[531] J. Jordan, Geometric preferential attachment in non-uniform metric spaces, *Electronic Journal of Probability* 18 (2013) 1-15.

[532] J. Jordan and A. Wade, Phase transitions for random geometric preferential attachment graphs, *Advances in Applied Probability* 47 (2015).

[533] F. Juhász, On the spectrum of a random graph, *in Algebraic methods in graph theory, (Lovász et al., eds) Coll. Math. Soc. J. Bolyai, North Holland* 25 (1981) 313-316.

[534] J. Kahn, E. Lubetzky and N. Wormald, The threshold for combs in random graphs, see arxiv.org.

[535] J. Kahn, E. Lubetzky and N. Wormald, Cycle factors and renewal theory, see arxiv.org.

[536] J. Kahn, B. Narayanan and J. Park, The threshold for the square of a Hamilton cycle, see arxiv.org.

[537] J. Kahn and G. Kalai, Thresholds and expectation thresholds, *Combinatorics, Probability and Computing* 16 (2007) 495-502.

[538] J. Kahn and E. Szemerédi, On the second eigenvalue in random regular graphs - Section 2, *Proceedings of the 21'st Annual ACM Symposium on Theory of Computing* (1989) 587-598.

[539] M. Kahle, Topology of random simplicial complexes: a survey, *AMS Contemporary Volumes in Mathematics*, To appear in AMS Contemporary Volumes in Mathematics. Nov 2014. arXiv:1301.7165.

[540] S. Kalikow and B. Weiss, When are random graphs connected? *Israel J. Math.* 62 (1988) 257-268.

[541] N. Kamčev, M. Krivelevich and B. Sudakov, Some Remarks on Rainbow Connectivity, see arxiv.org.

[542] D. Kang, T.Kelly, D. Kühn, D.Osthus, and V. Pfenninger, Perfect matchings in random sparsifications of Dirac hypergraphs, *Combinatorica*

[543] M. Kang, M. Karoński, C. Koch and T. Makai, Properties of stochastic Kronecker graphs, *Journal of Combinatorics*, to appear.

[544] M. Kang and T. Łuczak, Two critical periods in the evolution of random planar graphs, *Transactions of the American Mathematical Society* 364 (2012) 4239-4265.

[545] M. Kang, W. Perkins and J. Spencer, The Bohman-Frieze Process Near Criticality, see arxiv.org.

[546] D. Karger and C. Stein, A new approach to the minimum cut problem, *Journal of the ACM* 43 (1996) 601-640.

[547] M. Karoński, A review of random graphs, *Journal of Graph Theory* 6 (1982) 349-389.

[548] M. Karoński, Random graphs, In: *Handbook of Combinatorics, T. Graham, M. Grötschel and L. Lovás, Eds.*, Elsevier science B.V. (1995) 351-380.

[549] M. Karoński and T. Łuczak, The phase transition in a random hypergraph, *Journal of Computational and Applied Mathematics* 1 (2002) 125-135.

[550] M. Karoński and B. Pittel, Existence of a perfect matching in a random $(1+\varepsilon^{-1})$–out bipartite graph, *Journal of Combinatorial Theory B* (2003) 1-16.

[551] M. Karoński and A. Ruciński, On the number of strictly balanced subgraphs of a random graph, In: *Graph Theory, Proc. Łagów, 1981, Lecture Notes in Mathematics 1018, Springer, Berlin* (1983) 79-83.

[552] M. Karoński and A. Ruciński, Problem 4, In *Graphs and other Combinatorial Topics, Proceedings, Third Czech. Symposium on Graph Theory, Prague* (1983).

[553] M. Karoński and A. Ruciński, Poisson convergence and semi-induced properties of random graphs, *Math. Proc. Camb. Phil. Soc.* 101 (1987) 291-300.

[554] M. Karoński, E. Scheinerman and K. Singer-Cohen, On Random Intersection Graphs: The Subgraph Problem, *Combinatorics, Probability and Computing* 8 (1999) 131-159.

[555] R.M. Karp, An upper bound on the expected cost of an optimal assignment, *Discrete Algorithms and Complexity: Proceedings of the Japan-US Joint Seminar (D. Johnson et al., eds.)*, Academic Press, New York, 1987, 1-4.

[556] R.M. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* 1 (1990) 73-93.

[557] R.M. Karp, A. Rinnooy-Kan and R.V. Vohra, Average Case Analysis of a Heuristic for the Assignment Problem, *Mathematics of Operations Research* 19 (1994) 513-522.

[558] R.M. Karp and M. Sipser, Maximum matchings in sparse random graphs, *Proceedings of the 22nd IEEE Symposium on the Foundations of Computer Science* (1981) 364-375.

[559] Zs. Katona, Width of scale-free tree, *J. of Appl. Prob.* 42 (2005) 839-850.

[560] Zs. Katona, Levels of scale-free tree, manuscript

[561] L. Katz, Probability of indecomposability of a random mapping function, *Ann. Math. Statist.* 26 (1955) 512-517.

[562] T. Kelly, A. Müyesser and A. Pokrovskiy, Optimal spread for spanning subgraphs of Dirac hypergraphs, see arxiv.org

[563] G. Kemkes, X. Perez-Gimenez and N. Wormald, On the chromatic number of random d-regular graphs, *Advances in Mathematics* 223 (2010) 300-328.

[564] R. Keusch and A. Steger, The game chromatic number of dense random graphs, *Electronic Journal of Combinatorics* 21 (2014).

[565] J-H. Kim, The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures and Algorithms* (1995) 173-207.

[566] J-H. Kim and V. Vu, Sandwiching random graphs: universality between random graph models, *Advances in Mathematics* 188 (2004) 444-469.

[567] J.H. Kim and N.C. Wormald, Random matchings which induce Hamilton cycles, and hamiltonian decompositions of random regular graphs, *Journal of Combinatorial Theory* B 81 (2001) 20-44.

[568] J.F.C. Kingman, The population structure associated with the Ewens sampling formula, *Theoretical Population Biology* 11 (1977) 274-283.

[569] W. Kinnersley, D. Mitsche, P. Pralat, A note on the acquaintance time of random graphs, Electronic Journal of Combinatorics 20(3) (2013) #P52.

[570] M. Kiwi and D. Mitsche, A bound for the diameter of random hyperbolic graphs, see arxiv.org.

[571] B.J.K. Kleijn and S. Rizzelli, Contiguity and remote contiguity of some random graphs (2024), see arxiv.org.

[572] J. Kleinberg, The small-world phenomenon: An algorithmic perspective, *Proceedings of the 32nd ACM Symposium on Theory of Computing* (2000) 163-170.

[573] D. E. Knuth, R. Motwani and B. Pittel, Stable husbands, *Random Structures and Algorithms* 1 (1990) 1-14.

[574] F. Knox, D. Kühn and D. Osthus, Edge-disjoint Hamilton cycles in random graphs, to appear in *Random Structures and Algorithms*.

[575] L.M. Koganov, Universal bijection between Gessel-Stanley permutations and diagrams of connections of corresponding ranks, *Mathematical Surveys* 51 (1996) 333-335.

[576] Y. Kohayakawa, B. Kreuter and A. Steger, An extremal problem for random graphs and the number of graphs with large even girth, *Combinatorica* 18 (1998) 101-120.

[577] Y. Kohayakawa, T. Łuczak and V. Rödl, On K_4-free subgraphs of random graphs, *Combinatorica* 17 (1997) 173-213.

[578] Y. Kohayakawa, and B. Kreuter, Threshold functions for asymmetric Ramsey properties involving cycles. *Random Structures and Algorithms* 11 (1997), 245–276.

[579] J. Komlós and E. Szemerédi, Limit distributions for the existence of Hamilton circuits in a random graph, *Discrete Mathematics* 43 (1983) 55-63.

[580] V.F. Kolchin, A problem of the allocation of particles in cells and cycles of random permutations, *Theory Probab. Appl.* 16 (1971) 74-90.

[581] V.F. Kolchin, Branching processes, random trees, and a generalized scheme of arrangements of particles, *Mathematical Notes* 21 (1977) 386-394.

[582] V.F. Kolchin, Random Mappings, Optimization Software, Inc., New York (1986).

[583] V.F. Kolchin, Random Graphs, Cambridge University Press, Cambridge (1999).

[584] J. Komlós, G. N. Sárközy, and E. Szemerédi, On the square of a Hamiltonian cycle in dense graphs. *Random Structures Algorithms* 9 (1996) 193–211.

[585] W. Kordecki, On the connectedness of random hypergraphs, *Commentaciones Mathematicae* 25 (1985) 265-283.

[586] W. Kordecki, Normal approximation and isolated vertices in random graphs, In: *Random Graphs'87, Eds. M. Karoński, J. Jaworski and A. Ruciński*, John Wiley and Sons, Chichester (1990) 131-139.

[587] I.N. Kovalenko, Theory of random graphs, *Cybernetics and Systems Analysis* 7 (1971) 575-579.

[588] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, *Physics Review E* (2010) 82:036106.

[589] M. Krivelevich, The choice number of dense random graphs, *Combinatorics, Probability and Computing* 9 (2000) 19-26.

[590] M. Krivelevich, Component sizes in the supercritical percolation on the binary cube, (2023)2023, see arxiv.org.

[591] M. Krivelevich, M. Kwan, and B. Sudakov, Cycles and matchings in randomly perturbed digraphs and hypergraphs. *Combinatorics, Probability and Computing* 25 (2016) 909–927.

[592] M. Krivelevich, M. Kwan, and B. Sudakov, Bounded-degree spanning trees in randomly perturbed graphs, *Journal SIAM J. on Discrete Mathematics* 31 (2017) 155-171.

[593] M. Krivelevich, C. Lee and B. Sudakov, Resilient pancyclicity of random and pseudo-random graphs, *SIAM Journal on Discrete Mathematics* 24 (2010) 1-16.

[594] M. Krivelevich, C. Lee and B. Sudakov, Long paths and cycles in random subgraphs of graphs with large minimum degree, see arxiv.org.

[595] M. Krivelevich, P.-S. Loh and B. Sudakov, Avoiding small subgraphs in Achlioptas processes, *Random Structures and Algorithms* 34 (2009) 165-195.

[596] M. Krivelevich, E. Lubetzky and B. Sudakov, Hamiltonicity thresholds in Achlioptas processes, *Random Structures and Algorithms* 37 (2010) 1-24.

[597] M. Krivelevich, A. Lubetzky and B. Sudakov, Longest cycles in sparse random digraphs, *Random Structures and Algorithms* 43 (2013) 1-15.

[598] M. Krivelevich, E. Lubetzky and B. Sudakov, Cores of random graphs are born Hamiltonian, *Proceedings of the London Mathematical Society* 109 (2014) 161-188.

[599] M. Krivelevich and P. Michaeli, Small subgraphs in the trace of a random walk, *Electronic Journal of Combinatorics* 24 (2017).

[600] M. Krivelevich, D, Reichman, and W. Samotij, Smoothed analysis on connected graphs. *SIAM Journal on Discrete Mathematics* 29 (2015) 1654-1669.

[601] M. Krivelevich and W. Samotij, Optimal packings of Hamilton cycles in sparse random graphs, *SIAM Journal on Discrete Mathematics* 26 (2012) 964-982.

[602] M. Krivelevich and B. Sudakov, The chromatic numbers of random hypergraphs, *Random Structures and Algorithms* 12 (1998) 381-403.

[603] M. Krivelevich and B. Sudakov, Minors in expanding graphs, *Geometric and Functional Analysis* 19 (2009) 294-331.

[604] M. Krivelevich and B. Sudakov, The phase transition in random graphs - a simple proof, *Random Structures and Algorithms* 43 (2013) 131-138.

[605] M. Krivelevich, B. Sudakov, and P. Tetali, On smoothed analysis in dense graphs and formulas. *Random Structures and Algorithms* 29 (2006) 180-193.

[606] M. Krivelevich, B. Sudakov, V. Vu and N. Wormald, Random regular graphs of high degree, *Random Structures and Algorithms* 18 (2001) 346-363.

[607] M. Krivelevich, B. Sudakov, V.H. Vu and N.C. Wormald, On the probability of independent sets in random graphs, *Random Structures & Algorithms* 22 (2003) 1-14.

[608] M. Krivelevich, V.H. Vu, Choosability in random hypergraphs, *Journal of Combinatorial Theory B* 83 (2001) 241-257.

[609] G. Kronenberg and Y. Spinka, Independent sets in random subgraphs of the hypercube, see arxiv.org.

[610] R. Krueger, private communication.

[611] M. Kuba and A. Panholzer, On the degree distribution of the nodes in increasing trees, *J. Combinatorial Theory (A)* 114 (2007) 597-618.

[612] L. Kucera, Expected complexity of graph partitioning problems, *Discrete Applied Mathematics* 57 (1995) 193-212.

[613] D. Kühn and D. Osthus, On Pósa's conjecture for random graphs, *SIAM Journal on Discrete Mathematics* 26 (2012) 1440-1457.

[614] V. Kurauskas, On small subgraphs in random intersection digraphs, *Discrete Mathematics* 313 (2013) 872-885.

[615] V. Kurauskas and M. Bloznelis, Large cliques in sparse random intersection graphs, see arxiv.org.

[616] A.N. Lageras and M. Lindholm, A note on the component structure in random intersection graphs with tunable clustering, *The Electronic Journal of Combinatorics* 15 (2008) #N10.

[617] C. Lee and B. Sudakov, Dirac's theorem for random graphs, *Random Structures and Algorithms* 41 (2012) 293-305.

[618] N. Lefebvre , The first-order contiguity of sparse random graphs with prescribed degrees, *In: Jain R., Jain S., Stephan F. (eds) Theory and Applications of Models of Computation. TAMC 2015. LNCS vol 9076, Springer* (2015) 177-188.

[619] M. Lelarge, A new approach to the orientation of random hypergraphs, see arxiv.org.

[620] C. Lennon and B. Pittel, On a likely number of solutions for a stable marriage problem, *Combinatorics, Probability and Computing* (2009) 371-421.

[621] J. Leskovec, Dynamics of Large Networks, PhD in Computer Science, Carnegie Mellon University, 2008.

[622] J. Leskovec, D. Chakrabarti, J. Kleinberg and C. Faloutsos, Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication, in:*PKDD '05: Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases* (2005) 133-145.

[623] J. Leskovec, J. Kleinberg, and C. Faloutsos, Graphs over time:densification laws, shrinking diameters and possible explanations, in *Proc. of ACM SIGKDD Conf. on Knowledge Discovery in Data Mining* (2005) 177-187.

[624] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos and Z. Ghahramani, Kronecker Graphs: An Approach to Modeling Networks, *Journal of Machine Learning Research* 11 (2010) 985-1042.

[625] K. Lin and G. Reinert, Joint vertex degrees in the inhomogeneous random graph model $\mathscr{G}(n,\{p_{ij}\})$, *Adv. in Appl. Prob.* 44 (2012) 139-165.

[626] N. Linial and R. Meshulam, Homological connectivity of random 2-complexes, *Combinatorica* 26 (2006) 475-487.

[627] N. Linial, and E. Rozenman, Random Lifts of Graphs: Perfect Matchings, *Combinatorica* 25 (2005) 407-424.

[628] S. Linusson and J. Wästlund, A proof of Parisi's conjecture on the random assignment problem, *Probability Theory and Related Fields* 128 (2004) 419-440.

[629] J. Lu, and Q. Feng, (1998). Strong consistency of the number of vertices of given degrees in nonuniform random recursive trees, *Yokohama Math. J* 45 (1998) 61-69.

[630] E. Lubetzky and A. Sly, Cutoff phenomena for random walks on random regular graphs, *Duke Mathematical Journal* 153 (2010) 475-510.

[631] E. Lubetzky and Y. Zhao, On the variational problem for upper tails in sparse random graphs, see arxiv.org.

[632] T. Łuczak, On the equivalence of two basic models of random graphs, *Proceedings of Random Graphs'87, (Michal Karonski, Jerzy Jaworski, Andrzej Rucinski, eds.), Wiley, Chichester* (1990) 151-158.

[633] T. Łuczak, On the number of sparse connected graphs, *Random Structures and Algorithms* 1 (1990) 171-174.

[634] T. Łuczak, Component behaviour near the critical point, *Random Structures and Algorithms* 1 (1990) 287-310.

[635] T. Łuczak, Cycles in a random graph near the critical point, *Random Structures and Algorithms* 2 (1991) 421-440.

[636] T. Łuczak, Size and connectivity of the k-core of a random graph, *Discrete Mathematics* 91 (1991) 61-68.

[637] T. Łuczak, The chromatic number of random graphs, *Combinatorica* 11 (1991) 45-54.

[638] T. Łuczak, A note on the sharp concentration of the chromatic number of random graphs, *Combinatorica* 11 (1991) 295-297.

[639] T. Łuczak, The phase transition in a random graph, *Combinatorics, Paul Erdős is Eighty, Vol. 2, eds. D. Miklós, V.T. Sós, and T. Szőnyi, Bolyai Soc. Math. Stud. 2, J. Bolyai Math. Soc., Budapest* (1996) 399-422.

[640] T. Łuczak, Random trees and random graphs, *Random Structures and Algorithms* 13 (1998) 485-500.

[641] T. Łuczak, On triangle free random graphs, *Random Structures and Algorithms* 16 (2000) 260-276.

[642] T. Łuczak, B. Pittel and J. C. Wierman, The structure of a random graph at the point of the phase transition, *Transactions of the American Mathematical Society* 341 (1994) 721-748

[643] T. Łuczak and P. Pralat, Chasing robbers on random graphs: zigzag theorem, *Random Structures and Algorithms* 37 (2010) 516-524.

[644] T. Łuczak and A. Ruciński, Tree-matchings in graph processes *SIAM Journal on Discrete Mathematics* 4 (1991) 107-120.

[645] T. Łuczak, A. Ruciński and B. Voigt, Ramsey properties of random graphs, *Journal of Combinatorial Theory B* 56 (1992) 55-68.

[646] T. Łuczak, Ł. Witkowski and M. Witkowski, Hamilton Cycles in Random Lifts of Graphs, see arxiv.org.

[647] M. Mahdian and Y. Xu. Stochastic Kronecker graphs, *Random Structures and Algorithms* 38 (2011) 453-466.

[648] H. Mahmoud, Evolution of Random Search Trees, John Wiley & Sons, Inc., New York, 1992.

[649] H. Mahmoud, The power of choice in the construction of recursive trees, *Methodol. Comput. Appl. Probab.* 12 (2010) 763-773.

[650] H.M. Mahmoud and R.T. Smythe, A survey of recursive trees, *Theory Probability Math. Statist.* 51 (1995) 1-27.

[651] H.M. Mahmoud, R.T. Smythe and J. Szymański, On the structure of random plane-oriented recursive trees and their branches, *Random Structures and Algorithms*, 4 (1993) 151-176.

[652] F. Manzo, M. Quattropani and E. Scoppola, A probabilistic proof of Cooper and Frieze's "First Visit Time Lemma", *ALEA Latin American Journal of Probability and Mathematical Statistics* (2021)

[653] E. Marczewski, Sur deux propriétés des classes d'ensembles, *Fundamenta Mathematicae* 33 (1945) 303-307.

[654] C. Martinez and D. Mitsche, On the jump of the cover time in random geometric graphs, see arxiv.org.

[655] D. Matula, The largest clique size in a random graph, *Technical Report, Department of Computer Science, Southern methodist University*, (1976).

[656] D. Matula, Expose-and-merge exploration and the chromatic number of a random graph, *Combinatorica* 7 (1987) 275-284.

[657] R. Mauldin, The Scottish Book: mathematics from the Scottish Cafe, Birkhäuser Boston, 1981.

[658] G. McColm, First order zero-one laws for random graphs on the circle, *Random Structures and Algorithms* 14 (1999) 239-266.

[659] G. McColm, Threshold functions for random graphs on a line segment, *Combinatorics, Probability and Computing* 13 (2004) 373-387.

[660] C. McDiarmid, Determining the chromatic number of a graph, *SIAM Journal on Computing* 8 (1979) 1-14.

[661] C. McDiarmid, Clutter percolation and random graphs, *Mathematical Programming Studies* 13 (1980) 17-25.

[662] C. McDiarmid, General percolation and random graphs, *Annals of Applied Probability* 13 (1981) 40-60.

[663] C. McDiarmid, On the method of bounded differences, *in Surveys in Combinatorics, ed. J. Siemons, London Mathematical Society Lecture Notes Series* 141, Cambridge University Press, 1989.

[664] C. McDiarmid, Expected numbers at hitting times, *Journal of Graph Theory* 15 (1991) 637-648.

[665] C. McDiarmid, Random channel assignment in the plane, *Random Structures and Algorithms* 22 (2003) 187-212.

[666] C. McDiarmid and T. Müller, On the chromatic number of random geometric graphs, *Combinatorica* 31 (2011) 423-488.

[667] C. McDiarmid and B. Reed, Colouring proximity graphs in the plane, *Discrete Mathematics* 199 (1999) 123-137.

[668] C. McDiarmid, A. Steger, and D. Welsh, Random planar graphs, *Journal of Combinatorial Theory* B 93 (2005) 187-205.

[669] C. McDiarmid, A. Scott and P. Whiters, The component structure of dense random subgraphs of the hypercube, *Random Structures and Algorithms*, 59 (2021) 3-24.

[670] A. McDowell and R. Mycroft, Hamilton l-cycles in randomly perturbed hypergraphs. *Electronic Journal of Combinatorics* 25 (2018) P4.36.

[671] B.D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combinatoria* 19 (1985) 15-25.

[672] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica* 11 (1991) 369-382.

[673] B.D. McKay and N.C. Wormald, The degree sequence of a random graph I, The models, *Random Structures and Algorithms* 11 (1997) 97-117.

[674] N. Martin and J. England, Mathematical Theory of Entropy. Cambridge University Press, 2011.

[675] A. Mehrabian, Justifying the small-world phenomenon via random recursive trees, mauscript (2014).

[676] A. Meir and J.W. Moon, The distance between points in random trees, *J. Combinatorial Theory* 8 (1970) 99-103.

[677] A. Meir and J.W. Moon, Cutting down random trees, *J. Austral. Math. Soc.* 11 (1970) 313-324.

[678] A. Meir and J.W. Moon, Cutting down recursive trees, *Mathematical Biosciences* 21 (1974) 173-181.

[679] A. Meir and J.W. Moon, On the altitude of nodes in random trees, *Canadian J. Math.* 30 (1978) 997-1015.

[680] S.Micali and V.Vazirani, An $O(|V|^{1/2}|E|)$ algorithm for finding maximum matching in general graphs, *Proceedings of the 21st IEEE Symposium on Foundations of Computing* (1980) 17-27.

[681] M. Mihail and C. Papadimitriou, On the eigenvalue power law, *Randomization and Approximation Techniques, 6th International Workshop* (2002) 254-262.

[682] S. Milgram, The Small World Problem, *Psychology Today* 2 (1967) 60-67.

[683] I. M. Milne-Thomson, *The Calculus of Finite Differences.* Macmillian, London (1951).

[684] M. Molloy, Cores in random hypergraphs and Boolean formulas, *Random Structures and Algorithms* 27 (2005) 124-135.

[685] M. Molloy and B. Reed, The Size of the Largest Component of a Random Graph on a Fixed Degree Sequence, *Combinatorics, Probability and Computing* 7 (1998) 295-306.

[686] M. Molloy, H. Robalewska, R. W. Robinson and N. C. Wormald, 1- factorisations of random regular graphs, *Random Structures and Algorithms* 10 (1997) 305-321.

[687] R. Montgomery, Embedding bounded degree spanning trees in random graphs, see arxiv.org.

[688] R. Montgomery, Sharp threshold for embedding combs and other spanning trees in random graphs, see arxiv.org.

[689] R. Montgomery, Hamiltonicity in random graphs is born resilient, see arxiv.org.

[690] R. Montgomery, Spanning trees in random graphs, *Advances in Mathematics* (2019) .

[691] R. Montgomery, Hamiltonicity in random graphs is born resilient, see arxiv.org.

[692] R. Montgomery, Topics in random graphs, Lecture notes, University of Birmingham.

[693] J.W. Moon, The distance between nodes in recursive trees, *London Math. Soc. Lecture Notes Ser.* 13 (1974) 125-132.

[694] T.F. Móri, On random trees, *Studia Sci. Math. Hungar.* 39 (2003) 143-155.

[695] T.F. Móri The maximum degree of the Barabási-Albert random tree, *Combin.Probab. Comput.* 14 (2005) 339-348.

[696] K. Morris, A. Panholzer and H. Prodinger, On some parameters in heap ordered trees, *Combinatorics, Probability and Computing* 13 (2004) 677-696.

[697] E. Mossel and A. Sly, Gibbs rapidly samples colorings of $G_{n,d/n}$, *Journal of Probability Theory and Related Fields* 148 (2010) 37-69.

[698] T. Müller, Two point concentration in random geometric graphs, *Combinatorica* 28 (2008) 529-545.

[699] T. Müller, Private communication.

[700] T. Müller, X. Pérez and N. Wormald, Disjoint Hamilton cycles in the random geometric graph, *Journal of Graph Theory* 68 (2011) 299-322.

[701] T. Müller and M. Stokajović, A threshold for the Maker-Breaker clique game, *Random Structurs and Algorithms* 45 (2014) 318-341.

[702] L. Mutavchiev, A limit distribution related to a random mappings and its application to an epidemic process, *Serdica* 8 (1982) 197-203.

[703] H. S. Na and A. Rapoport, Distribution of nodes of a tree by degree, *Mathematical Biosciences* 6 (1970) 313-329.

[704] A. Nachmias and Y. Peres, Critical random graphs: diameter and mixing time, *Annals of Probability* 36 (2008) 1267-1286.

[705] A. Nachmias and Y. Peres, The critical random graph, with martingales, *Israel Journal of Mathematics* 176 (2010) 29-41.

[706] C. Nair, B. Prabhakar and M. Sharma, Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures, Random Structures and Algorithms 27 (2005) 413-444.

[707] D. Najock and C.C. Heyde (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology, *Journal of Applied Probability* 19 (1982) 675-680.

[708] C. St.J. Nash-Williams, Edge-disjoint spanning trees of finite graphs, *Journal of the London Mathematical Society* (1961) 445-450.

[709] R. Nenadov, Y. Person, N. Škoric and A. Steger, An algorithmic framework for obtaining lower bounds for random Ramsey problems, see arxiv.org.

[710] Short proof of the hypergraphcontainer theorem, see arxiv.org.

[711] R. Nenadov and A. Steger, A Short proof of the Random Ramsey Theorem, to appear in *Combinatorics, Probability and Computing*.

[712] R. Nenadov and N. Škorić, Powers of Hamilton cycles in random graphs and tight Hamilton cycles in random hypergraphs, see arxiv.org.

[713] R. Nenadov, A. Steger and Stojacović, On the threshold for the Maker-Breaker H-game, see arxiv.org.

[714] R. Nenadov, A. Steger and M. Trujić, Resilience of Perfect matchings and Hamiltonicity in Random Graph Processes, see arxiv.org.

[715] R. Nenadov and M. Trujic, Sprinkling a few random edges doubles the power. *SIAM Journal on Discrete Mathematics* 35 (2021) 998-1004.

[716] M.E.J. Newman and D.J. Watts, Scaling and percolation in the small-world network model,*Phys. Rev. E* 60 (1999) 7332–7342.

[717] S. Nicoletseas, C. Raptopoulos and P.G. Spirakis, The existence and efficient construction of large independent sets in general random intersection graph, In: *ICALP 2004, LNCS 3142, Springer* (2004) 1029-1040.

[718] S. Nikoletseas, C. Raptopoulos and P.G. Spirakis, On the independence number and Hamiltonicity of uniform random intersection graphs, *Theoret. Comput. Sci.* 412 (2011) 6750-6760.

[719] I. Norros and H. Reittu, On a conditionally Poissonian graph process, *Adv. in Appl. Probab.* 38 (2006) 59-75.

[720] D. Osthus, H. J. Prömel, and A. Taraz, On random planar graphs, the number of planar graphs and their triangulations, *Journal of Combinatorial Theory* B 88 (2003) 119-134.

[721] J. Palacios, A bound for the covering time of random walks on graphs, *Statistics and Probability Letters* (1992).

[722] Z. Palka, On the number of vertices of given degree in a random graph, *J. Graph Theory* 8 (1984) 167-170.

[723] Z. Palka, Extreme degrees in random graphs, *Journal of Graph Theory* 11 (1987) 121-134.

[724] E. Palmer and J. Spencer, Hitting time for k edge disjoint spanning trees in a random graph, *Period Math. Hungar.* 91 (1995) 151-156.

[725] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, *Random Structures and Algorithms* 31 (2007) 203-226.

[726] A. Panholzer and G. Seitz, Ordered increasing k-trees: Introduction and analysis of a preferential attachment network model, *Discrete Mathematics and Theoretical Computer Science (AofA 2010)* (2010) 549-564.

[727] [PKBnV10] F. Papadopoulos, D. Krioukov, M. Boguñá and A. Vahdat, Greedy forwarding in dynamic scale-free networks embedded in hyperbolic metric spaces, *Proceedings of the 29th Conference on Information Communications, INFOCOM'10* (2010) 2973-2981.

[728] O. Parczyk, 2-universality in randomly perturbed graphs, *European Journal of Combinatorics* 87 (2020) 103118.

[729] G. Parisi, A conjecture on Random Bipartite Matching, Physics e-Print archive (1998).

[730] J. Park and H. Pham, A proof of the Hahn-Kalai conjecture, see arxiv.org.

[731] E. Peköz, A. Röllin and N. Ross, Degree asymptotics with rates for preferential attachment random graphs, *The Annals of Applied Probability* 23 (2013) 1188-1218.

[732] M. Penrose, On k-connectivity for a geometric random graph, *Random Structures and Algorithms* 15 (1999) 145-164.

[733] M. Penrose, Random Geometric Graphs, *Oxford University Press*, 2003.

[734] N. Peterson and B. Pittel, Distance between two random k-out graphs, with and without preferential attachment, see arxiv.org.

[735] J. Pitman, Random mappings, forest, and subsets associated with the Abel-Cayley-Hurwitz multinomial expansions, *Seminar Lotharingien de Combinatoire* 46 (2001) 46.

[736] H. Pham, A. Sah, M. Sawhney, and M. Simkin, A toolkit for robust thresholds, see arxiv.org.

[737] B. Pittel, On distributions related to transitive closures of random finite mappings, *The Annals of Probability* 11 (1983) 428-441.

[738] B.Pittel, On tree census and the giant component in sparse random graphs, *Random Structures and Algorithms* 1 (1990), 311-342.

[739] B. Pittel, On likely solutions of a stable marriage problem, *Annals of Applied Probability* 2 (1992) 358-401.

[740] B. Pittel, Note on the heights of random recursive trees and random m-ary search trees, *Random Structures and Algorithms* 5 (1994) 337-347.

[741] B. Pittel, On a random graph evolving by degrees, *Advances in Mathematics* 223 (2010) 619-671.

[742] B. Pittel, J. Spencer and N. Wormald, Sudden emergence of a giant k-core in a random graph, *Journal of Combinatorial Theory B* 67 (1996) 111-151.

[743] B. Pittel, L. Shepp and E. Veklerov, On the number of fixed pairs in a random instance of the stable marriage problem, *SIAM Journal on Discrete Mathematics* 21 (2008) 947-958.

[744] B. Pittel and R. Weishar, The random bipartite nearest neighbor graphs, *Random Structures and Algorithms* 15 (1999) 279-310.

[745] D. Poole, On weak hamiltonicity of a random hypergraph, see arxiv.org.

[746] L. Pósa, Hamiltonian circuits in random graphs, *Discrete Mathematics* 14 (1976) 359-364.

[747] E. Powierski, Ramsey properties of randomly perturbed dense graphs, arXiv:1902.02197.

[748] P. Pralat, N. Wormald, Meyniel's conjecture holds for random d-regular graphs, see arxiv.org.

[749] P. Pralat, Cleaning random graphs with brushes, *Australasian Journal of Combinatorics* 43 (2009) 237-251.

[750] P. Pralat, Cleaning random d-regular graphs with Brooms, *Graphs and Combinatorics* 27 (2011) 567-584.

[751] V. M. Preciado and A. Jadbabaie, Spectral analysis of virus spreading in random geometric networks, *Proceedings of the 48th IEEE Conference on Decision and Control* (2009).

[752] H. Prodinger, Depth and path length of heap ordered trees, *International Journal of Foundations of Computer Science* 7 (1996) 293-299.

[753] H. Prodinger and F.J. Urbanek, On monotone functions of tree structures, *Discrete Applied Mathematics* 5 (1983) 223-239.

[754] H. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Archives of Mathematical Physics* 27 (1918) 742-744.

[755] J. Radakrishnan, Entropy and Counting, www.tcs.tifr.res.in/ jaikumar/Papers/EntropyAndCounting.pdf

[756] A. Rényi and G. Szekeres, On the height of trees,*Journal Australian Mathematical Society* 7 (1967) 497-507.

[757] M. Radcliffe and S.J. Young, Connectivity and giant component of stochastic Kronecker graphs, see arxiv.org.

[758] S. Rai, The spectrum of a random geometric graph is concentrated, *Journal of Theoretical Probability* 20 (2007) 119-132.

[759] A. Rényi, On connected graphs I, *Publ. Math. Inst. Hungar. Acad. Sci.* 4 (1959) 385-387.

[760] O. Riordan, Spanning subgraphs of random graphs, *Combinatorics, Probability and Computing* 9 (2000) 125-148.

[761] O. Riordan and A. Selby, The maximum degree of a random graph, *Combinatorics, Probability and Computing* 9 (2000) 549-572.

[762] O. Riordan, The small giant component in scale-free random graph, *Combin. Prob. Comput.* 14 (2005) 897-938.

[763] O. Riordan, Long cycles in random subgraphs of graphs with large minimum degree, see arxiv.org.

[764] O. Riordan and L. Warnke, Achlioptas process phase transitions are continuous, *Annals of Applied Probability* 22 (2012) 1450-1464.

[765] O. Riordan and N. Wormald, The diameter of sparse random graphs, *Combinatorics, Probability and Computing* 19 (2010) 835-926.

[766] H.D. Robalewska, 2-factors in random regular graphs, *Journal of Graph Theory* 23 (1996) 215-224.

[767] R.W. Robinson and N.C.Wormald, Almost all cubic graphs are Hamiltonian, *Random Structures and Algorithms* 3 (1992) 117-126.

[768] R.W. Robinson and N.C.Wormald, Almost all regular graphs are Hamiltonian, *Random Structures and Algorithms* 5 (1994) 363-374.

[769] V. Rödl and A. Ruciński, Lower bounds on probability thresholds for Ramsey properties, in *Combinatorics, Paul Erdős is Eighty, Vol.1, Bolyai Soc. Math. Studies* (1993), 317-346.

[770] V. Rödl, A. Ruciński and E. Szemerédi, An approximate Dirac-type theorem for k-uniform hypergraphs, *Combinatorica* 28 (2008) 220-260.

[771] V. Rödl and A. Ruciński, Random graphs with monochromatic triangles in every edge coloring. *Random Structures and Algorithms* 5 (1994), 253–270.

[772] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties *Journal of the American Mathematical Society* 8 (1995) 917-942.

[773] D. Romik, The Surprising Mathematics of Longest Increasing Subsequences, Cambridge University Press, New York (2014).

[774] S.M. Ross, A random graph, *J. Applied Probability* 18 (1981) 309-315.

[775] H. Rubin and R. Sitgreaves, Probability distributions related to random transformations on a finite set, *Tech. Rep. 19A, Applied Mathematics and Statistics Laboratory, Stanford University*, (1954).

[776] A. Ruciński, When a small subgraphs of a random graph are normally distributed, *Probability Theory and Related Fields* 78 (1988) 1-10.

[777] A. Ruciński and A. Vince, Strongly balanced graphs and random graphs, *Journal of Graph Theory* 10 (1986) 251-264.

[778] A. Ruciński, Matching and covering the vertices of a random graph by copies of a given graph,*Discrete Mathematics* 105 (1992) 185-197.

[779] A. Ruciński and N. Wormald, Random graph processes with degree restrictions, *Combinatorics, Probability and Computing* 1 (1992) 169-180.

[780] A. Rudas, B,Toth and B. Való, Random trees and general branching processes, *Random Structures and Algorithms* 31 (2007) 186-202.

[781] K. Rybarczyk and D. Stark, Poisson approximation of the number of cliques in a random intersection graph *Journal of Applied Probability* 47 (2010) 826 - 840.

[782] K. Rybarczyk and D. Stark, Poisson approximation of counts of subgraphs in random intersection graphs, submitted.

[783] K. Rybarczyk, Equivalence of the random intersection graph and $G(n,p)$, *Random Structures and Algorithms* 38 (2011) 205-234.

[784] K. Rybarczyk, Sharp threshold functions for random intersection graphs via a coupling method, *The Electronic Journal of Combinatorics* 18(1) (2011) #P36.

[785] K. Rybarczyk, Diameter, connectivity, and phase transition of the uniform random intersection graph, *Discrete Mathematic* 311 (2011) 1998-2019.

[786] K. Rybarczyk, Constructions of independent sets in random intersection graphs, *Theoretical Computer Science* 524 (2014) 103-125.

[787] K. Rybarczyk, The coupling method for inhomogeneous random intersection graphs, see arxiv.org.

[788] V.N. Sachkov, Combinatorial Methods in Discrete Mathematics, Cambridge University Press, (1996).

[789] A.A. Saposhenko, Metric properties of almost all boolean functions, (in Russian), *Diskretny Analiz* 10 (1967) 91-119.

[790] A.A. Saposhenko, Geometric structure of almost all boolean functions, (in Russian), *Problemy Kibernetiki* 30 (1975) 227-261.

[791] D. Saxton and A. Thomason, Hypergraph containers, see arxiv.org.

[792] D. Saxton and A. Thomason, Online containers for hypergraphs with applications to linear equations, see arxiv.org.

[793] M. Schacht, Extremal results for random discrete structures, see arxiv.org.

[794] J. Schmidt and E. Shamir, A threshold for perfect matchings in random d-pure hypergraphs, *Discrete Mathematics* 45 (1983) 287-295.

[795] J. Schmidt-Pruzan and E. Shamir, Component structure in the evolution of random hypergraphs, *Combinatorica* 5 (1985) 81-94.

[796] R. Schneider, Convex Bodies: The Brunn-Minkowski Theory, Cambridge University Press, 2013.

[797] A. Scott, On the concentration of the chromatic number of random graphs (2008) see arxiv.org.

[798] C. Seshadhri, A. Pinar, and T.G. Kolda, An in-depth analysis of stochastic Kronecker graphs, *The Journal of the ACM*, see arxiv.org.

[799] E. Shamir and J. Spencer, Sharp concentration of the chromatic number od random graphs $G_{n,p}$, *Combinatorica* 7 (1987) 121-129.

[800] C. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal* 27 (1948) 379–423.

[801] B. Selivanov, Pereschislenie odnorodnyckh hypergrafov c protstoǐ struktutoǐ, *Kombinatornyǐ Analiz.* 2 (1972) 60-67.

[802] S. Shelah and J. Spencer, Zero-One laws for sparse random graphs, *Journal of the American Mathematical Society* 1 (1988) 97-115.

[803] L.A. Shepp and S.P. Lloyd, Ordered cycle length in a random permutation, *Trans. Amer. Math. Soc.* 121 (1966) 340-357.

[804] L. Shi and N. Wormald, Coloring random regular graphs, *Combinatorics, Probability and Computing* 16 (2007) 459-494.

[805] L. Shi and N. Wormald, Coloring random 4-regular graphs, *Combinatorics, Probability and Computing* 16 (2007) 309-344.

[806] M. Šileikis and L. Warnke. A counterexample to the DeMarco-Kahn Upper Tail Conjecture, see arxiv.org.

[807] S. Sivasubramanian, Spanning trees in complete uniform hypergraphs and a connection to r-extended Shi hyperplane arrangements, see arxiv.org.

[808] G. Slade, The phase transition for random subgraphs of the n-cube, *Extended abstract for the 16th Annual International Conference on Formal Power Series and Algebraic Combinatorics*, (2004) Vancouver.

[809] B. Söderberg, General formalism for inhomogeneous random graphs, *Phys. Rev. E* 66 (2002) 066121.

[810] J. Spencer, Private Communication.

[811] J. Spencer, Enumerating graphs and Brownian motion, *Communications in Pure and Applied Mathematics* 50 (1997) 291-294.

[812] J. Spencer, The Strange Logic of Random Graphs, Springer, Berlin 2001.

[813] J. Spencer and N. Wormald, Birth control for giants, *Combinatorica* 27 (2007) 587-628.

[814] S. Spiro, A Smoother Notion of Spread Hypergraphs, *Combinatorics, Probability, and Computing* 22 (2021).

[815] D. Stark, The vertex degree distribution of random intersection graphs, *Random Structures Algorithms* 24 (2004) 249-258.

[816] J. M. Steele, On Frieze's $\zeta(3)$ limit forlengths of minimal spanning trees, *Discrete Applied Mathematics* 18 (1987) 99-103.

[817] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, In:*Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability , 1970*, Univ. of California Press, Berkeley, Vol. II (1972) 583-602.

[818] V.E. Stepanov, On the distribution of the number of vertices in strata of a random tree, *Theory of Probability and Applications* 14 (1969) 65-78.

[819] V.E. Stepanov, Limit distributions of certain characteristics of random mappings, *Theory Prob. Appl.* 14 (1969) 612-636.

[820] M. Stojaković and T. Szabó, Positional games on random graphs, *Random Structures and Algorithms* 26 (2005) 204-223.

[821] Ch. Su, J. Liu and Q. Feng, A note on the distance in random recursive trees, *Statistics & Probability Letters* 76 (2006) 1748-1755

[822] B. Sudakov and V. Vu, Local resilience of graphs, *Random Structures and Algorithms* 33 (2008) 409-433.

[823] E. Szemerédi, Regular partitions of graphs, *Colloques Internationaux C.N.R.S. No 260 – Problémes Combinatoires et Théorie des Graphes, Orsay* (1976), 399-401.

[824] J. Szymański, On the nonuniform random recursive tree, *Annals of Discrete Math.* 33 (1987) 297-306.

[825] J. Szymański, On the maximum degree and the height of random recursive tree, In:*Random Graphs'87*, Wiley, New York (1990) 313-324.

[826] J. Szymański, Concentration of vertex degrees in a scale free random graph process, *Random Structures and Algorithms* 26 (2005) 224-236.

[827] J. Tabor, On *k*-factors in stochastic Kronecker graphs, (2014) unpublished manuscript.

[828] M. Talagrand, Concentration of measures and isoperimetric inequalities in product spaces, *Publications Mathematiques de IT.H.E.S.* 81 (1996) 73-205.

[829] M.A. Tapia and B.R. Myers, Generation of concave node-weighted trees, *IEEE Trans. Circuits Systems* 14 (1967) 229-230.

[830] S-H. Teng and F.F. Yao, *k*-nearest-neighbor clustering and percolation theory, *Algorithmica* 49 (2007) 192-211.

[831] A.Thomason, A simple linear expected time algorithm for Hamilton cycles, *Discrete Mathematics* 75 (1989) 373-379.

[832] P. Turán, On an extremal problem in graph theory, *Matematikai és Fizikai Lapok (in Hungarian)* 48 (1941) 436-452.

[833] T.S. Turova, Phase transition in dynamical random graphs, *J. Stat. Phys.* 123 (2006) 1007-1032.

[834] T.S. Turova, Diffusion approximation for the componentsin critical inhomogeneous random graphsof rank 1, *Random Structures Algorithms* 43 (2013) 486-539.

[835] T.S. Turova, Asymptotics for the size of the largest component scaled to "$\log n$" in inhomogeneous random graphs, *Arkiv fur Matematik* 51 (2013) 371-403.

[836] W. T. Tutte, A census of planar triangulations, *Canadian Journal of Mathematics* 14 (1962) 21-38.

[837] W. T. Tutte, A census of planar maps, *Canadian Journal of Mathematics* 15 (1963) 249-271.

[838] W.F. de la Vega, Long paths in random graphs, *Studia Sci. Math. Hungar.* 14 (1979) 335-340.

[839] A. M. Vershik and A. A. Shmidt, Limit measures arising in the asymptotic theory of symmetric groups, I, *Theory of Probability and its Applications* 22 (1977) 70-85.

[840] E. Vigoda, Improved bounds for sampling colorings, *Journal of Mathematical Physics* 41 (2000) 1555-1569.

[841] O.V. Viskov, Some comments on branching process, *Mathematical Notes* 8 (1970) 701-705.

[842] V. Vu. On some degree conditions which guarantee the upper bound of chromatic (choice) number of random graphs, *Journal of Graph Theory* 31 (1999) 201-226.

[843] V. Vu, Spectral norm of random matrices, *Combinatorica* 27 (2007) 721-736.

[844] D.W. Walkup, Matchings in random regular bipartite graphs, *Discrete Mathematics* 31 (1980) 59-64.

[845] D.W. Walkup, On the expected value of a random asignment problem, *SIAM Journal on Computing* 8 (1979) 440-442.

[846] J. Wástlund, An easy proof of the $\zeta(2)$ limit in the random assignment problem, *Electronic Communications in Probability* 14 (2009) 261-269.

[847] D.J. Watts and S.H. Strogatz, Collective dynamics of ”small-world” networks, *Nature* 393 (1998) 440–442.

[848] K. Weber, Random graphs - a survey, *Rostock Math. Kolloq.* (1982) 83-98.

[849] K. Weber, On components of random graphs in the n-cube. *Elektron. Inf. verarb. Kybern.*, 22 (1986) 601 – 613.

[850] K. Weber, Poisson convergence in the n-cube, *Math. Nachr.*, 131 (1987) 49 – 57.

[851] K. Weber, On components of random subgraphs of the n-cube. *In: Random Graphs, Volume 2, A. Frieze and T. Łuczak eds, Wiley*, (1992) 263 – 278.

[852] D. West, Introduction to Graph Theory - Second edition, Prentice Hall 2001.

[853] E.P. Wigner, On the distribution of the roots of certain symmetric matrices, *Annals of Mathematics* 67 (1958) 325-327.

[854] H. Wilf, *Generatingfunctionology,* Academic Press (1990).

[855] L. B. Wilson, An analysis of the stable assignment problem, *BIT* 12 (1972) 569-575.

[856] E.M. Wright, The number of connected sparsely edged graphs, *Journal of graph Theory* 1 (1977) 317-330.

[857] E.M. Wright, The number of connected sparsely edged graphs II, *Journal of graph Theory* 2 (1978) 299-305.

[858] E.M. Wright, The number of connected sparsely edged graphs III, *Journal of graph Theory* 4 (1980) 393-407.

[859] N.C. Wormald, Models of random regular graphs, *In J.D. Lamb and D.A. Preece, editors, Surveys in Combinatorics, 1999, volume 267 of London Mathematical Society Lecture Note Series, Cambridge University Press* (1999) 239-298.

[860] N. C. Wormald, The differential equation method for random graph processes and greedy algorithms, Lectures on approximation and randomized algorithms (M. Karoński and H.J. Prömel, eds.), Advanced Topics in Mathematics, Polish Scientific Publishers PWN, Warsaw, 1999, pp. 73-155.

[861] O. Yagan and A. M. Makowski, Zero-one laws for connectivity in random key graphs, *IEEE Transations on Information Theory* 58 (2012) 2983-2999.

[862] B. Ycart and J. Ratsaby, The VC dimension of k-uniform random hypergraphs, *Random Structures and Algorithms* 30 (2007) 564-572.

[863] S.J. Young, Random Dot Product Graphs: A Flexible Model of Complex Networks, ProQuest (2008)

[864] S.J. Young and E.R. Scheinerman, Random dot product graph models for social networks, In: *Algorithms and Models for the Web-Graph, LNCS 4863, Springer* (2007) 138-149.

[865] Z. Zhang, L. Rong and F. Comellas, High dimensional random apollonian networks, *Physica A: Statistical Mechanics and its Applications* 364 (206) 610-618.

[866] J. Zhao, O. Yagan and V. Gligor, On k-connectivity and minimum vertex degree in random s-intersection graphs, *SIAM Meeting on Analytic Algorithmics and Combinatorics*, (2015), see arxiv.org.

[867] T. Zhou, G. Yan and B.H. Wang, Maximal planar networks with large clustering coefficient and power-law degree distribution, *Physics Reviews E* 71 (2005) 046141.

Chapter 32

Indices

Author Index

Main Index