RANDOM GRAPHS

Basic Models

$$G_{n,m} = ([n], E_{n,m})$$

Vertex set $[n] = \{1, 2, \dots, n\}$

Each graph with m edges has same probability

$$\frac{1}{\binom{N}{m}}, \qquad N = \binom{n}{2}$$

$$G_{n,p} =: ([n], E_{n,p})$$

Vertex set $[n]$

$$\Pr(G_{n,p} = G) = p^{|E(G)|} (1-p)^{N - |E(G)|}$$

i.e. each edge occurs independently with probability p.

Graph property $\mathcal{P}$.

$p = \frac{m}{N}$

$$\Pr(G_{n,p} \in \mathcal{P}) = \sum_{\mu=0}^{N} \Pr(G_{n,p} \in \mathcal{P} \mid |E_{n,p}| = \mu) \times \Pr(|E_{n,p}| = \mu)$$

$$= \sum_{\mu=0}^{N} \Pr(G_{n,\mu} \in \mathcal{P}) \Pr(|E_{n,p}| = \mu)$$

$$\geq \Pr(G_{n,m} \in \mathcal{P}) \Pr(|E_{n,p}| = m)$$

$$\Pr(|E_{n,p}| = m) = \binom{N}{m} p^m (1-p)^{N-m}$$

$$= (1+o(1)) \frac{N^N \sqrt{2\pi N}\; p^m (1-p)^{N-m}}{m^m (N-m)^{N-m}\, 2\pi \sqrt{m(N-m)}}$$

$m \to \infty$

$N-m \to \infty$

$$= (1+o(1)) \sqrt{\frac{N}{2\pi m (N-m)}}$$

$$\geq \frac{1}{10\sqrt{m}}$$

So,

$$\Pr(G_{n,m} \in \mathcal{P}) \leq 10 m^{1/2} \Pr(G_{n,p} \in \mathcal{P}).$$

Monotone Properties

A property is monotone increasing ↗ if

$$G \in \mathcal{P} \implies G + e \in \mathcal{P}$$

e.g. connectivity

Monotone decreasing ↘ if

$$G \in \mathcal{P} \implies G - e \in \mathcal{P}$$

e.g. planarity

Suppose $\mathcal{P}$ is $\nearrow$. $\qquad p = \frac{m}{N}$

$$\Pr(G_{n,p} \in \mathcal{P}) = \sum_{\mu=0}^{N} \Pr(G_{n,\mu} \in \mathcal{P}) \Pr(|E_{n,p}| = \mu)$$

$$\geq \Pr(G_{n,m} \in \mathcal{P}) \underbrace{\sum_{\mu=m}^{N} \Pr(|E_{n,p}| = \mu)}_{\geq \frac{1}{2} - o(1)}.$$

Central Limit Theorem $\Rightarrow$

$$\Pr(G_{n,m} \in \mathcal{P}) \leq 3 \Pr(G_{n,p} \in \mathcal{P}).$$

Graph Process:

$G_0 = ([n], \emptyset),\ G_1, G_2, \dots, G_m, \dots G_N = K_n$

$G_{m+1} = G_m$ plus random edge

G_m and $G_{n,m}$ have same distribution.

Markov Inequality

$X \geq 0$ is a random variable with finite mean μ.

$$\Pr(X \geq t) \leq \frac{\mu}{t}$$

Proof

$$E(X) = E(X \mid X < t)\Pr(X < t) + E(X \mid X \geq t)\Pr(X \geq t)$$

$$\geq t \Pr(X \geq t).$$

$\square$

Chebyshev Inequality

X is a random variable with finite mean μ and variance σ^2.

$$\Pr(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Proof

$$\Pr(|X-\mu| \geq t\sigma) = \Pr((X-\mu)^2 \geq t^2)$$

$$\leq \frac{E(|X-\mu|^2)}{t^2}$$

$$= \sigma^2/t^2.$$

$\square$

First Moment Method

Let X be a random variable with finite mean taking values in $\{0, 1, 2, \ldots\}$.

$$\Pr(X \neq 0) \leq E(X)$$

Proof

$$\Pr(X \neq 0) = \Pr(X \geq 1) \leq \frac{E(X)}{1}.$$

$\square$

Second Moment Method

Let X be a non-negative random variable with finite mean and variance. Then

$$\Pr(X > 0) \geq \frac{E(X)^2}{E(X^2)}$$

Proof

Let $$Y = \begin{cases} 0 & \text{if } X = 0 \\ 1 & \text{if } X > 0 \end{cases}$$

So $$XY = X$$

Cauchy-Schwartz inequality* implies

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\,\mathbb{E}(Y^2)$$

or

$$\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)\,\mathbb{P}(X>0).$$

$$\begin{aligned}
\mathbb{E}((X+tY)^2) &= \mathbb{E}(X^2) + 2t\mathbb{E}(XY) + t^2\mathbb{E}(Y^2) \\
&= (\mathbb{E}(X^2)^{1/2} + t\mathbb{E}(Y^2)^{1/2})^2 - 2t(\mathbb{E}(X^2)^{1/2}\mathbb{E}(Y^2)^{1/2} - \mathbb{E}(XY)) \\
&\geq 0 \text{ for all } t.
\end{aligned}$$

Put $t = -\mathbb{E}(X^2)^{1/2}/\mathbb{E}(Y^2)^{1/2}$ to obtain $\mathbb{E}(X^2)^{1/2}\mathbb{E}(Y^2)^{1/2} - \mathbb{E}(XY) \geq 0$.

* Consider quadratic $\mathbb{E}((X+tY)^2) \geq 0$, as a function of t.

Evolution of a random graph.

We look at how $G_0, G_1, \dots G_m, \dots$ evolves.

$\omega = \omega(n)$ denotes some slowly growing function e.g. $\omega = \log n$.

(1) $m \leq n^{1/2}/\omega$

G_m is a matching whp

whp: with high probability
i.e. with probability $1 - o(1)$
as $n \to \infty$.

Let $p = \frac{m}{N}$ and Let X_2 = number of paths of length 2 in $G_{n,p}$.

$$E_p(X_2) = 3\binom{n}{3} p^2$$

$$\leq \frac{n^3 \times n}{N^2 \times \omega^2}$$

$$\to 0.$$

$\Pr(G_{n,p}$ contains path of length 2$) = o(1)$

monotone property

$\Downarrow$

$\Pr(G_{n,m}$ contains a path of length 2$) = o(1)$.

(ii) $m = \omega n^{1/2}$, $m = o(n)$.

G_m contains a path of length 2 whp

Let $p = \frac{m}{N}$ and $X_2 = \#$ paths length 2.

$$E(X_2) = 3\binom{n}{3} p^2$$

$$\approx 2\omega^2$$

$$\to \infty$$

Does not imply $X_2 \neq 0$ whp.

Let $\mathcal{P}_2$ be the set of all paths of length two in K_n.

Let $\hat{X}_2$ = # of isolated paths of length 2 in $G_{n,p}$

$$\hat{X}_2 = \sum_{P \in \mathcal{P}_2} \mathbb{1}_{P \overset{i}{\subseteq} G_{n,p}}.$$

$$E(\hat{X}_2) = 3\binom{n}{3} p^2 (1-p)^{3(n-3)}$$

$$\geq (1-o(1)) \frac{n^3}{2} \cdot \frac{4w^2 n}{n^4} \cdot (1 - 6np)$$

$n = o(n)$

$\Downarrow$

$np = o(1)$

$$\longrightarrow \infty.$$

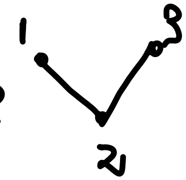
$$\hat{X}_2^2 = \sum_{P \in \mathcal{P}_2} \sum_{Q \in \mathcal{P}_2} \mathbb{1}_{P \overset{i}{\subseteq} G_{n,p}} \mathbb{1}_{Q \overset{i}{\subseteq} G_{n,p}}$$

$$= \sum_{P, Q \in \mathcal{P}_2} \mathbb{1}_{P \overset{i}{\subseteq} G_{n,p}} \mathbb{1}_{Q \overset{i}{\subseteq} G_{n,p}}$$

$P = Q$ or P, Q vertex disjoint

$$E(\hat{X}_2^2) = \sum_{P} \left\{ \sum_{Q} \Pr\left(G_{n,p} \overset{i}{\supseteq} Q \,\middle|\, G_{n,p} \overset{i}{\supseteq} P\right) \right\} \times \Pr\left(G_{n,p} \overset{i}{\supseteq} P\right)$$

Expression inside {} is same for all P.

$$= E(\hat{X}_2)\left(1 + \sum_{\substack{Q \cap \{1,2,3\} \\ = \emptyset}} \Pr\left(G_{n,p} \overset{i}{\supseteq} Q \,\middle|\, G_{n,p} \overset{i}{\supseteq}\right.\right.$$  $$\left.\left.\right)\right)$$

$$\leq E(\hat{X}_2)\left(1 + \binom{n}{3} p^2 (1-p)^{3(n-6)^{*}+1}\right)$$

$$\leq E(\hat{X}_2)\left(1 + (1-p)^{-3} E(\hat{X}_2)\right)$$

* conditioning means no edge to {1, 2, 3}

So

$$\Pr(\hat{X}_2 \neq 0) \geq \frac{E(\hat{X}_2)^2}{E(\hat{X}_2)\left(1 + (1-p)^{-3} E(\hat{X}_2)\right)}$$

$$= \frac{1}{(1-p)^{-3} + E(\hat{X}_2)^{-1}}$$

$$\to 1.$$

Thus

$$\Pr(G_{n,p} \supseteq \text{isolated 2-path}) \to 1$$

not monotone

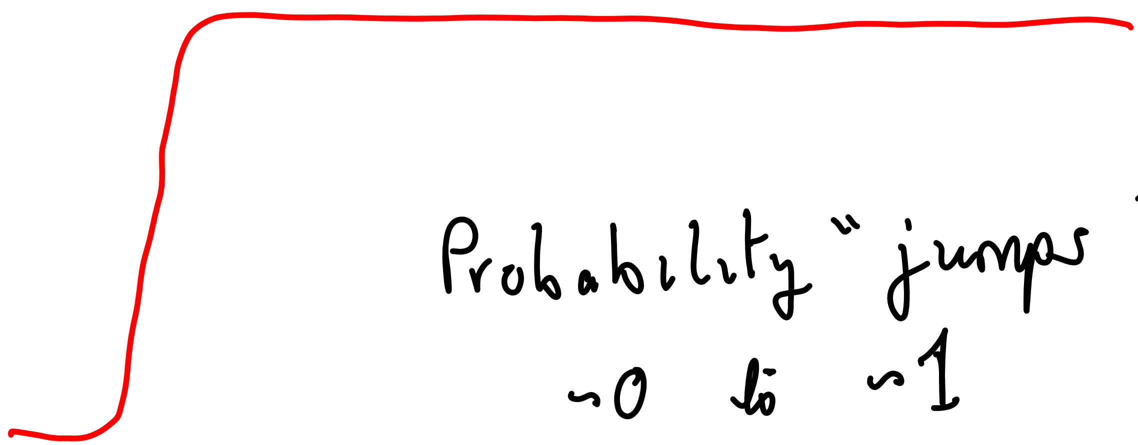
$$\Pr(G_{n,p} \supseteq \text{2-path}) \to 1$$

monotone

$$\Pr(G_m \supseteq \text{2-path}) \to 1$$

$$\Pr\left(G_m \supseteq \text{2-path}\right) = \begin{cases} o(1) & m \ll n^{1/2} \\ 1 - o(1) & m \gg n^{1/2} \end{cases}$$

We say that $n^{1/2}$ is the threshold for the existence of a 2-path in $G_{n,m}$



Probability "jumps" from ~ 0 to ~ 1

Note Title

1/3/2006

Small Trees

Fix $k \geq 3$.

$$m \leq \frac{n^{\frac{k-2}{k-1}}}{\omega} \Rightarrow G_m \text{ contains no tree with } k \text{ vertices.}$$

$$p = \frac{m}{N} \approx \frac{2}{\omega n^{k/(k-1)}}$$

Let

X_k = # of trees with k vertices in $G_{n,p}$

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

$$\leq \left(\frac{ne}{k}\right)^{k} k^{k-2} \left(\frac{3}{\omega n^{k/(k-1)}}\right)^{k-1}$$

$$< \left(\frac{3e}{\omega}\right)^{k-1}$$

$$\to 0.$$

$$\Pr(G_{n,p} \underbrace{\text{contains tree with } k \text{ vertices}}_{\text{monotone}}) \to 0$$

$$\Downarrow$$

$$\Pr(G_m \text{ contains a tree with } k \text{ vertices}) \to 0.$$

$m = \omega\, n^{\frac{k-2}{k-1}}$, $m = o(n)$

$\Rightarrow$ G_m contains a copy of <u>every</u> tree with k vertices.

$$p = \frac{m}{N}.$$

Fix some tree T with k vertices.

X_T = # of isolated copies of T in $G_{n,p}$.

$$E(X_T) = \binom{n}{k} \frac{k!}{\mathrm{aut}(T)^{*}} p^{k-1} (1-p)^{k(n-k)}$$

$$\simeq^{**} \frac{(2w)^{k-1}}{\mathrm{aut}(T)}$$

$$\to \infty.$$

*aut(H) = no. of automorphisms of H

$**$ = $(1+o(1))$ times .....

Let $\mathcal{T}$ be the set of copies of T in K_n.

$$E(X_T^2) = \sum_{T_1, T_2 \in \mathcal{T}} \Pr(T_2 \overset{i}{\subseteq} G_p \mid T_1 \overset{i}{\subseteq} G_p) \times \Pr(T_1 \overset{i}{\subseteq} G_p)$$

$$= E(X_T)\Big(1 + \sum_{\substack{T_2 \in \mathcal{T} \\ V(T_2) \cap [k] = \emptyset}} \Pr(T_2 \subseteq G_p \mid \hat{T} \subseteq G_p)\Big)$$

fixed copy of T on $[k]$.

$$\leq E(X_T)\big(1 + (1-p)^{-k} E(X_T)\big).$$

$$\Pr(X_T \neq 0) \geq \frac{E(X_T)^2}{E(X_T)\left(1+(1-p)^{-k} E(X_T)\right)}$$

$$\longrightarrow 1.$$

$\Pr(G_{n,p}$ contains isolated copy of $T) \to 1$

$\Downarrow$

$\Pr(G_{n,p}$ contains copy of $T) \to 1$

$\Downarrow$

$\Pr(G_m$ contains copy of $T) \to 1.$

Cycles

$m = O(n) \Longrightarrow G_m$ is a forest, whp

Suppose $m = n/\omega$

$$p = \frac{m}{N} \leq \frac{3}{\omega n}$$

X = # of cycles in $G_{n,p}$

$$E(X) = \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p^k$$

$$\leq \sum_{k=3}^{n} \frac{n^k}{2k} \cdot \frac{3^k}{\omega^k n^k}$$

$$= O(\omega^{-3})$$

$$\to 0.$$

$$\Pr(G_{n,p} \text{ is not a forest}) = o(1)$$

$$\Downarrow$$

$$\Pr(G_m \text{ is not a forest}) = o(1).$$

Poisson Convergence.

What happens if

$$m = c n^{(k-2)/(k-1)}$$

where $c > 0$ is constant?

Inclusion - Exclusion.

<u>Lemma</u>

Suppose $A_1, A_2, \cdots, A_r$ are events in some probability space Ω.

Suppose that $f_1, f_2, \cdots, f_s$ are boolean functions of $A_1, A_2, \cdots, A_s$

Suppose $\alpha_1, \alpha_2, \cdots \alpha_s$ are reals. Then if

$$\sum_{i=1}^{s} \alpha_i \Pr(f_i(A_1, A_2, \cdots, A_r)) \geq 0 \qquad (1)$$

whenever $\Pr(A_i) = 0$ or 1 then $(*)$ holds in general.

Write

$$f_i = \bigcup_{S \in T_i} \left(\left(\bigcap_{i \in S} A_i \right) \cap \left(\left(\bigcap_{i \notin S} \bar{A}_i \right) \right) \right)$$

so that

$$\Pr(f_i) = \sum_{S \in T_i} \Pr\left(\left(\bigcap_{i \in S} A_i \right) \cap \left(\bigcap_{i \notin S} \bar{A}_i \right) \right)$$

and then LHS (1) becomes

$$\sum_{S \subseteq [r]} \beta_S \Pr\left(\left(\bigcap_{i \in S} A_i \right) \cap \left(\bigcap_{i \notin S} \bar{A}_i \right) \right)$$

for some real β_S.

If (1) holds then $\beta_S \geq 0$, $\forall S$ since we can choose $A_i = \Omega$, $i \in S$, $A_i = \emptyset$ $i \notin S$.

$\square$

For $X \subseteq [r]$ let $A_X = \bigcap_{i \in X} A_i$

$$S_t = \sum_{|X|=t} \Pr(A_X)$$

$$\mathcal{E} = \{\text{none of } A_1, A_2, \dots, A_r \text{ occur}\}$$

Lemma

$$\Pr(\mathcal{E}) - \sum_{t=0}^{r} (-1)^t S_t \quad \begin{matrix} \leq 0 & r \text{ even} \\ \geq 0 & r \text{ odd} \end{matrix}$$

We only need to check when

$Pr(A_i) = 1 \quad 1 \leq i \leq l$

$Pr(A_i) = 0 \quad l < i \leq r$

$$Pr(\mathcal{E}) = \begin{cases} 1 & l = 0 \\ 0 & l \neq 0 \end{cases}$$

$$S_t = \binom{l}{t}$$

$l = 0$ trivial.

$\underline{l > 0}$

$$0 - \sum_{t=0}^{r} (-1)^t \binom{l}{t}$$

$$= \begin{cases} 0 & r \geq l \\ (-1)^r \binom{l-1}{r} & r < l \end{cases}$$

$\square$

Back to random graphs

Let $T_1, T_2, \dots, T_M$ be the list of copies of some fixed k vertex tree T.

$A_i = \{ T_i$ occurs as a component in $G_m \}$

Suppose $X \subseteq [M]$ with $|X| = t$, t fixed.

$\Pr(A_X) = 0$ if $\exists\, i, j \in X$ such that T_i, T_j share a vertex.

Suppose T_i, $i \in X$ are vertex disjoint.

$$\Pr(A_X) = \frac{\binom{\binom{n-kt}{2}}{m-(k-1)t}}{\binom{N}{m}}$$

Numerator = # ways of choosing m edges so that A_X occurs

Now

$$\frac{A^B}{B!} \geq \binom{A}{B} = \frac{A^B}{B!}\left(1-\frac{1}{A}\right)\left(1-\frac{2}{A}\right)\cdots\left(1-\frac{B-1}{A}\right)$$

$$\geq \frac{A^B}{B!}\left(1-\frac{B^2}{2A}\right)$$

So if A, B are functions of n and

$\frac{B^2}{A} \longrightarrow$ [redacted] 0 as $n \longrightarrow \infty$

then

$$\binom{A}{B} = (1+o(1))\frac{A^B}{B!}$$

Consider $\binom{\binom{n-kt}{2}}{m-(k-1)t}$ for $t \leq \log n$ say.

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

$$= N(1-O(kt/n))$$

So

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

$$\binom{\binom{n-kt}{2}}{m-(k-1)t} = (1+o(1))\,\frac{\left(N\left(1-O(kt/n)\right)\right)^{m-(k-1)t}}{\left(m-(k-1)t\right)!}$$

$$= \frac{(1+o(1))\;N^{m-(k-1)t}\;\left(1-O(mkt/n)\right)}{\left(m-(k-1)t\right)!}$$

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

Similarly

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

and so

$$\Pr(A_x) = \frac{\binom{\binom{n-kt}{2}}{m-(k-1)t}}{\binom{N}{m}}$$

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

So

$$S_t \approx \frac{1}{t!}\binom{n}{k,k,k,\dots,k}\left(\frac{k!}{\operatorname{aut}(T)}\right)^t\left(\frac{m}{N}\right)^{(k-1)t}$$

$$\approx \frac{n^{kt}}{t!(k!)^t}\cdot\left(\frac{k!}{\operatorname{aut}(T)}\right)^t\cdot\left(\frac{cn^{(k-2)/(k-1)}}{N}\right)^{(k-1)t}$$

$$\approx \frac{\lambda^t}{t!}$$

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

Fix r large

$$\Pr(\nexists \text{ component copy of } T) =$$

$$\sum_{t=0}^{r} (-1)^t S_t \quad + \theta_r$$

theta_r is non-positive if r is even
and non-negative if r is odd

$$= \sum_{t=0}^{r} (-1)^t (1+o(1)) \frac{\lambda^t}{t!} \quad + \theta_r$$

$$= (1+o(1)) \sum_{t=0}^{r} (-1)^t \frac{\lambda^t}{t!} \quad + \theta_r$$

[Here r can be thought of as a large constant while $n \to \infty$.]

$$(1+o(1))\sum_{t=0}^{2r-1}(-1)^t\frac{\lambda^t}{t!}\leq$$

$$\Pr(\nexists \text{ component copy of } T)$$

$$\leq (1+o(1))\sum_{t=0}^{2r}(-1)^t\frac{\lambda^t}{t!}$$

Letting $r \to \infty$

$$\Pr(\nexists \text{ component copy of } T) \Rightarrow e^{-\lambda}$$

If there is a copy of T which is not a component then either

(i) $\exists$ cycle — $\Pr((i)) = o(1)$

(ii) T is part of a tree of size $> k$ — $\Pr((ii)) = o(1)$.

So

$$\Pr(\exists \text{ copy of } T) \geq 1 - e^{-\lambda}.$$

Structure of graph when

$m = \frac{1}{2} cn$, $0 < c < 1$ constant.

We will work in $G_{n,p}$

$$p = \frac{c}{n} \approx \frac{m}{N}$$

Cycles

Whp the are $\leq \log n$ edges on cycles.

Let X_k = # cycles of length k.

$$E(X_k) = \binom{n}{k} \frac{(k-1)!}{2} p^k$$

$$< \frac{n^k}{k!} \frac{(k-1)!}{2} \left(\frac{c}{n}\right)^k$$

$$= \frac{c^k}{2k}.$$

So if $X = 3X_3 + 4X_4 + \dots + nX_n$

$\Rightarrow$ # edges on cycles

then

$$E(X) \le \sum_{k=3}^{n} k \cdot \frac{c^k}{2k} < \frac{1}{1-c}.$$

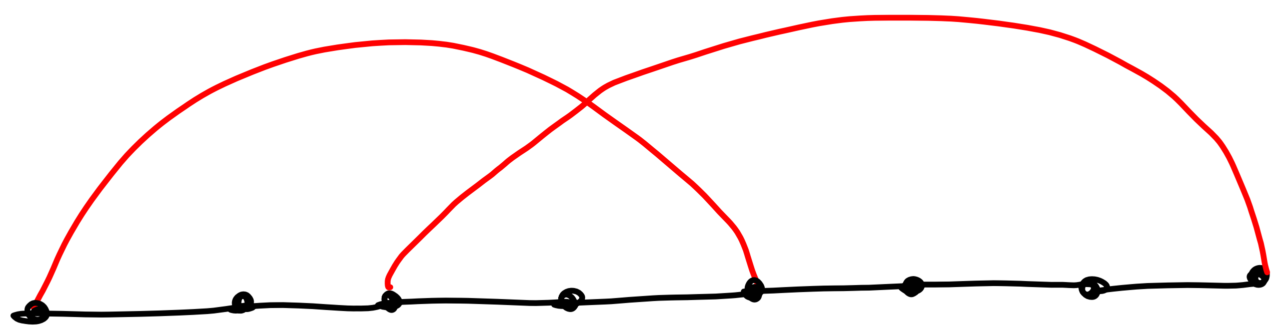
Applying the Markov inequality gives

$$\Pr(X > \log n) \le \frac{1}{(\log n)(1-c)} = o(1).$$

Claim: whp $\nexists$ a pair of cycles that are in the same component

Proof

If a pair exists then there is a minimal pair C_1, C_2



$$E(\# C_1, C_2) \leq \sum_{k \geq 3} \binom{n}{k} \cdot \frac{k!}{2} \; k^2 \, p^{k+1}$$

$$\leq \frac{1}{n} \sum_{k \geq 3} c^{k+1} k^2$$

$$\longrightarrow 0.$$

So whp every component contains at most one cycle.

We now show that whp size of largest component is $O(\log n)$.

Let X_k be the number of components of size k that are unicyclic

$E(X_k)$

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

$$\leq \frac{n^k}{k!} e^{-\frac{k(k-1)}{2n}} k^k \frac{c^k}{n^k} e^{-ck+ck(k-1)/2n+ck/2n}$$

$$\binom{n}{k} = \frac{n^k}{k!} \prod_{i=0}^{k-1} \left(1-\frac{i}{n}\right) \quad \& \quad 1-x \leq e^{-x}$$

$$\leq \frac{n^k}{k!} e^{-\frac{k(k-1)}{2n}} k^k \frac{c^k}{n^k} e^{-ck + \frac{ck(k-1)}{2n} + c/2}$$

$$\leq (ce^{1-c})^k e^{c/2}.$$

So if $\omega \to \infty$,

Pr($\exists$ unicyclic component of size $\geq \omega$)

$$\leq \sum_{k=\omega}^{n} e^{c/2} (ce^{1-c})^k$$

$\to 0$

since $ce^{1-c} < 1$ for $c \neq 1$.

Now let X_k be the number of isolated trees.

Let

$$\alpha = c - 1 - \log c$$

<u>Theorem</u>

Suppose $\omega \to \infty$

(i) Whp $\exists$ an isolated tree of size

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

(ii) Whp $\nexists$ an isolated tree of size

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

Now let X_k = number of isolated trees of size k.

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

(i) Suppose $k = O(\log n)$. Then

$$E(X_k) \approx \frac{(1+o(1))}{\sqrt{2\pi k}} \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck}$$

$$= \frac{(1+o(1))}{\sqrt{2\pi}} \frac{n}{k^{5/2}} \left(ce^{1-c}\right)^k.$$

Putting $k = k_-$

we see that

$$E(X_k) = \frac{(1+o(1))}{\sqrt{2\pi}} \frac{n}{k^{5/2}} \left(ce^{1-c}\right)^k.$$

$$= \frac{(1+o(1))}{\sqrt{2\pi}} \cdot \frac{n}{k^{5/2}} \cdot \frac{(\log n)^{5/2} e^{\alpha w}}{n}$$

$$\geqslant A e^{\alpha w}.$$

We continue via second moment method.

$$E(X_k^2) \le E(X_k)\left(1 + (1-p)^{-k} E(X_k)\right)$$

[Same argument as for fixed tree T of size k]

Thus

$$\frac{E(X_k)^2}{E(X_k^2)} \ge 1 - \frac{1}{2Ae^{\alpha w}} \longrightarrow 1.$$

and we have (i).

For (ii) we go back to

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

$$\leq \frac{A}{\sqrt{k}} \left(\frac{ne}{k}\right)^k e^{-k^2/2n} k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck+ck^2/2n}$$

$$\leq \frac{An}{k^{5/2}} (ce^{1-c})^k$$

and then

$$\sum_{k=k_+}^{n} E(X_k) \leq An \sum_{k=k_+}^{n} \frac{(ce^{1-c})^k}{k^{5/2}} = o(1).$$

Useful Identity

$0 \le c \le 1$ implies $\frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k = 1.$

Proof

Assume first $c < 1$.

Let σ = number of vertices of $G_{n,p}$ that lie on unicyclic components.

$$n = \sum_{k=1}^{n} k X_k + \sigma$$

# of trees of size k.

so

$$n = \sum_{k=1}^{n} k E(X_k) + E(\sigma)$$

(i) $E(\sigma) \leq \log n$

(ii) $$\sum_{k \geq k_+} k E(X_k) \leq \frac{1}{c} \sum_{k=k_+}^{n} \frac{(ce^{1-c})^k}{k^{3/2}} = O(1).$$

(iii) If $k < k_+$ then

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

$$= (1+o(1)) \frac{n}{c} \frac{k^{k-1}}{k!} (ce^{-c})^k$$

So

$$n = \sum_{k=1}^{n} k E(X_k) + E(\sigma)$$

$$= O(n) + \frac{n}{c} \sum_{k=1}^{k^+} \frac{k^{k-1}}{k!} (c e^{-c})^k$$

$$= O(n) + \frac{n}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (c e^{-c})^k$$

Now divide through by n.

$\square$

Structure of graph when

$m = \frac{1}{2} cn$, $c > 1$ constant.

We will work in $G_{n,p}$

$$p = \frac{c}{n} \approx \frac{m}{N}$$

Suppose now that X_k is the number of components of size k. Then

$$E(X_k) \leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}$$

$$\leq \frac{A}{\sqrt{k}} \left(\frac{ne}{k}\right)^k e^{-k^2/2n} k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck + ck^2/n}$$

$$\leq \frac{An}{k^{5/2}} \left(ce^{1-c+ck/n}\right)^k .$$

Now let $B_1 = B_1(c)$ be small enough so that

$$ce^{1-c+B_1} < 1.$$

and let $B_0 = B_0(c)$ be large enough so that

$$\left(ce^{1-c+o(1)}\right)^{B_0 \log n} < \frac{1}{n^2}.$$

It follows that whp $\nexists$ a component of size

$$k \in [B_0 \log n, B_1 n]$$

Our calculations for $c<1$ can be repeated to show that if

$$\alpha = c - 1 - \log c$$

Theorem

Suppose $\omega \to \infty$

(i) Whp $\exists$ an isolated tree of size

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

(ii) Whp $\nexists$ an isolated tree of size

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

provided $\omega = O(\log n)$.

We can say a little more about components of size k, $k = O(\log n)$.

If we repeat the calculations for $c < 1$ then we find that if Y_k is the number of isolated trees of size

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

then

$$E(Y_k) \geq A e^{\alpha \omega}$$

for some $A = A(c) > 0$.

$$E(Y_k^2) \leqslant E(Y_k) + E(Y_k)^2 (1-p)^{-k^2}$$

So

$$\operatorname{Var}(Y_k) \leqslant E(Y_k) + E(Y_k)^2\left((1-p)^{-k^2} - 1\right)$$

$$\leqslant E(Y_k) + 2E(Y_k)^2 c k^2/n.$$

So

$$\Pr\left(\left|Y_k - E(Y_k)\right| \geqslant \epsilon E(Y_k)\right)$$

$$\leqslant \frac{1}{\epsilon^2 E(Y_k)} + \frac{2ck^2}{\epsilon^2 n}. \quad (*)$$

We now estimate the total number of vertices on small tree components i.e. size $\leq B_0 \log n$.

(i) $1 \leq k \leq k_0 = \frac{1}{2\alpha} \log n$

$$E\left(\sum_{k=1}^{k_0} k Y_k\right) \approx \frac{n}{c} \sum_{k=1}^{k_0} \frac{k^{k-1}}{k!} (ce^{-c})^k$$

$$\approx \frac{n}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k$$

Since $\frac{k^{k-1}}{k!} < e^k$ and $ce^{1-c} < 1$.

Putting $\epsilon = \frac{1}{\log n}$ we see that the probability that any Y_k deviates from its mean by more than $1 \pm \epsilon$ is at most (see (*) on p 6)

$$\sum_{k=1}^{k_0} \left[\frac{(\log n)^2}{n^{1/3}} + O\left(\frac{(\log n)^4}{n} \right) \right] = o(1).$$

Thus whp

$$\sum_{k=1}^{k_0} k Y_k \approx \frac{n}{c} \sum_{k=1}^{k_0} \frac{k^{k-1}}{k!} \left(ce^{-c}\right)^k$$

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

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

where $0 < x < 1$ and $xe^{-x} = ce^{-c}$

$$= \frac{nx}{c}.$$

Now consider $k_0 < k \leq B_0 \log n$.

$$E\left(\sum_{k=k_0+1}^{B_0 \log n} k Y_k\right) \leq$$

$$\frac{n}{c} \sum_{k=k_0+1}^{B_0 \log n} \frac{An}{k^{3/2}} \left(ce^{1-c} + ck/n\right)^k$$

$$= o\left(n / (\log n)^{3/2}\right).$$

So, by the Markov inequality, whp,

$$\sum_{k=k_0+1}^{B_0 \log n} k Y_k = o(n).$$

Now consider the number of vertices Z_k on non-tree components with k vertices, $1 \leq k \leq B_0 \log n$.

$$E\left(\sum_{k=1}^{B_0 \log n} Z_k\right) \leq \sum_{k=1}^{B_0 \log n} \binom{n}{k} k^{k-2} \binom{k}{2} \left(\frac{c}{n}\right)^k \left(1-\frac{c}{n}\right)^{k(n-k)}$$

$$\leq \sum_{k=1}^{B_0 \log n} \left(c e^{1-c+k/n}\right)^k$$

$$= O(1).$$

So, by the Markov inequality, whp

$$\sum_{k=1}^{B_0 \log n} Z_k = o(n).$$

So far: whp

$xe^{-x} = ce^{-c}$

there are $\approx \frac{nx}{c}$

vertices on components of size k,

$1 \leq k \leq \beta_0 \log n$.

The giant component.

Let $c_1 = c - \frac{\log n}{n^2}$ and $p_1 = \frac{c_1}{n}$

and define p_2 by

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

Then

$$G_{n,p} = G_{n,p_1} \cup G_{n,p_2}$$

since probability e is not included in is

$(1-p_1)(1-p_2)$.

Note that

$$p_2 \geq \frac{\log n}{n^2}.$$

If $x_1 e^{-x_1} = c_1 e^{-c_1}$ then $x_1 \approx x$ and so by our previous analysis, whp, G_{n,p_1} has no components of size in the range $[B_0 \log n, B_1 n]$.

Suppose there are components $C_1, C_2, \dots, C_\ell$ with $|C_i| > B_0 n$. Thus $\ell \leq \frac{1}{B_0}$.

Now we add in the edges of G_{n,p_2}.

$$\Pr(\exists\, i,j : \text{no } G_{n,p_2} \text{ edge joins } C_i, C_j)$$

$$\leq \binom{l}{2} (1-p_2)^{(B_0 n)^2}$$

$$\leq l^2 e^{-B_0^2 (\log n)^2}$$

$$= o(1).$$

So whp $G_{n,p}$ has a unique component of size greater than $B_0 \log n$, and it is of size $\left(1 - \frac{x}{c}\right) n$.

Duality

Let $N = \frac{n x}{c} \approx$ # vertices outside giant whp.

Let $q = \frac{x}{N}$ ($= p$).

Note that $x < 1$ and # of isolated trees of size k is whp

$$\approx \frac{n}{c} \cdot \frac{k^{k-2}}{k!} (c e^{-c})^k$$

$$= \frac{N}{x} \cdot \frac{k^{k-2}}{k!} (x e^{-x})^k.$$

Thus graph outside of giant component is asymptotically equal to $G_{N, \frac{x}{N}}$ in distribution.

Branching Processes

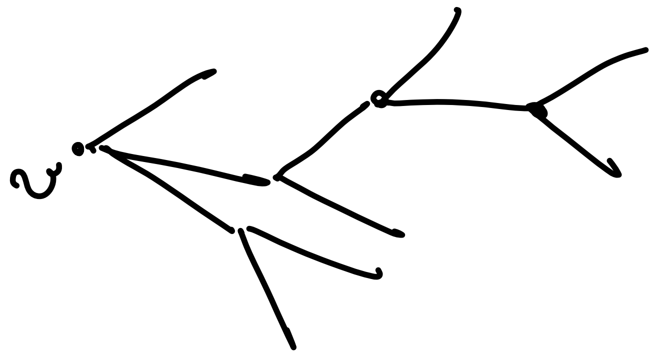
If $p = c/n$ and $d(v)$ is the degree of vertex v then

$$\Pr(d(v)=k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$= (1+o(1)) \frac{c^k e^{-c}}{k!}$$

i.e. the degree distribution is asymptotically <u>Poisson</u> with mean c.

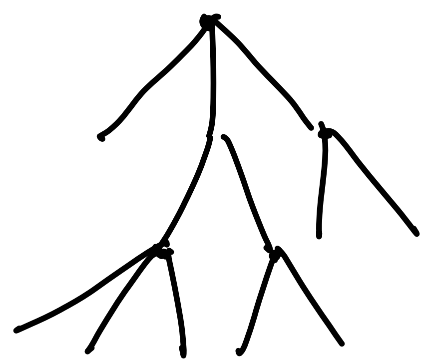
Since there are few "small" cycles, locally, $G_{n,p}$ should look like



and this has led to a comparison with **Branching Processes.**

It is not really so useful a method for here, but it can be the right approach for other models of a random graph.

In a simple branching process there is an initial individual who "gives birth" to X_1 children and then dies. Each of the X_1 individuals give birth and die and so on.



The number of children X produced by an individual is a random variable independent of the number produced by any other.

Let

$$p_k = P_r(X = k), \qquad k = 0, 1, 2, \ldots$$

and

$$G(z) = \sum_{k=0}^{\infty} p_k z^k$$

is the probability generating function (p.g.f.) of X.

Let

$$\mu = E(X)$$
$$= G'(1).$$

Let X_t be the number of individuals in generation t. Thus

$$X_0 = 1$$

$$E(X_{t+1}) = \sum_{k=0}^{\infty} E(X_{t+1} \mid X_t = k)\, P_r(X_t = k)$$

$$= \sum_{k=0}^{\infty} k\mu\, P_r(X_t = k)$$

$$= \mu E(X_t)$$

and so

$$E(X_t) = \mu^t.$$

Let T denote the total size of the set of individuals produced. $T = \infty$ is allowed and $P_r(T=\infty)$ is one of the important parameters of the process.

Theorem

$P_r(T < \infty) = y$ where y is the smallest non-negative root of

$$y = G(y).$$

In particular, $y=1$ if $\mu \leq 1$.

Before proving this, let us consider the case where X has Poisson distribution with mean c.

$$G(z) = \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} z^k$$

$$= e^{c(z-1)}$$

From the theorem, the "extinction probability" y satisfies

$$y = e^{c(y-1)}$$

But then

$$cy\, e^{-cy} = c\, e^{-c}$$

Assume $c > 1$ and then $x = cy < 1$.

If we choose a vertex v and look at the BFS tree grown from v then (as we will check) this looks like our branching process.

If $T = \infty$ corresponds to being in the giant and v is chosen randomly, then

$$\Pr(v \in \text{Giant}) \approx 1 - y = 1 - \frac{x}{c}.$$

Proof of Theorem

Let G_t be the p.g.f. for X_t. Thus

$$G_{t+1}(z) = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} P_r(X_{t+1}=k \mid X_t=\ell) P_r(X_t=\ell)\, z^k$$

$$= \sum_{\ell=0}^{\infty} G_t(z)^{\ell}\, P_r(X_t=\ell) \quad {**}$$

$$= G(G_t(z))$$

** If X, Y have p.g.f.'s f, g then X + Y has p.g.f. f × g.

Let $y_t = \Pr(X_t = 0)$ so that

$$y_t = G_t(0) = G(G_{t-1}(0)) = G(y_{t-1}).$$

Now y_t is monotone increasing to $\Pr(T<\infty)$ and so the continuity of G implies

$$y = G(y).$$

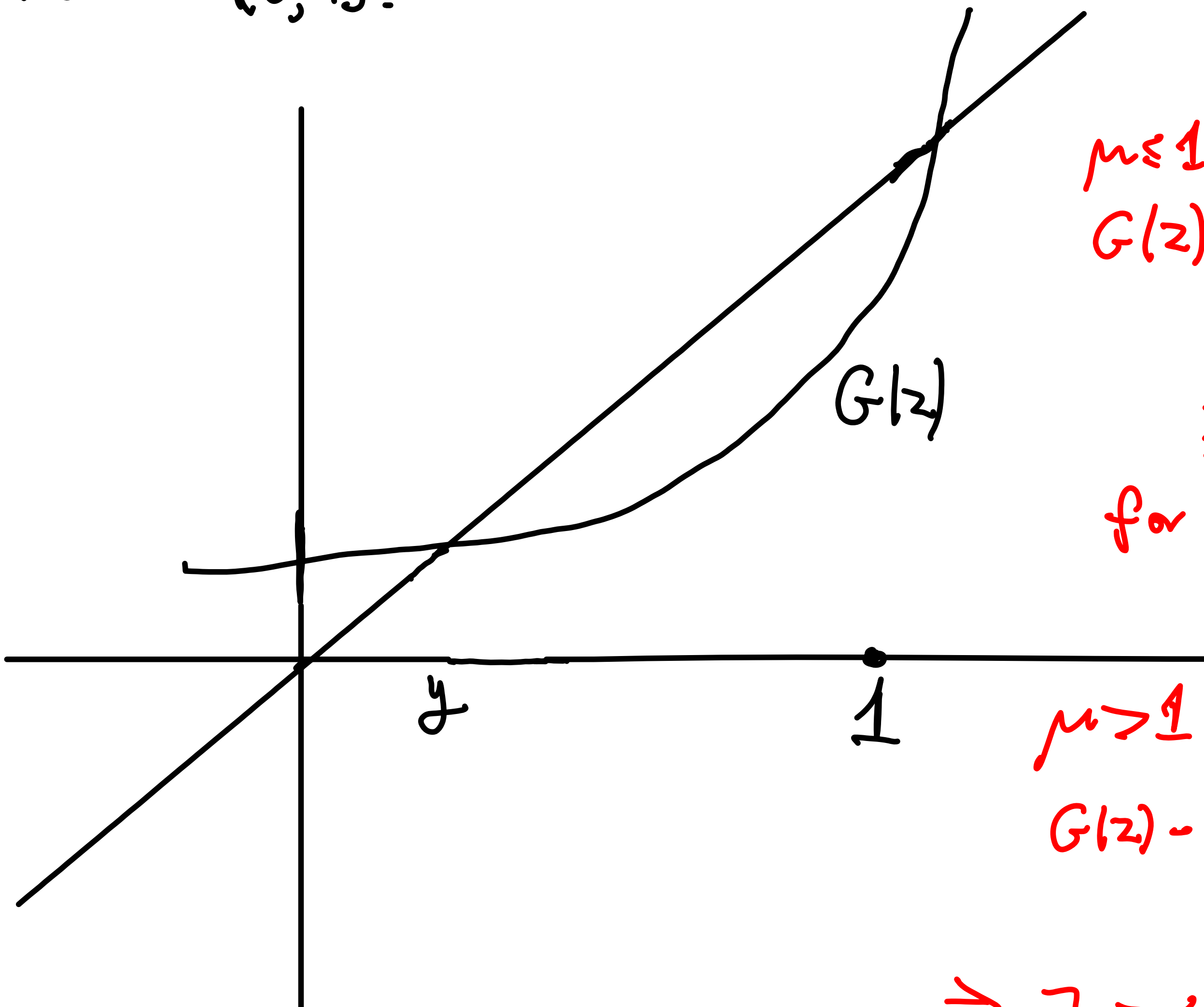
If ξ is any non-negative root of $z = G(z)$ then

$$y_1 = G(0) \le G(\xi) = \xi$$

and

$$y_t \le \xi \implies y_{t+1} = G(y_t) \le G(\xi) = \xi.$$

G is strictly convex on $[0,1]$ — $G''(z) = \sum_{k=2}^{\infty} k(k-1) p_k z^k > 0$ for $z \in (0,1]$.



$\mu \leq 1$:

$$G(z) > G(1) - G'(1)(1-z)$$
$$= 1 - \mu(1-z)$$
$$\geq z$$

for $0 \leq z < 1$

$\mu > 1$:

$$G(z) - z \quad \begin{array}{ll} < 0 & z = 1-\epsilon \\ \geq 0 & z = 0 \end{array}$$

$\Rightarrow \exists\, z < 1-\epsilon$ s.t.
$G(z) - z = 0$. $\square$

Thus

$$y = \Pr(T < \infty) = \lim_{t \to \infty} \Pr(T \leq t)$$

and we can write

$$\Pr(T \leq t) = y - \sigma(t)$$

where $\sigma(t) \geq 0$ and $\lim_{t \to \infty} \sigma(t) = 0.$

Back to $G_{n,p}$, $p = c/n$ $c > 1$.

Suppose we choose a vertex a and do a BFS from a until either

(i) we have explored the component C_a containing a

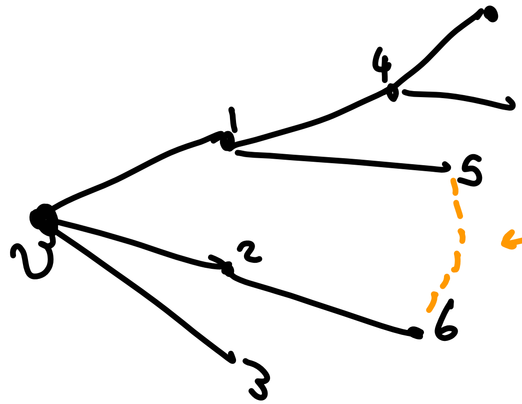
or

(ii) explored $\omega \to \infty$ vertices.*

Let T_a be the (partial) BFS tree produced.

We are going for ease of proof rather than best possible

Now fix a tree H with $\leq \omega = n^{1/2}(\log n)^3$ vertices and maximum degree $(\log n)^2$.



We do not include these edges in def. of T_a

Let d_i = degree of i, $i = 0, 1, 2, \dots, \ell$

$$\Pr(H = T_a) = \prod_{i=0}^{\ell} \binom{n_i}{d_i} p^{d_i} (1-p)^{n_i - d_i - (s_i)}$$

where

$$n_i = n - 1 - d_1 - \dots - d_{i-1}$$

$$= \left(\prod_{i=0}^{\ell} \frac{c^{d_i} e^{-c}}{d_i!} \right) \left(1 + O\left(\frac{\omega}{n} \right) \right)$$

$$= \left(\prod_{i=0}^{\ell} \frac{c^{d_i} e^{-c}}{d_i!} \right) \left(1 + O\left(\frac{\omega}{n} \right) \right)$$

$$= \Pr(H \text{ is branching process tree}) \times (1 + o(1))$$

Thus,

$$\Pr(|C_a| < \omega) =$$

$$\Pr(|C_a| < \omega \wedge \Delta \geq \overbrace{(\log n)^2}^{L}) +$$

$$\le n \binom{n-1}{L} \left(\frac{c}{n}\right)^L$$

$$\le n \left(\frac{ce}{L}\right)^L = o(1)$$

$$\Pr(|C_a| < \omega \wedge \Delta \le (\log n)^2)$$

$$= o(1) + \sum_{H:|H|<\omega} \Pr\left((T_a = H) \wedge \left(\Delta(G \setminus C_a) \leq (\log n)^2\right)\right)$$

$$= o(1) + \sum_{H:|H|<\omega} \Pr(T_a = H) \Pr\left(\Delta(G \setminus C_a) \leq (\log n)^2\right)$$

$$= o(1) + (1+o(1)) \sum_{H:H<\omega} \Pr(T_a = H)$$

$$= o(1) + (1+o(1)) \sum_{H:|H|<\omega} \Pr(H \text{ is branching process tree})$$

$$= o(1) + (1+o(1)) \Pr(T_a < \omega) \sim y.$$

Thus if

$X_0 = \# v : |C_a| < \omega$, $\omega \to \infty$

then

$$E(X) = ny\left(1 - O(\omega/n) - o(\omega)\right).$$

We next show, via Chebyshef, that X_0 is concentrated around its mean.

C_v

In constructing C_v we do not look at edges here i.e. they are unconditioned

We claim that for $b \neq a$

$$\Pr(|C_b| < \omega \mid |C_a| < \omega) \qquad (*)$$

$$\le \frac{|C_a|}{n} + (1+o(1)) \Pr(|C_b| < \log n)$$

$\Pr(w \in C_a)$

Fixing C_a, we replace n by $n - (\le \omega)$ in computing $\Pr(|C_a| < \omega)$.

It follows from ⊛ on previous page that

$$E(X_0^2) \leq$$
$$E(X_0) + E(X_0) \times \frac{\omega}{n} + (1+o(1)) E(X_0)^2$$

i.e.

$$Var(X_0) \leq 2E(X_0) + \eta E(X_0)^2$$

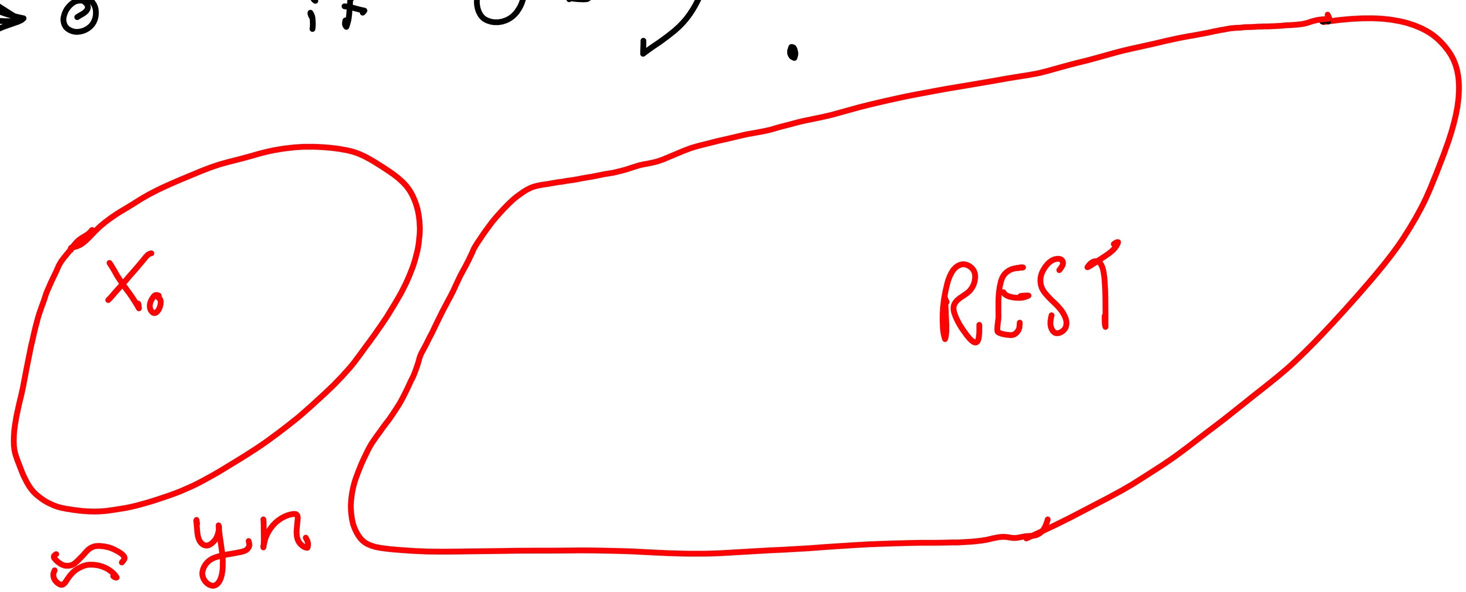
where $\eta \longrightarrow 0$.

Then

$$\Pr\left(|X_0 - E(X_0)| \geq \theta n\right)$$

$$\leq \frac{2E(X_0)}{\theta^2 n^2} + \frac{\eta\, E(X_0)^2}{\theta^2 n^2}$$

$\longrightarrow 0$ if $\theta = \eta^{1/3}$.



Our aim now is to show that REST is connected, without using previous analysis.

Suppose $|C_v| \geq n^{1/2}(\log n)^3$ and we stop our DFS from v when we reach $n^{1/2}(\log n)^3$.

size $= n^{1/2}(\log n)^3$

We argue next that whp

$$|N(S_a)| \geq n^{1/2}(\log n)$$

Indeed

$$\Pr\Big(\exists S,T : |S| = \underbrace{n^{1/2}(\log n)^3}_{k},\ |T| = \underbrace{n^{1/2}\log n}_{\ell} :$$

S induces a connected subgraph and

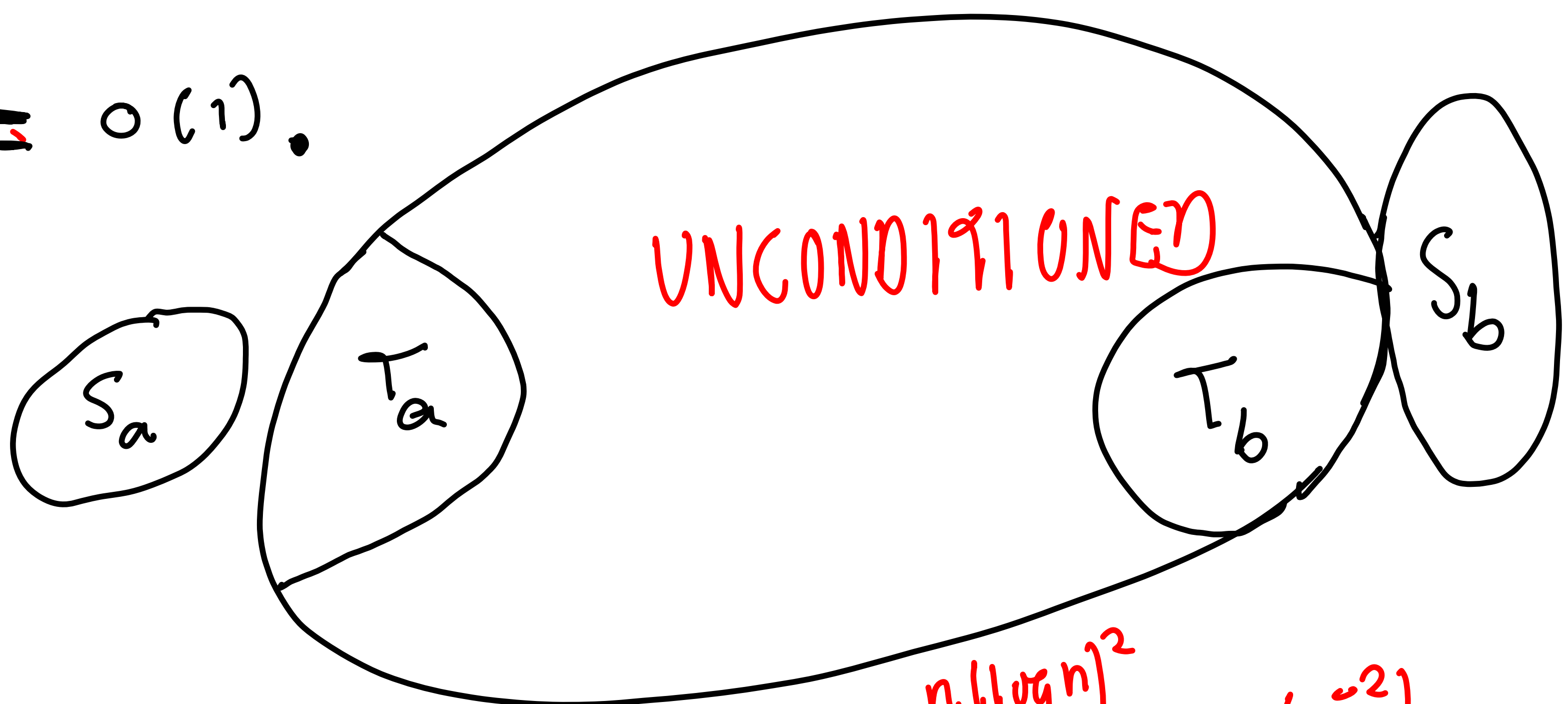
there are no $S : [n] \setminus S \cup T\Big)$

$$\leq \binom{n}{k}\binom{n}{\ell} k^{k-2} p^{k-1} (1-p)^{k(n-k-\ell)}.$$

$$\leq \left(\frac{ne}{k}\right)^k \cdot \left(\frac{ne}{\ell}\right)^\ell \cdot k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck(1-o(1))}$$

$$\leq n\left(ce^{1-c} \cdot n^{1/(\log n)^2}\right)^k \qquad \ell \leq \frac{k}{(\log n)^2}$$

$$= o(1).$$



$$\Pr(\text{no } T_a, T_b \text{ edges}) \leq (1-p)^{n(\log n)^2} = o(n^{-2}).$$

This shows that vertices a, $|C_a| \geq n^{\frac{1}{2}}(\log n)^3$ form a connected component.

Connectivity of random graphs

Let $p = \frac{\log n + c_n}{n}$. We prove

$$\lim_{n\to\infty} \Pr(G_{n,p} \text{ is connected}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-e^{-c}} & c_n \to c \\ 1 & c_n \to +\infty \end{cases}$$

If $p_1 > p_2$ then we can write

$$G_{n,p_1} = G_{n,p_2} \cup G_{n,p_3}$$

where $$(1-p_1) = (1-p_2)(1-p_3)$$

and so

$$\Pr(G_{n,p_1} \text{ is connected})$$

$$\geq \Pr(G_{n,p_2} \text{ is connected})$$

can replace "is connected" by any monotone ↑ property.

It suffices to prove that

$$\Pr(G_{n,p} \text{ is connected}) \longrightarrow e^{-e^{-c}}$$

when $p = \frac{\log n + c}{n}$.

Now

$$\Pr(G_{n,p} \text{ is } \textbf{not} \text{ connected})$$

$$= \Pr\left(\bigcup_{i=1}^{n/2} \exists \text{ a component of size } i \right)$$

So we have

$\Pr(\exists \text{ isolated vertex}) \leq$

$$\Pr(G_{n,p} \text{ is not connected}) \leq$$

$$\Pr(\exists \text{ isolated vertex}) + \sum_{k=2}^{n/2} \Pr\left(\exists \begin{array}{c}\text{component of} \\ \text{size } k\end{array}\right)$$

Now

$$\sum_{k=2}^{n/2} \Pr(\exists \text{ component of size } k)$$

$$\leq \sum_{k=2}^{n/2} E(\#\text{ of components of size } k)$$

$$\leq \sum_{k=2}^{n/2} \underbrace{\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}}_{u_k}$$

For $2 \leq k \leq 10$

$$u \leq e^k n^k \cdot \left(\frac{\log n + c}{n}\right)^{k-1} \cdot e^{-k(n-10)\frac{\log n + c}{n}} \leq (1+o(1)) \frac{e^{k(1-c)} (\log n)^{k-1}}{n^{k-1}}$$

and for $k \geq 10$

$$u_k \leq \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{\log n + c}{n}\right)^{k-1} e^{-k(\log n + c)/2}$$

$$\leq n\left(\frac{e^{1-c/2+o(1)} \log n}{n^{1/2}}\right)^k$$

so

$$\sum_{k=2}^{n/2} u_k \leq (1+o(1)) \frac{e^{-c} \log n}{n} + \sum_{k=10}^{n/2} n^{1+o(1)-k/2}$$

$$= O\left(n^{o(1)-1}\right).$$

It follows that

$\Pr(G_{n,p}$ is connected$) =$

$\Pr(\nexists$ an isolated vertex$) + o(1)$.

So now let

X_0 = the number of isolated vertices in $G_{n,p}$.

Then

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

$$= n \exp\{(n-1)\log(1-p)\}$$

$$= n \exp\left\{-(n-1)\sum_{k=1}^{\infty} \frac{p^k}{k}\right\}$$

$$= n \exp\left\{-(\log n + c) + O\left(\frac{(\log n)^3}{n}\right)\right\}$$

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

If we let

A_i be the event {vertex i is isolated}

and if
$$S_t = \sum_{\substack{X \subseteq [n] \\ |X| = t}} \Pr(A_X)$$

then
$$S_t = \binom{n}{t} (1-p)^{t(n-t) + \binom{t}{2}}$$

$$\approx e^{-tc}/t! \qquad t = O(1).$$

Thus we deduce, as in our study of isolated trees,

that
$$\lim_{n\to\infty} \Pr(X_0 = 0) = e^{-e^{-c}}.$$

Hitting Time Version in Graph Process

Let
$$m_1^* = \min\{m : \delta(G_m) \geq 1\}$$
$$m_c^* = \min\{m : G_m \text{ is connected}\}$$

We show

$$m_1^* = m_c^* \quad \text{whp}$$

Let

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

and

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

We first show that whp

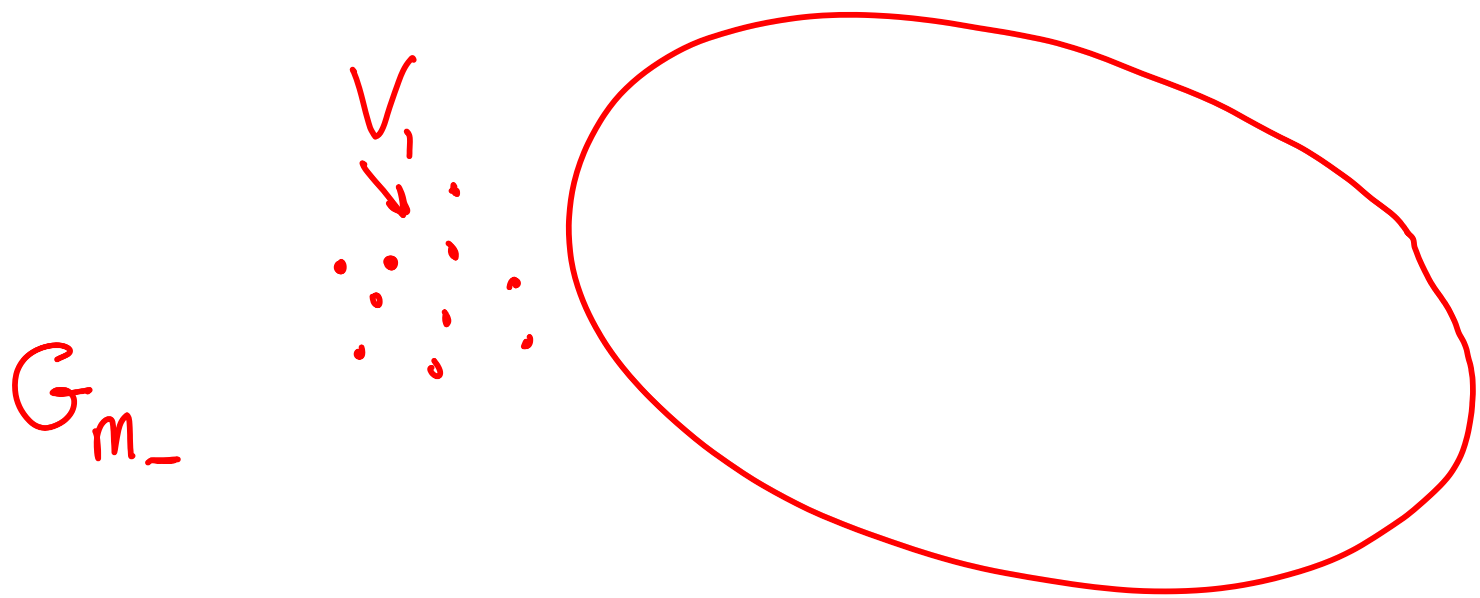
(i) G_{m_-} consists of a giant connected component plus a set V_1 of $\leq 2\log n$ vertices.

(ii) G_{m_+} is connected.

Assume (i) and (ii).

It follows that whp

$$m_- \le m_1^* \le m_c^* \le m_+$$



To create G_{m_+} we add $m_+ - m_-$ random edges.

$m_1^* = m_c^*$ if none of these edges is contained in V_1

Thus

$$\Pr(m_1^* < m_c) \leq O(1) + (m_+ - m_1)\,\frac{\frac{1}{2}|V_1|^2}{N - m_+}$$

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

$$= O(1).$$

(1) Let $p_- = \frac{m_-}{N} \approx \frac{\log n - \log\log n}{n}$

and let X_1 = # isolated vertices in G_{n,p_-}.

Then

$$E(X_1) = n(1-p_-)^{n-1}$$

$$= n e^{-np + O(np^2)}$$

$$\approx \log n.$$

$$E(X_1^2) = E(X_1) + n(n-1)(1-p)^{2n-3}$$

$$\leq E(X_1) + E(X_1)^2 (1-p)^{-3}$$

so

$$\mathrm{Var}(X_1) \leq E(X_1) + 4E(X)^2 p$$

$$\Pr(X_1 \geq 2\log n) = \Pr\big(|X_1 - E(X_1)| \geq (1+o(1))E(X_1)\big)$$

$$\leq (1+o(1))\left(\frac{1}{E(X_1)} + 4p\right)$$

$$= o(1).$$

Having $\geq 2\log n$ isolated vertices is a monotone property and so whp G_{m_-} has $< 2\log n$ isolated vertices.

To show that the rest of G_{m_-} is a single component we let X_k, $2 \leq k \leq \frac{n}{2}$ be the number of components with k vertices in G_{p_-}.

Repeating the calculation on p5

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

Let $\mathcal{E} = \{\exists \text{ component of size } 2 \leq k \leq \frac{1}{2}n\}$

$$\Pr(G_{m_-} \in \mathcal{E}) \leq O(\sqrt{n}) \Pr(G_{n,p_-} \in \mathcal{E})$$

$$= o(1)$$

and this complete proof of (i).

(ii) G_{m_+} is connected whp.

This follows from $G_{n,p}$ is connected whp for $np - \log n \to \infty$

or by implication G_m is connected whp if

$$n \cdot \frac{m}{N} - \log n \to \infty$$

$$\frac{n\, m_+}{N} = \frac{n\left(\frac{1}{2} n \log n + \frac{1}{2} n \log\log n\right)}{N}$$

$$\approx \log n + \log\log n .$$

k - connectivity.

Here we will prove that if $k = O(1)$ and

$$m = \frac{1}{2} n \left(\log n + (k-1) \log\log n + c_n \right)$$

then

$$\lim_{n \to \infty} \Pr\left(G_{n,m} \text{ is } k\text{-connected}\right) = \begin{cases} 0 & c_n \to -\infty \\ e^{-\frac{e^{-c}}{(k-1)!}} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$$

Let $p = \dfrac{\log n + (k-1)\log\log n + c}{n}$

We will prove

(i) $E(\#\text{ vertices of degree} \leq k-2) = o(1)$

(ii) $E(\#\text{ vertices of degree } k-1) \approx \dfrac{e^{-c}}{(k-1)!}$

It then a simple matter to verify that

$$\Pr(\delta(G_{n,p}) \geq k) \approx e^{-\frac{e^{-c}}{(k-1)!}}$$

$E(\#\text{ vertices of degree } t \leq k-1)$

$$= n\binom{n-1}{t} p^t (1-p)^{n-1-t}$$

$$\simeq n \cdot \frac{n^t}{t!} \cdot \frac{(\log n)^t}{n^t} \cdot \frac{e^{-c}}{n(\log n)^{k-1}}$$

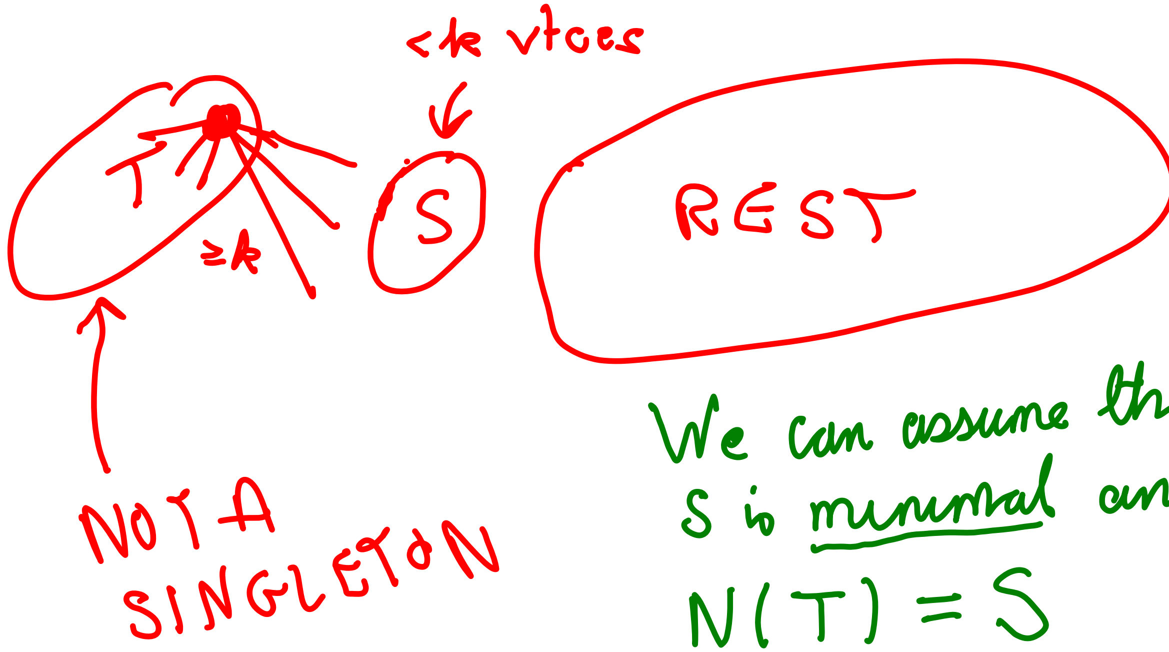
and (i) and (ii) follow immediately.

We now show that,

$\Pr(\exists\, S,\ |S| < k$ and $T,\ k - |S| + 1 \le |T| \le \frac{1}{2}(n-s)$

T is a component of $G_{n,p} \setminus S) = o(1)$.

This implies that if $\delta(G_{n,p}) \ge k$ then it is k-connected whp



We can assume that S is <u>minimal</u> and then

$N(T) = S$

First moment:

$$E(\#S,T) \leq$$

Case 1: $s+2 \leq t \leq \log n$

$$\sum_{s=0}^{k-1} \sum_{t=s+2}^{\log n} \binom{n}{s}\binom{n}{t} t^{t-2} p^{t-1} (1-p)^{t(n-s-t)}$$

$$\leq \sum_{s=0}^{k-1} \sum_{t=s+2}^{\log n} n^s \cdot \left(\frac{ne}{t}\right)^t \cdot t^{t-2} \cdot \left(\frac{e^{o(1)} \log n}{n}\right)^{t-1} \frac{e^{O((\log n)^2/n)}}{n^t (\log n)^{(k-1)t}}$$

$$\leq \sum_{s=0}^{k-1} \sum_{t=s+2}^{\log n} \left(e^{1+o(1)} \log n\right)^t n^{s-t}$$

$$= O(1).$$

Case 2 : $t > \log n$

$$\sum_{s=0}^{k-1} \sum_{t=\log n}^{\frac{1}{2}(n-s)} \binom{n}{s}\binom{n}{t} t^{t-2} p^{t-1} (1-p)^{t(n-s-t)}$$

$$\leq \sum_{s=0}^{k-1} \sum_{t=\log n}^{\frac{1}{2}(n-s)} n^{s} \left(\frac{ne}{t}\right)^{t} t^{t-2} \left(\frac{e^{o(1)} \log n}{n}\right)^{t-1} n^{-t/2}$$

$$\leq \sum_{s=0}^{k-1} \sum_{t=\log n}^{\frac{1}{2}(n-s)} n^{1+s-\frac{1}{2}t} \left(e^{1+o(1)} \log n\right)^{t}$$

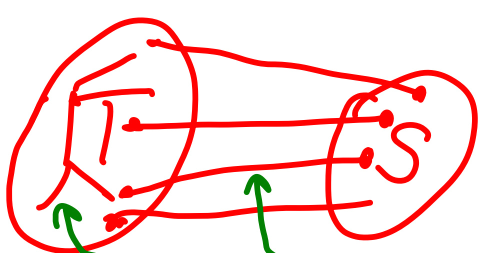
$= o(1)$.

Case 3 : $k-s+1 \le t \le s+1$

$$\sum_{s=0}^{k-1} \sum_{\substack{t \ge 2 \\ t \ge k-s+1}}^{s+1} \binom{n}{s}\binom{n}{t} t^{t-2} \binom{st}{s} p^{t-1+s} (1-p)^{t(n-s-t)}$$

$$\le \sum_{s=0}^{k-1} \sum_{t} n^{s+t} \, 2^{st} \left(\frac{e^{o(1)} \log n}{n}\right)^{t-1+s} \frac{1+o(1)}{n^{t}}$$

$$= o(1).$$



$\#\text{edges} \ge t-1+s$

$\square$

Note Title 2/7/2006

Perfect Matchings in Random Graphs

Let $K_{n,n,p}$ be the random bipartite graph with vertex bipartition $A = B = [n]$ in which each of the n^2 possible edges appears independently with probability p.

Theorem

Let $p = \frac{\log n + c_n}{n}$.

$$\lim_{n\to\infty} \Pr(K_{n,n,p} \text{ has a perfect matching}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-2e^{-c}} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$$

$$= \lim_{n\to\infty} \Pr(\delta(K_{n,n,p}) \geq 1).$$

Let X_0 = # isolated vertices.

$$E(X_0) = 2n(1-p)^n$$
$$\approx 2e^{-c}.$$

By previously used techniques we

$$\Pr(X_0 = 0) \approx e^{-2e^{-c}}.$$

We will now use Hall's condition.

$G = K_{n,n,p}$ contains a perfect matching iff

$$\forall S \subseteq A, \quad |N(S)| \geq |S|. \qquad (*)$$

It is convenient to replace $(*)$ by

$$\begin{aligned} &\forall S \subseteq A, \ |S| \leq \tfrac{1}{2}n, \ |N(S)| \geq |S| \\ &\forall T \subseteq B, \ |T| \leq \tfrac{1}{2}n, \ |N(T)| \geq |T|. \end{aligned} \qquad (**)$$

$$\Pr(\exists v : v \text{ isolated})$$

$$\leq \Pr(\nexists \text{ a perfect matching}) \leq$$

$$\Pr(\exists v : v \text{ isolated}) \quad +$$

$$\Pr(\exists k, S \subseteq A, T \subseteq B, |S| = k \geq 2, |T| = k-1$$

$$N(S) \subseteq T \text{ and } \underbrace{e(S:T)}_{\#S:T \text{ edges}} \geq 2k-2 \}$$

? Why $e(S:T) \geq 2k-2$?

Take a pair S, T with $|S|+|T|$ as small as possible.

(i) If $|S| > |T|+1$, remove $|S|-|T|-1$ vertices from S

(ii) Suppose $\exists\, w \in T$ such that w has < 2 nbs in S. Remove w and its (unique) nbr in S.

Repeat until (i) & (ii) do not hold. $|S|$ will stay at least 2 if $S \geq 1$.

$$E(\#\text{ sets } S, T) \le$$

$$2\sum_{k=2}^{n/2} \binom{n}{k}\binom{n}{k-1}\binom{k(k-1)}{2k-2} p^{k} (1-p)^{k(n-k)}$$

$$\le 2\sum_{k=2}^{n/2} \left(\frac{ne}{k}\right)^{k} \left(\frac{ne}{k-1}\right)^{k-1} \left(\frac{ke(\log n + c)}{2n}\right)^{2k-2} e^{-npk\left(1-\frac{k}{n}\right)}$$

$$\le 8\sum_{k=2}^{n/2} \underbrace{n\left(\frac{e^{O(1)}(\log n)^{2} n^{k/n}}{n}\right)^{k}}_{u_k}$$

Case 1: $2 \leq k \leq n^{3/4}$

$$u_k = n\left(\frac{e^{O(1)}(\log n)^2 n^{k/n}}{n}\right)^k$$

$$= e^{O(k)} n^{1+o(1)-k}.$$

So $\displaystyle\sum_{k=2}^{n^{3/4}} u_k = O(1/n).$

Case 2: $n^{3/4} < k \leq n/2$.

$$u_k = n\left(\frac{e^{O(1)}(\log n)^2 n^{k/n}}{n}\right)$$

$$\leq n^{1-k/3}$$

So
$$\sum_{k=n^{3/4}}^{n/2} u_k = O(n^{-n^{3/4}/4}).$$

So,

$$\Pr(\nexists \text{ a perfect matching}) = \Pr(\exists \text{ isolated vertex}) + o(1).$$

We now consider $G_{n,p}$.

We could try to replace Hall's Theorem by Tutte's theorem, but it is simpler to use Hall's theorem.

Theorem

Let $p = \frac{\log n + c_n}{n}$.

$$\lim_{\substack{n\to\infty \\ n \text{ even}}} \Pr(G_{n,p} \text{ has a perfect matching}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-e^{-c}} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$$

$$= \lim_{n\to\infty} \Pr(\delta(G_{n,p}) \geq 1).$$

First of all if

X_0 = # of isolated vertices.

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

$$\approx e^{-c}$$

By previously used techniques we

$$\Pr(X_0 = 0) \approx e^{-e^{-c}}.$$

Suppose $n = 2m$ and $A = \{1, 2, \dots, m\}$

$$B = \{m+1, \dots n\}$$

We will choose $A^* \subseteq A$, $B \subseteq B$ $\quad |A^*| = |B^*| = s$

where s is small, such that whp $G_{n,p}$

contains a perfect matching between

$$\hat{A} = (A \setminus A^*) \cup B^*$$

and $\hat{B} = (B \setminus B^*) \cup A^*$.

Let

$$V_0 = \{ v : |N(v)| \leq \frac{\log n}{100} \}$$

$$A_0 = \{ v \in A \cap V_0 : |N(v) \cap A| > |N(v) \cap B| \}$$

$$B_0 = \{ w \in B \cap V_0 : |N(w) \cap B| > |N(w) \cap A| \}$$

$$A_1 = \{ v \in A \setminus A_0 : |N(v) \cap B| < \frac{\log n}{200} \}$$

$$B_1 = \{ w \in B \setminus B_0 : |N(w) \cap A| < \frac{\log n}{200} \}$$

Suppose

$$|A_0 \cup A_1| = |B_0 \cup B_1| + r$$

where $r \geq 0$.

Choose $R \subseteq B \setminus (B_0 \cup B_1)$

with $|R| = r$.

$$A^* = A_0 \cup A_1$$

$$B^* = B_0 \cup B_1 \cup R$$

We show that, conditional on $\delta \geq 1$, there is whp a perfect matching between $\hat{A}$ and $\hat{B}$.

Lemma

Whp $|V_0| \leq n^{1/10}$.

Proof

$$E(|V_0|) \leq n \sum_{k=0}^{\frac{1}{100}\log n} \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$\leq 2n \binom{n-1}{\frac{1}{100}\log n} p^{\frac{1}{100}\log n} \frac{e^{-c}}{n}$$

$$\leq 2\left(100 e^{1+o(1)}\right)^{\frac{\log n}{100}} e^{-c}.$$

Now use Markov inequality. $\square$

Similarly, Whp

$$|A_1|, |B_1| \leq n^{2/3}.$$

Lemma

Whp $|A_1 \cup B_1| \leq n^{6/10}$.

Proof

$$\mathbb{E}(|V_0|) \leq n \sum_{k=0}^{\frac{1}{200}\log n} \binom{\frac{1}{2}n-1}{k} p^k (1-p)^{\frac{1}{2}n-1-k}$$

$$\leq 2n \binom{\frac{1}{2}n-1}{\frac{1}{200}\log n} p^{\frac{1}{200}\log n} \frac{e^{-c}}{n^{1/2}}$$

$$\leq 2n^{1/2} \left(200e^{1+o(1)}\right)^{\frac{\log n}{200}} e^{-c}.$$

Now use Markov inequality. $\square$

Lemma

Why, $v \in V_0$, $w \in A_1 \cup B_1 \Rightarrow N(v) \cap N(w) = \emptyset$

Proof

$$\Pr(\exists\, v, w : N(v) \cap N(w) \neq \emptyset)$$

$$\leq 3 \binom{n}{3} p^2 \left(\sum_{k=0}^{\frac{1}{100}\log n} \binom{n-3}{k} p^k (1-p)^{n-3-k} \right) \leftarrow \frac{1}{n^{1-\epsilon}}$$

$$\times \left(\sum_{k=0}^{\frac{1}{200}\log n} \binom{\frac{1}{2}n-3}{k} p^k (1-p)^{\frac{1}{2}n-3-k} \right) \leftarrow \frac{1}{n^{\frac{1}{2}-\epsilon}}$$

$$\leq 3 \binom{n}{3} \left(\frac{\log n + c}{n} \right)^2 n^{-4/3}$$

$$= o(1).$$

$\square$

Lemma

Whp $\nexists\, v : |N(v) \cap (A_1 \cup B_1 \cup V_0)| \geq 3$

Proof

$$\Pr(\exists v) \leq n \binom{n}{3} p^3$$

$$n \binom{n}{3} p^3 \left(\sum_{k=0}^{\frac{1}{100}\log n} \binom{\frac{1}{2}n-5}{k} p^k (1-p)^{\frac{1}{2}n-5-k} \right)^3$$

$$\leq n (\log n)^3 \cdot n^{-6/5}$$

$$= o(1).$$

$\square$

Lemma

Whp $S \subseteq A \setminus (A_0 \cup A_1)$ implies

$$|N_B(S)| \geq \frac{\log n}{500}|S| \text{ for } |S| \leq \frac{n}{(\log n)^3}$$

Proof

We first show that whp

$$|S| \leq \frac{n}{10(\log n)^2} \quad \text{implies} \quad e(S) < 2|S|.$$

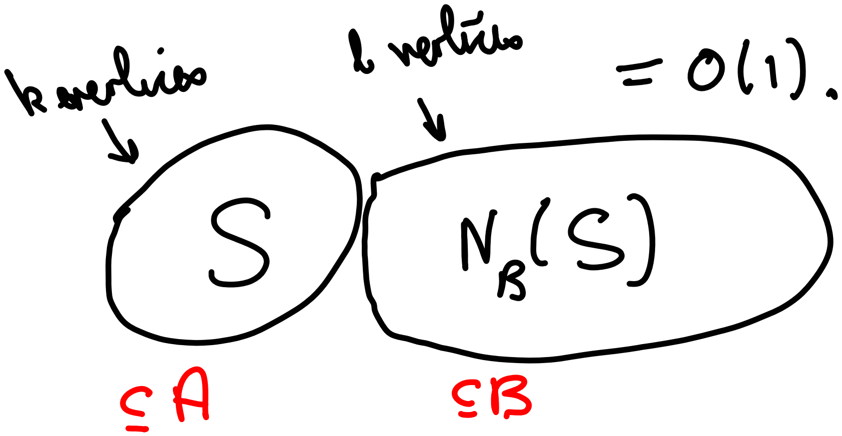
#edges inside S

$$\Pr(\exists S: e(S) \geq 2|S|) \leqslant \frac{n}{10(\log n)^2} \leftarrow n_0$$

$$\sum_{k=4} \binom{n}{k} \binom{\binom{k}{2}}{2k} p^{2k} \leq \sum_{k=4}^{n_0} \left(\frac{ne}{k}\right)^k \left(\frac{ke}{2} \frac{(\log n + c)}{n}\right)^{2k}$$

$$\doteq \sum_{k=4}^{n_0} \left(\frac{k}{n} \cdot \frac{e^3}{4} \cdot (\log n + c)^2 \right)^k$$

$= O(1)$.



$$k + l \leqslant \frac{n}{10(\log n)^2}$$

$$\Rightarrow \frac{k \log n}{200} < 2(k+l)$$

$$\Rightarrow l > \frac{k \log n}{500}$$

So $k \leq \dfrac{n}{(\log n)^3}$

Lemma

Whp $\nexists\, v \in A,\ u_1, u_2, u_3 \in B$ such that $u_i \in N(A_1 \cup V_0) \cap N(v)$ for $i = 1, 2, 3$.

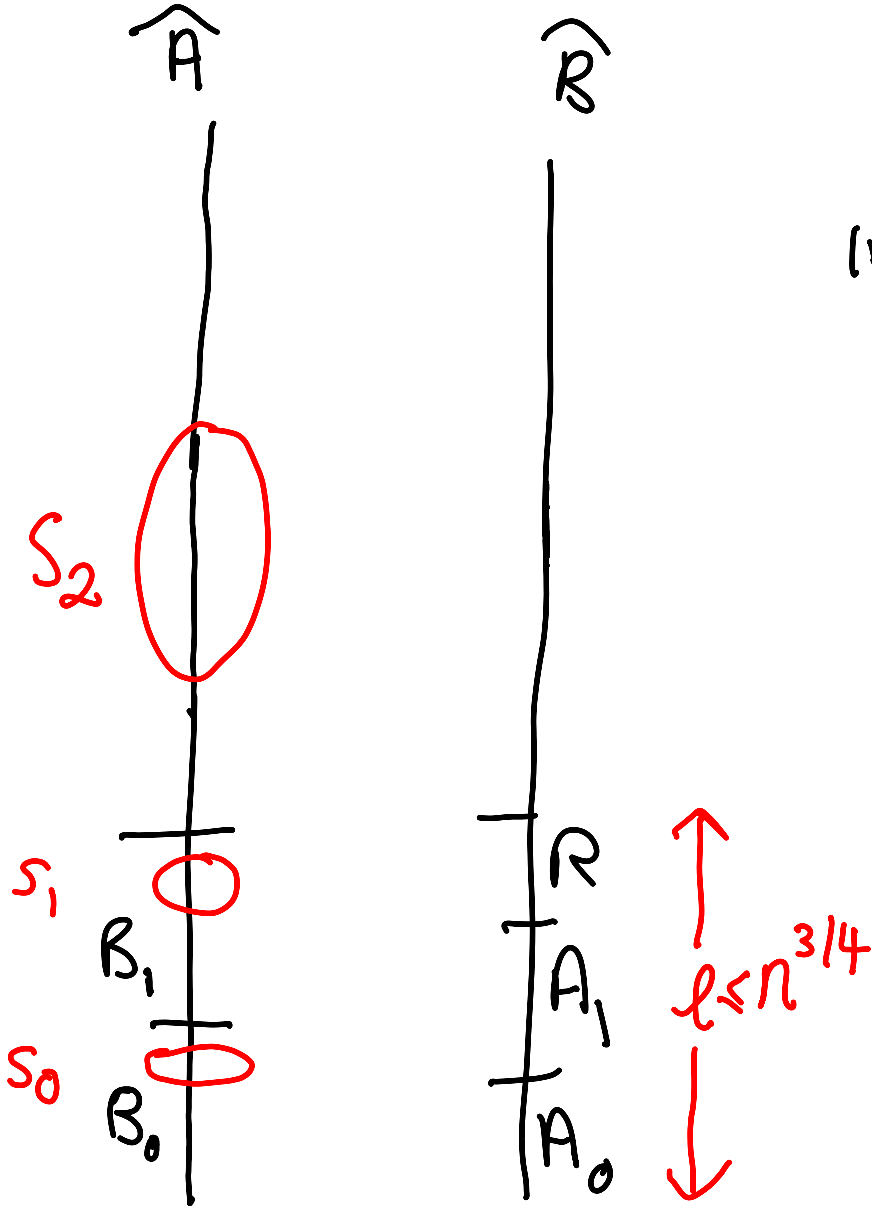
Proof

$$\Pr(\neg) \leq n \binom{n}{3} p^3 \left(np \sum_{k=0}^{\log n/100} \binom{\frac{1}{2}n}{k} p^k (1-p)^{n-k} \right)^3$$

$$\leq n (\log n)^2\, n^{-6/5}$$

$$= o(1).$$

$\square$



Try Hall Condition

(1) $|S| \leq \frac{n}{(\log n)^3}$

$$S = \underset{S \cap B_0}{S_0} \cup \underset{S \cap B_1}{S_1} \cup S_2$$

$$|N_B(S)| \geq |S_0| + \frac{\log n}{600}|S_1| + \frac{\log n}{500}|S_2| - 2|S_2|$$

$$\geq |S|.$$

(ii) $\frac{n}{2(\log n)^3} < |S| \leq n/4$

<u>Lemma</u>

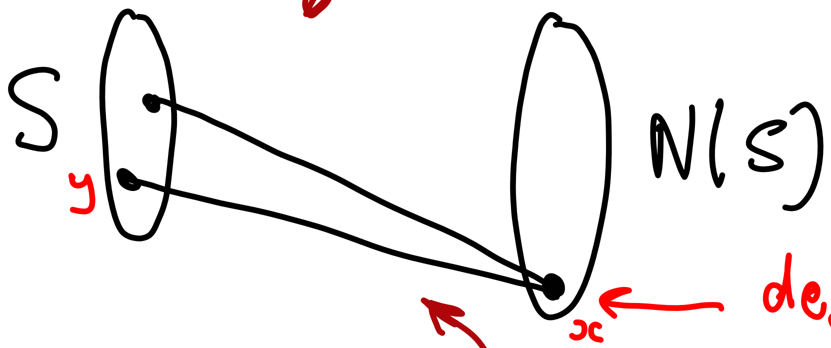
Whp $\nexists$ $S \subseteq A$, , such that $|N_B(S)| \leq |S| + 2\ell$

[This completes proof that Hall's condition holds whp.

$N_{\hat{B}}(S) \geq N_B(S \setminus (A_0 \cup A_1)) - \ell \geq |S| - \ell - \ell + 2\ell$.]

<u>Proof</u>

As before we actually consider $S \subseteq A$, $|S| \leq n/4$ and double our estimate.



#edges $\geq 2k+2l$

degree ≥ 2, else we can remove y from S. Can't go below n/ by previous analysis

$|N(S)| = k-1$

Probability $\leq$

$$2\sum_{k=\frac{n}{2(\log n)^3}}^{\frac{1}{4}n}\binom{\frac{1}{2}n}{k}\binom{\frac{1}{2}n}{k+2l}\binom{k(k+2l)}{2(k+2l)}p^{2k+4l}(1-p)^{k(\frac{1}{2}n-k-2l)}$$

$$\leq 2\sum_{k}\frac{n^k e^k}{2^k k^k}\cdot\frac{n^{k+2l}e^{k+2l}}{2^{k+2l}(k+2l)^{k+2l}}\left(\frac{ke(\log n+c)}{2n}\right)^{2k+4l} n^{-k/5}$$

$$\leq 2\sum_{k} \frac{n^k e^k}{2^k k^k} \cdot \frac{n^{k+2l} e^{k+2l}}{2^{k+2l} (k+2l)^{k+2l}} \left(\frac{ke(\log n + c)}{2n}\right)^{2k+4l} n^{-k/5}$$

$$\leq \sum_{k} \left(\frac{e^{O(1)} (\log n)^2}{n^{1/5}}\right)^k$$

$$= o(1).$$

$\square$

Check: $k^{2l} = \left(k^{2l/k}\right)^k = \left(e^{O(1)}\right)^k.$

Hamilton Cycles in Random Graphs

Theorem

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

$$\lim_{n \to \infty} \Pr(G_{n,m} \text{ is Hamiltonian}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-e^{-c}} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$$

$$= \lim_{n \to \infty} \Pr(\delta(G_{n,m}) \geq 2).$$

The proof of this is complicated and so we start by proving a weaker theorem.

Let $p = \frac{25 \log n}{n}$. Then

$G_{n,p}$ is Hamiltonian whp.

Write

$$G_{n,p} = G_{n,p_1} \cup G_{n,p_2}$$

where

$$p_1 = \frac{20 \log n}{n}$$

and $1 - p = (1-p_1)(1-p_2)$, $p_2 \approx \frac{5 \log n}{n}$

We first show that whp $G_1 = G_{n,p_1}$ has a Hamilton path.

Let $\lambda(G)$ denote the length of a longest path in G.

Let $\mathcal{E}_v$ be the event

$$\lambda(G_1 \setminus v) = \lambda(G_1)$$

Then

G_1 does not have a Hamilton path

$\Rightarrow \exists v : \mathcal{E}_v$ occurs.

G, not Hamiltonian

longest path P

v

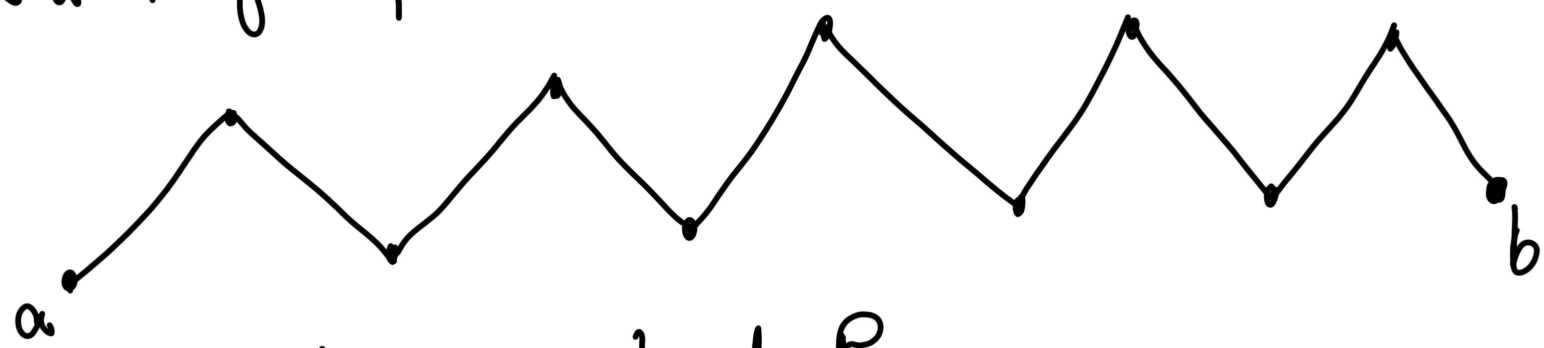
$\mathcal{E}_v$ occurs.

We show now that

$$\Pr\left(\bigcup_v \mathcal{E}_v\right) \le n \Pr(\mathcal{E}_n) = o(1).$$

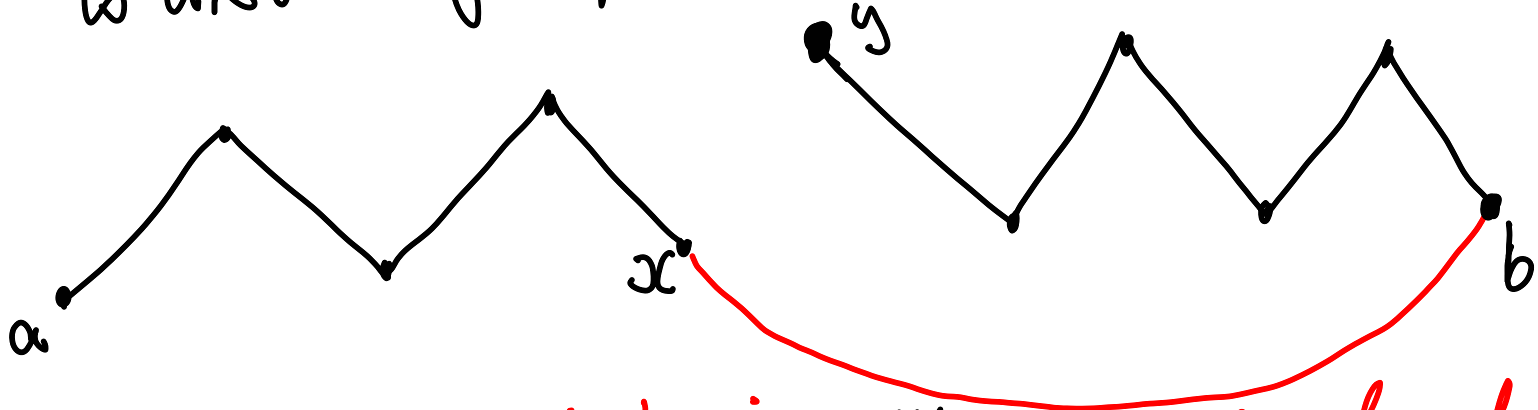
Posá Lemma

P is a longest path



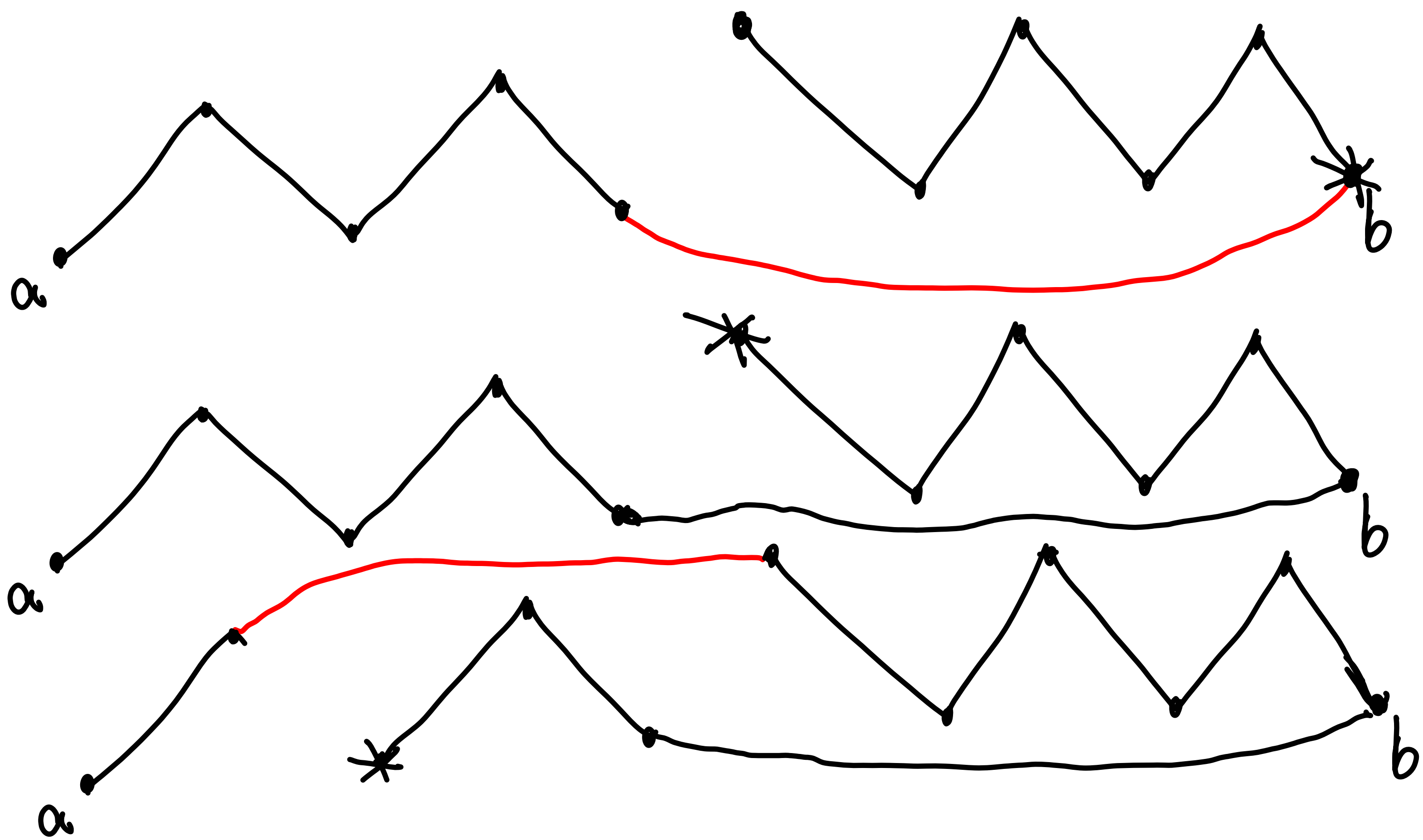
No edge from b go outside P.

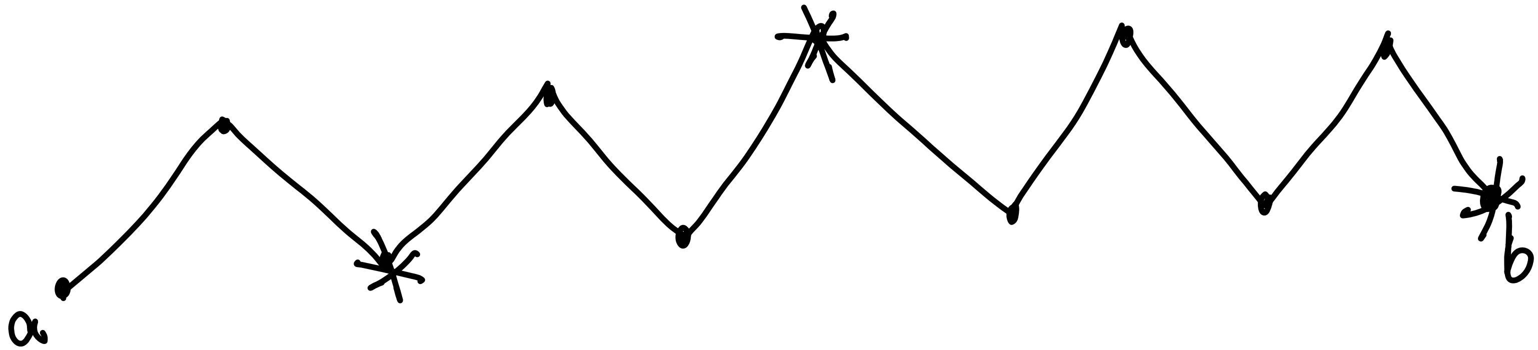
P′ is also longest path:



P′ is obtained by a rotation with a as fixed endpoint.

Now let END denote the set of v such that $\exists$ longest path P_v from a to v such that P_v is obtained from P by a sequence of rotations with a fixed.





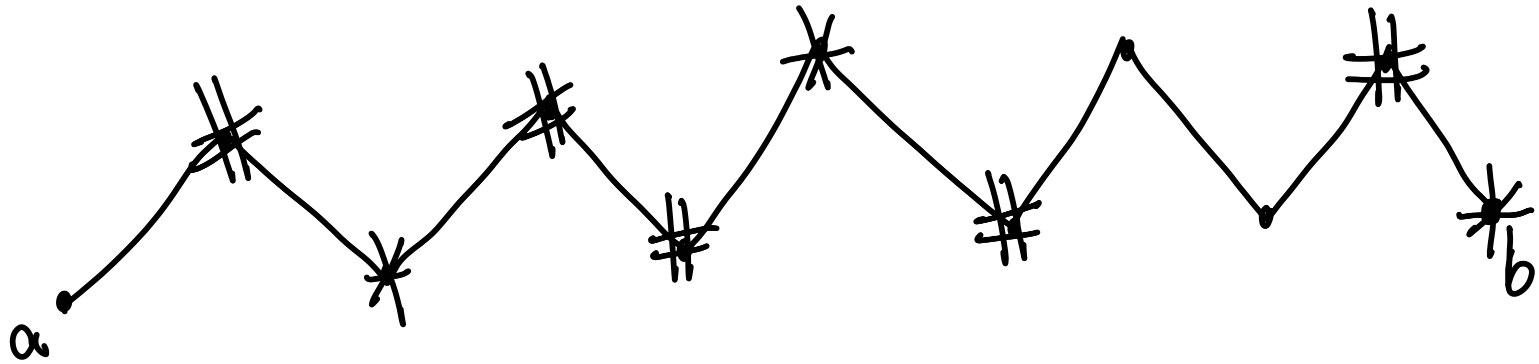
$END = \{*\}$

Lemma

If $v \in P \setminus END$ and v is adjacent to $w \in END$ then there exist $x \in END$ such that the edge $(x, v) \in P$ or $(v, x) \in P$.

Corollary

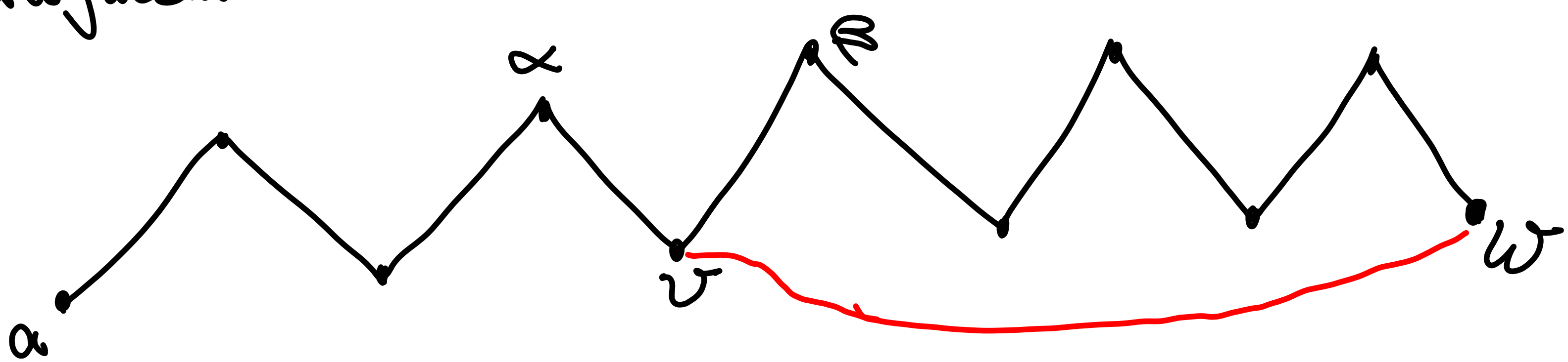
$|N(END)| < 2|END|$.

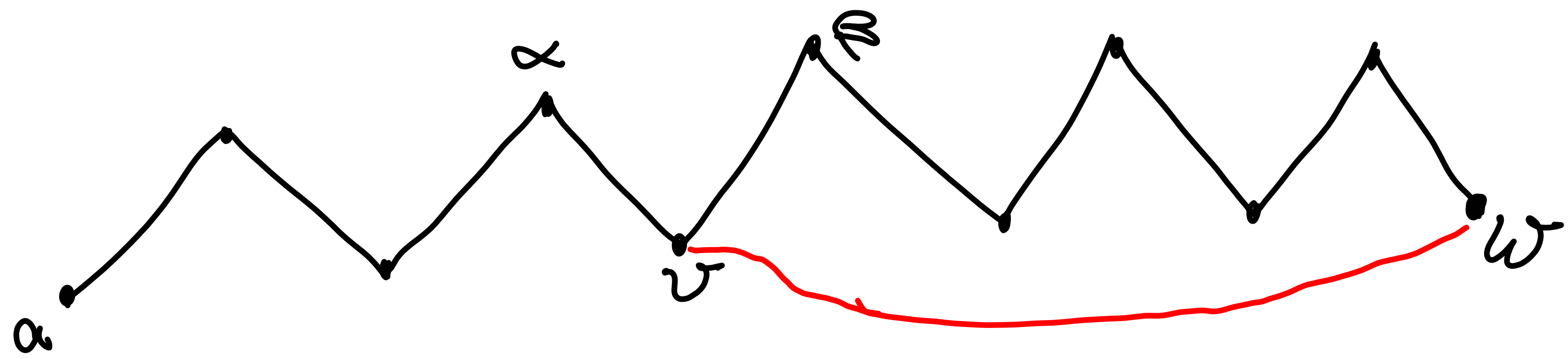


$N(\text{END}) = \{\#\}$

Proof of Lemma

Suppose that x, y are the neighbours of v on P and that $v, x, y \notin \text{END}$ and that v is adjacent to $w \in \text{END}$. Consider P_w





Now $\{\alpha, \beta\} = \{x, y\}$ because if a rotation deleted (x, v) say then x or v becomes an endpoint.

But then $\beta \in END$.

$\square$

Lemma

Whp $S \subseteq [n-1], |S| \leq \frac{1}{4}n \Longrightarrow$

$|N(S)| \geq 2|S|$

in $G_1 \setminus \{n\}$

Proof

$P_1(\exists S: |S| \leq \frac{1}{4}n$ and $|N(S)| < 2|S|) \leq$

$$\sum_{k=1}^{\frac{1}{4}n-1} \binom{n-1}{k}\binom{n-1}{2k} \underbrace{(1-p_1)}_{\leq e^{-kp_1/4} \leq n^{-5k}}^{k(n-1-3k)} \leq$$

$$\sum_{k=1}^{n/4} \left[\frac{ne}{k} \cdot \frac{n^2e^2}{k^2} \cdot n^{-5}\right]^k = O(n^{-2}).$$

It follows that if P is a longest path in $G_1\setminus\{n\}$ and END is defined w.r.t. P then

$$\Pr(|END| \le n/4) = O(n^{-2}).$$

Now the edges incident with n are unconditioned by $G_1\setminus\{n\}$ and (see p3)

$$\mathcal{E}_n \Rightarrow \nexists \text{ edge from } n \text{ to END.}$$

So

$$\Pr(\mathcal{E}_n) \le O(n^{-2}) + (1-p_1)^{n/4} = O(n^{-2}).$$

So $\Pr(G_1 \text{ does not have a Hamilton path}) = O(n^{-1})$.

Now use the G_{n,p_2} edges.

Let P be a Hamilton path in G_1 and let END be defined w.r.t. P.

By arguing as for $G_1 \setminus \{n\}$ we see that $|END| \geq \frac{n}{4}$ whp.

Let a be the fixed endpoint of P.

Then

$G_{n,p}$ not Hamiltonian $\Longrightarrow$ $\nexists$ a G_{n,p_2} edge from a to END.

Thus

$$Pr(G_{n,p} \text{ is not Hamiltonian}) = o(1) + (1-p_2)^{n/4} = o(1).$$

Let us now go to $G = G_{n,m}$, $m = \frac{1}{2}n(\log n + \log\log n + c)$

and $G_{n,p}$, $p = \frac{m}{N}$.

Let a vertex of G be large if its degree is at least $\lambda = \frac{\log n}{100}$, and small otherwise.

<u>Lemma</u>

Whp $v, w \in SMALL \Rightarrow dist(v,w) \geq 5$

<u>Proof</u>

$$\Pr_p(\neg) \leq \binom{n}{2}\left(\sum_{l=0}^{3}\binom{n}{l}p^{l+1}\right)\left(\sum_{k=0}^{\lambda}\binom{n}{k}p^k(1-p)^{n-k}\right)^2$$

$$\lesssim \frac{1}{2}n(\log n)^4\left(\sum_{k=0}^{\lambda}\frac{(\log n)^k}{k!}\cdot\frac{e^{-c}}{n\log n}\right)^2$$

$$\lesssim \frac{1}{2} n (\log n)^4 \Bigg(\underbrace{\sum_{k=0}^{\lambda} \frac{(\log n)^k}{k!} \cdot \frac{e^{-c}}{n \log n}}_{u_k} \Bigg)^2$$

$$\frac{u_{k+1}}{u_k} > 100$$

$$\leq n (\log n)^4 \left(\frac{(\log n)^\lambda}{\lambda!} \frac{e^{-c}}{n \log n} \right)^2$$

$$\lambda! \geq \left(\frac{\lambda}{e} \right)^\lambda$$

$$= O\left(\frac{(\log n)^3}{n} \cdot (100e)^{\frac{2\log n}{100}} \right)$$

$$= O(n^{-3/4}).$$

So $$Pr_m(\neg) = O(m^{\frac{1}{2}} n^{-3/4}) = o(1).$$

$\square$

Lemma

Whp $|SMALL| \le n^{1/4}$.

Proof

$$\Pr_p\left(|SMALL| > n^{1/4}\right)$$

$$\le n \sum_{k=0}^{\log n/100} \underbrace{\binom{n-1}{k} p^k (1-p)^{n-1-k}}_{u_k}$$

$$u_{k+1} / u_k = \frac{n-1-k}{k+1} \cdot p \cdot \frac{1}{1-p} \ge 50.$$

$$\le 2n \left(\frac{nep \log n}{100\, n}\right)^{\frac{\log n}{100}} \cdot \frac{1}{n}$$

$$\le n^{1/5}.$$

Now apply Markov and monotonicity to go to $G_{n,m}$. $\square$

Lemma

Whp $\nexists$ a cycle C_4 containing a small vertex.

Proof

$$\Pr_p(\neg) \leq \frac{1}{2} n^4 p^4 \sum_{k=0}^{\log n/100} \binom{n}{k} p^k (1-p)^{n-1-k}$$

$$\leq (\log n)^4 \, n^{-3/4}.$$

So $\Pr_m(\neg) \leq \tilde{O}(m^{1/2} n^{-3/4}) = o(1).$

$\square$

Lemma

Whp, $\forall\, |S| \le \frac{n}{(\log n)^3}$, $e(S) \le 2|S|$

Proof

$\Pr_p(\exists S : |S| \le \frac{n}{(\log n)^3}$ and $e(S) > 2|S|) \le$

$$\sum_{s=4}^{n/(\log n)^3} \binom{n}{s} \binom{\binom{s}{2}}{2s} p^{2s}$$

$$\le \sum_s \left(\frac{ne}{s} \cdot \left(\frac{se\log n}{2n} \right)^2 \right)^s$$

$$= \sum_s \left(\frac{s}{n} \cdot \frac{e^3 (\log n)^2}{2} \right)^s$$

$= O(n^{-3})$. So $\Pr_m(\neg) = O(m^{1/2} n^{-3}) = o(1)$.

$\square$

Lemma

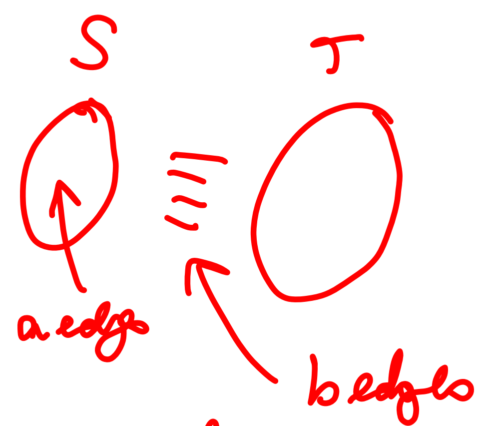
$S \subseteq LARGE$, $\quad |S| \le \frac{n}{\log n} \Rightarrow |N(S)| \ge \frac{\log n}{1000}|S|$.

{large vtces}

Proof

(a) $\quad 1 \le |S| \le \frac{n}{(\log n)^3}$ $\qquad s = |S|$

$T = N(S)$ $\qquad t = |T|$



$$\Pr_p(\exists S) \le \sum_{s=\sqrt{\lambda}}^{n/(\log n)^3} \sum_{t=0}^{\frac{\log n}{1000}s} \binom{n}{s}\binom{n}{t}\binom{\binom{s+t}{2}}{\frac{\log n}{200}s} p^{\frac{\log n}{200}s}$$

$$\times (1-p)^{s(n-s-t)}$$

$2a+b \ge \frac{\log n}{100} s$

$a \le 2s$

$\Downarrow$

$a+b \ge \frac{\log n}{200} s$

$$\leq \sum_{s,t} \left(\frac{ne}{s}\right)^s \left(\frac{ne}{t}\right)^t \left(\frac{(s+t)^2 e\,(200)}{2s\log n} \cdot \frac{(1+o(1))\log n}{n}\right)^{s\log n/200}$$

$$\times\, n^{-s(1-1/(\log n)^2)}$$

The summand increases with $t \ll n$ and so we can put $t = \frac{\log n}{1000}s$ and then very crudely

$$\leq n \sum_{s \geq \sqrt{\lambda}} \left(\frac{ne}{s} \cdot \left(\frac{1000\,ne}{s\log n}\right)^{\frac{\log n}{1000}} \left(\frac{e(\log n)^2 s}{1000\,n}\right)^{\log n/200} \left(\frac{1+o(1)}{n}\right)\right)^s$$

$$< n \sum_{s \geq \sqrt{\lambda}} \left(\frac{e^6 (\log n)^9 s^4}{10^{12}\, n^4} \right)^{\frac{\log n}{1000}s}$$

$= O(n^{-\Omega(\sqrt{\lambda}\log n)})$ and so $\Pr_m(\neg) = o(1)$.

(b) $\frac{n}{(\log n)^3} \leq |S| \leq \frac{n}{\log n}$

$$\Pr_p[\neg] \leq \sum_{s,t} \binom{n}{s}\binom{n}{t}\binom{st}{t} p^t (1-p)^{s(n-s-t)}$$

S:T edges

T = N(S)

$$\leq \sum_{s,t} \underbrace{\left(\frac{ne}{s}\right)^s \left(\frac{ne}{t}\right)^t (sep)^t \, n^{-s\left(1-\frac{s+t}{n}\right)}}_{u_{s,t}}$$

$$\frac{u_{s,t+1}}{u_{s,t}} = \frac{ne}{t+1}\cdot\left(\frac{t}{t+1}\right)^t\cdot(sep)\cdot n^{s/n}$$

$\geq 10.$

So

$$\Pr(\neg) \le 2\sum_{s}\left(\frac{ne}{s}\right)^{s}\left(10^{3}e^{2+o(1)}\right)^{\frac{s\log n}{1000}} n^{-s\left(1-\frac{s(1+\log n/1000)}{n}\right)}$$

$$= 2\sum_{s \ge \frac{n}{(\log n)^3}}\left(\frac{e}{s}\cdot\left(10^{3}e^{2+o(1)}\right)^{\frac{\log n}{1000}}\cdot n^{\frac{1}{1000}+o(1)}\right)^{s}$$

$$= O\left(n^{-\Omega(n^{1/2})}\right)$$

and so

$$\Pr_m(\neg) = o(1).$$

$\square$

Suppose now that $X \subseteq E(G)$ and

(i) $|X| = \log n$

(ii) X is a matching

(iii) X is not incident with a small vertex

(iv) X avoids the edges of some longest path of G.

We say that X is deletable.

Let $G_X = G \setminus X$

Lemma

Suppose that $\delta(G) \geq 2$ and

(i) $v, w \in SMALL \Rightarrow dist(v,w) \geq 5$ and $v \notin$ any C_4.

(ii) $S \subseteq LARGE$, $|S| \leq \frac{n}{\log n} \Rightarrow |N(S)| \geq \frac{\log n}{1000}|S|$.

(iii) X is deletable.

Then $S \subseteq [n]$, $|S| \leq 10^{-4} n \Rightarrow |N_X(S)| \geq 2|S|$

nbrs in G_X.

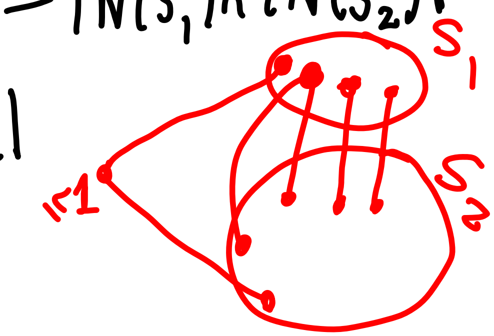
Proof

Let $S_1 = S \cap SMALL$ and $S_2 = S \setminus S_1$

$$|N(S)| \geq |N(S_1)| + |N(S_2)| - |N(S_1) \cap S_2| - |N(S_2) \cap S_1| - |N(S_1) \cap N(S_2)|$$

$$\geq |N(S_1)| + |N(S_2)| - 2|N(S_1) \cap S_2| - |S_2|$$

$$\geq |N(S_1)| + |N(S_2)| - 3|S_2|.$$



Now $|N(S_1)| \geq 2|S_1|$

and $|N(S_2)| \geq 9|S_2|$.

(i) $|S_2| \leq \frac{n}{\log n} \Rightarrow |N(S_2)| \geq \frac{\log n}{1000}|S_2|$

(ii) $|S_2| > \frac{n}{\log n}$. Take $S_2' \subseteq S_2$, $|S_2| = \frac{n}{\log n}$

$$|N(S_2)| \geq |N(S_2')| - |S_2|$$
$$\geq \frac{n}{1000} - |S_2|$$
$$\geq 9|S_2|.$$

So $|N(S)| \geq 2|S_1| + 6|S_2|$

and

$$|N_X(S)| \geq |N(S)| - |S_2|$$

X is a matching and it avoids SMAZZ

$$\geq 2|S_1| + 5|S_2|$$

$$\geq 2|S|.$$

Summary

(I) $\lim_{n\to\infty} \Pr(\delta(G_{n,m}) \geq 2) = e^{-e^{-c}}$.

(II) $G_{n,m}$ is connected whp

(III) $|SMALL| \leq n^{1/4}$, $v,w \in SMALL \Rightarrow dist(v,w) \geq 5$,
$\nexists\, C_4 : C_4 \cap SMALL \neq \emptyset$, whp.

(IV) If $\delta(G) \geq 2$ and X is deletable then whp

$$|N_X(S)| < 2|S| \Rightarrow |S| \geq 10^{-4} n.$$

$\mathcal{G} = \{$all graphs on $[n]$ with m edges$\}$

$\mathcal{G}_0 = \{G \in \mathcal{G} : \delta(G) \geq 2$ and (II) — (IV) hold$\}$

$\mathcal{G}_1 = \{G \in \mathcal{G}_0 : G$ is not Hamiltonian$\}$

Couloring Argument

Suppose $G \in \mathcal{G}_1$ and X be deletable

Let P be a longest path in G_X.

Then

$$|END_X| \geq 10^{-4} n$$ (add subscript X to END)

Now for each $b \in END$, start with P_b and do all possible rotations, starting from P_b, but with b as a fixed endpoint. Let $END_X(b)$ be the set of endpoints produced.

We do a bit of re-naming

$$END_X \leftarrow END_X \cup \{a\}$$

$$END_X(a) \leftarrow END_X / \{a\}$$

Now we can say that for $b \in END$, we have

$$|END_X(b)| \geq \frac{n}{1000}$$

Now for $G \in \mathcal{G}$ and $X \subseteq E(G)$, $|X| = \omega = \log n$ choose some fixed longest path P_X of G_X.

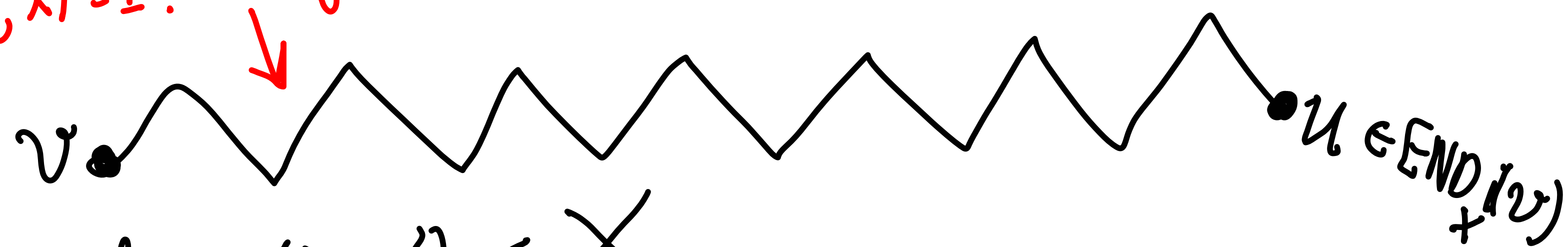
Furthermore choose so that if $G, G' \in \mathcal{G}$ and $G_X = G'_Y$ then $P_X = P'_Y$ i.e. path depends on G_X and not G, X.

$$\theta(G, X) = \begin{cases} 1 : & \text{(a) } G \in \mathcal{G}_1 \\ & \text{(b) } X \cap E(P_X) = \emptyset \\ & \text{(c) } X \text{ is deletable} \\ 0 : & \text{otherwise} \end{cases}$$

Note that $a(G, X) = 1$ implies

If $u \in END_X(v)$ then $(u, v) \notin X$

$a(G, X) = 1$. Longest in G_X and G



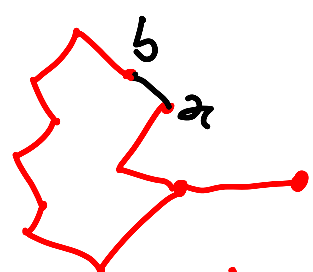
$\nexists$ edge $(u, v) \in X$

Either

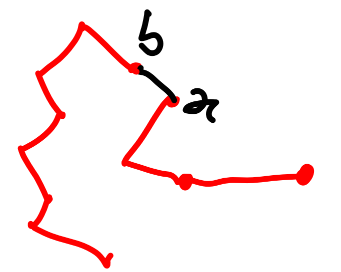
(i) $\ell(P_0) = n-1$

$\Rightarrow$ G is Hamiltonian

(ii)



G is connected



longer path

Now a double counting estimate for $\sum_G \sum_X a(G,X)$.

(i) Fix $G \in \mathcal{G}_1$.

$$\sum_X a(G,X) \geq \binom{m}{\omega}\left(1 - \frac{n + n^{1/4}\log n + \omega}{m - \omega}\right)^{\omega}$$

$$\geq \binom{m}{\omega} \Big/ 10.$$

$$\left(1 - \frac{\approx 2}{\omega}\right)^{\omega}$$

Random choose ω edges $e_1, e_2, \ldots$

$$\Pr(a(G,X)=1) \geq \prod_{i=0}^{\omega-1} \Pr(e_i \text{ avoids } P_X, \text{SMALL}, e_1, \ldots, e_{i-1} \mid e_1, \ldots, e_{i-1})$$

So

$$|\mathcal{G}_1| \leq \frac{10}{\binom{m}{\omega}} \sum_{G \in \mathcal{G}} \sum_X a(G,X).$$

(ii) Now fix a graph H with $m-w$ edges.

Let
$$S_H = \sum_{G,X:\, G_X = H} a(G,X).$$

If $S_H > 0$ then H has the expansion properties we expect and its END sets are large. Thus

$$S_H \le \binom{N-m+w}{w}\left(1-\frac{\binom{n/1000}{2}}{N}\right)^{w} \le \binom{N-m+w}{w} e^{-(10^{-6}-o(1))w}.$$

There $\binom{N}{m-w}$ ways to add w edges to H. bounds the probability that a randomly chosen set of w edges avoids joining a to $END_H(a)$ for $a \in END_H$.

Thus

$$\sum_{G\in\mathcal{G}}\sum_{X} a(G,X) \leqslant \binom{N}{m-w}\binom{N-m+w}{w} e^{-\beta w}$$

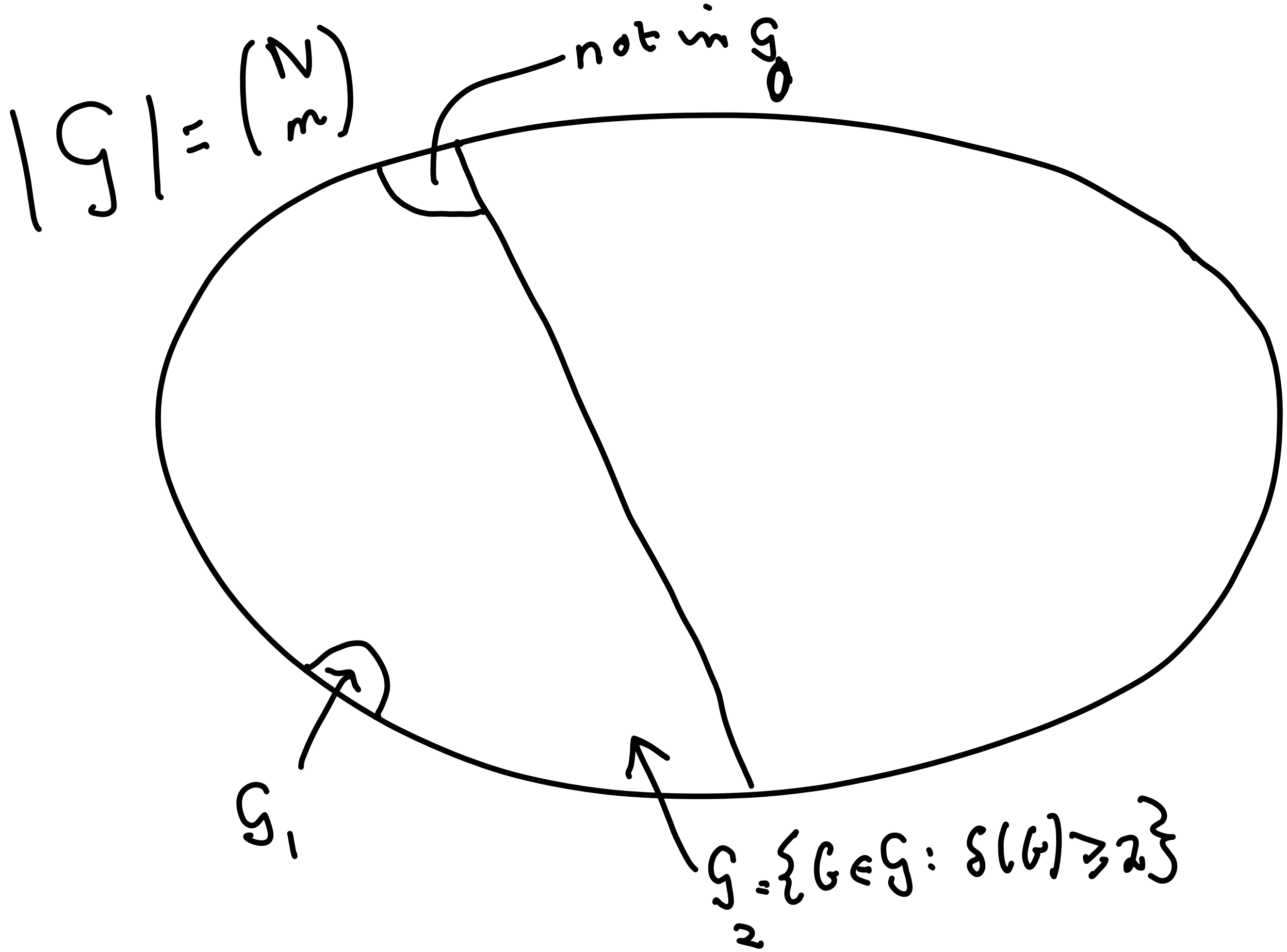
$\beta \approx 10^{-6}$

$$= \binom{N}{m}\binom{m}{w} e^{-\beta w}$$

and so

$$|\mathcal{G}_1| \leqslant \frac{10}{\binom{m}{w}} \sum_{G\in\mathcal{G}}\sum_{X} a(G,X).$$

$$\leqslant 10 e^{-\beta w}\binom{N}{m}.$$



$$\Pr(G \text{ is not Ham } \& \ \delta(G) \geq 2) = \Pr(G \in (\mathcal{G}_2 \setminus \mathcal{G}_0) \cup \mathcal{G}_1) = o(1).$$

$$\Pr(G \text{ is Ham } \& \ \delta(G) \geq 2) = e^{-e^{-c}} - o(1).$$

$\square$

Separation of largest degrees,
Graph isomorphism and
edge coloring.

Lemma

Let $k = (n-1)p + x\sqrt{(n-1)pq}$, p constant, $q = 1-p$,

where $x \le (\log n)^2$ (for convenience).

Then

$$B_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} = (1+o(1))\sqrt{\frac{1}{2\pi npq}}\, e^{-x^2/2}.$$

Proof

Stirling's Formula gives

$$B_k = (1+o(1))\sqrt{\frac{1}{2\pi npq}}\left(\left(\frac{(n-1)p}{k}\right)^{\frac{k}{n-1}}\left(\frac{(n-1)q}{n-1-k}\right)^{1-\frac{k}{n-1}}\right)^{n-1}.$$

Now

$$\left(\frac{k}{(n-1)p}\right)^{k/n-1} = \left(1 + x\sqrt{\frac{q}{p(n-1)}}\right)^{k/n-1}$$

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

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

$$\left(\frac{n-1-k}{(n-1)q}\right)^{1-k/n-1} = \left(1 - x\sqrt{\frac{p}{q(n-1}}\right)^{1-k/n-1}$$

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

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

So

$$\left(\frac{k}{(n-1)p}\right)^{k/n-1} \left(\frac{n-1-k}{(n-1)q}\right)^{1-k/n-1} =$$

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

Substituting into

$$(1+o(1))\sqrt{\frac{1}{2\pi npq}}\left(\left(\frac{(n-1)p}{k}\right)^{\frac{k}{n-1}}\left(\frac{(n-1)q}{n-1-k}\right)^{1-\frac{k}{n-1}}\right)^{n-1}$$

gives required expression.

$\square$

Lemma

Let $\epsilon = \frac{1}{10}$ and p be constant

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

Then whp

(i) $\Delta(G_{n,p}) \le k_+$

(ii) There are $\Omega(n^{2\epsilon(1-\epsilon)})$ vertices of degree at least k_-

(iii) $\nexists$ $u \neq v$ such that $d(u), d(v) \geq k_-$ and $|d(u) - d(v)| \le 10$.

We first prove that as $x \to \infty$

$$\frac{1}{x} e^{-x^2/2}\left(1 - \frac{1}{x^2}\right) \leq \int_x^\infty e^{-y^2/2}\,dy \leq \frac{1}{x} e^{-x^2/2}. \qquad (***)$$

<u>Proof</u>

$$\int_x^\infty e^{-y^2/2}\,dy = -\int_x^\infty \frac{1}{y}\left(e^{-y^2/2}\right)'\,dy$$

$$= -\left[\frac{1}{y} e^{-y^2/2}\right]_x^\infty - \int_x^\infty \frac{1}{y^2} e^{-y^2/2}\,dy$$

$$= \frac{1}{x} e^{-x^2/2} + \left[\frac{1}{y^3} e^{-y^2/2}\right]_x^\infty + 3\int_x^\infty \frac{1}{y^4} e^{-y^2/2}\,dy$$

$\square$

(1) Let X be the number of vertices of degree k.

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

assuming that $k \leq k_2 = (n-1)p + (\log n)^2\sqrt{(n-1)pq}$.

But if $k > k_2$ then

$E(X_k) \leq E(X_{k_2})$ – binomial $\searrow$ after mean

$$\approx n\exp\{-\Omega((\log n)^4)\}$$

$$= o(1).$$

So'y $Y_k = X_k + X_{k+1} + \cdots$

$$E(Y_k) \simeq \sum_{l=k}^{k_L} \sqrt{\frac{n}{2\pi pq}} \exp\left\{-\frac{1}{2}\left(\frac{l-(n-1)p}{\sqrt{(n-1)\,pq}}\right)^2\right\}$$

$$\simeq \sum_{l=k}^{\infty} \sqrt{\frac{n}{2\pi pq}} \exp\left\{-\frac{1}{2}\left(\frac{l-(n-1)p}{\sqrt{(n-1)\,pq}}\right)^2\right\}$$

$$\simeq \sqrt{\frac{n}{2\pi pq}} \int_{\lambda=k}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{\lambda-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\} d\lambda$$

If $k = (n-1)p + x\sqrt{(n-1)pq}$ then

$$\sqrt{\frac{n}{2\pi pq}} \int_{\lambda=k}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{\lambda-(n-1)p}{\sqrt{(n-1)pq}}\right)^2\right\} d\lambda$$

$$= \sqrt{\frac{n}{2\pi pq}} \cdot \sqrt{(n-1)pq} \cdot \int_{y=x}^{\infty} e^{-y^2/2}\, dy$$

$$\simeq \frac{n}{\sqrt{2\pi}} \cdot \frac{1}{x} \cdot e^{-x^2/2}$$

When $k = k_+$, $x = (1+\epsilon)\sqrt{2\log n}$ and (1) follows.

When $k = k_*$, $x = (1-\epsilon)\sqrt{2\log n}$

and $$E(Y_k) = \Omega\left(n^{2\epsilon(1-\epsilon)}\right) \to \infty.$$

We use the second moment method to show concentration.

$$E(Y_k(Y_{k-1}-1)) = n(n-1)\sum_{k\le k_1,k_2}^{\le k_*} \Pr(d(1)=k_1 \wedge d(2)=k_2)$$

$$= n(n-1)\left[\sum_{k_1,k_2} p\left(\hat{d}(1)=k_1-1 \wedge \hat{d}(2)=k_2-1\right) + (1-p)P\left(\hat{d}(1)=k_1 \wedge \hat{d}(2)=k_2\right)\right]$$

where $\hat{d}$ = # nbrs in $\{3,4,\dots,n\}$.

$$= n(n-1)\sum_{k_1,k_2}\Big[p\,P(\hat{d}(1)=k_1-1)\,P(\hat{d}(2)=k_2-1)$$
$$+ (1-p)\,P(\hat{d}(1)=k_1)\,P(\hat{d}(2)=k_2)\Big]$$

$$\frac{P(\hat{d}(1)=k_1-1)}{P(\hat{d}(1)=k_1)} = \frac{\binom{n-2}{k_1-1}(1-p)}{\binom{n-2}{k_1}\,p} = \frac{k_1(1-p)}{(n-2-k_1)\,p}$$
$$= 1+\tilde{O}(n^{-1/2}).$$

$$= n(n-1)\sum_{k_1,k_2}\Big[P(\hat{d}(1)=k_1)\,P(\hat{d}(2)=k_2)\,(1+\tilde{O}(n^{-1/2}))\Big]$$

$$= n(n-1) \sum_{k_1, k_2} \left[P(\hat{d}(1) = k_1)\, P(\hat{d}(2) = k_2) \left(1 + \tilde{O}(n^{-1/2})\right) \right]$$

$$\frac{P(\hat{d}(1) = k_1)}{P(d(1) = k_1)} = \frac{\binom{n-2}{k_1}}{\binom{n}{k_1}} (1-p)^{-2} = 1 + \tilde{O}(n^{1/2})$$

$$= n(n-1) \sum_{k_1, k_2} \left[P(d(1) = k_1)\, P(d(2) = k_2) \left(1 + \tilde{O}(n^{-1/2})\right) \right]$$

$$= E(Y_k)\left(E(Y_k) - 1\right)\left(1 + \tilde{O}(n^{-1/2})\right)$$

So, with $k = k_-$,

$$\Pr\left(Y_k \le \tfrac{1}{2} E(Y_k)\right)$$

$$\le \frac{E(Y_k(Y_k - 1)) + E(Y_k) - E(Y_k)^2}{E(Y_k)^2/4}$$

$$= \tilde{O}\left(\frac{1}{n^{2\epsilon(1-\epsilon)}}\right)$$

$$= o(1).$$

This completes the proof of the second part.

$$\Pr(\neg(\text{iii})) \leq o(1) + \binom{n}{2} \sum_{k_1 = k_-}^{k_+} \sum_{|k_2 - k_1| \leq 10} \Pr(d(1) = k_1 \wedge d(2) = k_2)$$

$$= o(1) + \binom{n}{2} \sum_{k_1, k_2} \Big[p P(\hat{d}(1) = k_1 - 1) P(\hat{d}(2) = k_2 - 1) + (1-p) P(\hat{d}(1) = k_1) P(\hat{d}(2) = k_2) \Big]$$

Now

$$\sum_{k_1, k_2} P(\hat{d}(1) = k_1 - 1) P(\hat{d}(2) = k_2 - 1)$$

$$\leq 21 (1 + \tilde{O}(n^{-1/2})) \sum_{k_1} \Pr(\hat{d}(1) = k_1 - 1)^2$$

and

$$\sum_{k_1} \Pr\left(\hat{d}(1) = k_1 - 1\right)^2 \lesssim \frac{1}{2\pi p q n} \int_{y=x}^{\infty} e^{-y^2}\, dy ,$$

where $x = \dfrac{k_1 - (n-1)p}{\sqrt{(n-1)pq}} \gtrsim (1-\epsilon)\sqrt{2}\log n$

$$= \frac{1}{\sqrt{8}\pi p q n} \int_{z = x\sqrt{2}}^{\infty} e^{-z^2/2}\, dz$$

$$\lesssim \frac{1}{\sqrt{8}\,\pi p q n} \cdot \frac{1}{x\sqrt{2}} \cdot n^{-2(1-\epsilon)^2} .$$

We get a similar bound for $\sum_{k_1} \Pr\left(\hat{d}(1) = k_1\right)^2$.

Thus

$$\Pr(\neg(iii)) = o\left(n^{2-1-2(1-\varepsilon)^2}\right)$$

$$= o(1).$$

$\square$

Edge Colouring

The Chromatic Index $\chi'(G)$ of graph G is the minimum number of colors that can be used to color the edges of G so that if 2 edges share a vertex, 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 $\chi'(G_{n,p}) = \Delta(G_{n,p})$ whp.

Graph Isomorphism

In this section we describe a procedure for ordering the vertices of a graph G. If it succeeds then it is possible to quickly tell if $G \cong H$, for any H.

Algorithm

Input G. Parameter L.

Step 1

Re-label vertices so that degrees satisfy

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

If $\exists\, i \leq L$ such that $d_G(v_i) = d_G(v_{i+1})$: FAIL

Step 2

For $i > L$ let

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

Re-label vertices so that these sets satisfy

$$X_{L+1} \succeq X_{L+2} \succeq \cdots \succeq X_n$$ — lexicographic ordering.

If $\exists\, i > L$ such that $X_i = X_{i+1}$: FAIL.

Suppose now that the above algorithm succeeds for G.

Given an n-vertex graph H we run the algorithm on H.

(i) If algorithm fails $G \not\cong H$.

(ii) Suppose ordering of $V(H)$ is $w_1, w_2, \dots, w_n$. Then

$G \cong H \iff$ $v_i \to w_i$ is an isomorphism.

Claim

Let $\rho = p^2 + q^2$ and $L = 3 \log_{1/\rho} n$.

Then whp the algorithm succeeds on $G = G_{n,p}$.

Proof

We have already proved that Step 1 succeeds whp.

We must now show that $X_i \neq X_j \ \forall i,j$ whp but there is slight problem because edges (v_i, v_j) are conditioned due to us knowing v_i has a high degree.

Fix i, j and let $\hat{G}_{i,j} = G \setminus \{i, j\}$.

Now if i, j are not high degree vertices then the L largest degree vertices in G, $\hat{G}_{i,j}$ will coincide, whp.

This is because there is whp, a gap of ≥ 10 between high vertex degrees in G.

Thus

$$\Pr(\text{Step 2 fails}) \leq$$

$$o(1) + \sum_{1 \leq i < j \leq n} \Pr\big(i, j \text{ have same nbrs among } L \text{ largest degree vertices in } \hat{G}_{i,j}\big)$$

$$= o(1) + \binom{n}{2} p^L$$

$$= o(1).$$

$\square$

Automorphisms

It follows from the previous section that whp, $G_{n,p}$ has no non-trivial automorphisms.

For if $\sigma : [n] \to [n]$ is an automorphism, then

(i) $\sigma(v_i) = v_i$, $1 \le i \le L$

where v_i is the vertex with the ith largest degree.

(ii) $\sigma(v) = v$ for $v \notin \{v_1, v_2, \ldots, v_L\}$.

This is because all of the sets X_v are distinct.

Janson's Inequality

Suppose that $0 \leq p_i \leq 1$ for $i = 1, 2, \dots, M$.

Let X be a random subset of $[M]$ where

$$\Pr(i \in X) = p_i$$

independently for $i \in [M]$

Let $S_1, S_2, \dots, S_\ell \subseteq [M]$ and let

$$Z_i = \begin{cases} 1 & S_i \subseteq X \\ 0 & S_i \nsubseteq X \end{cases}$$

Let $Z = Z_1 + Z_2 + \cdots + Z_\ell$

Count the number of S_i that occur.

Let

$$\mu = E(Z) = \sum_{i=1}^{\ell} p_i$$

and

$$\Delta^* = \frac{1}{2} \sum_{S_i \cap S_j \neq \emptyset} E(Z_i Z_j).$$

<u>Theorem</u>

$$\Pr(Z \leq \mu - t) \leq e^{-t^2/2\Delta^*}.$$

Proof

Let $\Psi(s) = E(e^{-sZ})$, $s \geq 0$.

$$\Pr(Z \leq \mu - t) = \Pr(e^{s(\mu - t - Z)} \geq 1)$$

$$\leq E(e^{s(\mu - t - Z)})$$

$$= e^{s(\mu - t)} \Psi(s).$$

Write

$$\log \Pr(Z \le \mu - t) \le \log \tilde{\Psi}(s) + s(\mu - t)$$

TO BE SHOWN $\le -\frac{\mu^2}{\Delta^*}\left(1 - e^{-s\Delta^*/\mu}\right) + s(\mu - t).$

If $t = \mu$ we simply let $s \to \infty$. Otherwise

to minimise RHS we take

$$s = -\frac{\mu}{\Delta^*} \log\left(1 - \frac{t}{\mu}\right)$$

and then

$$\log \Pr(Z \le \mu - t) \le -\frac{\mu}{\Delta^*}\left(t + (\mu - t)\log\left(1 - \frac{t}{\mu}\right)\right)$$

$$\log \Pr(Z \leq \mu - t) \leq -\frac{\mu}{\Delta^*}\left(t + (\mu - t)\log\left(1 - \frac{t}{\mu}\right)\right)$$

$$\leq -\frac{\mu}{\Delta^*}\left(t - (\mu - t)\frac{t}{\mu}\right)$$

$$= -\frac{t^2}{2\Delta^*},$$

TO BE SHOWN

$$\log \Psi(s) \leq -\frac{\mu^2}{\Delta^*}\left(1-e^{-s\Delta/\mu}\right).$$

Now

$$\log \psi(s) = \int_{u=0}^{s} \frac{\psi'(u)}{\psi(u)}\, du$$

and

$$\psi'(u) = -E\left(Ze^{-uZ}\right)$$

$$= -\sum_{i=1}^{M} E\left(Z_i e^{-uZ}\right).$$

For each $i \in [M]$ we write

$$Z = X_i + Y_i$$

where
$$X_i = \sum_{j : S_j \cap S_i \neq \emptyset} Z_j .$$

Then
$$E(Z_i e^{-uZ}) = p_i E\left(e^{-uX_i} e^{-uY_i} \mid Z_i = 1\right)$$

FKG inequality $$\geq p_i E\left(e^{-uX_i} \mid Z_i = 1\right) E\left(e^{-uY_i} \mid Z_i = 1\right)$$

$$= p_i E\left(e^{-uX_i} \mid Z_i = 1\right) E\left(e^{-uY_i}\right)$$

$$= p_i E(e^{-uX_i} | Z_i = 1) E(e^{-uY_i})$$

$$\geq p_i E(e^{-uX_i} | Z_i = 1) \psi(u).$$

So

$$\frac{\psi'(s)}{\psi(s)} \leq -\mu \sum_{i=1}^{M} \frac{p_i}{\mu} E(e^{-sX_i} | Z_i = 1)$$

$$\leq -\mu \sum_{i=1}^{M} \frac{p_i}{\mu} e^{-E(sX_i | Z_i = 1)}$$ Jensen

$$\leq -\mu \exp\left\{ -\sum_{i=1}^{M} E\left(\frac{s p_i}{\mu} X_i \middle| Z_i = 1 \right) \right\}$$

$$= -\mu \exp\left\{ -\frac{s}{\mu} \sum_{i=1}^{M} E(X_i | Z_i = 1) P(Z_i = 1) \right\}$$

$$= -\mu \exp\left\{ -\frac{s}{\mu} \sum_{i=1}^{M} E(X_i \mid Z_i = 1) P(Z_i = 1) \right\}$$

$$= -\mu e^{-s\Delta^*/\mu}$$

$$\sum_{i=1}^{M} E(X_i \mid Z_i = 1) P(Z_i = 1)$$

$$= \sum_{i=1}^{M} \sum_{j : S_j \cap S_i \neq \emptyset} P(Z_j = 1 \mid Z_i = 1) P(Z_i = 1)$$

$$= \Delta^* .$$

So
$$\frac{\psi'(s)}{\psi(s)} \leq -\mu e^{-s\Delta^*/\mu}.$$

$$\log \psi(s) = \int_{u=0}^{s} \frac{\psi'(u)}{\psi(u)}\, du$$

$$\leq -\int_{u=0}^{s} \mu e^{-u\Delta^*/\mu}\, du$$

$$= -\frac{\mu^2}{\Delta^*}\left(1 - e^{-s\Delta^*/\mu}\right). \qquad \square$$

Now let

$$\Delta = \frac{1}{2} \sum_{\substack{S_i \cap S_j \neq \emptyset \\ S_i \neq S_j}} E(Z_i Z_j).$$

<u>Theorem</u>

(i) $\Pr(Z=0) \leq e^{-\mu+\Delta}$

(ii) $\Pr(Z=0) \leq e^{-\frac{\mu^2}{\mu+\Delta}}$

((ii) follows directly from first theorem.)

For (1) we have (see p 8)

$$\frac{\psi'(s)}{\psi(s)} \leq -\sum_i p_i E\left(e^{-sX_i} \mid Z_i = 1\right).$$

So

$$\log(P(Z=0)) = \int_0^\infty (\log \psi(s))' \, ds$$

$$\leq -\int_0^\infty \sum_i p_i E\left(e^{-sX_i} \mid Z_i = 1\right) ds$$

$$= -\sum_i p_i \int_{s=0}^{\infty} E\left(e^{-sX_i} \mid Z_i = 1\right) ds$$

$$= -\sum_i p_i \, E\left(\frac{1}{X_i} \,\middle|\, Z_i = 1\right)$$

$$\leq -\sum_i p_i \, E\left(1 - \tfrac{1}{2}(X_i - Z_i) \,\middle|\, Z_i = 1\right)$$

$$\frac{1}{X_i} = \frac{1}{(X_i - Z_i) + Z_i} = \frac{1}{\underset{\uparrow \text{ integer}}{(X_i - Z_i)} + 1} \geq 1 - \tfrac{1}{2}(X_i - Z_i)$$

$$\leq -\sum_i p_i \, E\left(1 - \tfrac{1}{2}(X_i - Z_i) \,\middle|\, Z_i = 1\right)$$

$$= -\mu + \Delta.$$

$\square$

The diameter of Random Graphs

Theorem

Let $d \geq 2$ be a fixed positive integer. Suppose that $c > 0$ and

$$p^d n^{d-1} = \ln(n^2/c).$$

Then

$$\lim_{n\to\infty} \Pr(\text{diameter } G_{n,p} = k) = \begin{cases} e^{-c/2} & : k = d \\ 1 - e^{-c/2} & : k = d+1. \end{cases}$$

(a) Whp $\mathrm{diam}(G) \geq d$.

Fix $v \in V$ and let

$$N_k(v) = \{ w : \mathrm{dist}(v, w) = k \}.$$

We show that whp, for $0 \leq k < d$,

$$|N_k(v)| \leq (2np)^k$$

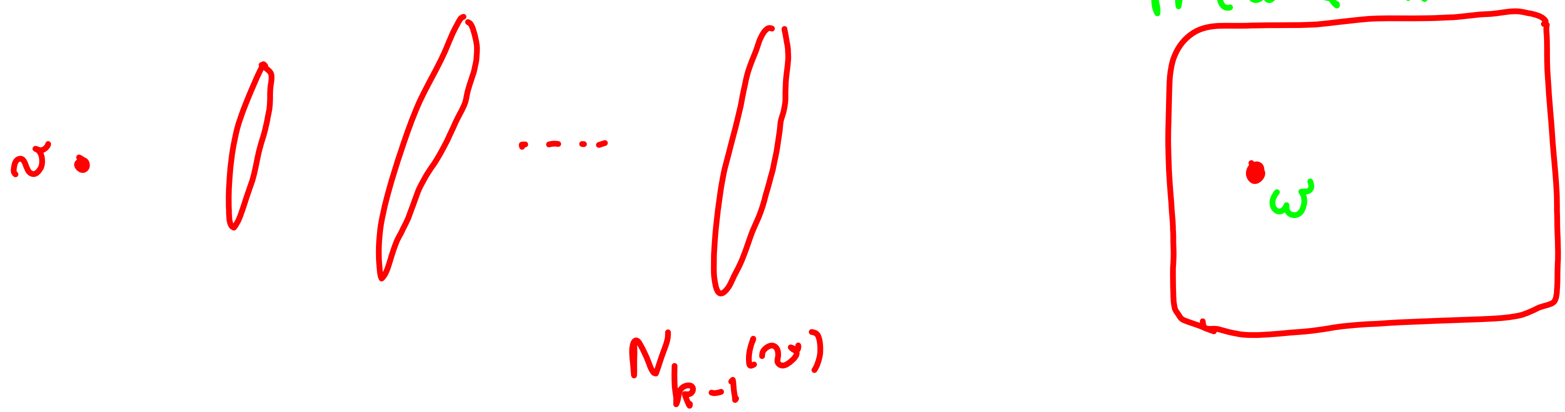
$$\approx (2n \ln n)^{k/d}$$

$$= o(n).$$

We observe that given $N_i(v)$, $i = 0, 1, \dots, k-1$,

that $|N_k(v)|$ is distributed as

$$\mathrm{Bin}\left(n - \sum_{i=0}^{k-1} |N_i(v)|,\quad 1 - \underbrace{(1-p)^{|N_{k-1}(v)|}}_{\Pr(w \notin N_k(v))}\right)$$



Let $\mathcal{E}_i = \{ |N_i(v)| \leq (2np)^i \}$.

Condition on $\mathcal{E}_0, \mathcal{E}_1, \dots \mathcal{E}_{k-1}$.

Does not condition on edges from $N_{k-1}(v)$ to
$V \setminus \bigcup_{i=0}^{k-1} N_i(v)$.

$|N_k(v)|$ is distributed as $\mathrm{Bin}(\nu, q)$
where $\nu < n$ and

$$q = 1 - (1-p)^{|N_{k-1}(v)|}$$
$$\leq |N_{k-1}(v)|\, p.$$
$$\leq (2np)^{k-1} p$$
$$< 1.$$

Thus

$$E\left(|N_k(v)| \;\middle|\; \mathcal{E}_0, \mathcal{E}_1, \dots, \mathcal{E}_{k-1}\right)$$

$$= vq$$

$$\leq np\,|N_{k-1}(v)|.$$

Chernoff bound gives

$$\Pr\left(|N_k(v)| \geq (2np)^k \;\middle|\; \mathcal{E}_0, \mathcal{E}_1, \dots, \mathcal{E}_{k-1}\right)$$

$$\leq \Pr\left(\mathrm{Bin}(n, |N_{k-1}(v)|p) \geq (2np)^k \;\middle|\; \mathcal{E}_{k-1}\right)$$

$$\leq \Pr\left(\mathrm{Bin}(n, (2np)^{k-1}p) \geq (2np)^k\right)$$

$$\leq \Pr\left(\mathrm{Bin}\left(n, (2np)^{k-1} p\right) \geq (2np)^{k}\right)$$

$$\leq e^{-(2np)^{k-1} np/3}$$

$$\ll n^{-2}.$$

So

$$\Pr\left(\bigcup_{i=0}^{d-1} N_i(v) = [n]\right) \leq$$

$$\sum_{k=1}^{d-1} \Pr\left(\overline{\mathcal{E}_k} \mid \mathcal{E}_1, \dots \mathcal{E}_{k-1}\right) =$$

$$O(n^{-2}).$$

(b) Whp $\operatorname{diam}(G) \le d+1$.

__Proof__

Fix $v, w \in [n]$. Then for $1 \le k < d$,

Let $\mathcal{F}_k = \left\{ |N_k(v)| \ge \left(\frac{np}{2}\right)^k \right\}$.

$$\Pr\left(\overline{\mathcal{F}_k} \,\middle|\, \mathcal{E}_1, \mathcal{F}_1, \ldots, \mathcal{E}_{k-1}, \mathcal{F}_{k-1}\right) =$$

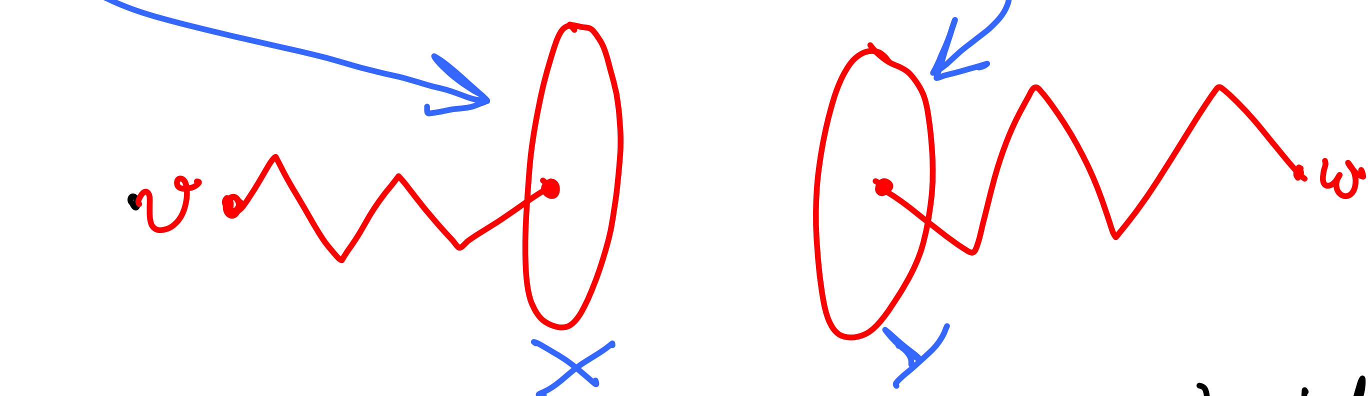
$$\Pr\left(\operatorname{Bin}\Big(\underbrace{n - \sum_{i=0}^{k-1} |N_i(v)|}_{n - o(n)},\ \underbrace{1 - (1-p)^{|N_{k-1}(v)|}}_{\ge \frac{3}{4}\cdot\left(\frac{np}{2}\right)^{k-1} p}\Big) \le \left(\frac{np}{2}\right)^k\right)$$

$$\le e^{-\Omega\left(\left(\frac{np}{2}\right)^k\right)} = O(n^{-3}).$$

So with probability $1 - O(n^{-3})$,

$$|N_{\lfloor d/2 \rfloor}(v)| \geq \left(\frac{np}{2}\right)^{\lfloor d/2 \rfloor} \quad \text{and}$$

$$|N_{\lceil d/2 \rceil}(w)| \geq \left(\frac{np}{2}\right)^{\lceil d/2 \rceil}.$$



Either $X \cap Y \neq \emptyset$ and $dist(v,w) \leq \lfloor d/2 \rfloor + \lceil d/2 \rceil = d$

or

$$\Pr(\nexists\, X:Y \text{ edge}) \leq (1-p)^{\left(\frac{np}{2}\right)^d}$$

$$\Pr(\nexists\, X:Y \text{ edge}) \leq (1-p)^{\left(\frac{np}{2}\right)^d}$$

$$\leq \exp\left\{-\left(\frac{np}{2}\right)^d p\right\}$$

$$\leq \exp\left\{-(2-o(1))\, np \ln n\right\}$$

$$= o(n^{-3}).$$

So

$$\Pr(\exists\, v,w : \operatorname{dist}(v,w) > d+1) = o(n^{-1}).$$

$$p^d n^{d-1} = \ln(n^2/c).$$

Now consider d or $d+1$ as diameter.

We use Janson's inequality.

For $v, w \in [n]$ let

$$\mathcal{A}_{v,w} = \{ v, w \text{ are not joined by a path of length } d \}$$

For $x_1, x_2, \dots, x_{d-1}$ let

$$\mathcal{B}_{\underbrace{x_1, x_2, \dots, x_{d-1}}_{X}} = \{ (v, x_1, x_2, \dots, x_{d-1}, w) \text{ is a path in } G_{n,p} \}.$$

Let $$Z = \sum_{x = x_1, \dots, x_{d-1}} Z_x \leftarrow \begin{cases} 1 : \mathcal{B}_x \text{ occurs} \\ 0 : \neg\mathcal{B}_x \text{ occurs} \end{cases}$$

$$\mu = E(Z) = (n-2)(n-3)\dots(n-d)\,p^d$$
$$\approx \ln(n^2/c).$$

Let
$$\Delta = \sum_{\substack{x = x_1, x_2, \dots x_{d-1} \\ y = y_1, y_2, \dots y_{d-1}}} Pr(\mathcal{B}_x \cap \mathcal{B}_y)$$

v, x, w and v, y, w share an edge

$x \neq y$

$$\leq \sum_{c=1}^{d-1} n^{2(d-1)-c} \; p^{2d-c}$$

edges in common between x and y

$$= O\left(\sum_{c=1}^{2d-2} n^{2(d-1)-c-\frac{d-1}{d}(2d-c)} (\log n)^{\frac{2d-c}{d}} \right)$$

$$= \tilde{O}(n^{-c/d})$$

$$= o(1).$$

Applying $\Pr(Z=0) \leqslant e^{-\mu+\Delta}$ we get

$$\Pr(Z=0) \leqslant (1+o(1))\, c/n^2 .$$

On the other hand the FKG inequality implies

$$\Pr(Z=0) \geqslant (1-p^d)^{(n-2)(n-3)\cdots(n-d)}$$

$$= (1-o(1))\, c/n^2$$

So

$$\Pr(A_{v,w}) = \Pr(Z=0) = (1+o(1))\, c/n^2 .$$

So $$E(\# v,w : A_{v,w} \text{ occurs}) \approx c/2$$

and we should expect that

$$P_r(\nexists\, v,w : A_{v,w} \text{ occurs}) \approx e^{-c/2}. \quad (1)$$

Indeed, if we choose $v_1, w_1, v_2, w_2, \ldots, v_k, w_k$, k constant, we find that

$$P_r(A_{v_1,w_1} \cap A_{v_2,w_2} \cap \cdots \cap A_{v_k,w_k}) \quad (2)$$

$$\approx \left(c/n^2\right)^k$$

and (1) follows as in previous arguments.

For (2) we define

$$Z = Z_1 + Z_2 + \cdots + Z_k$$

where

Z_i = # paths of length d from v_i to w_i.

We need to show that the corresponding $\Delta = o(1)$

and then we need to show that $1 \le r < s \le k$

$$\Delta_{r,s} = \sum_{\substack{x = x_1, x_2, \ldots x_{d-1} \\ y = y_1, y_2, \ldots y_{d-1}}} \Pr\left(\mathcal{B}^r_x \cap \mathcal{B}^s_y\right) = o(1)$$

v_r, x, w_r and v_s, y, w_s share an edge

$x \neq y$

But

$$\Delta_{r,s} \leq \sum_{c=1}^{d-1} n^{2(d-1)-c} \; p^{2d-c}$$

$$= o(1)$$

as before. $\square$

Independence and Chromatic Number

Theorem

Suppose $0 < p < 1$ is constant and $b = \frac{1}{1-p}$.

Then whp

$$\alpha(G_{n,p}) \approx 2 \log_b n$$

$\alpha(G)$ = size of largest independent set in G.

Proof

Let X_k = # of independent sets of size k.

(1) Let $k = \lceil 2 \log_b n \rceil$

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

$$\leq \left(\frac{ne}{k(1-p)^{1/2}} \cdot (1-p)^{k/2} \right)^k$$

$$\leq \left(\frac{e}{k(1-p)^{1/2}} \right)^k$$

$$= o(1).$$

(ii) Let now

$$k = \lfloor 2\log_b n - 3\log_b \log_b n \rfloor$$

Let $$\Delta^* = \sum_{\substack{i,j \\ S_i \sim S_j}} \Pr(S_i \wedge S_j \text{ are independent in } G_{n,p})$$

where $S_1, S_2, \ldots S_{\binom{n}{k}}$ are all k-subsets of $[n]$

and $S_i \sim S_j$ iff $|S_i \cap S_j| \geq 2$.

$$\Pr(X_k = 0) \leq \exp\left\{ -\frac{E(X_k)^2}{\Delta^*} \right\}$$

Janson's Inequality

$$\frac{\Delta^*}{E(X_k)^2} = \frac{\binom{n}{k}(1-p)^{\binom{k}{2}} \sum_{j=2}^{k} \binom{n-k}{k-j}\binom{k}{j}(1-p)^{\binom{k}{2}-\binom{j}{2}}}{\left(\binom{n}{k}(1-p)^{\binom{k}{2}}\right)^2}$$

$$= \sum_{j=2}^{k} \underbrace{\frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}}(1-p)^{-\binom{j}{2}}}_{u_j}$$

$$\frac{u_j}{u_2} \le \left[\frac{k}{n-2k}\cdot\frac{ke}{j-2}\cdot(1-p)^{-\frac{j+1}{2}}\right]^{j-2} \qquad j>2$$

$$< 1.$$

So

$$\frac{E(X_k)^2}{\Delta^*} \geq \frac{1}{k u_2} \geq \frac{n^2(1-p)}{k^5}$$

So

$$\Pr(X_k = 0) \leq e^{-\Omega(n^2/(\log n)^5)}, \quad (*)$$

$\square$

Theorem

$$\chi(G_{n,p}) \approx \frac{n}{2\log_b n}$$

Proof

(1) $$\chi(G_{n,p}) \geq \frac{n}{\alpha(G_{n,p})}$$

$$\approx \frac{n}{2\log_b n}.$$

(ii)

Let $\nu = \frac{n}{(\log n)^2}$. It follows from $(*)$ on p5 that

$$\Pr(\exists S : |S| \geq \nu \text{ and } S \not\supseteq \text{independent set of size } \geq k_0 = 2\log_b n - 3\log_b\log_b n)$$

$$\leq \binom{n}{\nu} \exp\left\{-\Omega\left(\frac{\nu^2}{(\log n)^5}\right)\right\}$$

$$= o(1).$$

So assume that every set of size $\geq \nu$ contains an independent set of size $\geq k_0$.

So we repeatedly

Choose an independent set of size k_0.

Give it a new colour.

Repeat until number of uncoloured vertices is $\leq \nu$.

Give each remaining vertex its own colour.

Number of colours used

$$\leq \frac{n}{k_0} + \nu \approx \frac{n}{2\log_b n}.$$

$\square$

Performance of Greedy Algorithm

Algorithm (GREEDY)

k is current colour

A is current set of vertices that might get colork in current round.

U is current set of uncoloured vertices.

begin

$k \leftarrow 0$; $A \leftarrow [n]$; $U \leftarrow [n]$;

while $U \neq \emptyset$;

begin

START ITERATION

$k \leftarrow k+1$; $A \leftarrow U$;

Choose $v \in A$ and give it colour k;

$U \leftarrow U \setminus \{v\}$

$A \leftarrow A \setminus (\{v\} \cup N(v))$

If $A \neq \emptyset$

otherwise

end

Theorem

Whp GREEDY uses $\approx \frac{n}{\log_b n}$ colours

(about twice as many as it "should").

Proof

At the start of an iteration the edges inside U are unexamined. Suppose that $|U| \geq \nu = \frac{n}{(\log n)^2}$. We show that $\approx \log_b n$ vertices get colour k.

Each iteration choose a maximal independent set from the remaining uncolored vertices.

$$\Pr\Big(\exists S : |S| \leq \underbrace{\log_b n - 3\log_b \log_b n}_{k_0} \text{ and } S \text{ is maximal independent}\Big)$$

$$\leq \sum_{s=1}^{k_0} \binom{n}{s} (1-p)^{\binom{s}{2}} \left(1-(1-p)^s\right)^{n-s}$$

$$\leq \sum_{s=1}^{k_0} \left[\frac{ne}{s}(1-p)^{\frac{s-1}{2}}\right]^s e^{-(n-s)(1-p)^s}$$

$$\leq \sum_{s=1}^{k_0} \left(ne^{1+(1-p)^s}(1-p)^{\frac{s-1}{2}}\right)^s e^{-n(1-p)^s}$$

$$\leq \sum_{s=1}^{k_0} \left(n e^{1+(1-p)^s} (1-p)^{\frac{s-1}{2}} \right)^s e^{-n(1-p)^s}$$

$$\leq \sum_{s=1}^{k_0} (n e^2)^s e^{-n(1-p)^s}$$

$$\leq k_0 (n e^2)^{k_0} e^{-(\log_b n)^3}$$

$$\leq e^{-(\log_b n)^3/2}.$$

So the probability that we fail to use $\geq k_0$ colours while $|U| \geq \nu$ is at most $n e^{-(\log_b \nu)^3/2}$

$= o(1)$.

On the other hand let

$$k_1 = \log_b n + 2\log_b \log_b n$$

Consider one round. Let $U_0 = V$ and suppose $u_1, u_2, \dots$ get colour k and $U_{i+1} = U_i \setminus (\{u_i\} \cup N(u_i))$.

Then
$$E(|U_{i+1}| \mid U_i) \leq |U_i|(1-p)$$

and so
$$E(|U_k|) \leq n(1-p)^k .$$

So
$$\Pr(k_1 \text{ vertices coloured in a round}) \leq \frac{1}{(\log_b n)^2}$$

$$\Pr(2k_1 \text{ vertices coloured in a round}) \leq \frac{1}{n^3}$$

So let

$$\delta_i = \begin{cases} 1 & \le k_1 \text{ colours used in Round } i \\ 0 & \text{otherwise} \end{cases}$$

We see that

$$\Pr(\delta_i = 1 \mid \delta_1, \delta_2 \dots \delta_{i-1}) = 1 - O(1/(\log n)^2)$$

and deduce that whp $O(n/(\log_b n)^2)$ rounds colour more than k_1 vertices and no round colours more than $2k_1$ vertices.

$\square$

Concentration

Theorem

$$\Pr\left(\left|X(G_{n,p}) - E(X(G_{n,p}))\right| \geq t\right) \leq 2e^{-\frac{t^2}{2n}}$$

Proof

Write $X = Z(Y_1, Y_2, \dots, Y_n)$ where

$$Y_j = \{(i,j) \in E(G_{n,p}) : i < j\}.$$

Then $$|Z(Y_1, \dots, Y_i, \dots, Y_n) - Z(Y_1, \dots, \hat{Y}_i, \dots, Y_n)| \leq 1$$

and the theorem follows from a martingale inequality.

$\square$

Concentration from Martingales

A sequence of random variables $X_0, X_1, \dots X_n, \dots$ where $X_i = X_i(A_0, A_1, \dots, A_i)$ is called a **martingale w.r.t. $A_0, A_1, \dots, A_n, \dots$**

Nb: $\underbrace{E(X_{i+1} \mid A_0, A_1, \dots, A_i)}_{}$ $= X_i$

this is a random variable.

$$X_{i+1}(\omega) = \sum_{\substack{\hat{\omega} \,:\, A_j(\hat{\omega}) = A_j(\omega) \\ 0 \le j \le i}} X_{i+1}(\hat{\omega}) \Pr(\hat{\omega}).$$

Theorem

Suppose that $X_0, X_1, \ldots, X_n$ is a martingale, w.r.t. $A_0, A_1, \ldots, A_n$ and

$$a_i \leq X_{i+1} - X_i \leq b_i \qquad , \quad i = 1, 2, \ldots, n.$$

Then

$$P_r(|X_n - X_0| \geq t) \leq 2e^{-2t^2/\sum_i (b_i - a_i)^2}.$$

Proof

We first consider $P_r(X_n - X_0 \geq t)$

Fix $\lambda > 0$. Then

$$\Pr(X_n - \mu \geq t) = \Pr\left(e^{\lambda(X_n - X_0 - t)} \geq 1 \right)$$

$$\leq E\left(e^{\lambda(X_n - X_0 - t)} \right)$$

$$= e^{-\lambda t} E\left(e^{\lambda(X_n - X_0)} \right)$$

$$= e^{-\lambda t} E\left(\exp\left\{ \sum_{i=0}^{n} \lambda Y_i \right\} \right)$$

where $Y_i = X_i - X_{i-1}$.

$$= e^{-\lambda t} E\left(\prod_{i=0}^{n} e^{\lambda Y_i} \right)$$

We show that

$$E\left(\prod_{i=0}^{n} e^{\lambda Y_i}\right) \leq e^{\frac{\lambda^2}{8}(b_n-a_n)^2} E\left(\prod_{i=0}^{n-1} e^{\lambda Y_i}\right) \quad (1)$$

and then induction gives

$$E\left(\prod_{i=0}^{n} e^{\lambda Y_i}\right) \leq \exp\left\{\lambda^2 \sum_{i=1}^{n} (b_i-a_i)^2/8\right\}$$

and

$$Pr(X_n - X_0 \geq t) \leq \exp\left\{-\lambda t + \lambda^2 \sum_{i=1}^{n} (b_i-a_i)^2/8\right\}$$

Now choose

$$\lambda = \frac{4t}{\sum_{i=1}^{n} (b_i-a_i)^2}.$$

Proof of (1).

$$E\left(\prod_{i=0}^{n} e^{\lambda Y_i}\right)$$

$$= E\left(\prod_{i=0}^{n-1} e^{\lambda Y_i} \, E\left(e^{\lambda Y_n} \mid A_0, A_1, \ldots, A_{n-1}\right)\right)$$

and we obtain (1) from

$$E\left(e^{\lambda Y_n} \mid A_0, A_1, \ldots, A_{n-1}\right) \leq e^{\lambda^2 (b_n - a_n)^2 / 8} \quad (2)$$

Proof of (2)

Y_n satisfies

(i) $E(Y_n) = 0$ and (ii) $a_n \leq Y_n \leq b_n$

$\left(E(Y_n) = E_{A_0, \dots A_{n-1}}\left(E_{A_n}(X_n - X_{n-1} \mid A_0, A_1, \dots, A_{n-1})\right) = 0 \right)$

Now if $a_n \leq Y_n \leq b_n$

$$e^{\lambda Y_n} \leq \frac{Y_n - a_n}{b_n - a_n} e^{\lambda b_n} + \frac{b_n - Y_n}{b_n - a_n} e^{\lambda a_n}$$

and so

$$E\left(e^{\lambda Y_n}\right) \leq \frac{b_n}{b_n - a_n} e^{\lambda a_n} - \frac{a_n}{b_n - a_n} e^{\lambda b_n}$$

$$E(e^{\lambda Y_n}) \le \frac{b_n}{b_n - a_n} e^{\lambda a_n} - \frac{a_n}{b_n - a_n} e^{\lambda b_n}$$

$$= e^{f(y)}$$

where if $p = -a_n/(b_n - a_n)$, $y = (b_n - a_n)\lambda$,

$$f(y) = -py + \ln(1 - p + pe^{y})$$

$$f''(y) = \frac{p(1-p)e^{-y}}{(p + (1-p)e^{-y})^2} \le \frac{1}{4} \qquad \left[\frac{AB}{(A+B)^2} \le \frac{1}{4}\right]$$

and so

$$f(y) \le y^2/8$$

proving (2).

For the lower bound

$$\Pr(X_n - X_0 \leq -t) = \Pr(-X_n + X_0 \geq t)$$

$$\leq e^{-2t^2 / \sum_i (b_i - a_i)^2}.$$

$\square$

Sometimes our sequence is a submartingale or a supermartingale.

$$E(X_{i+1} \mid \cdots) \leq X_i \qquad E(X_{i+1} \mid \cdots) \geq X_i$$

To bound $\Pr(X_n - X_0 \geq t)$ we used

$$e^{\lambda Y_n} \leq \frac{Y_n - a_n}{b_n - a_n} e^{\lambda b_n} + \frac{b_n - Y_n}{b_n - a_n} e^{\lambda a_n}$$

$$= Y_n \Bigg[\underbrace{\frac{e^{\lambda b_n} - e^{\lambda a_n}}{b_n - a_n}}_{\geq 0} \Bigg] + \frac{b_n}{b_n - a_n} e^{\lambda a_n} - \frac{a_n}{b_n - a_n} e^{\lambda b_n}$$

$E(Y_n) \leq 0$ for a supermartingale.

So our estimate for $\Pr(X_n - X_0 \geq t)$ is valid for supermartingales. For $\Pr(X_n - X_0 \leq -t)$, it is valid for submartingales.

We now prove a similar, but slightly different version:

<u>Theorem</u>

Suppose that $X_0, X_1, \ldots, X_n$ is a martingale, w.r.t. $A_0, A_1, \ldots, A_n$ and

for $0 \le a_i \le 1$, $i = 1, 2, \ldots, n$,

$$-a_i \le X_{i+1} - X_i \le 1 - a_i \quad , \quad i = 1, 2, \ldots, n.$$

Let $a = \frac{1}{n}(a_1 + \cdots + a_n)$ and $\bar{a} = 1 - a$. Then

$$\Pr(|X_n - X_0| \ge nt) \le \left(\left(\frac{a}{a+t}\right)^{a+t} \left(\frac{\bar{a}}{\bar{a} - t}\right)^{\bar{a} - t} \right)^n$$

for $t \le \bar{a}$.

We will first observe that

$$E(e^{\lambda Y_n}) \leq (1-a_n)e^{-\lambda a_n} + a_n e^{\lambda(1-a_n)}$$

so that

$$P_t(X_n - X_0 \geq nt) \leq e^{-\lambda nt} \prod_{k=1}^{n} \left[(1-a_k)e^{-\lambda a_k} + a_k e^{\lambda(1-a_k)} \right]$$

$$= e^{-\lambda n(a+t)} \prod_{k=1}^{n} \left(1 - a_k + a_k e^{\lambda}\right)$$

$$\leq e^{-\lambda n(a+t)} \left(1 - a + a e^{\lambda}\right)^n$$

Now put

$$e^{\lambda} = \frac{(a+t)(1-a)}{a(1-a-t)} .$$

Corollary

Under the conditions above:

(i)

$$\Pr(|X_n - X_0| \geq t) \leq 2e^{-2t^2/n}$$

(ii)

$$\Pr(X_n - X_0 \geq \epsilon a n) \leq \left((1+\epsilon)^{1+\epsilon} e^{-\epsilon}\right)^{an}$$

$$\leq e^{-\frac{\epsilon^2 a n}{2(1+\epsilon/3)}}$$

(iii)

$$\Pr(X_n - X_0 \leq -\epsilon a n) \leq e^{-\epsilon^2 a n/2}$$

(1)

Let

$$f(t) = \ln\left(\left(\frac{a}{a+t}\right)^{a+t} \left(\frac{\bar{a}}{\bar{a}-t}\right)^{\bar{a}-t} \right)$$

$$f'(t) = \ln\left(\frac{a(\bar{a}-t)}{(a+t)\bar{a}} \right)$$

$$f''(t) = -\left((a+t)(\bar{a}-t) \right)^{-1} \leq -4$$

$f(0) = f'(0) = 0$ and so

$$f(t) \leq -2t^2.$$

(ii)

$$\Pr(X_n - X_0 \geq \epsilon a n) \leq \left[e^{-\lambda a(1+\epsilon)}(1 - a + a e^{\lambda})\right]^n$$

Now let $e^{\lambda} = 1+\epsilon$

$$\leq \left[(1+\epsilon)^{-a(1+\epsilon)}(1 + a\epsilon)\right]^n$$

$$\leq \left[(1+\epsilon)^{-(1+\epsilon)} e^{\epsilon}\right]^{an}$$

and now use

$$(1+\epsilon)\ln(1+\epsilon) - \epsilon \geq \frac{3\epsilon^2}{6+2\epsilon}$$

to get second inequality in (ii).

(iii)

$$f(t) = \ln\left(\left(\frac{a}{a+t}\right)^{a+t}\left(\frac{\bar{a}}{\bar{a}-t}\right)^{\bar{a}-t}\right)$$

$h(x) = f(-ax)$ for $0 \leq x \leq 1$.

$$h''(x) = a^2 f''(-ax) = -\frac{a}{(1-x)(\bar{a}+xa)} \leq -a$$

and so

$$f(-ax) \leq -ax^2/2.$$

Doob Martingale

$\Omega = \{(A_1, A_2, \ldots, A_n)\}$

Let $Z = Z(A_1, A_2, \ldots, A_n)$ be a random variable with $E(Z) = 0$.

Define random variables, $X_0 = 0$ and

$$X_i = E(Z \mid A_1, A_2, \ldots, A_i), \quad 1 \leq i \leq n$$

Claim

$X_0, X_1, \ldots, X_n$ is a martingale w.r.t. $A_0, A_1, \ldots, A_n$ with $X_0 = E(Z) = 0$ and $X_n = Z$.

Proof

$$E(X_{i+1} \mid A_0, A_1, \ldots, A_i)$$

A_{i+1}

$$= E\big(E(Z \mid A_0, A_1, \ldots, A_{i+1}) \mid A_0, A_1, \ldots, A_i\big)$$

given $A_0, \ldots, A_i$ we are averaging X_{i+1} over A_{i+1}

$$= X_i .$$

$\square$

Case 1

$Z = Z_1 + Z_2 + \cdots + Z_n$

where $Z_1, Z_2, \ldots, Z_n$ are independent.

Put

$$X_i = \sum_{j=1}^{i} \left(Z_j - E(Z_j) \right).$$

$$X_{i+1} = X_i + Z_{i+1} - E(Z_{i+1})$$

$$E(X_{i+1} \mid X_i, X_{i-1}, \ldots X_1) = X_i$$

and all the derived inequalities apply.

In particular if $0 \leq Z_i \leq 1$ and $E(Z_i) = a_i$

then we get bound on

$$\Pr\left(\left| Z - \sum a_i \right| \geq t \right)$$

by considering

$$\hat{Z}_i = Z_i - a_i \in [-a_i, 1 - a_i].$$

Case 2

$Z = Z(A_1, \dots, A_n)$ and $A_0, A_1, \dots A_n$ are independent.

Theorem

If

$$a_i \leq Z(A_1, \dots, A_i, \dots A_n) - Z(A_1, \dots, \hat{A_i}, \dots, A_n) \leq b_i$$

for all $i, A_1, A_2, \dots, A_n, \hat{A_i}$ then

$$\Pr(|Z - E(Z)| \geq t) \leq 2e^{-2t^2 / \sum_{i=1}^{n} (b_i - a_i)^2}.$$

We can assume w.l.o.g. that $E(Z)=0$.

Now define $X_i = E(Z \mid A_1,\dots,A_i)$ as before

$$X_{i+1}-X_i = \sum_{\hat{A}_{i+1},\dots,\hat{A}_n} \underbrace{\left(Z(A_1,\dots,A_{i+1},\hat{A}_{i+2},\dots,\hat{A}_n) - Z(A_1,\dots,A_i,\hat{A}_{i+1},\dots,\hat{A}_n)\right)}_{\in[a_i,b_i]} * \Pr(\hat{A}_{i+1}) * \Pr(\hat{A}_{i+2}) * \dots * \Pr(\hat{A}_n)$$

So $a_i \le X_{i+1}-X_i \le b_i$.

$\square$

In $G_{n,p}$ we can take

(i) $A_1, A_2, \ldots, A_{\binom{n}{2}}$ as independent 0-1 random variables defining G.

(ii) $\tilde{A}_i = \{ (v,i) : v \le i \text{ and } (v,i) \in E(G_{n,p}) \}$

Case 3

For $G_{n,m}$ we need a slight modification.

Suppose

$$Z = Z(u_1, u_2, \dots, u_m)$$

where $u_1, u_2, \dots, u_N$ is a random permutation of $\{1, 2, \dots, N\}$.

Suppose that

one interchange

$$a_i \le Z(u_1, \dots u_i, \dots, u_m) - Z(u_1, \dots, \hat{u}_i, \dots, u_m) \le b_i$$

then

$$\Pr(|Z - E(Z)| \ge t) \le 2e^{-t^2/\sum_{i=1}^{m}(b_i - a_i)^2}$$

Now define $X_i = E[Z \mid u_1, u_2, \ldots, u_i)$

$$X_{i+1} - X_i =$$

$$\sum_{\hat{u}_{i+1}, \ldots, \hat{u}_N} \Big(Z(u_1, \ldots, u_i, \hat{u}_{i+1}, \ldots, \hat{u}_j = u_{i+1}, \ldots \hat{u}_N) - Z(u_1, \ldots u_i, u_{i+1}, \ldots, \hat{u}_{i+1}, \ldots \hat{u}_N) \Big)$$

$$\times \frac{1}{(N-i)!}$$

BWE INTERCHANGE

$\square$

Small Subgraphs

Let H be a fixed graph.

We use the notation n_H, e_H for the number of vertices and edges of G.

Also let

$$\rho_H = \frac{e_H}{v_H}.$$

Lemma

Let X_H denote the number of copies of H in $G_{n,p}$.

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

where

$\mathrm{aut}(H)$ is the number of automorphisms of H.

__Proof__

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

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

and all we need to show is that

$$a_H \times \operatorname{aut}(H) = v_H!$$

Each permutation σ of $[v_H]$ 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 $\{(x_i, y_i) : 1 \le i \le e_H\}$ is some fixed copy of H in K_{v_H}.

But $H_\sigma = H_{\tau\sigma}$ iff $\forall i \; \exists j$ such that $(x_{\tau\sigma(i)}, y_{\tau\sigma(i)}) = (x_{\sigma(j)}, y_{\sigma(j)})$ i.e. τ is an automorphism of H.

$\square$

Theorem

Suppose $p = o(n^{-1/\rho_H})$. Then whp, $G_{n,p}$ contains no copies of H.

Proof

Suppose that $p = \frac{1}{\omega} n^{-1/\rho_H}$ where $\omega(n) \to \infty$. Then

$$E(X_H) \le n^{v_H} \omega^{-e_H} n^{-e_H/\rho_H}$$

$$= \omega^{-e_H}$$

$$\to 0.$$

$\square$

Now consider the case where $n^{1/\rho_H} p \to \infty$.

Suppose $p = \omega n^{-1/\rho_H}$ where $\omega \to \infty$.

Then for some constant $c_H > 0$

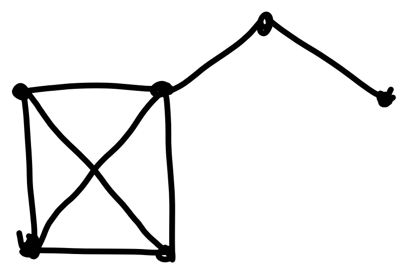
$$E(X_H) \geq c_H n^{v_H} \omega^{e_H} n^{-e_H/\rho_H}$$

$$= c_H \omega^{e_H}$$

$$\to \infty.$$

This is not enough to show that whp $G_{n,p}$ contains a copy of H.

Suppose

$H =$  $\quad n_H = 6$, $e_H = 8$

Let $p = n^{-5/7}$. Then $1/\rho_H > \frac{5}{7}$ and so

$$E(X_H) \approx c_H n^{6 - 8 \times 5/7} \longrightarrow \infty.$$

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

$$E(X_{\hat{H}}) \leq n^{4 - 6 \times 5/7} \longrightarrow 0$$

and so whp there are no copies of $\hat{H}$ and hence no copies of H.

Theorem

Let $\rho_H^* = \max\limits_{\substack{H' \subseteq H \\ v_{H'} > 0}} \rho_{H'}$.

(a) If $n^{-1/\rho_H^*} p \longrightarrow 0$

then whp $X_H = 0$.

(b) If $n^{-1/\rho_H^*} p \longrightarrow \infty$

then whp $X_H > 0$.

Proof

(a) follows from p5 because in this case there is whp, an $H' \subseteq H$ such that $X_{H'} = 0$.

(b) We use the second moment method:

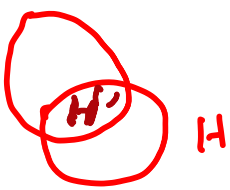
$$\Pr(X_H > 0) \geq \frac{E(X_H)^2}{E(X_H^2)}$$

$$E(X_H^2) = \sum_{i,j=1}^{N_H} \Pr(H_i \wedge H_j)$$

$H_1, H_2, \ldots$ are all copies of H in K_n.

$$= E(X_H) \sum_{j=1}^{N_H} \Pr(H_j \mid H_1)$$

$$\leq E(X_H)^2 + E(X_H) \sum_{\substack{H' \subseteq H \\ H' \neq H \\ e_{H'} > 0}} n^{v_H - v_{H'}} \, p^{e_H - e_{H'}}$$



So

constant

$$\frac{E(X_H^2)}{E(X_H)^2} - 1 \leq c_H \sum_{\substack{H' \subseteq H \\ H' \neq H \\ e_{H'} > 0}} n^{-v_{H'}} \, p^{-e_{H'}}$$

But

$$\sum_{\substack{H' \subseteq H \\ H' \neq H}} n^{-v_{H'}} p^{-e_{H'}} = \sum_{\substack{H' \subseteq H \\ H' \neq H \\ e_{H'} > 0}} n^{e_{H'}\left(\frac{1}{\rho^*_H} - \frac{1}{\rho_{H'}}\right)} \omega^{-e_{H'}}$$

$p = \omega n^{-1/\rho^*_H}$

$$= O(\omega^{-1}).$$

Thus

$$\frac{E(X_H^2)}{E(X_H)^2} = 1 + o(1).$$

$\square$

Every monotone property has a threshold

Let $\mathcal{G}$ be a monotone increasing property of graph. Assume $\overline{K_n} \notin \mathcal{G}$ and $K_n \in \mathcal{G}$.

Given $0 < \epsilon < 1$ we define $p(\epsilon)$ by

$$\Pr\left(G_{n,p(\epsilon)} \in \mathcal{G} \right) = \epsilon .$$

$p(\epsilon)$ exists because
$\Pr(G_{n,p} \in \mathcal{G})$ is a polynomial in p that
increases from 0 $(p=0)$ to 1 $(p=1)$.

Theorem

$p^* = p(1/2)$ is a threshold for $\mathcal{G}$.

Proof

Suppose $G_1, G_2, \dots, G_k$ are independent copies of $G_{n,p}$. Then

(i) $$G_1 \cup G_2 \cup \dots \cup G_k \overset{\text{"same distribution as"}}{\cong} G_{n,\, \underbrace{1-(1-p)^k}_{\leq\, kp}} \underset{\text{coupling}}{\leq} G_{n,kp}.$$

(ii) With this coupling

$$G_{n,kp} \notin \mathcal{G} \implies G_1, G_2, \dots, G_k \notin \mathcal{G}.$$

So

$$\Pr(G_{n,kp} \notin \mathcal{G}) \leq \Pr(G_{n,p} \notin \mathcal{G})^k.$$

(i) Suppose now $p = p^*$ and $k = \omega \to \infty$

$$\Pr(G_{n,\omega p^*} \notin \mathcal{G}) \leq 2^{-\omega} = o(1).$$

(ii) Now suppose $p = p^*/\omega$.

$$\frac{1}{2} = \Pr(G_{n,p^*} \notin \mathcal{G}) \leq \Pr(G_{n,p^*/\omega} \notin \mathcal{G})^{\omega}$$

So
$$\Pr(G_{n,p^*/\omega} \notin \mathcal{G}) \geq 2^{-1/\omega} = 1 - o(1).$$

$\square$

Expected Length of 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.

Let L_n be the length of the minimum spanning tree of K_n.

Theorem

$$\lim_{n\to\infty} E(L_n) = \zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3} = 1.202\cdots$$

Proof

Suppose that $T = T(\{X_e\})$ is the minimum spanning tree, unique with probability one.

$$L_n = \sum_{e \in T} X_e$$

$$= \sum_{e \in T} \int_{p=0}^{1} \mathbb{1}_{p \leq X_e} \, dp$$

$$a = \int_0^1 \mathbb{1}_{x \leq a} \, dx$$

$$= \int_{p=0}^{1} \sum_{e \in T} \mathbb{1}_{p \leq X_e} \, dp$$

$$= \int_{p=0}^{1} \sum_{e \in T} \mathbb{1}_{p \le X_e} \, dp$$

$$= \int_{p=0}^{1} \left| \{ e \in T : X_e \ge p \} \right| dp$$

$$= \int_{p=0}^{1} \left(\kappa(G_p) - 1 \right) dp .$$

$\kappa(G)$ = # components of G.

G_p = graph induced by edge e with $X_e \le p$

$\equiv G_{n,p}$

$$E(L_n) = \int_{p=0}^{1} \left(E(\kappa(G_p)) - 1 \right) dp .$$

So we estimate $E(\kappa(G_p))$.

(i) $p \geq \frac{6\log n}{n} \implies E(\kappa(G_p)) = 1 + o(1)$.

$$E(\kappa(G_p)) \leq 1 + n\Pr(G_p \text{ is not connected})$$

$$\leq 1 + n \sum_{k=1}^{\frac{1}{2}n} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}$$

$$\leq 1 + n^2 \sum_{k=1}^{\frac{1}{2}n} \left(ne \cdot \frac{6\log n}{n} \cdot \frac{1}{n^3} \right)^k$$

$$= 1 + o(1).$$

$$E(L_n) = \int_{p=0}^{\frac{6\log n}{n}} E(\kappa(G_p) - 1)\,dp + O(1)$$

$$= \int_{p=0}^{\frac{6\log n}{n}} E(\kappa(G_p))\,dp + O(1).$$

Write

$$\kappa(G_p) = \sum_{k=1}^{(\log n)^2} \underbrace{A_k}_{\substack{\#\text{ of } k\text{-components} \\ \text{that are trees}}} + \sum_{k=1}^{(\log n)^2} \underbrace{B_k}_{\substack{\#\text{ of } k\text{-components} \\ \text{that are not trees}}} + \overbrace{C}^{\substack{\#\text{ components of} \\ \text{size} \geq (\log n)^2.}}$$

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

$$= (1 + o(1)) \; n^k \cdot \frac{k^{k-2}}{k!} p^{k-1} (1-p)^{kn}$$

uniform in p

$$E(B_k) \leq \binom{n}{k} k^{k-2} \binom{k}{2} p^k (1-p)^{k(n-k)}$$

$$\leq (1+o(1)) \left(np e^{1-np} \right)^k$$

$$\leq (1+o(1))$$

$$C \leq \frac{n}{(\log n)^2}.$$

$$\int_{p=0}^{\frac{6\log n}{n}} \sum_{k=1}^{(\log n)^2} E(B_k)\, dp \leq \frac{6\log n}{n} \cdot (\log n)^2 \cdot (1 + o(1))$$

$$= o(1).$$

$$\int_{p=0}^{\frac{6\log n}{n}} C\, dp \leq \frac{6\log n}{n} \cdot \frac{n}{(\log n)^2} = o(1)$$

$$E(L_n) =$$

$$o(1) + (1 + o(1)) \sum_{k=1}^{(\log n)^2} n^k \cdot \frac{k^{k-2}}{k!} \int_{p=0}^{\frac{6\log n}{n}} p^{k-1} (1-p)^{kn}\, dp$$

$$\sum_{k=1}^{(\log n)^2} n^k \cdot \frac{k^{k-2}}{k!} \int_{p=\frac{6\log n}{n}}^{1} p^{k-1} (1-p)^{kn} \, dp$$

$$\leq \sum_{k=1}^{(\log n)^2} n^k e^k \int_{p=\frac{6\log n}{n}}^{1} n^{-6k} \, dp \quad = o(1).$$

$E(L_n) =$

$$o(1) + (1+o(1)) \sum_{k=1}^{(\log n)^2} n^k \cdot \frac{k^{k-2}}{k!} \int_{p=0}^{1} p^{k-1} (1-p)^{kn} \, dp$$

$$= O(1) + (1 + o(1)) \sum_{k=1}^{(\log n)^2} n^k \cdot \frac{k^{k-2}}{k!} \frac{(k-1)!\,(k(n-k))!}{(k(n-k+1))!}$$

$$= O(1) + (1 + o(1)) \sum_{k=1}^{(\log n)^2} n^k k^{k-3} \prod_{i=1}^{k} \frac{1}{k(n-k)+i}$$

$$= O(1) + (1 + o(1)) \sum_{k=1}^{(\log n)^2} \frac{1}{k^3}$$

$$= O(1) + (1 + o(1)) \sum_{k=1}^{\infty} \frac{1}{k^3} .$$

$\square$

Random Graphs with a Fixed Degree Sequence.

Let $\underline{d} = (d_1, d_2, \ldots, d_n)$

where $d_1 + d_2 + \cdots + d_n = 2M$ is even.

Let $\mathcal{G}_{n,\underline{d}} = \{$simple graphs with vertex set $[n]$ such that degree $d(i) = d_i,\ i \in [n]\}$

$G_{n,\underline{d}}$ is chosen randomly from $\mathcal{G}_{n,\underline{d}}$.

We assume that $d_1, d_2, \ldots, d_n \geq 1$ and that $\sum d_i(d_i - 1) = \Omega(n)$.

Configuration model

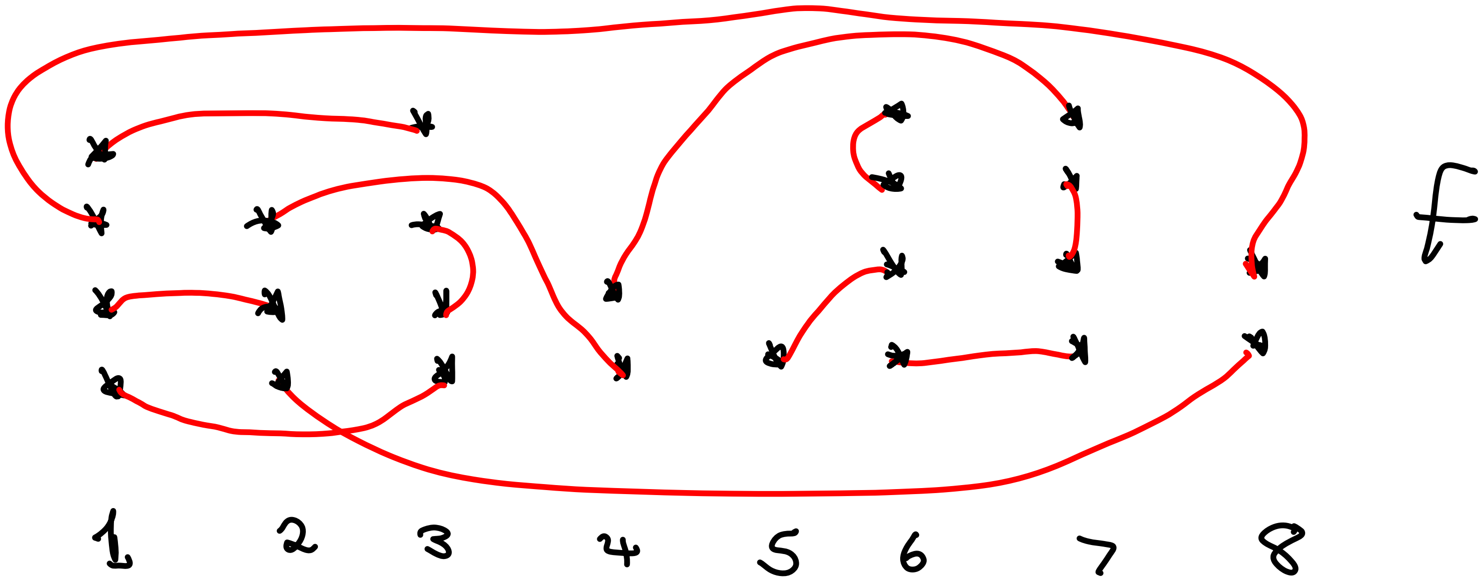
Let $W_1, W_2, \dots, W_n$ be a partition of W, where $|W_i| = d_i$ for $1 \le i \le n$.

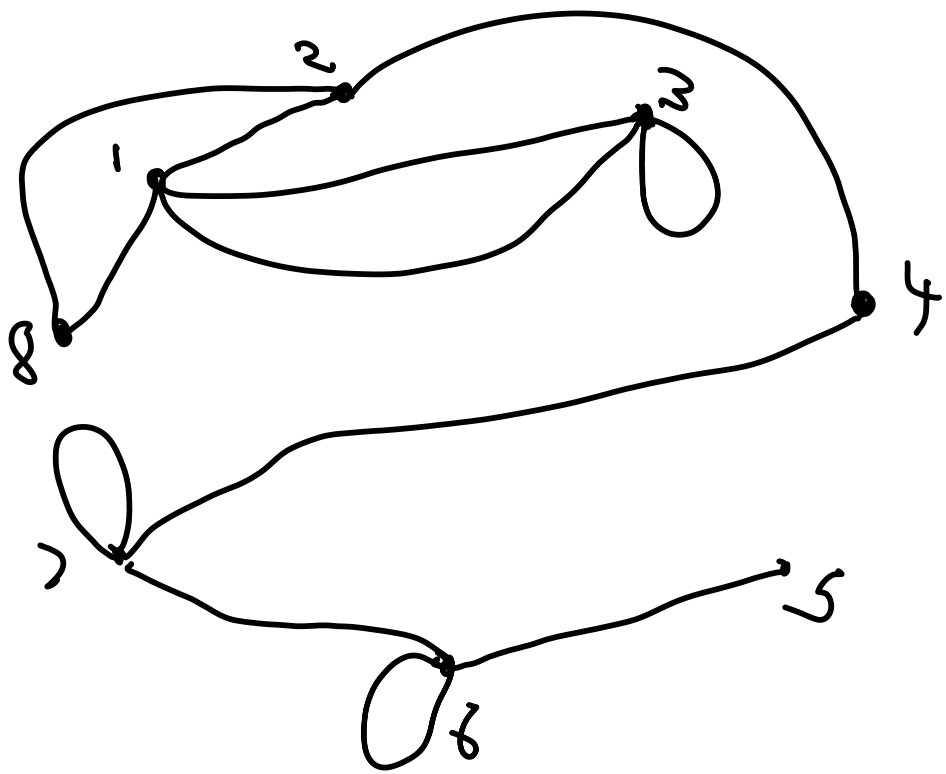
For $x \in W$ define $\phi(x)$ by $x \in W_{\phi(x)}$.

Let F be a partition of W into m pairs (a configuration).

Given F we define the (multi) graph $\gamma(F)$

$$\gamma(F) = \left([n], \{(\phi(x), \phi(y)) : (x, y) \in F\}\right).$$


1 2 3 4 5 6 7 8
F


1
2
3
4
5
6
7
8

Lemma

If $G \in \mathcal{G}_{n,\underline{d}}$, 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 get all F for which $\gamma(F) = G$.

$\square$

Corollary

If F is chosen uniformly at random from Ω (the set of configurations) and $G_1, G_2 \in \mathcal{G}_{n,d}$ then

$$\Pr(\gamma(F) = G_1) = \Pr(\gamma(F) = G_2).$$

So we can choose a random F and accept $\gamma(F)$ iff there are no loops or multiple edges.

The next question is: What is

$$\Pr(\gamma(F) \text{ is simple}) \tag{i}$$

Since

$$|\Omega| = \frac{(2m)!}{m!\, 2^m} \tag{ii}$$

Take d_i "distinct" copies of i for $i = 1, 2, \dots, n$ and take a permutation of these $2m$ symbols. Read off F, pair by pair.

Each distinct F arises in $m!\, 2^m$ ways.

(i) & (ii) will tell us how big is $\underset{\smile}{\mathcal{G}}_{n,d}$.

Alternative Construction of F

Begin

$U = W$; $F = \emptyset$;

For $i = 1, 2, \ldots, m$ **do**

Begin

Choose x arbitrarily from U;

Choose y randomly from $U \setminus \{x\}$;

$F := F \cup \{(x, y)\}$;

$U := U \setminus \{x, y\}$

End

Each F arises with probability $\dfrac{1}{(2m-1)(2m-3)\cdots 1}$

$= |\Omega|^{-1}$.

Let $\Delta = \max\{d_1, d_2, \dots, d_n\}$.

Notation: $F \overset{\text{ran}}{\in} \Omega \equiv$ F is chosen uniformly from Ω.

Lemma

Assume that $\Delta \leq n^{1/6}$ ← not best possible and $F \overset{\text{ran}}{\in} \Omega$. Then whp

(a) $\gamma(F)$ has no double loops

(b) $\gamma(F)$ has $\leq \Delta \log n$ loops

(c) $\gamma(F)$ has no triple edges.

(d) $\gamma(F)$ has no adjacent double edges.

(e) $\gamma(F)$ has $\leq \Delta^2 \log n$ double edges.

Proof

(a)

$\Pr(F \text{ contains a double loop})$

$$\leq \sum_{i=1}^{n} \binom{d_i}{4} \cdot 3 \cdot \left(\frac{1}{2m-3}\right)^2$$

$$\leq n \Delta^4 m^{-2}$$

$$= O(1),$$

(b)

Let $k_1 = \Delta \log n$.

$$\Pr(F \text{ has} \geq k_1 \text{ loops}) \leq O(1) + \sum_{\substack{x_1 + \cdots + x_n = k_1 \\ x_i = 0,1}} \prod_{i=1}^{n} \binom{d_i}{2}^{x_i} \left(\frac{1}{2m - 2k_1} \right)^{k_1}$$

$$\leq O(1) + \left(\frac{\Delta}{2m} \right)^{k_1} \sum_{x_1 + x_2 + \cdots + x_n = k_1} \prod_{i=1}^{n} d_i^{x_i}$$

$$\leq O(1) + \left(\frac{\Delta}{2m} \right)^{k_1} \frac{(d_1 + \cdots + d_n)^{k_1}}{k_1!} \leq O(1) + \left(\frac{\Delta e}{k_1} \right)^{k_1} = O(1).$$

(c)

$$\Pr(F \text{ contains a triple edge})$$

$$\leq \sum_{1 \leq i < j \leq n} \binom{d_i}{3}\binom{d_j}{3} \cdot 6 \cdot \left(\frac{1}{2m-5}\right)^3$$

$$\leq \Delta^4 \left(\sum d_i\right)^2 n^{-3}$$

$$= o(1).$$

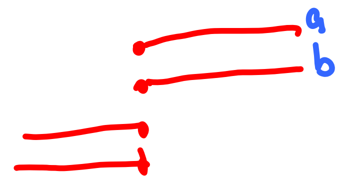
(d)

$$\Pr(F \text{ contains 2 adjacent double edges}) \leq$$

$$\sum_{i=1}^{n} \binom{d_i}{2}^2 \left(\frac{\Delta}{2m-8}\right)^2$$

$$\leq \frac{\Delta^3}{(2m-8)^2} \sum_{i=1}^{n} d_i$$

$$= O(1).$$



$\Pr \leq \dfrac{\Delta}{2m-8}$ that a, b are in same W_i.

(e) Let $k_2 = \Delta^2 \log n$.

Pr (F has $\geq k_2$ double edges)

$$O(1) + \sum_{\substack{x_1 + \cdots + x_n = k_2 \\ x_i = 0,1}} \prod_{i=1}^{n} \left[\binom{d_i}{2} \frac{\Delta}{2m - 4k_2} \right]^{x_i}$$

$$O(1) + \left(\frac{\Delta^2}{m} \right)^{k_2} \sum_{\substack{x_1 + \cdots + x_n = k_2 \\ x_i = 0,1}} \prod_{i=1}^{n} d_i^{x_i}$$

$$\leq O(1) + \left(\frac{\Delta^2}{m} \right)^{k_2} \cdot \frac{(2m)^{k_2}}{k_2!}$$

$$= O(1).$$

$\square$

Switchings

Let now

$\Omega_{i,j} = \{ F \in \Omega :$ F has i loops, j double edges and no double loops or triple edges and no vertex incident with 2 double edges $\}$

Lemma (Switchings) Let $M_1 = 2m$ and $M_2 = \sum_i d_i(d_i - 1)$

For $i \le k_1$ and $j \le k_2$

$$\frac{|\Omega_{i-1,j}|}{|\Omega_{i,j}|} = \frac{2iM_1}{M_2}\left(1 + \tilde{O}\left(\frac{\Delta^3}{n}\right)\right)$$

$$\frac{|\Omega_{0,j-1}|}{|\Omega_{0,j}|} = \frac{4jM_1^2}{M_2^2}\left(1 + \tilde{O}\left(\frac{\Delta^3}{n}\right)\right).$$

Corollary

$$\frac{|\Omega_{0,0}|}{|\Omega|} = (1+o(1))\, e^{-\lambda(\lambda+1)}$$

where $\lambda = \dfrac{M_2}{2M_1}$.

Thus

$$|G_{n,\underline{d}}| \simeq e^{-\lambda(\lambda+1)} \cdot \frac{1}{\prod_{i=1}^{n} d_i!} \cdot \frac{(2m)!}{m!\, 2^m}.$$

Proof

It follows from the switching lemma that $i \le k_1$ and $j \le k_2$ implies

$$\frac{|\Omega_{i,j}|}{|\Omega_{0,0}|} = (1+o(1))\,\frac{\lambda^{i+2j}}{i!\; j!}$$

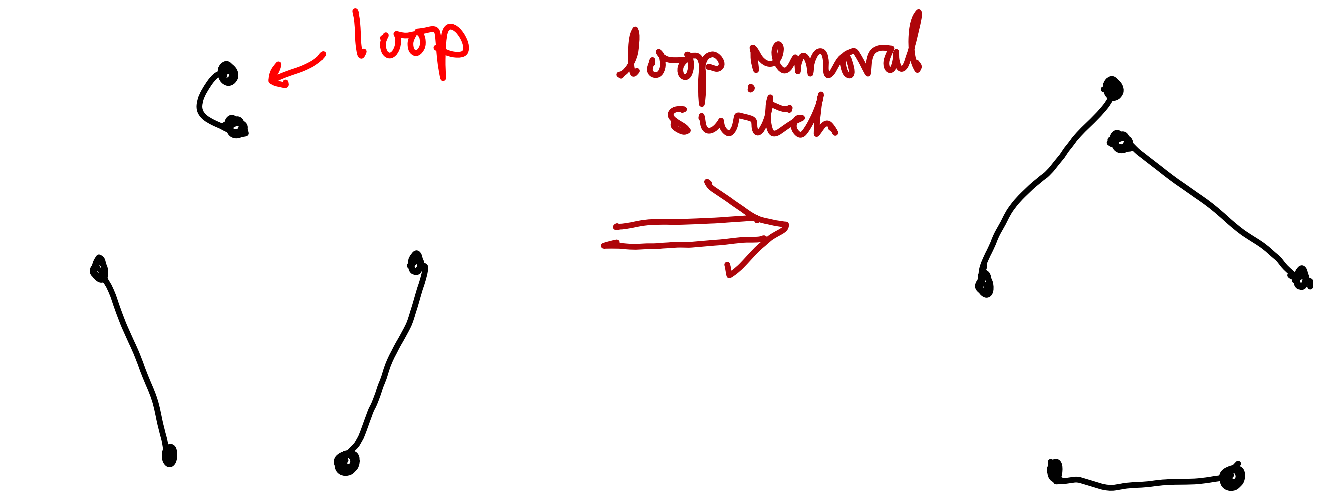
Therefore

$$(1-o(1))\,|\Omega| = (1+o(1))\,|\Omega_{0,0}| \sum_{i=0}^{\lambda_1} \sum_{j=0}^{\lambda_2} \frac{\lambda^{i+2j}}{i!\; j!}$$

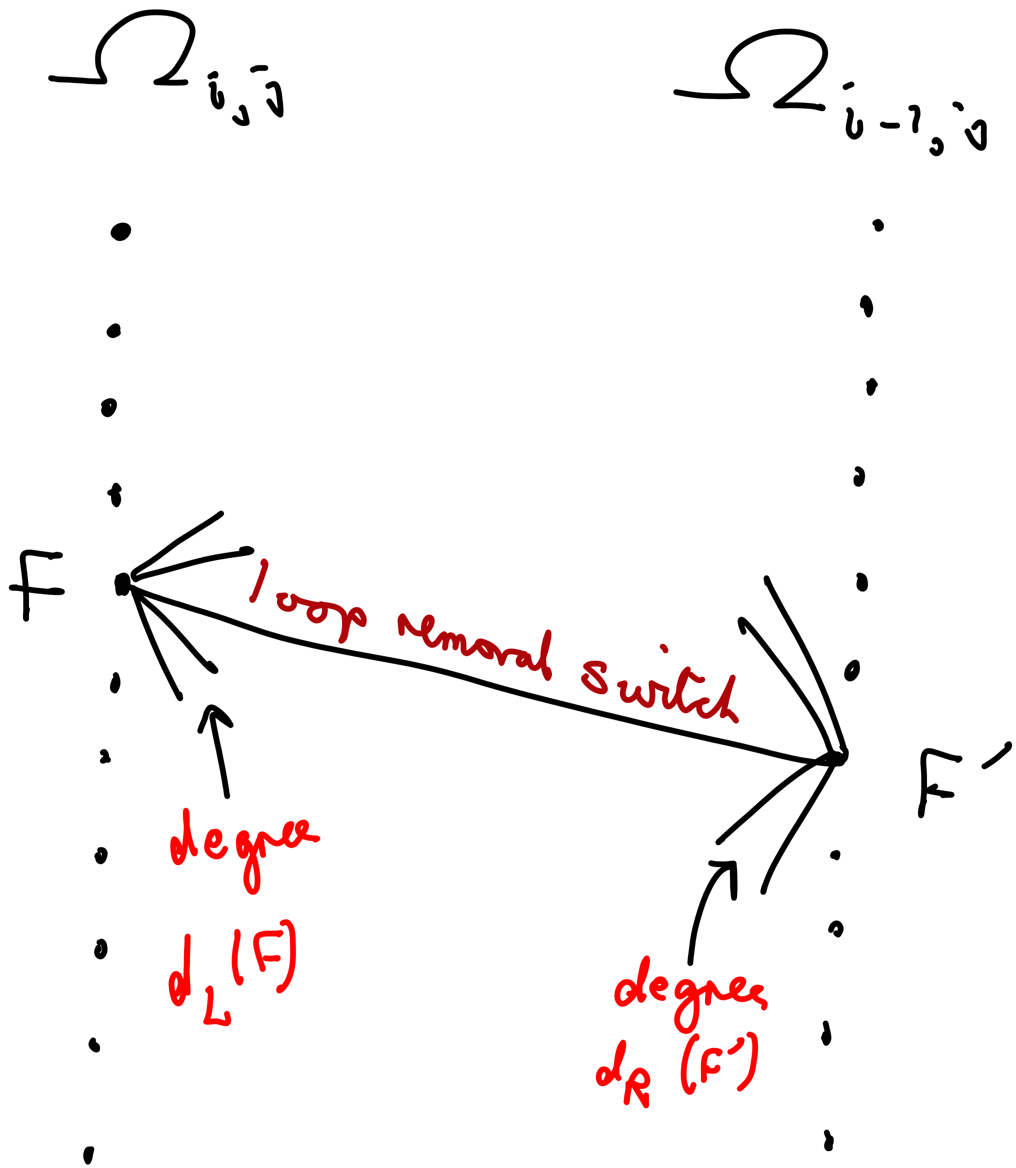
$$= (1+o(1))\,|\Omega_{0,0}|\, e^{\lambda(\lambda+1)}.$$

$\square$

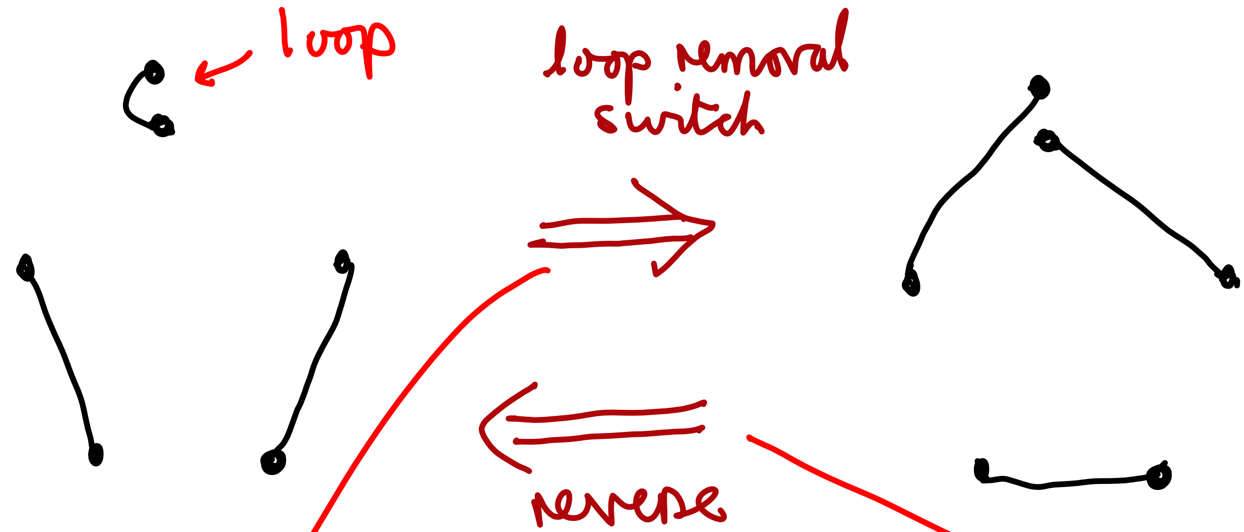
Proof of switching lemma



In general this operation takes a member F of $\Omega_{i,j}$ to a member F' of $\Omega_{i-1,j}$ unless it creates new loops or multiple edges.



$\Omega_{i,j}$
$\Omega_{i-1,j}$
F
F'
loop removal switch
degree $d_L(F)$
degree $d_R(F')$



#choices : $\le i \cdot M_1^2$

$\ge i \cdot M_1^2 - \tilde{O}(i M_1 \Delta^2)$

#choices $\le M_1 M_2 / 2$

$\ge M_1 M_2 / 2 - \tilde{O}((M_1+M_2)\Delta^3)$

Now

$$\sum_{F \in \Omega_{i,j}} d_L(F) = \sum_{F' \in \Omega_{i-1,j}} d_R(F')$$

$\leq i M_1^2 |\Omega_{i,j}|$

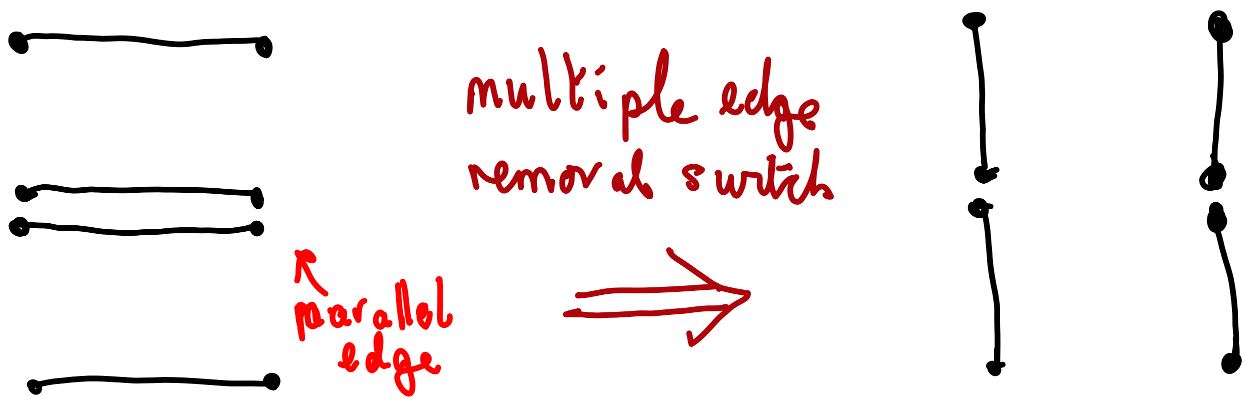
$\geq i M_1^2 |\Omega_{i,j}| \times (1 - \tilde{O}(i\Delta^2/M_1))$

$\leq \frac{1}{2} M_1 M_2 |\Omega_{i-1,j}|$

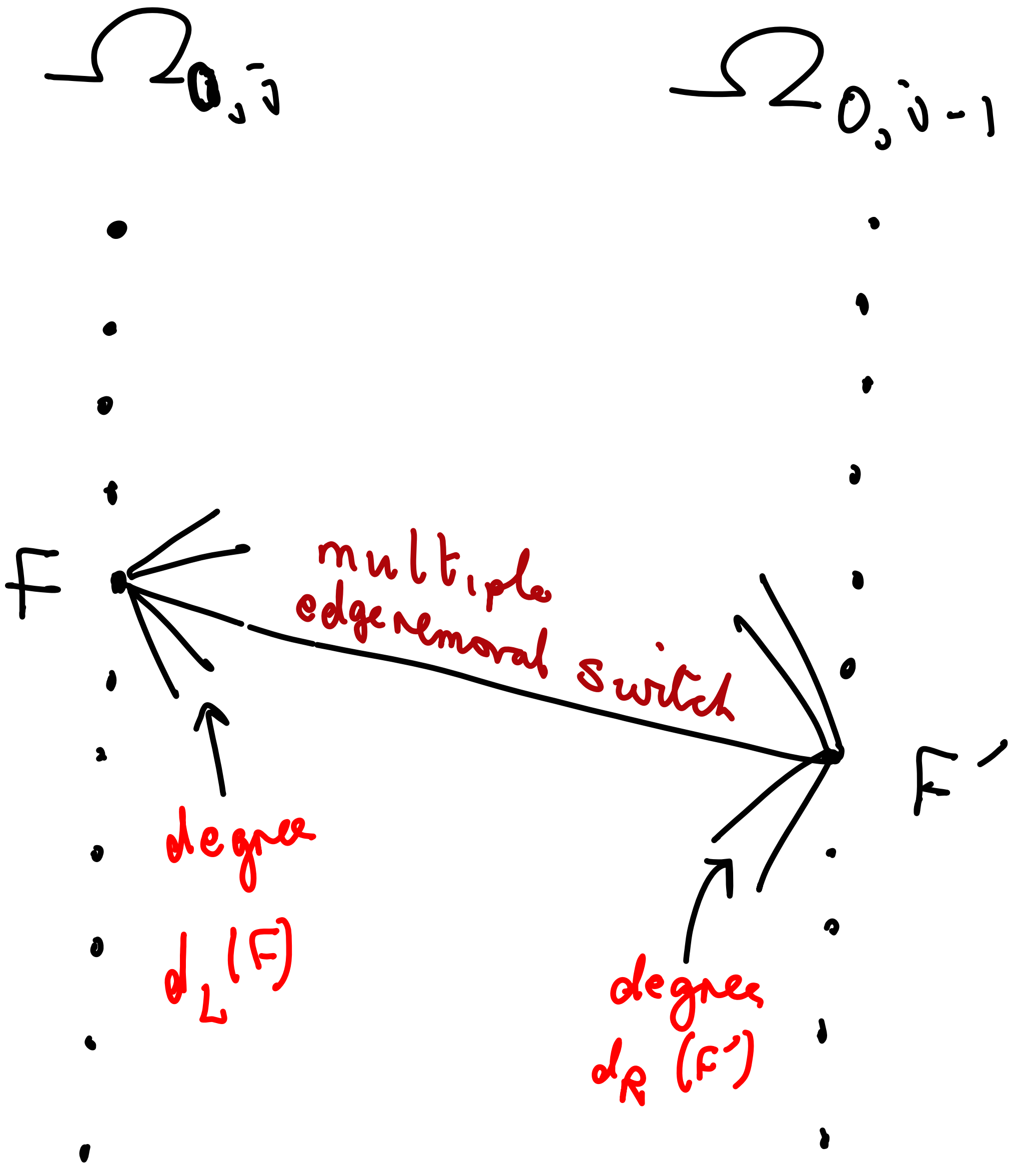
$\geq \frac{1}{2} M_1 M_2 |\Omega_{i-1,j}| \times \left(1 - \tilde{O}\left(\frac{\Delta^3}{M_1} + \frac{\Delta^3}{M_2}\right)\right)$

So

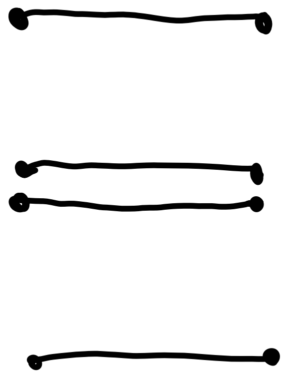
$$\frac{|\Omega_{i-1,j}|}{|\Omega_{i,j}|} = \frac{2iM_1}{M_2}\left(1 + \tilde{O}\left(\frac{\Delta^3}{n}\right)\right)$$



In general this operation takes a member F of $\Omega_{i,j}$ to a member F' of $\Omega_{i,j-1}$ unless it creates new loops or multiple edges.



$\Omega_{0,i}$
$\Omega_{0,i-1}$
F
F'
multiple edge removal switch
degree $d_L(F)$
degree $d_R(F')$



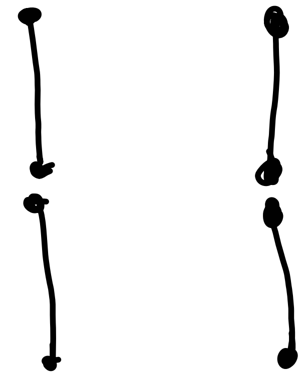
parallel edge

multiple edge removal & switch

$\Longrightarrow$

$\Longleftarrow$

reverse



#choices : $\le j \times M_1^2$

$\ge j \times M_1^2 - \tilde{O}(j M_1 \Delta^2)$

#choices $\le M_2^2/4$

$\ge M_2^2/4 - \tilde{O}(M_2 \Delta^3)$

Now

$$\sum_{F\in\Omega_{0,j}} d_L(F) = \sum_{F'\in\Omega_{0,j-1}} d_R(F')$$

$\le jM_1^2|\Omega_{0,j}|$

$\ge jM_1^2|\Omega_{0,j}|$
$\times\left(1-\tilde{O}(j\Delta^2/M_1)\right)$

$\le \frac{1}{4}M_2^2|\Omega_{0,j-1}|$

$\le \frac{1}{4}M_2^2|\Omega_{0,j-1}|$
$\times\left(1-\tilde{O}\left(\frac{\Delta^3}{M_2}\right)\right)$

So

$$\frac{|\Omega_{0,j-1}|}{|\Omega_{0,j}|} = \frac{4jM_1^2}{M_2^2}\left(1+\tilde{O}\left(\frac{\Delta^3}{n}\right)\right)$$

$\square$

So if $F \overset{\text{ran}}{\in} \Omega$,

$$\Pr(\gamma(F) \text{ is simple}) \approx e^{-\lambda(\lambda+1)}$$

where $$\lambda = \frac{\sum d_i(d_i - 1)}{2 \sum d_i},$$

So for any (multi) graph property $\mathcal{P}$

$$\Pr(G_{n,\underline{d}} \in \mathcal{P}) \leq (1+o(1))\, e^{\lambda(\lambda+1)} \Pr(\gamma(F) \in \mathcal{P})$$

assuming $\Delta \leq n^{1/6}$ [Not best known.]

$$\Pr(G_{n,d} \in \mathcal{P}) \leq (1+o(1))\, e^{\lambda(\lambda+1)} \Pr(\gamma(F) \in \mathcal{P})$$

This is particularly useful if $\lambda = O(1)$

e.g. random r-regular graphs where

r is a constant. Here $\lambda = \frac{r-1}{2}$.

Theorem

Let $G_{n,r}$ denote a random r-regular graph, $r \geq 3$ constant, vertex set $[n]$. Then whp $G_{n,r}$ is r-connected.

Corollary

If n is even then whp $G_{n,r}$ has a perfect matching.

[An r-edge connected, r-regular graph, with n even, has a perfect matching.]

Proof

$$\#\{\text{edges} \subseteq K \cup L \text{ incident with } K\} \geq \frac{rk+l}{2}$$

$l = |L| \leq r-1$

$L = N(K)$

(i) $\underbrace{\frac{r}{r-2}}^{r_0} \leq k = |K| \leq n e^{-10}$

$$\Pr(\exists\, K, L) \leq \sum_{k,l} \binom{n}{k}\binom{n}{l}\binom{rk}{\frac{rk+l}{2}} \left(\frac{r(k+l)}{rn}\right)^{\frac{rk+l}{2}}$$

$$\leq \sum_{k,l} n^{-\left(\frac{r}{2}-1\right)k + l/2} \frac{e^{k+l}}{k^k l^l} 2^{rk} (k+l)^{\frac{rk+l}{2}}$$

$$\le \sum_{k,l} n^{-(\frac{r}{2}-1)k + l/2} \frac{e^{k+l}}{k^k l^l} 2^{rk} (k+l)^{\frac{rk+l}{2}}$$

$$\left(\frac{k+l}{l}\right)^{l/2} \le e^{k/2} \qquad \left(\frac{k+l}{k}\right)^{k/2} \le e^{l/2}$$

$$(k+l)^{rk/2} \le k^{rk/2} e^{lr/2}$$

$$\frac{r-1}{2k} < \frac{r}{2}-1$$

$$k > \frac{r-1}{r-2}$$

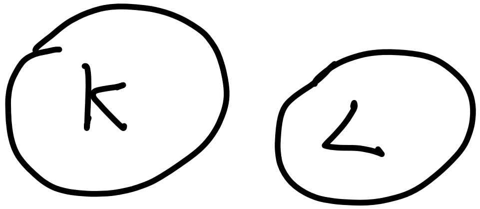
$$\le C_r \sum_{k=r}^{n/10} \sum_{l=0}^{r-1} n^{-(\frac{r}{2}-1)k + l/2} e^{3k/2} k^{(r-2)k/2}$$

constant

$$= C_r \sum_{k=r_0}^{n/10} \sum_{l=0}^{r-1} \left(n^{-(\frac{r}{2}-1)+\frac{l}{2k}} e^{3/2} k^{\frac{r}{2}-1} 2^r \right)^k$$

$$= o(1).$$

(ii) $2 \le k \le \frac{r}{r-2} \le 3$



KvL contains

$\frac{rk}{2} + l \ge k+l+1$ edges

$$\Pr(\exists S : s = |S| \le r-1+3, \text{ contains } s+1 \text{ edges})$$

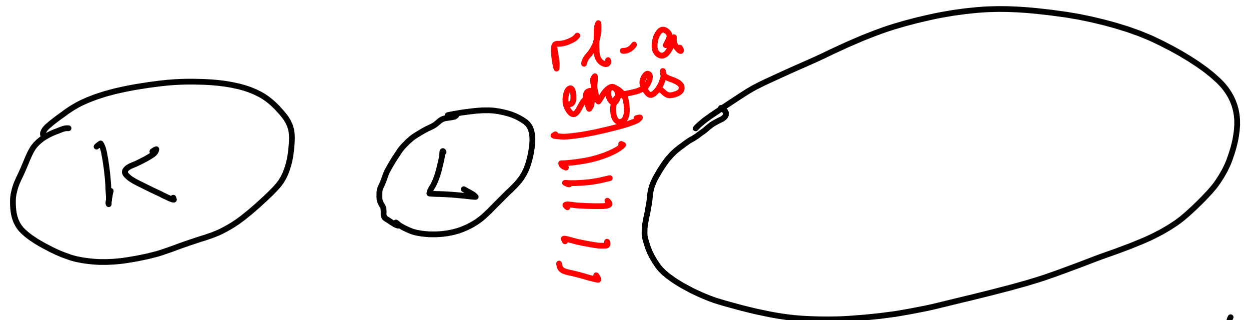
$$\le \sum_{s=4}^{r+2} \binom{n}{s} \binom{rs/2}{s+1} \left(\frac{rs}{rn}\right)^{s+1}$$

$$\le \sum_{s=4}^{r+2} n^{s} \cdot 2^{rs/2} \cdot s^{s+1} \cdot n^{-s-1}$$

$$= o(1).$$

(iii) $ne^{-10} < k \le \frac{n}{2}$

$\phi(2m) = \dfrac{(2m)!}{m!\,2^m} \approx 2^{\frac{1}{2}}\left(\dfrac{2m}{e}\right)^m$



$$\Pr(\exists K, L) \le \sum_{k,l,a} \binom{n}{k}\binom{n}{l}\binom{rl}{a} \frac{\phi(rk+rl-a)\,\phi(r(n-k-l)+a)}{\phi(rn)}$$

$$\le C_r \sum_{k,l,a} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{l}\right)^l \frac{(rk+rl-a)^{rk+rl-a}\,(r(n-k-l)+a)^{r(n-k-l)+a}}{(rn)^{rn}}$$

$$\le C_r \sum_{k,l,a} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{l}\right)^l e^{O(1)} \left(\frac{k}{n}\right)^{rk} \left(1-\frac{k}{n}\right)^{r(n-k)}$$

$$= C_r \sum_{k,l,a} \left(\frac{ne}{k}\right)^k \left(\frac{ne}{l}\right)^l e^{O(1)} \left(\frac{k}{n}\right)^{rk} \left(1-\frac{k}{n}\right)^{r(n-k)}$$

$$\leq C_r \sum_{k,l,a} \left(\left(\frac{k}{n}\right)^{r-1} \cdot e^{1-r/2} \cdot n^{r/k} \right)^k$$

$$= O(1).$$

$$\square$$

Differential Equations Method

Consider the following simple process:

We start with n isolated vertices $1, 2, \dots, n$.

At a general step, we choose a (still) isolated vertex v and add an edge to a randomly chosen w.

Question: how long before there are no isolated vertices?

Let

$$X(t) = \# \text{ isolated vertices after } t \text{ steps.}$$

$$X(0) = n$$

$$E\left(X(t+1) - X(t) \mid X(t)\right) = -1 - \frac{X(t)}{n-1}. \qquad (*)$$

Now put $t = \tau n$, $0 \leq \tau \leq 1$

and $n x(\tau) = X(t)$.

(*) on p2 suggest that

$$x'(\tau) = -1 - x(\tau)$$

given

$$x(\tau) = 2e^{-\tau} - 1.$$

In which case we would expect that the process ends when $t \approx n \ln 2$.

We now consider the following greedy algorithm for finding an independent set in a graph.

<u>GREEDY</u>

<u>begin</u>

$I \leftarrow \emptyset$; $A \leftarrow V$;

<u>While</u> $A \neq \emptyset$ <u>do</u>

Choose $v \in A$; [Random Choice]

$I \leftarrow I \cup \{v\}$; $A \leftarrow A \setminus (\{v\} \cup N(v))$

<u>od</u>

Output I

<u>end</u>

Greedy produces an independent set.

We begin by studying the likely size of the output, if G is a random r-regular graph.

We use the configuration model of r-regular graphs i.e. $W = W_1 \cup W_2 \cup \cdots \cup W_n$ where $W_i = [(i-1)r+1, ir]$

We will expose the random pairing of W as the algorithm progresses i.e. not before.

If vertex i is placed in the independent set I, then and only then, do we expose the pairs involving W_i.

Let the degree of a vertex j at a general step of the algorithm be the number of exposed pairs involving W_j

Thus a general step of GREEDY involves

(i) Choose a vertex i of degree zero.

(ii) Expose the pairs involving W_i.

Let $t = |I|$ be the number of steps taken so far and let P_t refer to the current set of exposed pairs.

Let $X(t)$ be the number of vertices of degree zero.

The number of vertices in the set chosen by GREEDY is t_0, where $X(t_0) = 0$.

$$E\left(X(t+1)-X(t)\mid P_t\right)=$$

$$-1-\frac{X(t)r}{n-2t}+O\left(\frac{1}{\alpha n}\right) \qquad (*)$$

$v\in I$

assuming $t\le\left(\frac{1}{2}-\alpha\right)n$

We expose r pairs associated with v. For first pair there are still $r(X(t)-1)$ points associated with vertices of degree zero, (excluding v). There are $rn-2rt$ points unpaired altogether. So the probability of pairing with vertex of degree zero is $\frac{r(X(t)-1)}{rn-2rt-1}=\frac{X(t)}{n-2t}+O\left(\frac{1}{n}\right)$.

Repeat r times to get $(*)$.

Putting $t = \tau n$ and $X(t) = n x(\tau)$, this suggests that we solve

$$x'(\tau) = -1 - \frac{r\, x(\tau)}{1-2\tau}$$

$$x(0) = 1.$$

Solution: $x(\tau) = \dfrac{(r-1)(1-2\tau)^{r/2} - (1-2\tau)}{r-2}$

The smallest positive solution to $x(\tau) = 0$ is

$$\tau_0 = \frac{1}{2}\left(1 - \left(\frac{1}{r-1}\right)^{2/(r-2)}\right)$$

and then number of vertices in independent set chosen by GREEDY is whp, $\approx \tau_0 n$.

For the following:

$q_0, q_1, \ldots\ q_t, \ldots q_n \in S$ is a random process.

$H_t = (q_0, q_1, \ldots, q_t)$ is the history to time t.

$X(0), X(1), \ldots X(t), \ldots$ are random variables where

$X(t) = X_t(H_t)$.

$D \subseteq \mathbb{R}^2$ is open and connected and

$$\left(0, \frac{X_0(q_0)}{n}\right) \in S$$

[We can assume

[q_0 is fixed

We further assume

(i) $|X(t)| \leq C_0 n$, $\forall\, t < T_D$ where C_0 is constant.

(ii) $|X(t+1) - X(t)| \leq \beta = \beta(n) \geq 1$, $\forall\, t < T_D$

(iii) $|E(X(t+1) - X(t) \mid H_t) - f(t/n, X(t)/n)| \leq \lambda_0$, $\forall\, t < T_D$

(iv) $f(t, x)$ is continuous and satisfies a Lipschitz condition on $D \cap \{(t,x) : t \geq 0\}$

i.e. $|f(\underset{\sim}{x}) - f(\underset{\sim}{x}')| \leq L \, \|\underset{\sim}{x} - \underset{\sim}{x}'\|_\infty$.

Example 1

$$H_t = (i_1, i_2), (i_3, i_4), \dots (i_{2t-1}, i_{2t}) \quad 1 \le i_k \le n$$

$$X_t(H_t) = n - |\{i_1, i_2, \dots, i_{2t}\}|$$

$$f(t, x) = -1 - x$$

$$D = (-1, 1)^2$$

$c_0 = 1$

$\lambda_0 = \frac{1}{n-1}$

$L = 1$

Example 2

$W = [rn]$

$H_t = (i_1, i_2), (i_3, i_4), \ldots (i_{2t-1}, i_{2t}) \quad 1 \le i_k \le rn$

$X_t(H_t) = n - |\{a : \exists s \text{ s.t. } i_s \in W_a\}|$

$f(t, x) = -1 - \frac{rx}{1-2t}$

$c_0 = 1$

$\lambda_0 = \frac{1}{2n}$

$L = r/2\alpha$

$D = (-1, \frac{1}{2} - \alpha) \times (0, 1)$

Theorem

Suppose $\lambda > \lambda_0$ and C is sufficiently large and $\sigma = \inf\{\tau : (\tau, z(\tau)) \notin D_0 = \{(t,z) \in D : \ell^\infty$ distance of (t,z) to boundary of $D \geq C\lambda\}\}$

Where $z(t)$, $0 \leq \tau \leq \sigma$ be the unique solution to

$$\dot{z}(\tau) = f(\tau, x) \qquad (*)$$

$$z(0) = \frac{X_0(q_0)}{n}$$

With probability $1 - O\left(\frac{\beta}{\lambda}\exp\left(-\frac{n\lambda^3}{\beta^3}\right)\right)$

$$X(t) = n\, z(t/n) + O(\lambda n)$$

uniformly in $0 \leq t \leq \sigma n$.

Proof

Let $\omega = \left\lceil \frac{n\lambda}{\beta} \right\rceil$.

We can assume that $\lambda/\beta \geq n^{-1/3}$ else there is nothing to prove.

We study the concentration of $X(t+\omega) - X(t)$,

So assume that $(t/n, X(t/n)) \in D_0$.

For $0 \leq k \leq \omega$ we have

Note that $\left| \frac{X(t+k)}{n} - \frac{X(t)}{n} \right| \leq \frac{k\beta}{n} \leq 2\lambda$

So $\left\| \underbrace{\left(\frac{t+k}{n}, \frac{X(t+k)}{n} \right)} - \left(\frac{t}{n}, \frac{X(t)}{n} \right) \right\|_\infty \leq 2\lambda$

and so [the bracketed point] is in D, assuming $C \geq 2$.

$$E\left(X(t+k+1) - X(t+k) \mid H_{t+k} \right) =$$

$$f\left(\frac{t+k}{n}, \frac{X(t+k)}{n} \right) + \theta_k =$$

$|\theta_k| \leq \lambda$

$$f\left(\frac{t}{n}, \frac{X(t)}{n} \right) + \psi_k + \theta_k =$$

$|\psi_k| \leq \frac{L\beta k}{n}$

$$f\left(\frac{t}{n}, \frac{X(t)}{n} \right) + \rho$$

where $|\rho| \leq 2L\lambda$.

Now, given H_t, let

$$Z_k = X(t+k) - X(t) - k f\left(t/n, \frac{X(t)}{n}\right) - 2kL\lambda.$$

Then

$$E(Z_k - Z_{k-1} \mid Z_0, \dots Z_{k-1}) \leq 0$$

i.e. $Z_0, Z_1, \dots, Z_w$ is a supermartingale.

Also

$$|Z_k - Z_{k-1}| \leq \beta + \left|f\left(t/n, \frac{X(t)}{n}\right)\right| + 2L\lambda$$

$$\leq K_0 \beta$$

where $K_0 = O(1)$.

O(1) by continuity and boundedness of S.

So, conditional on H_β,

$$\Pr\left(X(t+w) - X(t) - w f(t/n, X(t/n)) \geq 2Lw\lambda + K_0 \beta \sqrt{2\alpha w} \right) \leq e^{-\alpha}.$$

Similarly,

$$\Pr\left(X(t+w) - X(t) - w f(t/n, X(t/n)) \leq -2Lw\lambda - K_0 \beta \sqrt{2\alpha w} \right) \leq e^{-\alpha}.$$

Here we produce a supermartingale or equivalently consider $-X(t)$.

Thus

$$\Pr\left(\left|X(t+w) - X(t) - w\, f(t/n, X(t/n))\right| \geq \overbrace{2Lw\lambda + K_0 \beta \sqrt{2\alpha w}}^{err}\right) \leq 2e^{-\alpha}.$$

We will choose

$$\alpha = \frac{n\lambda^3}{\beta^3}$$

so that $w\lambda$ and $\beta\sqrt{2\alpha w}$ are both $\Theta(n\lambda^2/\beta)$ giving

$$err \leq K_1 \frac{n\lambda^3}{\beta}$$

Now let $k_i = i\,w$ for $i = 0, 1, \ldots, i_0 = \lfloor \sigma n / w \rfloor$.

We will show by induction that

$$\Pr\left(\exists j \leq i : |X(k_j) - z(k_j/n)\,n| \geq B_j\right) \leq 2 i e^{-\alpha}$$

where

$$B_j = B\left(\left(1 + \frac{Lw}{n}\right)^j - 1\right)\frac{n\lambda^2}{\beta}$$

and where B is another constant.

The induction begins with $z(0) = \dfrac{X(0)}{n}$.

Note that $B_{i_0} = O\left(\dfrac{n\lambda^3}{\beta}\right) = O(\lambda n)$.

Now write

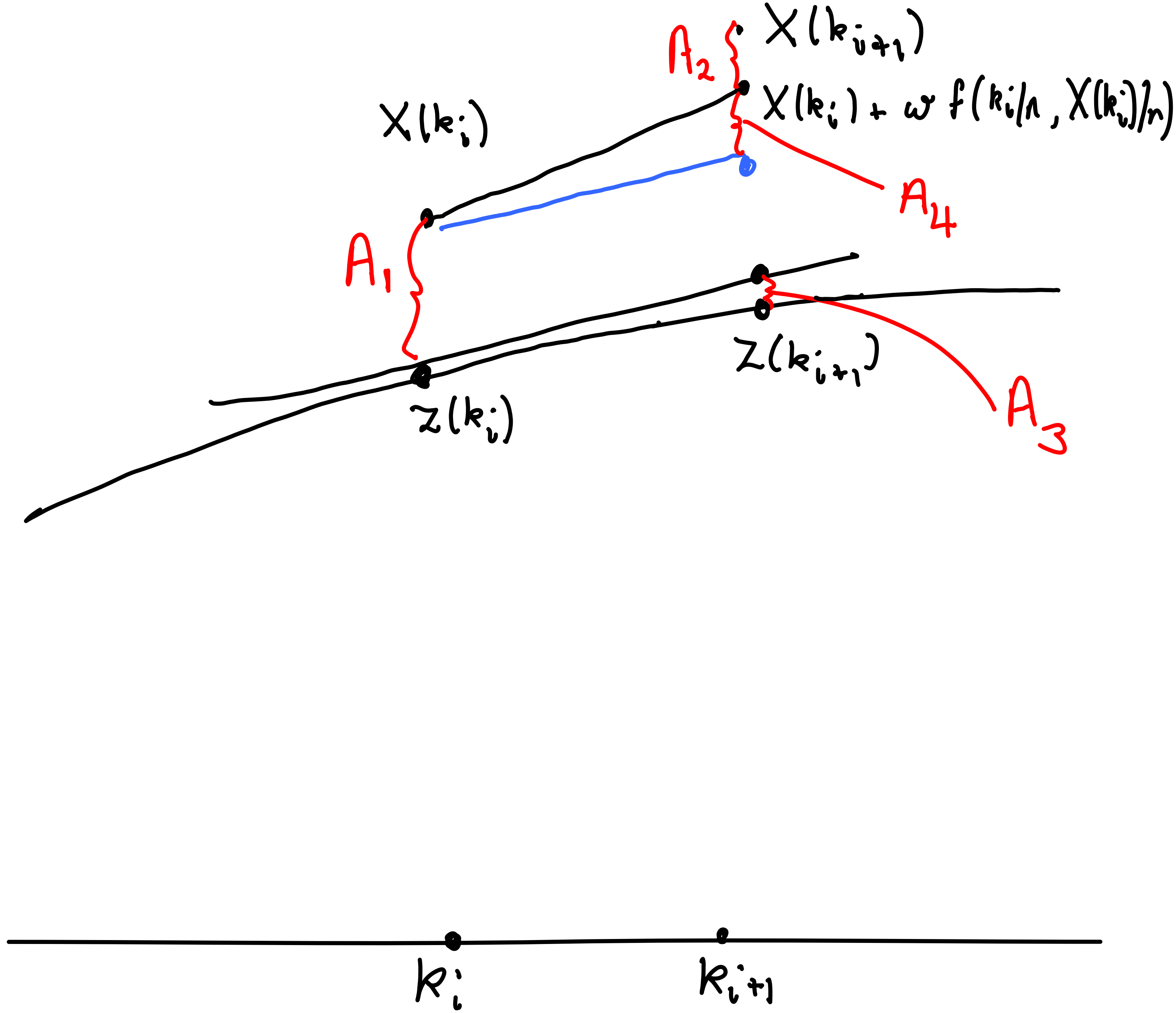
$$|X(k_{i+1}) - z(k_{i+1}/n)\,n| = |A_1 + A_2 + A_3 + A_4|$$

$$A_1 = X(k_i) - z(k_i/n)\,n$$

$$A_2 = X(k_{i+1}) - X(k_i) - w f(k_i/n, X(k_i)/n))$$

$$A_3 = w z'(k_i/n) + z(k_i/n)\,n - z(k_{i+1}/n)\,n$$

$$A_4 = w f(k_i/n, X(k_i)/n) - w z'(k_i/n)$$



$X(k_{i+1})$
$X(k_i)$
$X(k_i) + w\, f(k_i/n, X(k_i)/n)$
A_1
A_2
A_3
A_4
$Z(k_i)$
$Z(k_{i+1})$
k_i
k_{i+1}

$$A_1 = X(k_i) - Z(k_i/n)\,n$$

The induction gives

$$|A_1| \leq B_i .$$

$$A_2 = X(k_{i+1}) - X(k_i) - w f(k_i/n, X(k_i/n))$$

$$|A_2| \le K_1 \frac{n\lambda^3}{\beta}$$

with probability $1 - 2e^{-\alpha}$.

$$A_3 = w z'(k_i/n) + z(k_i/n)\, n - z(k_{i+1}/n)\, n$$

$$|A_3| \le L \frac{w^2}{n^2} \cdot n = L \frac{w^2}{n} \le 2Ln \frac{\lambda^3}{\beta^2}$$

$$A_4 = \omega f(k_i/n, X(k_i)/n) - \omega Z'(k_i/n)$$

$$|A_4| \le \frac{\omega L A_1}{n} \le \frac{\omega L}{n} B_i .$$

Thus, for some $B > 0$,

$$B_{i+1} \le |A_1| + |A_2| + |A_3| + |A_4|$$

$$\le \left(1 + \frac{\omega L}{n}\right) B_i + B n \lambda^2/\beta .$$

Finally consider $k_i \le t < k_{i+1}$.
From "time" k_i to t, the change in X and nZ is at most $\omega\beta = O(n\lambda)$.

$\square$

The above proof generalises easily to the case where

(i) $X(t)$ is replaced by $X_1(t), X_2(t), \ldots X_a(t)$ where $a = O(1)$.

(ii) Condition (iii) on P11 holds with probability $1-\gamma$.

This adds $O(n\gamma)$ to the error probability.

We simply condition on (iii) always holding.

Eigenvalues of Random Graphs

Theorem

Suppose $(\ln n)^5 \leq np \leq n - (\ln n)^5$.
Let A denote the adjacency matrix of $G_{n,p}$. Let the eigenvalues of A be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then whp

(i) $\lambda_1 \approx np$

(ii) $|\lambda_i| \leq 2(\log n)^2 \sqrt{np(1-p)}$ $\qquad 2 \leq i \leq n$.

With more work $2(\log n)^2$ can be replaced by $2 + o(1)$.

Main Lemma

Let J be the all 1's matrix and $M = pJ - A$. Then whp

$$\|M\| \leq 2(\log n)^2 \sqrt{np(1-p)}$$

$$\|M\| = \max_{|x|=1} |Mx| = |\lambda_1(M)|$$

We first show that the lemma implies the theorem.

Let $\underline{e}$ denote the all 1's vector.

(a) $$|A\underline{e} - np\underline{e}| = |M\underline{e}| \leq \|M\| \cdot |\underline{e}| \leq 2(\log n)^2 n\sqrt{p(1-p)}$$

(b) Now suppose that $|\xi| = 1$ and $\xi \perp \underline{e}$. Then $J\xi = 0$ and

$$|A\xi| = |M\xi| \leq \|M\| \leq 2(\log n)^2 \sqrt{np(1-p)}$$

Now let $|x|=1$ and let $x=\alpha u+\beta y$ where $u=\frac{1}{\sqrt{n}}\mathbf{1}$ and $y\perp \underline{e}$ and $|y|=1$. Then

$$|Ax|\leq|\alpha|\,|Au|+|\beta|\,|Ay|$$

We have

$$|Au|=\frac{1}{\sqrt{n}}|A\underline{e}|\leq\frac{1}{\sqrt{n}}\left(np|\underline{e}|+\|M\|\cdot|\underline{e}|\right)$$
$$\leq np+2(\log n)^2\sqrt{np(1-p)}$$

$$|Ay|\leq 2(\log n)^2\sqrt{np(1-p)}$$

Thus

$$|Ax|\leq|\alpha|np+2(|\alpha|+|\beta|)(\log n)^2\sqrt{p(1-p)}$$
$$\leq np+3(\log n)^2\sqrt{p(1-p)}\,.$$

This implies that $\lambda_1 \leq (1+o(1))\,np$

But

$$|Au| \geq |(A+M)u| - |Mu|$$

$$= |pJu| - |Mu|$$

$$\geq np - 2(\log n)^2\sqrt{np(1-p)}$$

implying $\lambda_1 \geq (1-o(1))\,np$.

Now

$$\lambda_2 = \min_{\eta} \max_{0 \neq \xi \perp \eta} \frac{|A\xi|}{|\xi|}$$

$$\leq \max_{0 \neq \xi \perp e} \frac{|A\xi|}{|\xi|}$$

$$\leq 2(\log n)^2 \sqrt{np(1-p)}$$

$$\lambda_n = \min_{|\xi|=1} \xi^T A \xi \geq \min_{|\xi|=1} \xi^T A \xi - p \overbrace{\xi^T J \xi}^{\geq 0}$$

$$= \min_{|\xi|=1} -\xi^T M \xi \geq -\|M\| \geq -2(\log n)^2 \sqrt{np(1-p)}$$

Proof of Main Lemma

Putting $\hat{M} = M - pI_n$ (zero on diagonal)
we see that
$$\|M\| \leq \|\hat{M}\| + \|pI_n\| = \|\hat{M}\| + p$$
and so we bound $\|\hat{M}\|$.

Letting m_{ij} denote (i,j) entry of $\hat{M}$ we have

(i) $E(m_{ij}) = 0$

(ii) $Var(m_{ij}) \leq p(1-p) \leftarrow \sigma^2$.

(iii) $m_{ij}, m_{i'j'}$ are independent, unless $(i',j') = (j,i)$.

Now let $k \geq 2$ be an even integer.

$$\mathrm{Trace}(\hat{M}^k) = \sum_{i=1}^{n} \lambda_i(\hat{M})^k$$

$$\geq \max\{\lambda_1(\hat{M})^k, \lambda_n(\hat{M})^k\}$$

$$= \|\hat{M}\|^k.$$

We estimate

$$\|\hat{M}\| \leq \mathrm{Trace}(\hat{M}^k)^{1/k}$$

where

$$k = (\log n)^2.$$

$$E\left(\mathrm{Trace}\left(\hat{M}^{k}\right)\right) = \sum_{i_0=1}^{n} \sum_{i_1=1}^{n} \cdots \sum_{i_{k-1}=1}^{n} E\left(m_{i_0 i_1} m_{i_1 i_2} \cdots m_{i_{k-2} i_{k-1}} m_{i_{k-1} i_0}\right)$$

So

$$\|\hat{M}\|^{k} \le \sum_{\rho=2}^{k+1} E_{n,k,\rho}$$

where

$$E_{n,k,\rho} = \sum_{\substack{i_0=1 \\ |\{i_0, i_1, \ldots, i_{k-1}\}|=\rho}}^{n} \sum_{i_1=1}^{n} \cdots \sum_{i_{k-1}=1}^{n} \left| E\left(\prod_{j=0}^{k-1} m_{i_j i_{j+1}} \right) \right|$$

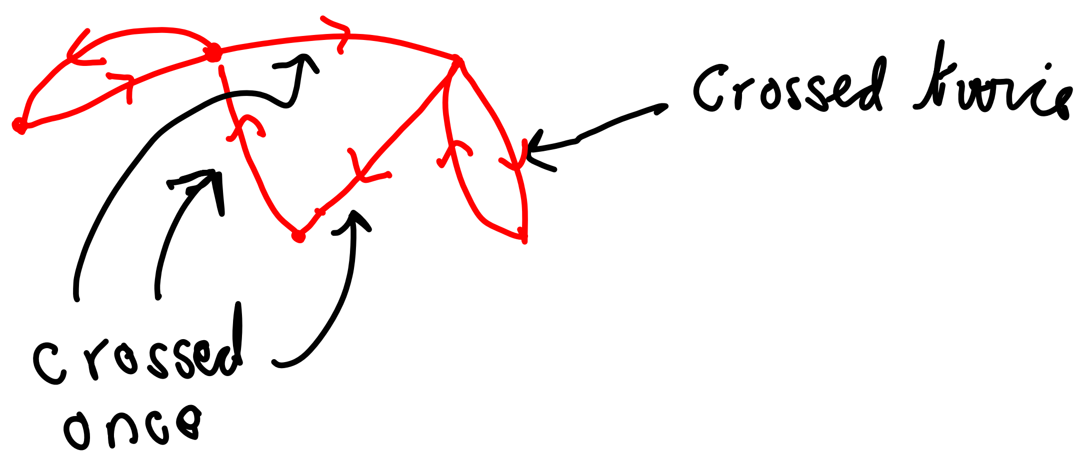
Note that $m_{i,i} = 0$ implies $E_{n,k,1} = 0$.

Each sequence $\underline{i} = i_0 i_1, \ldots, i_{k-1}, i_0$ corresponds to a walk on $W(\underline{i})$ on K_n, with n loops added.

Note that

$$E\left(\prod_{j=0}^{k-1} m_{i_j i_{j+1}}\right) = 0$$

if the walk $W(\underline{i})$ contains an edge that is crossed exactly once



On the other hand, $|m_{i_j}| \leq 1$ and so

$$\left| E\left(\prod_{j=0}^{k-1} m_{i_j i_{j+1}} \right) \right| \leq \sigma^{2(p-1)}$$

if each edge of $W(\underline{i})$ is crossed at least twice and if $|\{i_0, i_1, \ldots, i_{k-1}\}| = p$.

Let $R_{k,p}$ denote the number of (k,p)-walks.

We use the following trivial estimates:

(i) $\rho > \frac{k}{2}+1$ implies $R_{k,\rho} = 0$

(ii) $\rho \leq \frac{k}{2}+1$ implies

$$R_{k,\rho} \leq n^{\rho} \, k^{k}$$

choose the ρ distinct vertices

number of walks of length k

We have

$$\|\hat{M}\|^k \le \sum_{\rho=2}^{\frac{1}{2}k+1} R_{k,\rho}\, \sigma^{2(\rho-1)}$$

$$\le \sum_{\rho=2}^{\frac{1}{2}k+1} n^{\rho} k^{k} \sigma^{2(\rho-1)}$$

$$\le 2 n^{\frac{1}{2}k+1} k^{k} \sigma^{k}.$$

Thus

$$E(\|\hat{M}\|^k) \leq 2n^{\frac{1}{2}k+1} k^k \sigma^k$$

Then

$$\Pr(\|\hat{M}\| \geq 2k\sigma n^{\frac{1}{2}})$$

$$= \Pr(\|\hat{M}\|^k \geq (2k\sigma n^{\frac{1}{2}})^k)$$

$$\leq \frac{E(\|M\|^k)}{(2k\sigma n^{\frac{1}{2}})^k}$$

$$\leq \frac{2n^{\frac{1}{2}k+1}\, k^{k}\, \sigma^{k}}{(2k\sigma n^{1/2})^{k}}$$

$$= \left(\frac{(2n)^{1/k}}{2}\right)^{k}$$

$$= \left(\frac{1}{2} + o(1)\right)^{k}$$

$$= o(1).$$

Preferential Attachment.

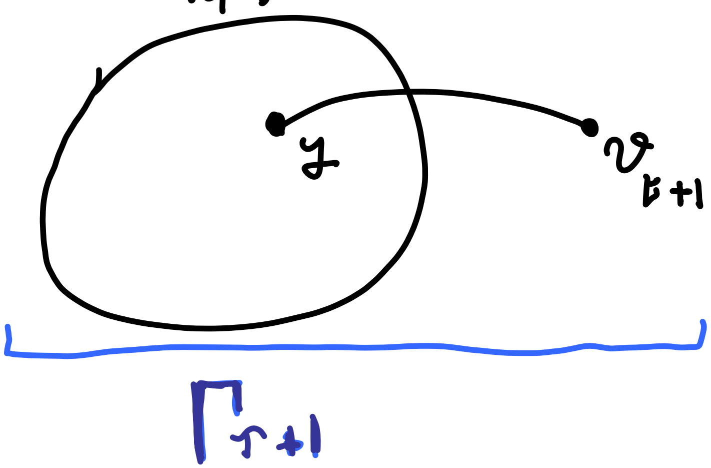
Fix $m > 0$, constant.

Sequence of graphs

$\Gamma_1, \Gamma_2, \ldots, \Gamma_{m-1}, G_1, \Gamma_{m+1}, \Gamma_{m+2}, \ldots, \Gamma_{2m-1}, G_2, \ldots \Gamma_{mt-1}, G_t, \ldots$

$\Gamma_{(m-1)t+1}, \ldots, \Gamma_{mt-1}, G_t$ have vertex set $v_1, v_2, \ldots, v_t$

$\Gamma_\tau, \ \tau \in [mt, mt+m-1]$



$$\Pr(y = v) = \begin{cases} \dfrac{\deg(v, \Gamma_\tau)}{2\tau + 1} & v \neq v_{t+1} \\[2ex] \dfrac{1}{2\tau + 1} & v = v_{t+1} \end{cases}$$

Expected Degree Sequence.

$D_k(t)$ = # of vertices of degree k in G_t, $m \le k = \tilde{O}(t^{1/2})$.

$\bar{D}_k(t) = E(D_k(t))$.

$$E(D_k(t+1) \mid G_t) = D_k(t) + 1_{k=m} + \epsilon(k,t)$$
$$+ m\left(\frac{(k-1)D_{k-1}(t)}{2mt} - \frac{kD_k(t)}{2mt}\right)$$

$$|\epsilon(k,t)| = O\left(\sum_{i=2}^{m} \frac{(k-i)^i D_{k-i}(t)}{(mt)^i}\right) = O\left(\frac{k}{t}\right) = \tilde{O}(t^{-1/2}).$$

$\ell D_\ell(t) \le 2mt$

To account for multiple edges and denominator being $2mt + (\le m)$.

Taking expectations over G_t,

$$\overline{D}_k(t+1) = \overline{D}_k(t) + 1_{k=m} + \tilde{O}(t^{-1/2})$$

$$+ m\left(\frac{(k-1)\overline{D}_{k-1}(t)}{2mt} - \frac{k\overline{D}_k(t)}{2mt}\right)$$

Under the assumption $\overline{D}_k(t) \sim d_k t$ we are led to the recurrence

$$d_k = 1_{k=m} + \left[(k-1)d_{k-1} - k d_k\right]/2$$

or

$$d_k = \frac{k-1}{k+2} d_{k-1} + \frac{1_{k=m}}{k+2} \times 2 \qquad k \geq m$$

$$= 0 \qquad k < m$$

$$d_k = \begin{cases} \dfrac{k-1}{k+2}\, d_{k-1} + \dfrac{1_{k=m}}{k+2} \times 2 & k \geq m \\ 0 & k < m \end{cases}$$

Therefore

$$d_m = \frac{2}{m+2}$$

$$d_k = d_m \prod_{l=m+1}^{k} \frac{l-1}{l+2}$$

$$= \frac{2m(m+1)}{k(k+1)(k+2)}.$$

Theorem

$$|\bar{D}_k(t) - d_k t| = \tilde{O}(t^{1/2})$$

Proof

Let $\Delta_k(t) = \bar{D}_k(t) - d_k t$. Then

$$\Delta_k(t+1) = \frac{k-1}{2t}\Delta_{k-1}(t) + \left(1 - \frac{k}{2t}\right)\Delta_k(t) + \underbrace{\tilde{O}(t^{-1/2})}_{\leq \alpha t^{-1/2}(\log t)^{\beta}}.$$

Now assume inductively on t that

$$|\Delta_k(t)| \leq A t^{1/2} (\log t)^{\beta} \qquad \forall k \geq 0$$

This is trivially true for small t (make A large) and $k < m$.

So

$$|\Delta_k(t+1)| \leq \frac{k-1}{2t}|\Delta_{k-1}(t)| + \left|\left(1-\frac{k}{2t}\right)\Delta_k(t)\right| + \alpha t^{-1/2}(\log t)^\beta$$

$$\leq \frac{k-1}{2t} A t^{1/2}(\log t)^\beta + \left(1-\frac{k}{2t}\right) A t^{1/2}(\log t)^\beta + \alpha t^{-1/2}(\log t)^\beta$$

$$\leq (\log t)^\beta \left(A t^{1/2} + \alpha t^{-1/2}\right)$$

$$(t+1)^{1/2} = t^{1/2}\left(1+\frac{1}{t}\right)^{1/2} \geq t^{1/2} + \frac{1}{3t^{1/2}} \qquad t \text{ large enough}$$

$$\leq (\log(t+1))^\beta \left(A\left[(t+1)^{1/2} - \frac{1}{3t^{1/2}}\right] + \frac{\alpha}{t^{1/2}}\right)$$

$$\leq A(\log(t+1))^\beta (t+1)^{1/2}.$$

$\square$

Concentration

$$\Pr\left(|D_k(t) - \bar{D}_k(t)| \geq u\right) \leq 2\exp\left\{-\frac{u^2}{8mt}\right\}.$$

Proof

Let $Y_1, Y_2, \ldots, Y_{mt}$ be the sequence of choices made in the construction of G_t.

$$Z_i = Z_i(Y_1, Y_2, \ldots, Y_i)$$

$$= E(D_k(t) \mid Y_1, Y_2, \ldots, Y_i).$$

Result follows from

$$|Z_i - Z_{i-1}| \leq 4.$$

Fix $y_1, y_2, \dots, y_i$ and $\hat{y}_i \neq y_i$. We define map

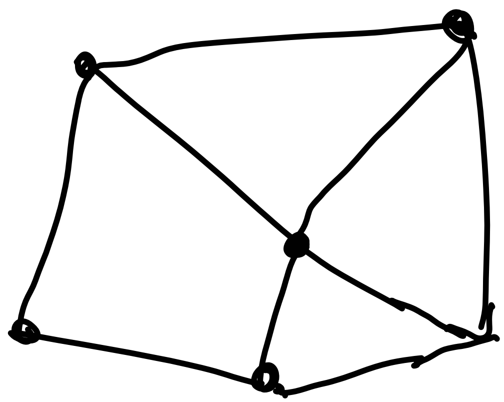
$$y_1, y_2, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_{mb}$$

$\Downarrow$ measure preserving projection Φ

$$y_1, y_2, \dots, y_{i-1}, \hat{y}_i, \hat{y}_{i+1}, \dots, \hat{y}_{mb}$$

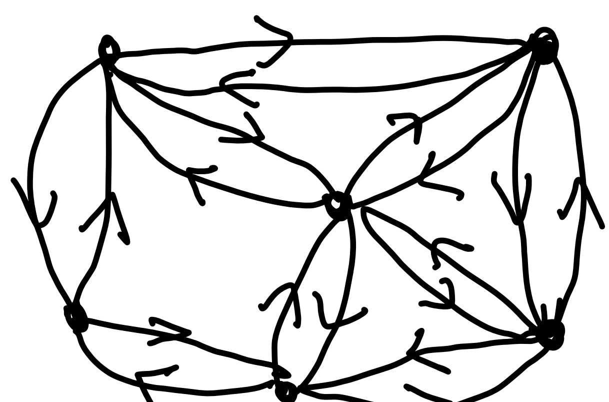
D_k changes by at most 4.

In preferential attachment we can view vertex choices as choices of a random arc

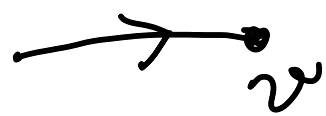


Choose vertex v according to degree

$\Downarrow$



choose random arc



So $Y_1, Y_2, \ldots$ can be viewed as a sequence of arc choices.

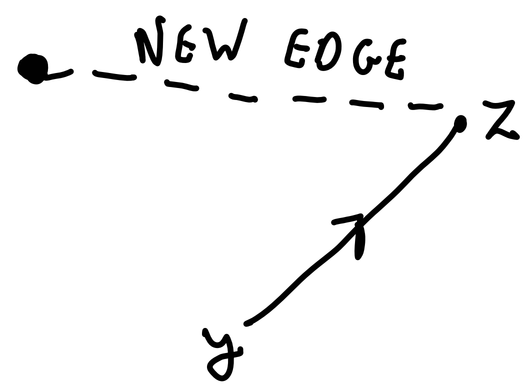
Let

$$Y_i = (x, v) \qquad x > v$$

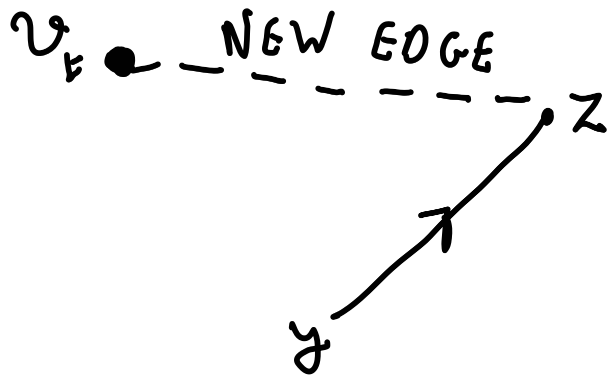
$$\hat{Y}_i = (\hat{x}, \hat{v}) \qquad \hat{x} > \hat{v}$$

$[\; x = \hat{x} \quad \text{if} \quad i \bmod m \neq 1 \;]$

Now suppose $j > i$ and $Y_j = (y, z)$. Then

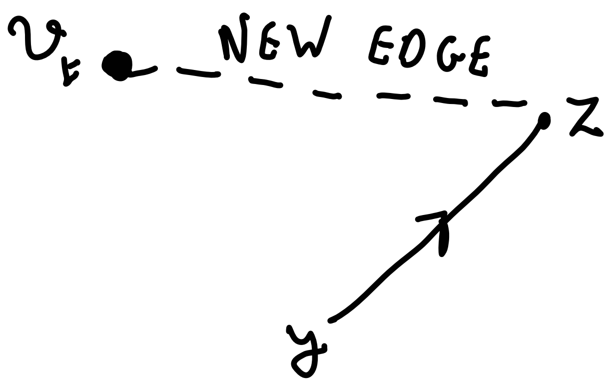


Now suppose $j > i$ and $Y_j = (y, z)$. Then

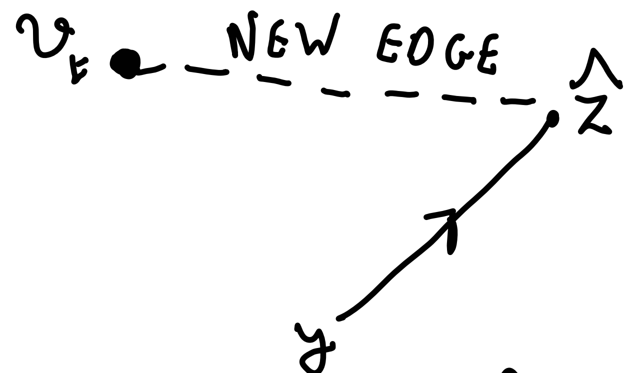


Only $x, \hat{x}, v, \hat{v}$ change degree in transformation.

In $\hat{}$ world



or



If (y, z) exists

else

$z = v \Rightarrow \hat{z} = \hat{v}$

$z = x \Rightarrow \hat{z} = \hat{x}$

Maximum Degree

Fix $k \le t$ and let X_ℓ = degree of v_k in Γ_ℓ.

Lemma

$$\Pr\left(X_{mt} \ge A (t/k)^{1/2} (\log t)^2 \right) = O\left(t^{-A/2}\right).$$

Proof

$$X_{mk} \le 2m.$$

If $0 < \lambda < \frac{1}{\log t}$ then

$$E\left(e^{\lambda X_{\ell+1}} \mid X_\ell\right) = e^{\lambda X_\ell}\left(1 - \frac{X_\ell}{2\ell} + \frac{X_\ell}{2\ell} e^{\lambda}\right)$$

$$\leq e^{\lambda X_\ell}\left(1 - \frac{X_\ell}{2\ell} + \frac{X_\ell}{2\ell}\left(1 + \lambda(1+\lambda)\right)\right)$$

$$\leq e^{\lambda\left(1 + \frac{1+\lambda}{2\ell}\right) X_\ell}$$

So if we define a sequence

$\lambda = \lambda_{m_s}, \lambda_{m_s+1}, \dots \lambda_{m_t}$

where

$$\lambda_{j+1} = \left(1 + \frac{1+\lambda_j}{2j}\right)\lambda_j < 1/\log t$$

then

$$
\begin{aligned}
E\left(e^{\lambda X_{mt}}\right) &\leq E\left(e^{\lambda_{m\ell+1} X_{mt-1}}\right) \\
&\ \ \vdots \\
&\leq E\left(e^{\lambda_{mt} X_{m\ell}}\right) \\
&\leq e^{2m/\log t}.
\end{aligned}
$$

$$\lambda_{j+1} \leq \left(1 + \frac{1 + 1/\log t}{2j}\right) \lambda_j$$

implies that

$$\lambda_{mt} \leq \lambda_{m\ell} \prod_{j=m\ell}^{mt} \left(1 + \frac{1+1/\log t}{2j}\right)$$

$$\leq \lambda_{m\ell} \exp\left\{ \sum_{j=m\ell}^{mt} \frac{1+1/\log t}{2j} \right\}$$

$$\leq 2(t/\ell)^{1/2} \lambda_{m\ell}.$$

So argument works for

$$\lambda_{m\ell} = \frac{(\ell/t)^{1/2}}{2\log t}.$$

This gives

$$E\left(\exp\Big\{\underbrace{\frac{(\ell\backslash t)^{1/2}}{2\log t}}_{\lambda} X_{mt}\Big\}\right) \leq e^{2m/\log t}$$

Finally,

$$\Pr\left(X_{mt} \geq A\,(t/\ell)^{1/2}(\log t)^2 \right)$$

$$\leq e^{-\lambda A (t/\ell)^{\frac{1}{2}}(\log t)^2}\, E(e^{\lambda X_{mt}})$$

$$\leq t^{-A/2}\, e^{2m/\log t}.$$

$\square$

Largest component in $G_{n,p}$ near $p = 1/n$.

Theorem

Let $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$ where $|\lambda| = O(1)$.

Let $C_1, C_2, \ldots$ denote the connected components of $G_{n,p}$ where $|C_1| \geq |C_2| \geq \cdots$. Then

(i) $$E\left(\sum_j |C_j|^2\right) \leq \begin{cases} 3n^{4/3} & \lambda = 0 \\ 4n^{4/3} & 0 < |\lambda| \leq 1/10 \\ n^{4/3}\left[2 + 5|\lambda|^5\right]. & |\lambda| \geq 1/10 \end{cases}$$

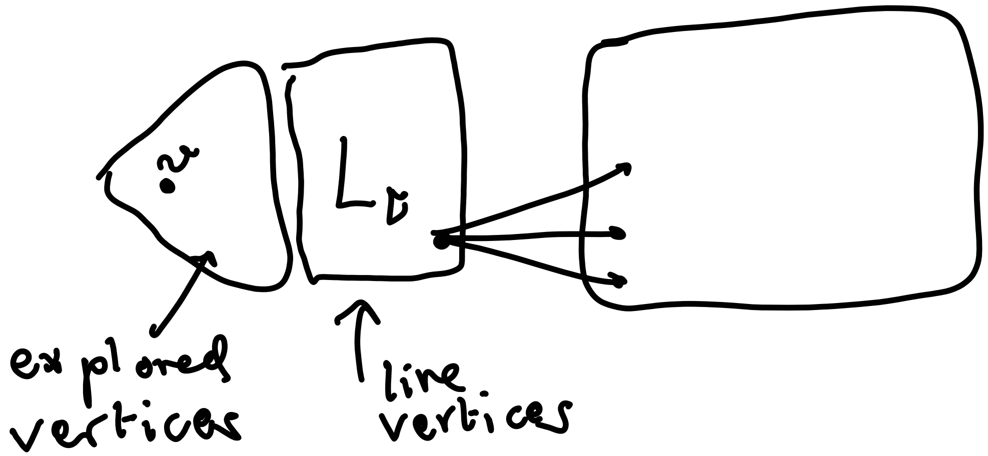
Markov

(ii) $$\Pr(|C_1| \geq An^{2/3}) \leq A^{-2}(4 + 5\sqrt{|\lambda|}).$$

(iii) $$\Pr(|C_1| \leq \delta n^{2/3}) \leq (33 + 2|\lambda|)\,\delta^{8/5}.$$

if δ is sufficiently small and n sufficiently large.

Fix vertex v. In BFS from v we construct sequence of sets



$Y_t = |L_t|$

$Y_0 = 1$

$$Y_t = \begin{cases} Y_{t-1} + \eta_t - 1, & Y_{t-1} > 0 \\ \eta_t & Y_{t-1} = 0 \end{cases}$$

where $\eta_t = B(n - Y_{t-1} - 1, p)$.

$\eta_1, \eta_2, \ldots$ are independent.

Note that if $C(v)$ is the component containing v then

$$|C(v)| = \min\{t : Y_t = 0\}.$$

$$\overset{d}{=} \tau.$$

$$S_t = 1 + \sum_{i=1}^{t} (\xi_i - 1)$$

$\xi_1, \xi_2 \ldots$ are indep. copies of $B(n,p)$.

We couple so that $\eta_1 \le \xi_1$, $\eta_2 \le \xi_2$, ...

It follows that

$$S_t \ge Y_t \quad \text{for} \quad t = 0, 1, 2, \ldots \Upsilon$$

$$E(S_{t+1} - S_t \mid S_t) \cdot np - 1.$$

Let

$$\hat{S}_t = S_t - t\,|np - 1|$$

Then

$$E(\hat{S}_{t+1} \mid \hat{S}_t) = (np - 1) - |np - 1| \leq 0$$

and so $(\hat{S}_t)$ is a super-martingale.

Now fix an integer $H > 0$ and let

$$\gamma = \min\{t \geq 1 : S_t \geq H \text{ or } S_t = 0\}$$

Note that

$$S_\gamma \geq H \Rightarrow Y_\gamma \leq S_\gamma$$

Let $\tau_0 = \min\{t \geq 0 : Y_{\gamma+t} = 0\}$

$$\tau \leq \gamma + \tau_0 \mathbb{1}_{\{S_\gamma \geq H\}}$$

$$[\, S_\gamma = 0 \Rightarrow \tau \leq \gamma \,].$$

$$E(T) \leq E(\gamma) + E(T_0 \mid S_\gamma \geq H)\, P(S_\gamma \geq H)$$

We prove

(i) $$P(S_\gamma \geq H) \leq \frac{1 + E(\gamma)\,|np - 1|}{H}$$

(ii) $$E(\gamma) \leq \frac{H+2}{npq - 4H\,|np-1|}$$

We make sure denominator is positive.

(iii) $$E(T_0 \mid S_\gamma \geq H) \leq \left(\frac{2(H+np)}{p}\right)^{1/2}$$

So,

$$E(T) \leqslant$$

$$\frac{H+2}{npq - 4H|np-1|} + \left(\frac{2(H+np)}{p}\right)^{\frac{1}{2}} \left(\frac{npq - 3H|np-1|}{npq - 4H|np-1|}\right) \cdot \frac{1}{H}$$

We choose H to (approximately) minimise the RHS.

If $\lambda = 0$

$$E(T) \leqslant \frac{H+2}{n-1} + \frac{\sqrt{2n(H+1)}}{H}$$

Put $H = n^{1/3} \implies E(T) \leqslant 3n^{1/3}$.

If $0 < |\lambda| < \frac{1}{10}$ then

$$E(T) \leq 2(H+2) + \frac{\sqrt{(2+o(1))n(H+1)} \times 7}{6H}$$

Putting $H = n^{1/3}$ gives

$$E(T) \leq 4n^{1/3}.$$

If $|\lambda| \geq \frac{1}{10}$ we put $H = \frac{n^{1/3}}{10|\lambda|}$ and then

$$E(T) \leq 2H + \frac{\sqrt{(2+o(1))nH} \times 7}{6H}$$

$$\leq n^{1/3}\left[2 + 5|\lambda|^{1/2}\right].$$

Now write

$$E(T) = E(|C(v)|)$$

$$= \frac{1}{n}\sum_{v=1}^{n} E(|C(v)|)$$

$$= \frac{1}{n} E\Big(\sum_j |C_j|^2\Big)$$

So $E\big(\sum_j |C_j|^2\big) \leq n\, E(T)$.

Main tool [OPTIONAL STOPPING]

Let $Z_0, Z_1, \ldots Z_t, \ldots$ be a random process.

T is a stopping time if the event $\{T = t\}$ depends only on $Z_0, Z_1, \ldots, Z_t$ and not on the future.

Optional Stopping

Suppose T is a stopping time.

(i) (Z_t) is a martingale $\Rightarrow E(Z_T) = E(Z_0)$.

(ii) (Z_t) is a supermartingale $\Rightarrow E(Z_T) \leq E(Z_0)$.

(iii) (Z_t) is a submartingale $\Rightarrow E(Z_T) \geq E(Z_0)$

We must also assume (Z_t) is bounded.

$$1 = E(\hat{S}_0) \geq E(\hat{S}_\gamma) = E(S_\gamma) - E(\gamma)(np-1)^+$$

$$\geq H\,P(S_\gamma \geq H) - E(\gamma)(np-1)^+$$

so

$$P(S_\gamma \geq H) \leq \frac{1 + E(\gamma)|np-1|}{H}$$

Lemma

Given $S_\gamma \geq H$, the conditional distribution of $S_\gamma - H \overset{d}{\leq} B(n,p)$.

Proof

$\xi = B(n,p) = I_1 + I_2 + \cdots + I_n$

Given $\xi \geq r$, $\xi - r \overset{d}{\leq} B(n,p)$.*

Suppose $r = \sum_{j=1}^{J} I_j$ so that $\xi - r$ has distribution $\overset{d}{\leq} B(n-J,p)$.

Conditioned on $\{\gamma = \ell\} \cap \{S_{\ell-1} = H - r\} \cap \{S_\gamma \geq H\}$,

$$S_\gamma - H \overset{d}{=} \xi_\ell - r \overset{d}{\leq} B(n,p).$$

Now average over ℓ, r.

$\square$

* $A \overset{d}{\leq} B$ if $\Pr(A \geq x) \leq \Pr(B \geq x)$, $\forall x$.

Write

$$S_\gamma^2 = H^2 + 2H(S_\gamma - H) + (S_\gamma - H)^2$$

Then, lemma on p10 implies

$$E(S_\gamma^2 \mid S_\gamma \geq H) \leq H^2 + 2Hnp + npq + (np)^2$$

$$\leq H^2 + 3H.$$

Define

$$t \wedge \gamma = \min\{t, \gamma\}.$$

and

$$A_t = S_{t\wedge\gamma}^2 - B(t\wedge\gamma)$$

where

$$B = npq - 2H|1-np|.$$

We claim that

(A_t) is a sub-martingale

$$E\left(S_{t+1}^2 - S_t^2 \mid S_t\right) =$$

$$2E\left(S_t(\xi_{t+1}-1)\right) + E\left((\xi_{t+1}-1)^2\right)$$

$$= 2S_t(np-1) + npq + 1 - np$$

$$\geq \underbrace{npq - 2H|np-1|}_{B}, \quad \forall\, t \leq \gamma.$$

$$E\left(\left[S_{t+1}^2 - B(t+1)\right] - \left[S_t^2 - Bt\right] \,\middle|\, S_t\right) \leq 0, \quad t \leq \gamma.$$

So

$$A_0 \leq E(A_\gamma)$$

or

$$1 \leq E(S_\gamma^2) - BE(\gamma)$$

So

$$1 + BE(\gamma) \leq E(S_\gamma^2) = E(S_\gamma^2 \mid S_\gamma \geq H) P_r(S_\gamma \geq H)$$

$$\leq (H+3)\big(1 + E(\gamma)\,|np-1|\big)$$

So

$$E(\gamma) \leq \frac{H+2}{B - (H+3)|np-1|} \leq \frac{H+2}{npq - 4H|np-1|}$$

We ensure this is positive.

Now consider $\qquad \tau_0 = \min\{t \geq 0 : Y_{\gamma+t} = 0\}$

$$Z_t = Y_{\gamma + t\wedge\tau_0} + \sum_{j=1}^{t\wedge\tau_0} jp$$

If $t < \tau_0$ then $\qquad \sigma = (t+1)\wedge\tau_0$

$$E(Z_{t+1} - Z_t \mid Z_t) = E(\eta_{\gamma+\sigma} + \sigma p)$$

$$= -1 + \left(n - Y_{\gamma+t\wedge\tau_0} - (\gamma + t\wedge\tau_0) + \sigma\right)p$$

$$\leq 0$$

and $Z_{t+1} = Z_t$ if $t \geq \tau_0$.

So (Z_t) is a supermartingale.

$$H + np \geq E(S_\gamma \mid S_\gamma \geq H)$$ Lemma on p10

$$\geq E(Z_0 \mid S_\gamma \geq H)$$ $S_\gamma \geq Y_\gamma$

$$\geq E(Z_{\tau_0} \mid S_\gamma \geq H)$$ Optional Stopping

$$\geq E(\tau_0^2 \mid S_\gamma \geq H)\, p/2$$ Take sum only.

By Cauchy-Schwarz

$$E(\tau_0 \mid S_\gamma \geq H) \leq E(\tau_0^2 \mid S_\gamma \geq H)$$

$$\leq \left(\frac{2(H+np)}{p}\right)^{1/2}$$

Proof of (iii)

Fix $h = An^{1/3}$, $A = O(1)$ to be determined.

Stage 1

$$T_h = \begin{cases} \min\{t \le \frac{n}{8h} : Y_t \ge h\} & \leftarrow \text{set non-empty} \\ \frac{n}{8h} & \text{otherwise} \end{cases}$$

If $Y_{t-1} > 0$ then

$$Y_t^2 - Y_{t-1}^2 = (\eta_t - 1)^2 + 2(\eta_t - 1)\, Y_{t-1}.$$

If $Y_{t-1} \le h$ then

$$E(Y_t^2 - Y_{t-1}^2 \mid Y_{t-1}) \ge (n-t-h)\,pq - 2(t+h)\,p\,h.$$

$$\ge \frac{1}{2}.$$

If $Y_{t-1} = 0$ then $E(Y_t^2 - Y_{t-1}^2) = E(\eta_t^2) \geq \frac{1}{2}$, under these assumptions.

So $Y^2_{t \wedge \tau_h} - \frac{1}{2}(t \wedge \tau_h)$ is a submartingale

and so

$$E(Y^2_{\tau_h}) - \frac{1}{2}\tau_h \geq 0.$$

Lemma on P13 $\Rightarrow$

$$E(Y^2_{\tau_h}) \leq h^2 + 3h \leq 2h^2.$$

So

$$2h^2 \geq E(Y^2_{\tau_h}) \geq \frac{1}{2}E(\tau_h) \geq \frac{T_1}{2}P_r\left(T_h = \frac{n}{8h}\right)$$

or

$$P_r\left(\tau_h = \frac{n}{8h}\right) \leq \frac{32h^3}{n}.$$

$$\tau_0 = \begin{cases} \min\{t \le \delta n^{2/3} : Y_{\tau_h + t} = 0\} & \leftarrow \text{set non-empty} \\ \delta n^{2/3} & \text{otherwise} \end{cases}$$

$$M_t = h - \min\{h, Y_{\tau_h + t}\}.$$

If $0 < M_{t-1} < h$ then

$$M_t^2 - M_{t-1}^2 = (\eta_{\tau_h + t} - 1)^2 + 2(1 - \eta_{\tau_h + t}) M_{t-1}$$

and so

$$E(M_t^2 - M_{t-1}^2 \mid M_{t-1}) \le npq + 2h\left(1 - \left(n - \frac{n}{8h} - \delta n^{2/3}\right)p\right)$$

$$\le 2(1 + A|\lambda|).$$

If $Y_{t-1} \geq h$ then $M_{t-1} = 0$ and $M_t \leq 1$.

So $Z_t = M^2_{t \wedge T_0} - 2(1 + A|\lambda|)(t \wedge T_0)$ is a supermartingale.

Now use P_h, E_h to denote conditioning on $\{Y_{\rho_h} \geq h\}$.

$Z_0 = 0$ and so

$$0 \geq E(Z_{T_0}) = E_h(M^2_{T_0}) - 2(1+A|\lambda|)E(T_0)$$

$$\geq E_h(M^2_{T_0}) - (1+A|\lambda|)\delta n^{2/3}.$$

So

$$P_h(T_0 < \delta n^{2/3}) \leq P_h(M_{T_0} \geq h) \leq \frac{E_h(M^2_{T_0})}{h^2} \leq \frac{(1+A|\lambda|)\delta n^{2/3}}{h^2}$$

implies

So

$$P(T_0 < \delta n^{2/3}) \leq P\left(T_h = \frac{n}{8h}\right) + P_h(T_0 < n^{2/3})$$

$$\leq \frac{32h^3}{n} + \frac{(1+A|\lambda|)\cdot \delta n^{2/3}}{h^2}$$

or

$$P(T_0 < \delta n^{2/3}) \leq 32A^3 + \frac{(1+A|\lambda|)\delta}{A^2}.$$

Putting $A = \delta^{1/5}$ for simplicity, we get

which gives

$$P(T_0 < \delta n^{2/3}) \leq (33 + 2|\lambda|)\,\delta^{3/5}.$$

We finally note that

$$|C_1| < \delta n^{2/3} \implies |C(v)| < \delta n^{2/3}$$

$$\implies T < \delta n^{2/3}$$

$$\implies T_0 < \delta n^{2/3}.$$

Perfect Matchings in Bipartite 2-OUT

$B_{k\text{-out}}$ is a random bipartite graph with vertex partition $X \cup Y$ where $|X| = |Y| = n$.

Each $x \in X$ chooses k random nbrs in Y

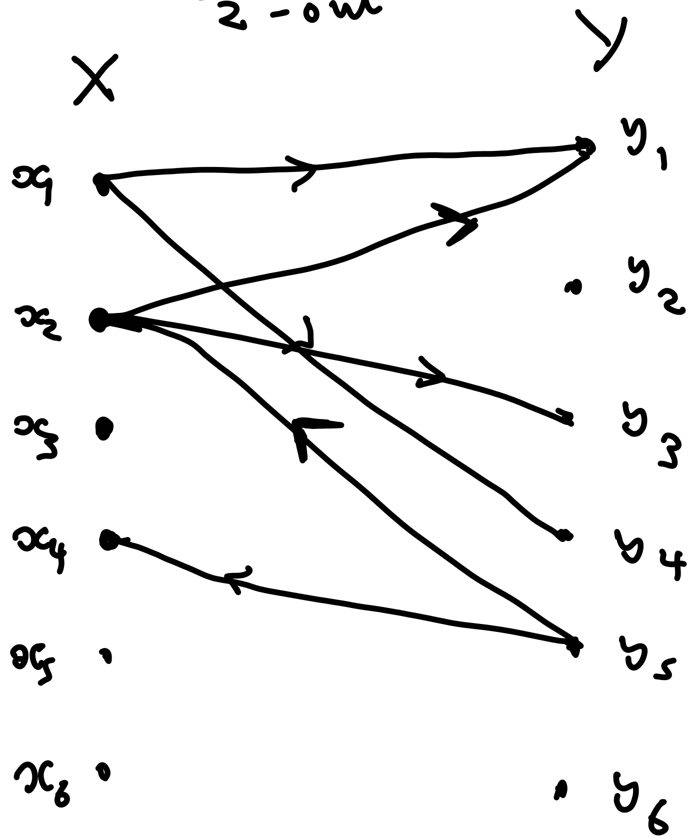
Each $y \in Y$ chooses k random nbrs in X.

Theorem

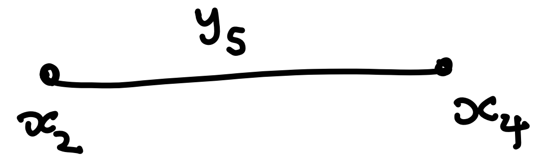
$B_{2\text{-out}}$ has a perfect matching whp.

Algorithmic Proof

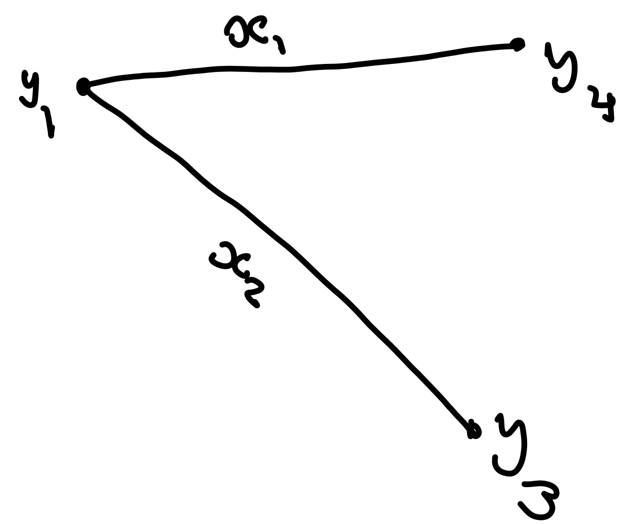
B_2-out



G_1: n random edges

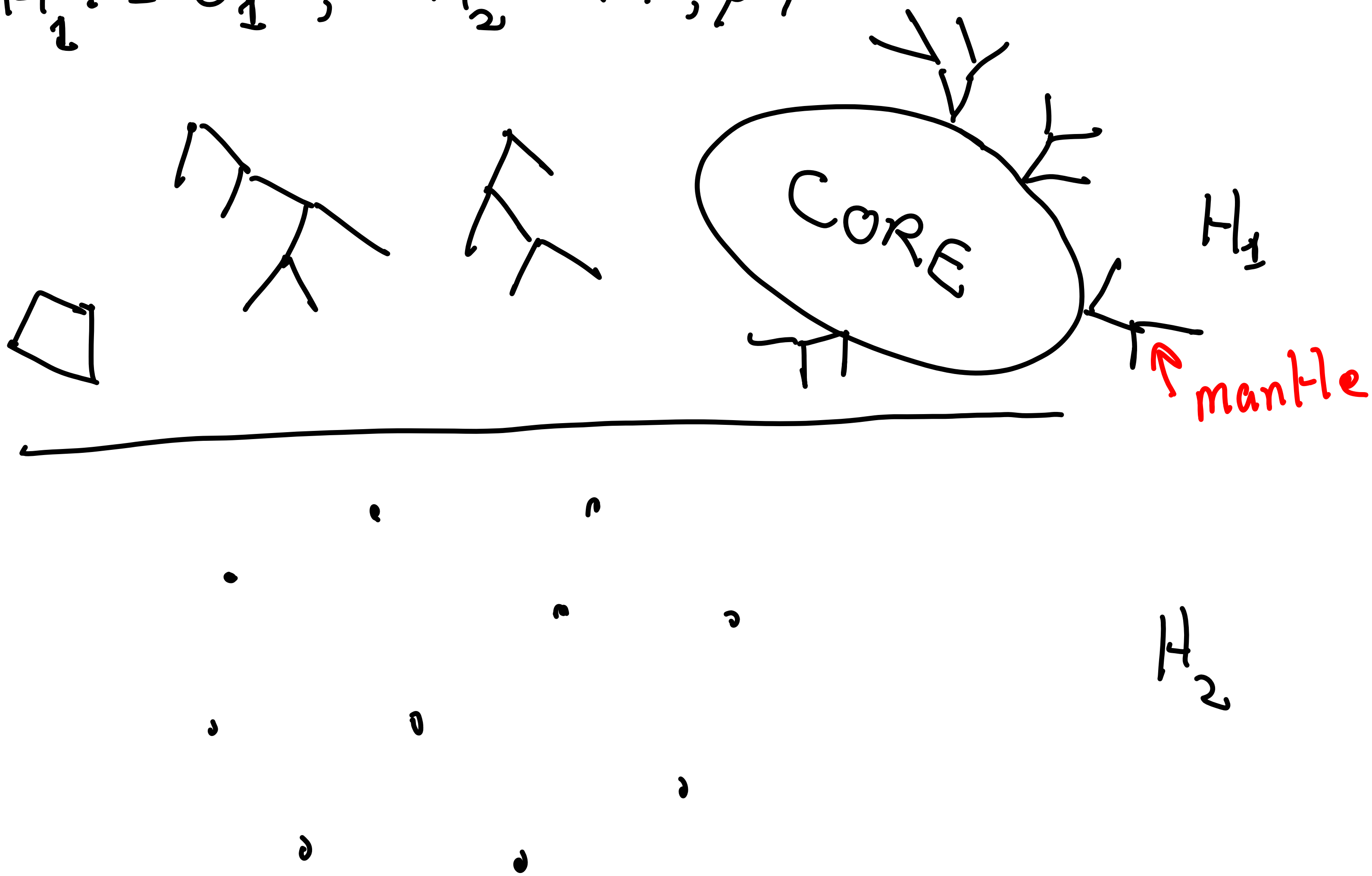


G_2: n random edges

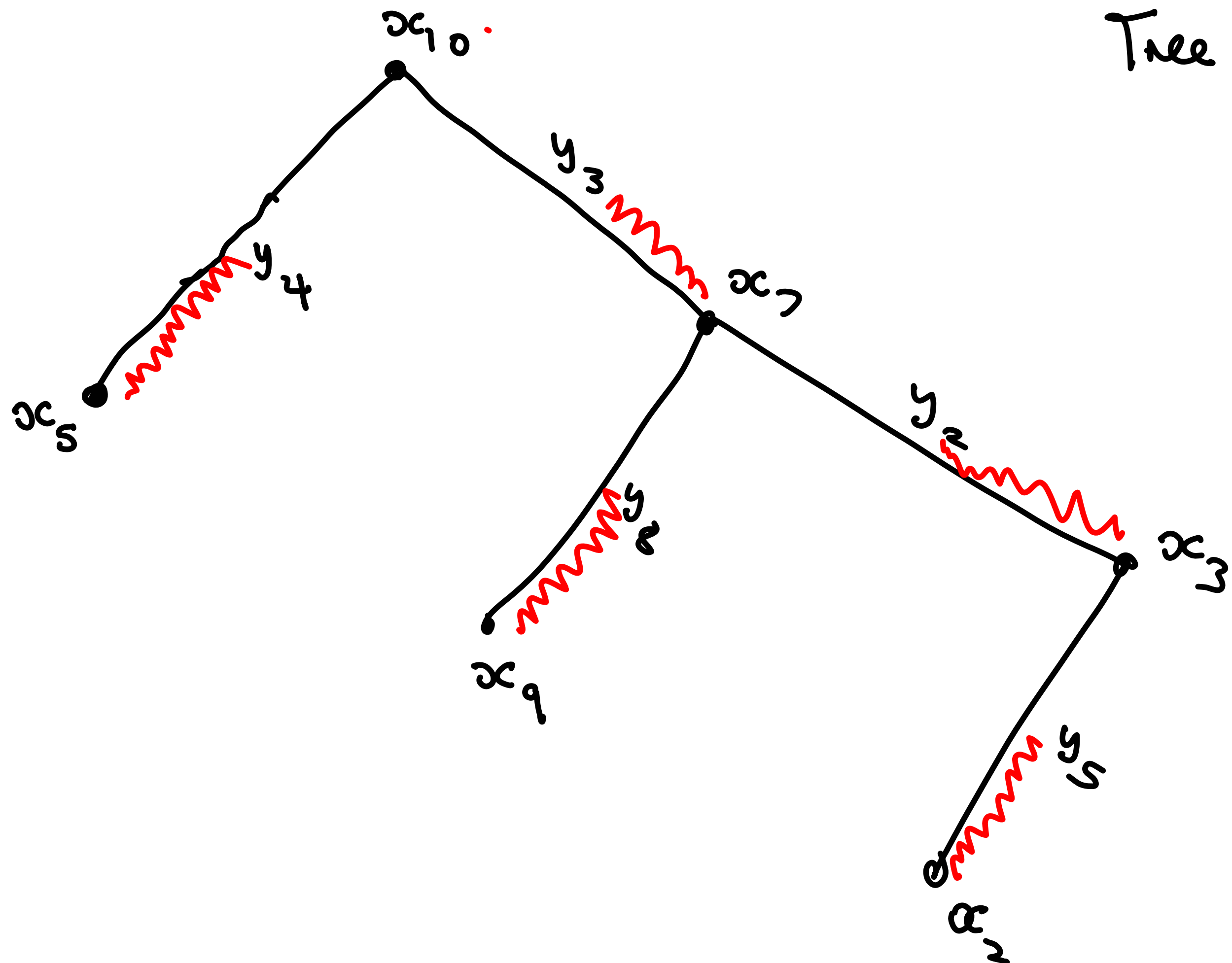


Algorithm

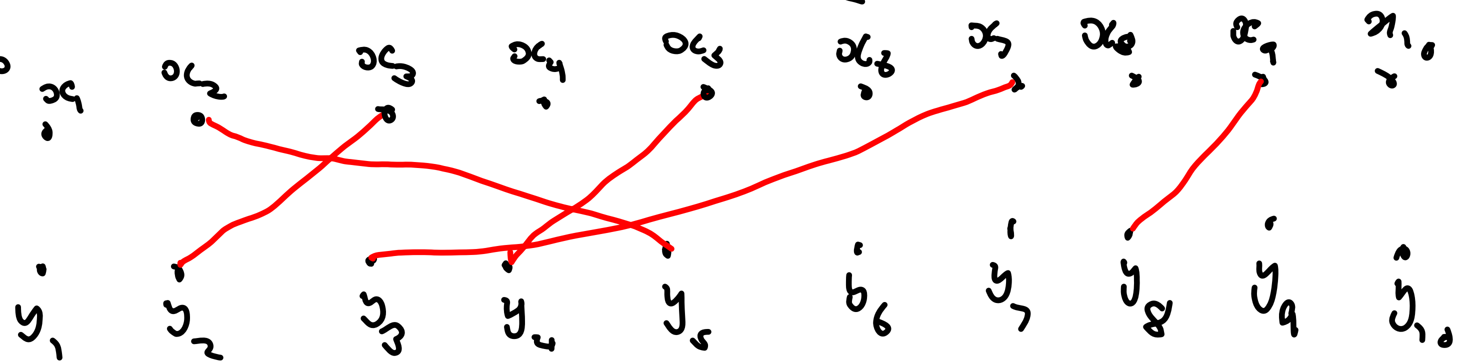
$H_1 := G_1$; $H_2 := (X, \emptyset)$

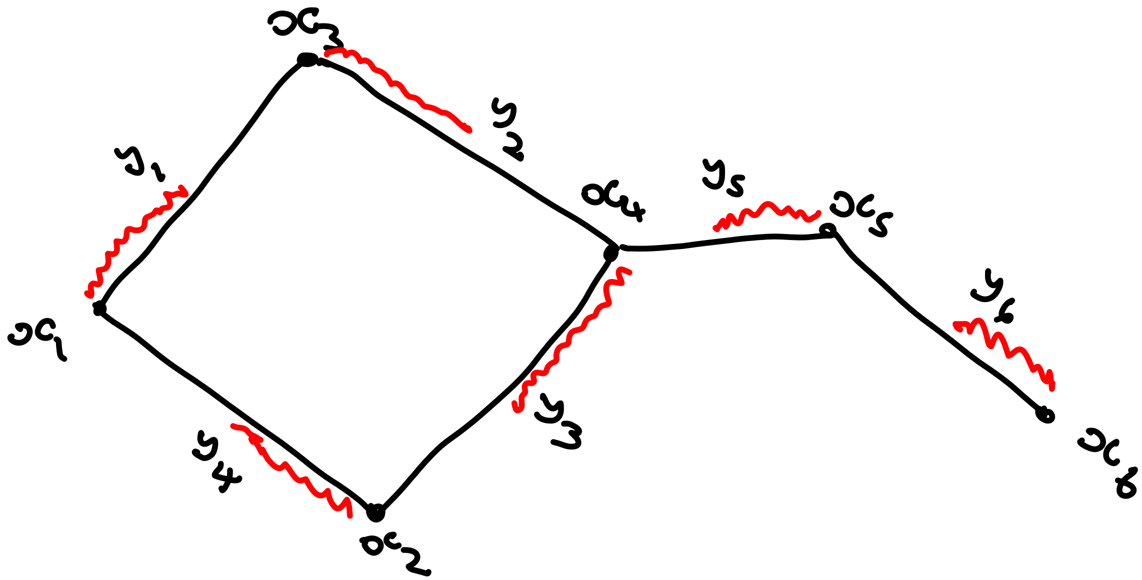


Tree in H_1

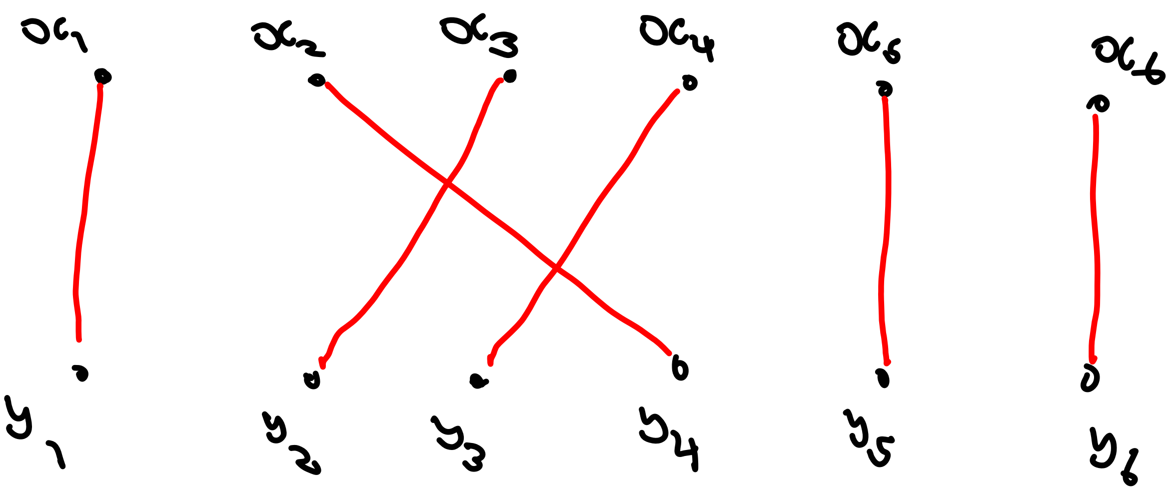


Yields





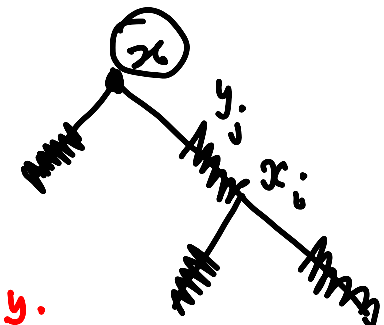
x1
x2
x3
x4
x5
x6
y1
y2
y3
y4
y5
y6



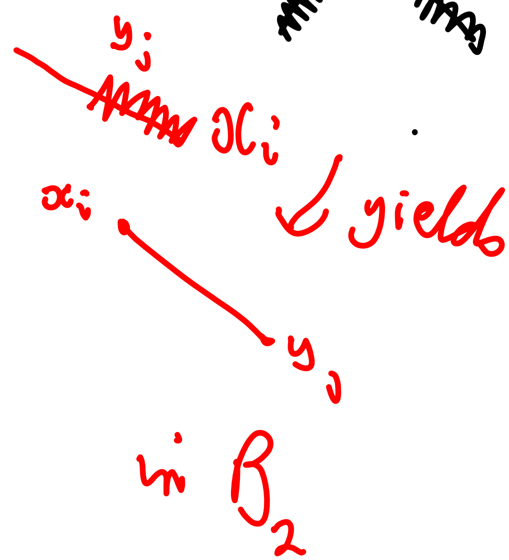
x1
x2
x3
x4
x5
x6
y1
y2
y3
y4
y5
y6

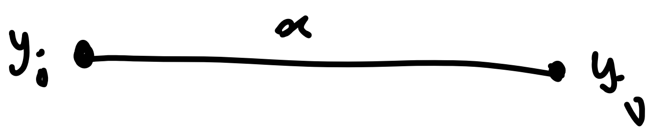
Step 1: If every isolated tree of H_1 contains a marked vertex: FOUND PERFECT MATCHING.

Step 2: Choose unmarked isolated tree T_j;
choose root x for T_j;
Mark x.



Step 3: Add edge with label x to H_2

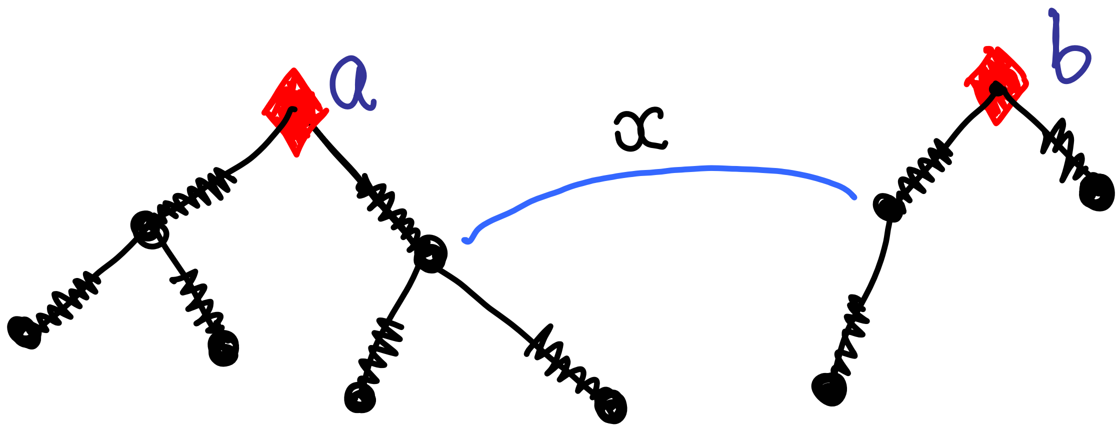


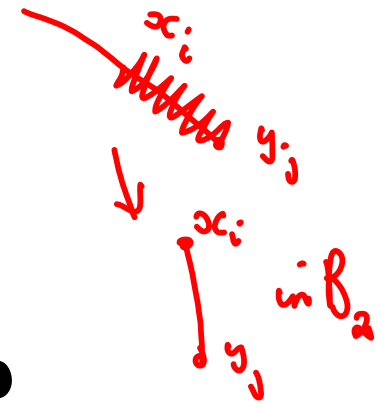


x chose y_i, y_j

Step 4 : Possibilities.

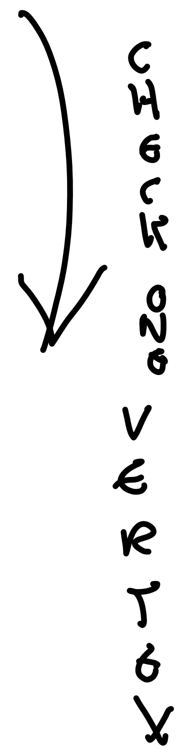
(1)

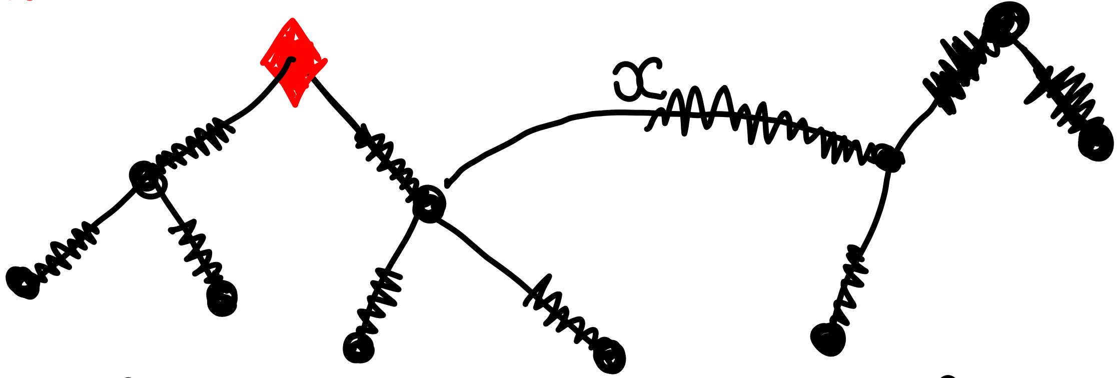




● - checked vertex

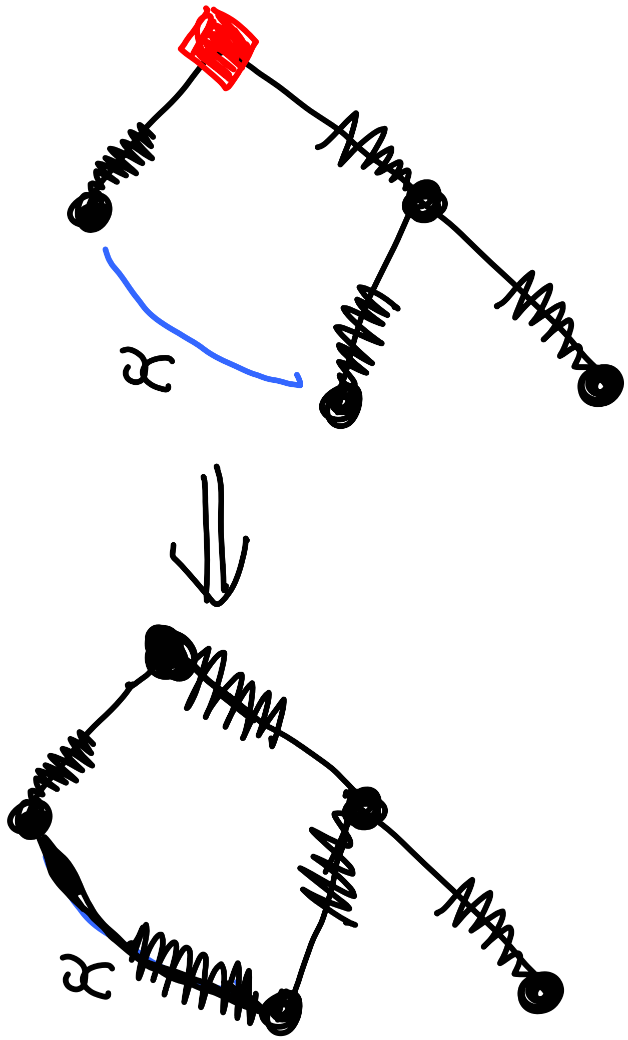
◆ - unchecked vertex



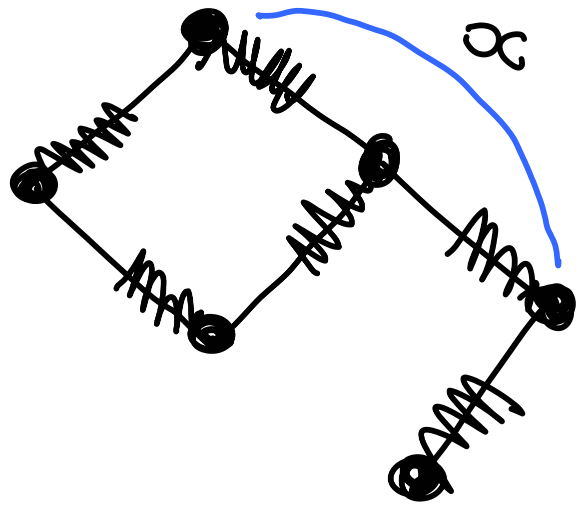


Give preference to vertex b if edge b in CORE of H_1 and edge a is not.

(ii)



(iii)



FAIL

In cases (i) & (ii) delete edge x from H_1.

Repeat from Step 1

Invariants

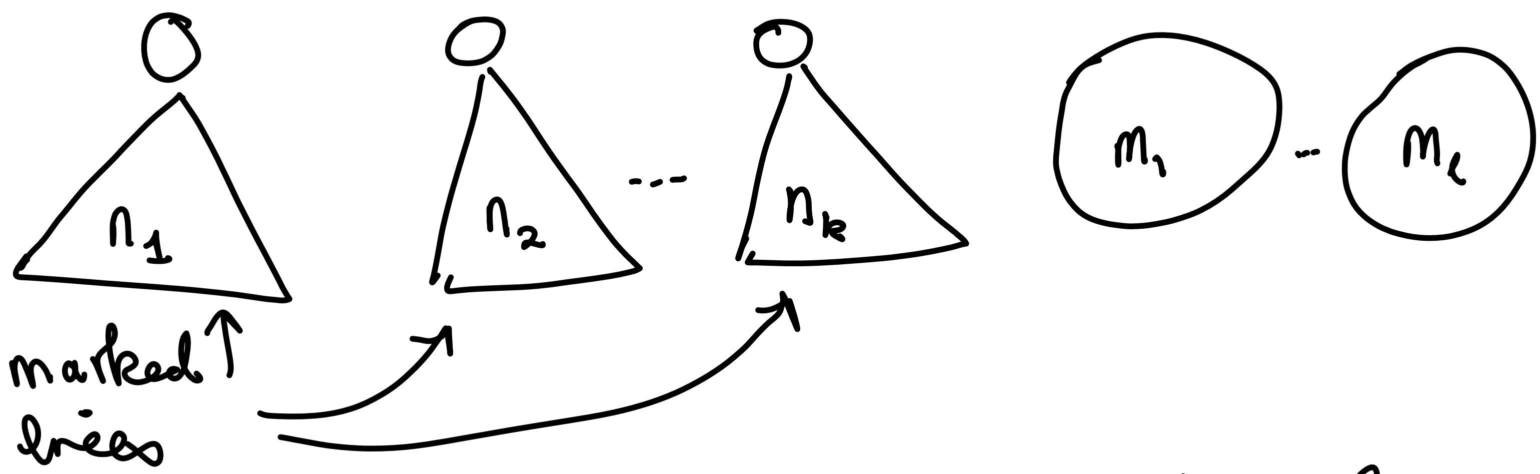
(i) # marked vertices $= n -$ # edges in H_1.

Each round marks one vertex and deletes one edge of H_1.

(ii) # checked vertices $=$ # edges of H_2.

Each round checks one vertex and adds one edge to H_2.

Suppose there are no unmarked trees.



$$\#\text{vertices} \quad n_1+n_2+\cdots+n_k+m_1+\cdots+m_l = n$$

$$\#\text{edges} \quad (n_1-1)+(n_2-1)+\cdots+(n_k-1)+(\geq m_1)+\cdots+(\geq m_l) = n-k$$

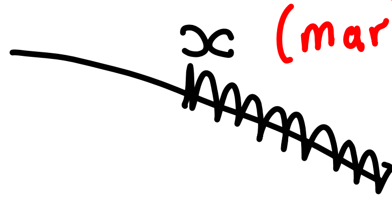
$\Rightarrow$ 

are unicyclic

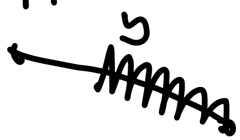
The edges give a matching M_1 in B_2 of size $n-k$ covering all unmarked OC's.

Note #rounds = #edges lost = k

H_2 contains k edges, also yielding matching M_2 in B_2.

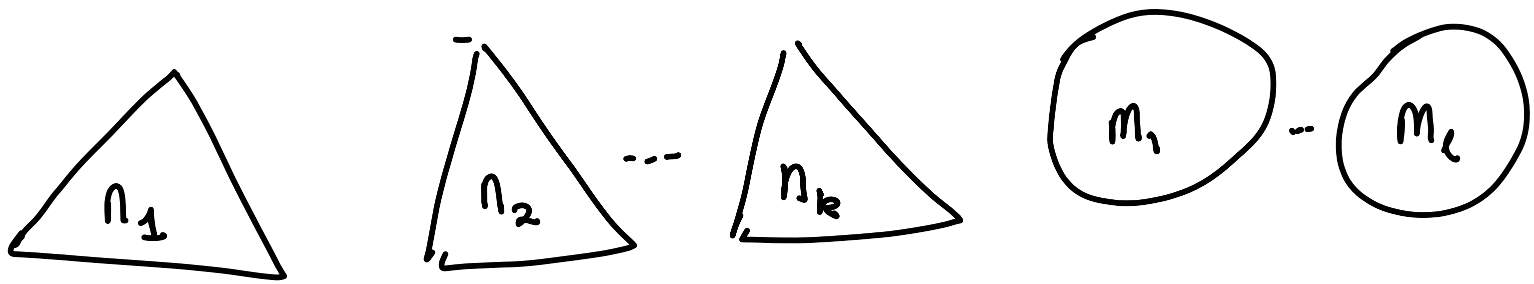


The ~~~ derived edge covers x.
M_1 does not cover x, but M_2 does.

Finally, suppose that M_1 does not cover y
i.e. $\nexists$ 

. This edge was deleted and so y is a checked vertex of H_2 and is covered by M_2.
Thus $M_1 \cup M_2$ = n edges covering $X \cup Y$, ie perfect matching.

Conversely, suppose H_1 consists of trees and unicyclic components



$$\# \text{edges} = n - k$$
$$= n - \# \text{marked vertices}$$

So every tree has a marked vertex and algorithm stops as soon as this happens.

Probability of Failure

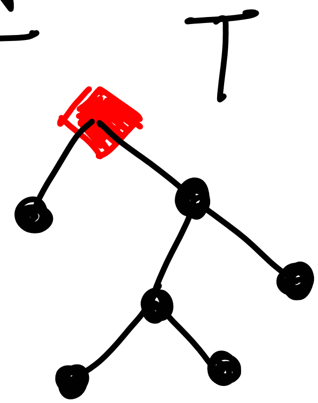
Claim (proved below)

Whp H_1 consists only of trees and unicyclic components before $.49n$ rounds.

Assume claim: H_2 consists of $\leq .49n$ random edges and so whp only contains trees and unicyclic components and so case (iii) of Step 4 does not happen.

Proof of Claim

Each H_2-tree has one unchecked vertex



Edge of H_1 corr. to unchecked vertices of H_2.

If ◆ corresponds to edge of CORE then our rule $\Rightarrow$ every vertex of T corresponds to edge of CORE.

So # vertices left in (what was) CORE = # trees of H_2 where every vertex corr. to an edge of CORE.

Size of CORE

Suppose $xe^{-x} = 2e^{-2}$, $0 < x < 1$.

Then CORE has $\approx (1 - x/2)^2 n$ edges.

$$.4 \leq x \leq .41$$

$$.63 \leq \left(1 - \frac{x}{2}\right)^2 \leq .64$$

Let Z = # trees in H_2 made up of vertices $y \in Y$ whose edge in H_1 belongs to CORE. $.49n$

$$E(Z) \leq o(1) + \sum_{k=1}^{(\log n)^2} \binom{n}{k} k^{k-2} \binom{.49n}{k-1} (k-1)! \left(\frac{1}{\binom{n}{2}}\right)^{k-1} \cdot (.64)^{k-1} \cdot \left(1 - \frac{k(n-k)}{\binom{n}{2}}\right)^{.49n}$$

$$\leq o(1) + .64n \sum_{k=1}^{(\log n)^2} \frac{k^{k-3}}{k!} (.64)^{k-1} \exp\left\{-\frac{.98k(n-k)}{(n-1)}\right\}$$

$$\leq .64n \sum_{k=1}^{(\log n)^2} \frac{k^{k-2}}{k!} (.64)^{k-1} \exp\left\{ - \frac{.98k(n-k)}{(n-1)} \right\} + o(n)$$

$$\leq (1+o(1))(.64)n \left(e^{-.98} \left(\frac{.6}{2} + \frac{3(.64)^3}{4} + \frac{16(.64)^3}{24} \right) + \sum_{k=5}^{(\log n)^2} \frac{1}{k^{5/2}} \cdot \frac{1}{.64} \cdot (.64e^{.02})^k \right)$$

$\leq \frac{1}{3}$

$\leq \frac{1}{5^{5/2} \times .64 \times (1-.65\times e^{.02})}$

$\leq \frac{1}{15}$

So, after .49 rounds, in expectation, # edges left in CORE, is $\leq \frac{2}{5}$ of original, and Chebyshev can be used to show this whp.

But deleting $\approx \frac{3}{5}$ of CORE's edges will whp leave just trees and unicyclic components:

Choose n random edges.

Build CORE

Delete $\approx \frac{3}{5}$ of edges.

(i) Whp $\approx \frac{3}{5}$ of edges of CORE are deleted.

(ii) Graph has $\approx \frac{2}{5} n$ edges and so has only trees plus unicyclic components.

So whp algorithm finishes before $.49n$ rounds with a perfect matching.

Random Mappings

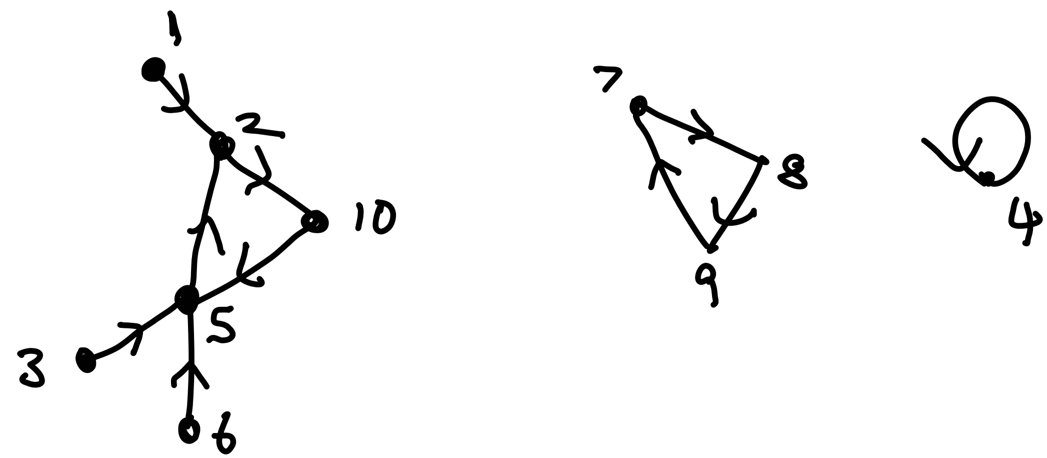
Let f be chosen uniformly at random from the set of all n^n mappings from $[n] \to [n]$.

Let D_f be the digraph $([n], (x, f(x)))$

and let G_f be obtained from D_f by ignoring orientation.

Ex:

x	1	2	3	4	5	6	7	8	9	10
$f(x)$	2	10	5	4	2	5	8	9	7	5



In general D_f consists of unicyclic components, where each such consists of a directed cycle C with trees rooted at each vertex of C.

Thm 1

$$\Pr(G_f \text{ is connected}) \approx \sqrt{\frac{\pi}{2n}}$$

Proof

Let $T(n,k)$ denote the number of forests with vertex set $[n]$, k trees, in which $1, 2, \ldots, k$ are in different trees. We show later that

$$T(n,k) = k n^{n-k-1}.$$

$$\Pr(G_f \text{ is connected}) =$$

$$n^{-n}\sum_{k=1}^{n} \underbrace{\binom{n}{k}(k-1)!}_{\substack{\text{choose cycle}\\ \text{of length } k}} \; \underbrace{T(n,k)}_{\text{finish off mapping}}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \underbrace{\prod_{j=0}^{k-1}\left(1-j/n\right)}_{u_k}$$

If $k \geq n^{3/5}$ then $u_k \leq \exp\left\{-\frac{k(k-1)}{2n}\right\} \leq e^{-\frac{1}{3}n^{1/5}}$.

If $k < n^{3/5}$ then $u_k = \exp\left\{-\frac{k^2}{2n} + O\left(\frac{k^3}{n^2}\right)\right\}$

So

$$\Pr(G_f \text{ is connected}) =$$

$$\frac{1+o(1)}{n}\sum_{k=1}^{n^{3/5}} e^{-k^2/2n} + O(n e^{-n^{1/5}/3})$$

$$= \frac{1+o(1)}{n}\int_0^\infty e^{-x^2/2n}\,dx + O(n e^{-n^{1/5}/3})$$

$$= \frac{1+o(1)}{\sqrt{n}}\int_0^\infty e^{-y^2/2}\,dy + O(n e^{-n^{1/5}/3})$$

$$\sim \sqrt{\frac{\pi}{2n}}.$$

Formula for $T(n,k)$:

$$T(n,1) = n^{n-2}$$ Cayley's Formula

$$T(n,k) = \sum_{\ell=0}^{n-k} \binom{n-k}{\ell} (\ell+1)^{\ell-1}\, T(n-\ell-1, k-1)$$

#vertices in k'tree is $\ell+1$

$$= \sum_{\ell=0}^{n-k} \binom{n-k}{\ell} (\ell+1)^{\ell-1} (k-1)(n-\ell-1)^{n-k-\ell-1}$$ induction

Abel's Formula

$$\sum_{\ell=0}^{m} \binom{m}{\ell} (x+\ell)^{\ell-1} (y+m-\ell)^{m-\ell-1} = \left(\frac{1}{x}+\frac{1}{y}\right)(x+y+m)^{m-1}$$

Take $m = n-k$, $x = 1$, $y = k-1$.

$\square$

Number of cycles:

Let Z_k = # of cycles of length k.

$$E(Z_k) = \binom{n}{k}(k-1)!\, n^{-k} = \frac{1}{k}\prod_{j=0}^{k-1}\left(1 - j/n\right)$$

If $Z = Z_1 + \cdots + Z_n$ then

$$E(Z) = \sum_{k=1}^{n} \frac{1}{k}\prod_{j=0}^{k-1}\left(1 - j/n\right)$$

$$\sim \int_1^\infty \frac{1}{x} e^{-x^2/2n}\, dx$$

$$\sim \log_e n .$$

Number of vertices on cycles:

$$E\left(\sum_{k=1}^{n} k Z_k\right) = \sum_{k=1}^{n} \prod_{j=1}^{k-1} \left(1 - j/n\right)$$

$$\sim \sqrt{\frac{\pi n}{2}} .$$

Shortest Paths

Let the arcs of the complete digraph D_n on $[n]$ be given independent lengths X_e, $e \in [n]^2$.

Here X_e is exponential with mean 1

i.e.

$$\Pr(X_e \geq t) = e^{-t}$$

for all $t \geq 0$.

Theorem

Let X_{ij} = distance from i to j. Then

(i) For any fixed i, j,

$$\Pr\left(\left|\frac{X_{i,j}}{\log n / n} - 1\right| \geq \epsilon\right) \to 0, \quad \forall \epsilon > 0$$

(ii) For any fixed i,

$$\Pr\left(\left|\frac{Z_i}{2\log n / n} - 1\right| \geq \epsilon\right) \to 0, \quad \forall \epsilon > 0$$

Here $Z_i = \max_j X_{ij}$.

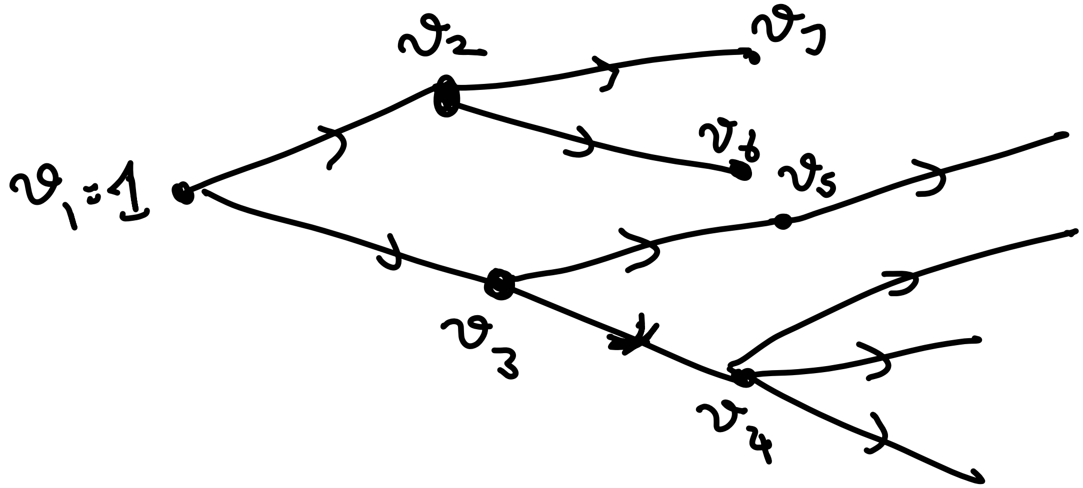
Proof

Two main properties of exponential X:

(P1) $\Pr(X > \alpha + \beta \mid X > \alpha) = \Pr(X > \beta)$.

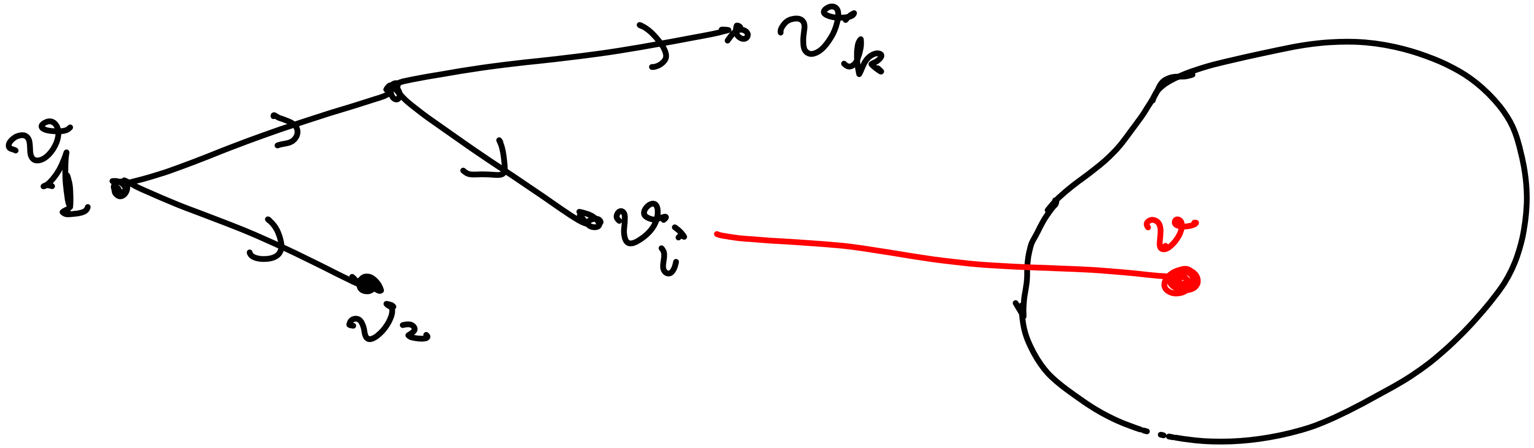
(P2) If $X_1, X_2, \ldots, X_m$ are independent exponentials then $\min\{X_1, X_2, \ldots, X_m\}$ is an exponential with mean $1/m$.

Fix $i = 1$ and consider Dijkstra's shortest path algorithm. This produces a tree



Suppose that vertices are added to the tree in the order $v_1, v_2, \dots, v_n$ and that $dist(v_1, v_i) = Y_i$.

It follows from P1 (p3) that



$$Y_{k+1} = \min_{\substack{i=1,\dots,k \\ v}} \Big[\underbrace{Y_i + X_{(v_i, v)}}_{\geq Y_k} \Big]$$

$$\stackrel{d}{=} Y_k + \text{Exponential}$$

So $\quad Y_{k+1} = Y_k + E_k \quad$ where E_k is exponential with mean $\frac{1}{k(n-k)}$ and is independent of Y_k.

So

$$E(Y_n) = \sum_{k=1}^{n-1} \frac{1}{k(n-k)}$$

$$= \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{1}{k} + \frac{1}{n-k} \right)$$

$$= \frac{2}{n} \sum_{k=1}^{n-1} \frac{1}{k}$$

$$\sim \frac{2 \log_e n}{n}.$$

Also $\operatorname{Var}(Y_n) = \sum_{k=1}^{n-1} \operatorname{Var}(E_k) = \sum_{k=1}^{n-1} \left(\frac{1}{k(n-k)} \right)^2$

$$\le 2 \sum_{k=1}^{n/2} \left(\frac{1}{k(n-k)} \right)^2 \le \frac{8}{n^2} \sum_{k=1}^{n/2} \frac{1}{k^2} = O(n^{-2})$$

and we can use Chebyshev to prove (ii).

Now fix $j=2$. Then if i is defined by $\sigma_i = 2$, we see that i is uniform over $\{2, 3, \ldots, n\}$.

So

$$E(X_{1,2}) = \frac{1}{n-1} \sum_{i=2}^{n} \sum_{k=1}^{i} \frac{1}{k(n-k)}$$

$$= \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{n-k}{k(n-k)}$$

$$\approx \frac{\log_e n}{n}.$$

For variance we have

$$X_{1,2} = \delta_2 Y_2 + \delta_3 Y_3 + \cdots + \delta_n Y_n$$

where

$\delta_i \in \{0,1\}$; $\delta_2 + \cdots + \delta_n = 1$; $\Pr(\delta_i = 1) = \frac{1}{n-1}$.

$$\mathrm{Var}(X_{1,2}) = \sum_{i=2}^{n} \mathrm{Var}(\delta_i Y_i) + \sum_{i \neq j} \mathrm{Covar}(\delta_i Y_i\, \delta_j Y_j)$$

$$\leq \sum_{i=2}^{n} \mathrm{Var}(\delta_i Y_i)$$

$$\mathrm{Covar}(\delta_i Y_i\, \delta_j Y_j) = E(\underset{\uparrow 0}{\delta_i Y_i\, \delta_j Y_j}) - E(\delta_i Y_i)\, E(\delta_j Y_j) \leq 0$$

$$\mathrm{Var}(X_{1,2}) \leq \sum_{i=2}^{n} \mathrm{Var}(\delta_i Y_i)$$

$$\leq \sum_{i=2}^{n} \frac{1}{n-1} \sum_{k=1}^{i-1} \left(\frac{1}{k(n-k)}\right)^2$$

$$= O\left(1/n^2\right).$$

Split sum at $n/2$

We can now use Chebyshev.

Digraphs

In this chapter we study the random digraph $D_{n,p}$. 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 $D_{n,p}$.

Case 1: $p = c/n$, $c < 1$

We will show that in this case

Theorem 1

Whp

(i) all strong components of $D_{n,p}$ are either cycles or single vertices.

(ii) The number of vertices on cycles is at most ω, for any $\omega = \omega(n) \to \infty$

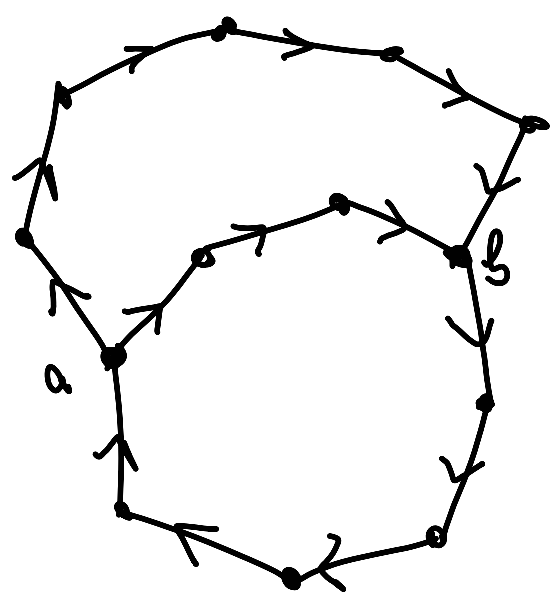
Proof

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 argue that if there is a component that is not a cycle or 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 sub-graphs is bounded by

$$\sum_{k=2}^{n} \sum_{\ell=1}^{n-k} \binom{n}{k} (k-1)! \left(\frac{c}{n}\right)^{k} \binom{n}{\ell} \ell! \left(\frac{c}{n}\right)^{\ell+1}$$

$$\leq \sum_{k=2}^{\infty} \sum_{\ell=1}^{\infty} \frac{c^{k+\ell+1}}{k\,n} = O\left(\frac{1}{n}\right).$$

Here ℓ is the number of vertices on the path P, excluding a, b.

$\square$

We now consider the case $p = c/n$ where $c > 1$.

We will prove the following theorem that is a directed analogue of the existence of a giant component in $G_{n,p}$.

Theorem 2

Let x be defined by $x < 1$ and $xe^{-x} = ce^{-c}$. Then whp $D_{n,p}$ contains a unique strong component of size $\approx (1 - \frac{x}{c})^2 n$. All other strong components are of logarithmic size.

General Strategy: For a vertex v let

$$D^+(v) = \{w : \exists \text{ path } v \text{ to } w \text{ in } D_{n,p}\}$$

$$D^-(v) = \{w : \exists \text{ path } w \text{ to } v \text{ in } D_{n,p}\}.$$

We will first prove

<u>Lemma 1</u>

There exist constants α, β (dependent only on c) such that whp

$\nexists v$ such that $|D^{\pm}(v)| \in [\alpha \log n, \beta n]$.

Proof

If there is a v such that $|D^+(v)| = s$ then $D_{n,p}$ 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

$$\binom{n}{s} s^{s-2} \left(\frac{c}{n}\right)^{s-1} \left(1-\frac{c}{n}\right)^{s(n-s)} \leq$$

$$\frac{n}{cs^2}\left(ce^{1-c+s/n}\right)^{s}.$$

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

$$\frac{n}{cs^2}\gamma^{s} \leq n^{-3} \quad \text{for} \quad s \geq \frac{4}{\log 1/\gamma}\log n.$$

$\square$

Fix a vertex $v \in [n]$ and consider a directed breadth first search from v.

Let $S_0^+ = \{v\}$ and given $S_0^+, S_1^+, \ldots, S_k^+ \subseteq [n]$ let $T_k^+ = \bigcup_{i=1}^{k} S_i^+$ and let

$$S_{k+1}^+ = \{ w \notin T_k^+ : \exists\, x \in T_k^+ \text{ s.t. } (x,w) \in E(D_{n,p}) \}.$$

Not surprisingly, we can show that the sub-graph 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 $Po(c)$ as the distribution of the number of offspring from a single parent.

Lemma 2

If $\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_k$ and $\hat{T}_k$ are defined with respect to the branching process and if $k \le k_0 = \log^3 n$ and $s_0, s_1, \ldots, s_k \le \log^4 n$ then

$$\Pr\left(|S_i^+| = s_i,\ 0 \le i \le k\right) = \left(1 + O\left(\frac{1}{n^{1-o(1)}}\right)\right) \Pr\left(|\hat{S}_i| = s_i,\ 0 \le i \le k\right).$$

<u>Proof</u>

$$\Pr\left(|\hat{S}_i| = s_i,\ 0 \le i \le k\right) = \prod_{i=1}^{k} \frac{(c s_{i-1})^{s_i} e^{-c s_{i-1}}}{s_i!},$$

Furthermore, putting $t_i := s_0 + s_1 + \dots + s_i$ we have

$$\Pr\left(|S_i^+| = s_i,\ 0 \le i \le k\right) = \prod_{i=1}^{k} \binom{s_{i-1}(n-t_i)}{s_i} \left(\frac{c}{n}\right)^{s_i} \left(1 - \frac{c}{n}\right)^{s_{i-1}(n-t_i) - s_i}$$

and the lemma follows by simple estimations. $\square$

Lemma 3

(a) $\Pr(|S_i^+| \geq s\log n \mid |S_{i-1}^+| = s) \leq n^{-10}$.

(b) $\Pr(|\hat{S}_i| \geq s\log n \mid |\hat{S}_{i-1}| = s) \leq n^{-10}$.

Proof

(a)

$$\Pr(|S_i^+| \geq s\log n \mid |S_{i-1}^+| = s) \leq$$

$$\Pr(B(sn, \tfrac{c}{n}) \geq s\log n) \leq \binom{sn}{s\log n}\left(\frac{c}{n}\right)^{s\log n}$$

$$\leq \left(\frac{sne\,c}{sn\log n}\right)^{s\log n}$$

$$\leq \left(\frac{ec}{\log n}\right)^{\log n}.$$

(b) is similar.

Next let

$$\mathcal{F} = \{ \exists i : |T_i^+| > \log^2 n \}$$

Lemma 4

$$\Pr(\mathcal{F}) = 1 - \frac{x}{c} + o(1)$$

Proof

$$\Pr(\mathcal{F}) = \Pr(\mathcal{F}_1) + o(1)$$

where

$$\mathcal{F}_1 = \{ \exists i \leq \log^2 n : |T_0^+|, \dots, |T_{i-1}^+| < \log^2 n \leq |T_i^+| \}$$

This follows from Lemma 3.

Applying Lemma 2 (on p12) we see that

$$\Pr(\mathcal{F}_1) = \Pr(\widehat{\mathcal{F}}_1) + o(1)$$

where $\widehat{\mathcal{F}}_1$ is defined w.r.t. the branching process.

Now let $\widehat{\mathcal{E}}$ be the event that the branching process becomes extinct.

We write

$$\Pr(\widehat{\mathcal{F}}_1) = \Pr(\widehat{\mathcal{F}}_1 \mid \neg\widehat{\mathcal{E}})\Pr(\neg\widehat{\mathcal{E}}) + \Pr(\widehat{\mathcal{F}}_1 \wedge \widehat{\mathcal{E}}). \quad (1)$$

To estimate (1) we first define

$$\rho = P_1(\hat{\mathcal{E}})$$

$$= \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} \rho^k .$$

This is if the origin of the process has k children then each of the processes spawned by them must become extinct for $\mathcal{E}$ to occur.

Thus

$$\rho = e^{c\rho - c} .$$

Substituting $\rho = \frac{\xi}{c}$ proves that

$$\Pr(\hat{\mathcal{E}}) = \frac{\xi}{c} \text{ where } \frac{\xi}{c} = e^{\xi - c}$$

and so $\xi = x$.

The lemma will follow from (1) [p16]

and this and $\Pr(\hat{\mathcal{F}} \mid \neg \mathcal{E}) = 1 - o(1)$

(see Lemma 3 [p14]) and

$$\Pr(\hat{\mathcal{F}} \wedge \mathcal{E}) = o(1). \tag{2}$$

Let us break the first $\log^2 n$ generations of the branching process into $\log n$ rounds of length $\log n$.

If $\neg\mathcal{E}$ occurs then we start each round with a non-zero population.

Claim 1

Each member of this population has a probability of at least $\epsilon > 0$ of producing $\log^2 n$ descendants at depth $\log n$. Here $\epsilon > 0$ depends only on c.

and so

$$\Pr(\overline{\mathcal{F}} \wedge \neg\mathcal{E}) \leq (1-\epsilon)^{\log n} = o(1).$$

If the current population of the process is s then the probability that it reach 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 object spawns at least 100 descendants and so there is a probability of at least

$$\epsilon_1 \left(1 - \sum_{s=100}^{\infty} e^{-\alpha s}\right)$$

that a single object spawns

$$\left(\frac{c+1}{2}\right)^{\log n} \gg \log^2 n$$

descendants at depth $\log n$.

This proves Claim 1 ([p19]) and completes the proof of Lemma 4.

$\square$

We state for future reference that the above argument supports the following claim.

Claim 2

$\Pr(\exists i : |S_i^+| \geq \log^2 n$ and $|T_i^+|$

We must now consider the probability that both $D^+(v)$ and $D^-(v)$ are large.

Lemma 5

$$\Pr\left(|D^-(v)| \geq \log^2 n \;\middle|\; |D^+(v)| \geq \log^2 n\right) = 1 - \frac{x}{c} + o(1).$$

Proof

Expose $S_0^+, S_1^+ \dots S_k^+$ until either $S_k^+ = \emptyset$ or we see that $|T_k^+| \geq \log^2 n$.

Now let S denote the set of edges/vertices defined by $S_0^+, S_1^+, \dots S_k^+$ we see that (see Lemma 2 [p12])

Let $\mathcal{C}$ 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 find that $|D^-(v)| < \log^2 n$ or $\geq \log^2 n$. Then

$$\Pr(\mathcal{C}) = 1 - \frac{1}{n^{1-o(1)}}$$

and

$$\Pr(|S_i^-| = s_i,\ 0 \leq i \leq k \mid \mathcal{C}) = \prod_{i=1}^{k} \binom{s_{i-1}(n'-t_i)}{s_i} \left(\frac{c}{n}\right)^{s_i} \left(1-\frac{c}{n}\right)^{s_{i-1}(n'-t_i)-s_i}$$

where $n' = n - |T_k^+|$.

Given this we can prove a conditional version of Lemma 2 and continue as before. $\square$

We have now shown that if

$$S = \{ v : |D^+(v)|, |D^-(v)| > \alpha \log n \}$$

then

$$E(|S|) = (1+o(1))\left(1 - \frac{x}{c}\right)^2 n.$$

We also claim that for any two vertices v, w

$$\Pr[v, w \in S] = (1+o(1)) \Pr(v \in S) \Pr(w \in S) \quad (3)$$

and therefore the Chebyshev inequality implies that whp

$$|S| = (1+o(1))\left(1 - \frac{x}{c}\right)^2 n.$$

But (3) follows in a similar manner to the proof of Lemma 5 (p22).

All that remains of the proof of Theorem 2 is to show that

$$\text{whp } S \text{ is a strong component.} \quad (4)$$

(Any $v \notin S$ is in a strong component of size $\leq 2\log n$).

We prove (4) by arguing that

$$\Pr\left(\exists v, w \in S : w \notin D^+(v)\right) = o(1). \qquad (5)$$

For this we expose $S_0^+, S_1^+, \ldots, S_k^+$ until we find that $|T_k^+(v)| \geq n^{\frac{1}{2}} \log n$.

At the same time we expose $S_0^-, S_1^-, \ldots, S_\ell^-$ until $|T_\ell^-(w)| \geq n^{\frac{1}{2}} \log n$.

If $w \notin D^+(v)$ then this experiment will have tried at least $(n^{\frac{1}{2}} \log n)^2$ times to find an edge from $D^+(v)$ to $D^-(w)$ and failed every time.

The probability of this is at most

$$\left(1 - \frac{c}{n}\right)^{n \log^2 n} = o(n^{-2}).$$

This completes the proof of Theorem 2.

$\square$

Strong Connectivity Threshold

Here we prove

Theorem 3

Suppose that $p = \frac{\log n + c_n}{n}$. Then

$$\lim_{n \to \infty} \Pr\left(D_{n,p} \text{ is strongly connected}\right) = \begin{cases} 0 & c_n \to -\infty \\ e^{-2e^{-c}} & c_n \to c \\ 1 & c_n \to +\infty \end{cases}$$

$$= \lim_{n \to \infty} \Pr\left(\nexists v \text{ such that } d^+(v) = 0 \text{ or } d^-(v) = 0\right).$$

__Proof__

We leave it as an exercise to prove that

$$\lim_{n\to\infty} \Pr(\nexists\, v \text{ such } d^+(v)=0 \text{ or } d^-(v)=0) = \begin{cases} 1 & c_n \to -\infty \\ 1-e^{-2e^{-c}} & c_n \to c \\ 0 & c_n \to \infty \end{cases}$$

Given this, one only has to show that if $c_n \not\to -\infty$ then whp there does not exist a vertex v such that $2 \leq |D^+(v)| \leq n/2$ or $2 \leq |D^-(v)| \leq n/2$.

But, here with $s+1 = |D^+(v)|$,

$$\Pr(\exists v) \le 2n \sum_{s=1}^{n/2} \binom{n}{s} (s+1)^{s-1} \left(\frac{c}{n}\right)^{s} (1-p)^{(s+1)(n-1-s)}$$

$$= o(1).$$ (Exercise)

Hamilton Cycles

Here we prove the following remarkable inequality:

Theorem 4

$$\Pr(D_{n,p} \text{ is Hamiltonian}) \geq \Pr(G_{n,p} \text{ is Hamiltonian})$$

Proof

Remark: This shows that if $p = \frac{\log n + \log\log n + \omega}{n}$ then $D_{n,p}$ is Hamiltonian whp. This result has been strengthened but it requires a much more difficult argument. The $\log\log n$ can be eliminated.

Proof

We consider a 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 K_n. Each $e_i = (v_i, w_i)$ gives rise to two directed edges $\vec{e}_i = (v_i, w_i)$ and $\overleftarrow{e}_i = (w_i, v_i)$.

In Γ_i we include $\vec{e}_j$ 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 $G_{n,p}$ with each edge (v,w) replaced by a pair of directed edges (v,w), (w,v) and $\Gamma_N = D_{n,p}$. Theorem 4 follows from

$$\Pr(\Gamma_i \text{ is Hamiltonian}) \geq \Pr(\Gamma_{i-1} \text{ is Hamiltonian})$$

To prove this we condition on the existence or otherwise of directed edges associated with $e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_N$.

Let $\mathcal{C}$ denote this conditioning.

Either $\mathcal{C}$ is such that

(a) $\mathcal{C}$ gives us a Hamilton cycle without arcs associated with e_i or there is no Hamilton cycle even if both $\vec{e}_i, \overleftarrow{e}_i$ occur

or $\mathcal{C}$ is such that

(b) $\exists$ a Hamilton cycle if at least one of $\vec{e}_i, \overleftarrow{e}_i$ occurs.

In Γ_{i-1} this happens with probability p

In Γ_i this happens with probability $1-(1-p)^2 > p$

[We will never require that <u>both</u> $\vec{e}_i, \overleftarrow{e}_i$ occur.]

$\square$