Random graphs

Alan Frieze

Carnegie Mellon University

Choosing a graph at random

Choosing a graph at random

$G_{n,p}$: Each edge e of the complete graph K_n is included independently with probability $p = p(n)$.

Whp $G_{n,p}$ has $\sim \binom{n}{2}p$ edges, provided $\binom{n}{2}p \to \infty$

$p = 1/2$, each subgraph of K_n is equally likely.

Choosing a graph at random

$G_{n,p}$: Each edge e of the complete graph K_n is included independently with probability $p = p(n)$.

Whp $G_{n,p}$ has $\sim \binom{n}{2}p$ edges, provided $\binom{n}{2}p \to \infty$

$p = 1/2$, each subgraph of K_n is equally likely.

$G_{n,m}$: Vertex set $[n]$ and m random edges.

If $m \sim \binom{n}{2}p$ then $G_{n,p}$ and $G_{n,m}$ have “similar” properties.

Random graphs first used to prove existence of graphs with certain properties:

Random graphs first used to prove existence of graphs with certain properties:

Erdős (1947): **Whp** the maximum size of a clique or independent set in $G_{n,1/2}$ is $\leq 2\log_2 n$.

Therefore

$$R(k,k) \geq 2^{k/2}.$$

Random graphs first used to prove existence of graphs with certain properties:

Mantel (1907): There exist triangle free graphs with arbitrarily large chromatic number.
Erdős (1959): There exist graphs of arbitrarily large girth and chromatic number.

$m = cn$, $c > 0$ is a large constant. **Whp** $G_{n,m}$ has $o(n)$ vertices on cycles of length $\leq \log\log n$ and no independent set of size more than $\frac{2\log c}{c}n$.

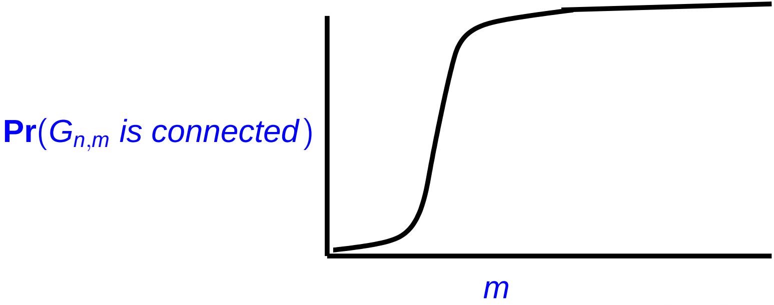
So removing the vertices on small cycles gives us a graph with girth $\geq \log\log n$ and chromatic number $\geq \frac{c+o(1)}{2\log c}$.

Erdős and Rényi began the study of random graphs in their own right.

On Random Graphs I (1959): $m = \frac{1}{2}n(\log n + c_n)$

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

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



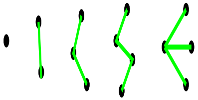
The evolution of a random graph, Erdős and Rényi (1960)

m	Structure of $G_{n,m}$ **whp**
$o(n^{1/2})$	Isolated edges and vertices

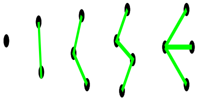
The evolution of a random graph, Erdős and Rényi (1960)

m	Structure of $G_{n,m}$ **whp**
$o(n^{1/2})$	Isolated edges and vertices
$n^{1/2} \log n$	Isolated edges and vertices and paths of length 2

The evolution of a random graph, Erdős and Rényi (1960)

m	Structure of $G_{n,m}$ **whp**
$o(n^{1/2})$	Isolated edges and vertices
$n^{1/2} \log n$	Isolated edges and vertices and paths of length 2
$n^{2/3} \log n$	Components are of the form 

The evolution of a random graph, Erdős and Rényi (1960)

m	Structure of $G_{n,m}$ **whp**
$o(n^{1/2})$	Isolated edges and vertices
$n^{1/2}\log n$	Isolated edges and vertices and paths of length 2
$n^{2/3}\log n$	Components are of the form 
$n^{\frac{k-1}{k}}\log n$	Components are trees of vertex size $1, 2, \ldots, k+1$. Each possible such tree appears.

m	Structure of $G_{n,m}$ **whp**
$\frac{1}{2}cn$ $c < 1$	Mainly trees. Some unicyclic components. Maximum component size $O(\log n)$
$\frac{1}{2}n$	Complicated. Maximum component size order $n^{2/3}$. Has subsequently been the subject of moreintensive study e.g. Janson, Knuth, Łuczak and Pittel (1993).
$\frac{1}{2}cn$ $c > 1$	Unique giant component of size $G(c)$n. Remainder almost all trees. Second largest component of size $O(\log n)$

m	Structure of $G_{n,m}$ **whp**
$\frac{1}{2}cn$ $c < 1$	Mainly trees. Some unicyclic components. Maximum component size $O(\log n)$
$\frac{1}{2}n$	Complicated. Maximum component size order $n^{2/3}$. Has subsequently been the subject of moreintensive study e.g. Janson, Knuth, Łuczak and Pittel (1993).
$\frac{1}{2}cn$ $c > 1$	Unique giant component of size $G(c)$n. Remainder almost all trees. Second largest component of size $O(\log n)$

Only very simple probabilistic tools needed. Mainly first and second moment method.

Connectivity threshold

$$p = (1+\epsilon)\frac{\log n}{n}$$

X_k = number of k-components, $1 \le k \le n/2$.
$X = X_1 + X_2 + \cdots + X_{n/2}$
$G_{n,p}$ is connected iff $X = 0$.

Connectivity threshold

$$p = (1+\epsilon)\frac{\log n}{n}$$

X_k = number of k-components, $1 \leq k \leq n/2$.
$X = X_1 + X_2 + \cdots + X_{n/2}$
$G_{n,p}$ is connected iff $X = 0$.

$$\begin{aligned}
\mathbf{Pr}(X \neq 0) &\leq \mathbf{E}(X) \\
&\leq \sum_{k=1}^{n/2} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\
&\leq \frac{n}{\log n} \sum_{k=1}^{n/2} \left(\frac{e \log n}{n^{(1+\epsilon)(1-k/n)}} \right)^k \\
&\to 0.
\end{aligned}$$

Hitting Time: Consider $G_0, G_1, \ldots, G_m, \ldots,$ where G_{i+1} is G_i plus a random edge.
Let m_k denote the minimum m for which $\delta(G_m) \geq k$.

Hitting Time: Consider $G_0, G_1, \ldots, G_m, \ldots,$ where G_{i+1} is G_i plus a random edge.
Let m_k denote the minimum m for which $\delta(G_m) \geq k$.

- **Whp** m_1 is the "time" when G_m first becomes connected.

Hitting Time: Consider $G_0, G_1, \ldots, G_m, \ldots,$ where G_{i+1} is G_i plus a random edge.
Let m_k denote the minimum m for which $\delta(G_m) \geq k$.

- **Whp** m_1 is the “time” when G_m first becomes connected.
- **Whp** m_1 is the “time” when G_m first has a perfect matching.
  Erdős and Rényi (1966).

Hitting Time: Consider $G_0, G_1, \ldots, G_m, \ldots,$ where G_{i+1} is G_i plus a random edge.
Let m_k denote the minimum m for which $\delta(G_m) \geq k$.

- **Whp** m_1 is the "time" when G_m first becomes connected.
- **Whp** m_1 is the "time" when G_m first has a perfect matching. Erdős and Rényi (1966).
- **Whp** m_2 is the "time" when G_m first has a Hamilton cycle. Ajtai, Komlós and Szemerédi (1985), Bollobás (1984).

Hitting Time: Consider $G_0, G_1, \dots, G_m, \dots,$ where G_{i+1} is G_i plus a random edge.
Let m_k denote the minimum m for which $\delta(G_m) \geq k$.

- **Whp** m_1 is the "time" when G_m first becomes connected.
- **Whp** m_1 is the "time" when G_m first has a perfect matching.
Erdős and Rényi (1966).
- **Whp** m_2 is the "time" when G_m first has a Hamilton cycle.
Ajtai, Komlós and Szemerédi (1985), Bollobás (1984).
- **Whp** m_k is the "time" when G_m first has $k/2$ edge disjoint Hamilton cycles.
Bollobás and Frieze (1985).

Some Open Problems

Some Open Problems

Is it true that **whp** G_m has $\delta(G_m)/2$ Hamilton cycles, for $m = 1, 2, \ldots, \binom{n}{2}$?

It is known to be true as long as $\delta(G_m) = O(1)$.

It is known that $G_{n,1/2}$ has $\sim n/4$ edge disjoint Hamilton cycles, Frieze and Krivelevich (2005).

Some Open Problems

Is it true that if we include the edges of the n-cube, Q^n with constant probability $p > 1/2$ then the resulting random subgraph is Hamiltonian **whp**?

It is known to have a perfect matching **whp** - Bollobás (1999).

Some Open Problems

If we randomly color the edges of $G_{n,Kn\log n}$ with Kn colors and K is sufficiently large, then **whp** there exists a Hamilton cycle with every edge a different color – Cooper and Frieze (2002).

If we only have $\sim \frac{1}{2}n\log n$ random edges, then how many colors do we need to get such a cycle **whp**?

If we only have n colors then how many edges do we need to get such a cycle **whp**?

Some Open Problems

If we randomly color the edges of $G_{n,Kn\log n}$ with Kn colors and K is sufficiently large, then **whp** there exists a Hamilton cycle with every edge a different color – Cooper and Frieze (2002).

If we only have $\sim \frac{1}{2}n\log n$ random edges, then how many colors do we need to get such a cycle **whp**?

If we only have n colors then how many edges do we need to get such a cycle **whp**?

If we replace Hamilton Cycle by Spanning Tree then the problem is solved: The hitting time for a multi-colored spanning tree is the maximum of the hitting time for connectivity and the appearance of $n-1$ colors – Frieze and McKay (1994).

Some Open Problems

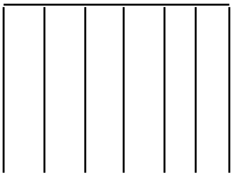
If we consider digraphs and ask for a multi-colored Hamilton cycle or spanning arborescence then nothing(?) is known.

Some Open Problems

Is it true that if T is a degree bounded tree with n vertices then **whp** $G_{n,Kn\log n}$ contains a **spanning** copy of T, for sufficiently large $K = K(T)$. Problem posed by Jeff Kahn.

True if T has a linear number of leaves.

The tree below seems to be a difficult one:



$n^{1/2}$ paths of length $n^{1/2}$

Small Subgraphs

Given a **fixed** graph H, one can ask when does $G_{n,p}$ contain a copy of H.

If X_H is the number of copies of H in $G_{n,p}$ then

$$\mathbf{E}(X_H) \sim C_H n^{v_H} p^{e_H}$$

where C_H is a constant, v_H, e_H are the number of vertices and edges in H.

Small Subgraphs

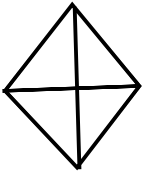
Given a **fixed** graph H, one can ask when does $G_{n,p}$ contain a copy of H.

If X_H is the number of copies of H in $G_{n,p}$ then

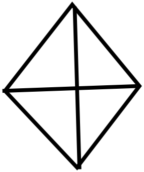
$$\mathbf{E}(X_H) \sim C_H n^{v_H} p^{e_H}$$

where C_H is a constant, v_H, e_H are the number of vertices and edges in H.

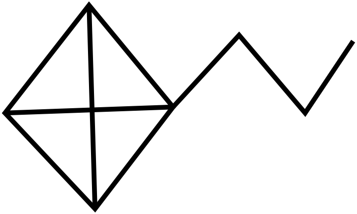
Does $\mathbf{E}(X_H) \to \infty$ imply that there is a copy of H **whp**?



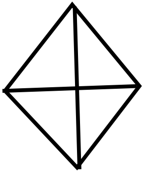
If $p = o(n^{-2/3})$ then $\mathbf{E}(X_H) \to 0$.
If $p = \omega n^{-2/3}$ then $\mathbf{E}(X_H) \to \infty$ and a copy of H exists **whp**.



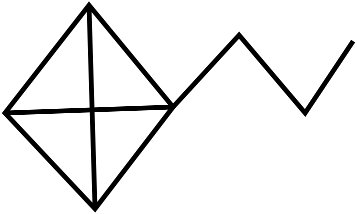
If $p = o(n^{-2/3})$ then $\mathbf{E}(X_H) \to 0$.
If $p = \omega n^{-2/3}$ then $\mathbf{E}(X_H) \to \infty$ and a copy of H exists **whp**.



If $p = n^{-3/4}$ then $\mathbf{E}(X_H) \to \infty$ but **whp** there is no copy of H.

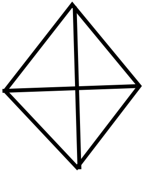


If $p = o(n^{-2/3})$ then $\mathbf{E}(X_H) \to 0$.
If $p = \omega n^{-2/3}$ then $\mathbf{E}(X_H) \to \infty$ and a copy of H exists **whp**.

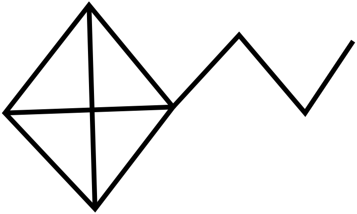


If $p = n^{-3/4}$ then $\mathbf{E}(X_H) \to \infty$ but **whp** there is no copy of H.

What we need is that $\mathbf{E}(X_{H'}) \to \infty$ for all subgraphs $H' \subseteq H$.

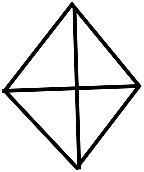


If $p = o(n^{-2/3})$ then $\mathbf{E}(X_H) \to 0$.
If $p = \omega n^{-2/3}$ then $\mathbf{E}(X_H) \to \infty$ and a copy of H exists **whp**.

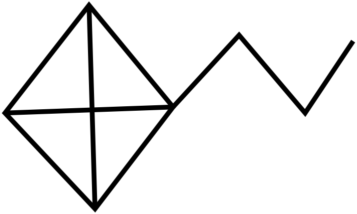


If $p = n^{-3/4}$ then $\mathbf{E}(X_H) \to \infty$ but **whp** there is no copy of H.

What we need is that $\mathbf{E}(X_{H'}) \to \infty$ for all subgraphs $H' \subseteq H$.
Bollobás (1981), Karoński and Ruciński (1983).



If $p = o(n^{-2/3})$ then $\mathbf{E}(X_H) \to 0$.
If $p = \omega n^{-2/3}$ then $\mathbf{E}(X_H) \to \infty$ and a copy of H exists **whp**.



If $p = n^{-3/4}$ then $\mathbf{E}(X_H) \to \infty$ but **whp** there is no copy of H.

What we need is that $\mathbf{E}(X_{H'}) \to \infty$ for all subgraphs $H' \subseteq H$.
Bollobás (1981), Karoński and Ruciński (1983).
Study of this problem has led to important probabilistic tools: Suen's inequality (1980), Janson's Inequality (1990) and the concentration inequality for multivariate polynomials by Kim and Vu (2004).

Graph Coloring

Graph Coloring

Matula (1970) showed using the second moment method that **whp** the maximum size $\alpha(G_{n,1/2})$ of an independent set is

$$2\log_2 n - 2\log_2\log_2 n + O(1).$$

Thus, **whp** $\chi(G_{n,1/2}) \geq\sim \frac{n}{2\log_2 n}$

Bollobás and Erdős (1976) and Grimmett and McDiarmid (1975) showed that **whp** a simple greedy algorithm uses $\sim \frac{n}{\log_2 n}$ colors.

Graph Coloring

A simple first moment calculation shows that **whp** $\alpha(G_{n,d/n})$ is

$$\leq 2\frac{\log d}{d}n$$

for d sufficiently large.

Thus, **whp**

$$\chi(G_{n,d/n}) \gtrsim \frac{d}{2\log d}$$

Shamir and Upfal (1984) showed that a slight modification of the greedy algorithm uses $\sim \frac{d}{\log d}$ colors.

Graph Coloring

It seemed “impossible “ to make any progress on this problem until the random graph community discovered

Graph Coloring

It seemed “impossible “ to make any progress on this problem until the random graph community discovered

Martingale Tail Inequalities

Azuma/Hoeffding

Graph Coloring

It seemed “impossible “ to make any progress on this problem until the random graph community discovered

Martingale Tail Inequalities

Azuma/Hoeffding

Let $Z = Z(X_1, \ldots, X_N)$ where $X_1, \ldots, X_N$ are independent. Suppose that changing one X_i only changes Z by ≤ 1. Then

$$\mathbf{Pr}(|Z - \mathbf{E}(Z)| \geq t) \leq e^{-t^2/(2n)}.$$

“Discovered” by Shamir and Spencer (1987) and by Rhee and Talagrand (1988).

Bollobás (1988) showed that $\chi(G_{n,1/2}) \sim \frac{n}{2\log_2 n}$.

Bollobás (1988) showed that $\chi(G_{n,1/2}) \sim \frac{n}{2\log_2 n}$.

Let Z be the maximum number of independent sets in a collection $S_1, \ldots, S_Z$ where each $|S_i| \sim 2\log_2 n$ and $|S_i \cap S_j| \leq 1$.

$$\mathbf{E}(Z) = n^{2-o(1)} \textit{ and changing one edge changes } Z \textit{ by } \leq 1$$

So,

$$\mathbf{Pr}\Big(\exists S \subseteq [n] : \ |S| \geq \frac{n}{(\log_2 n)^2} \textit{ and } S \textit{ doesn't contain a } (2 - o(1))\log_2 n \textit{ independent set}\Big) \leq 2^n e^{-n^{2-o(1)}} = o(1).$$

So, we color $G_{n,1/2}$ with color classes of size $\sim 2\log_2 n$ until there are $\leq n/(\log_2 n)^2$ vertices uncolored and then give each remaining vertex a new color.

Sparse random graphs: $p = d/n$.

Sparse random graphs: $p = d/n$.

$$\alpha(G_{n,d/n}) = \frac{(2 \pm \epsilon)\log d}{d} n$$

for large d, Frieze (1990).

Sparse random graphs: $p = d/n$.

$$\alpha(G_{n,d/n}) = \frac{(2 \pm \epsilon)\log d}{d} n$$

for large d, Frieze (1990).

Suppose $k \sim \frac{2\log d}{d} n$ and X_k is the number of independent k-sets in $G_{n,d/n}$

$$\mathbf{Pr}(X_k \neq 0) \geq \frac{\mathbf{E}(X_k)^2}{\mathbf{E}(X_k)^2} \geq e^{-a_1 n}.$$

Sparse random graphs: $p = d/n$.

$$\alpha(G_{n,d/n}) = \frac{(2 \pm \epsilon)\log d}{d} n$$

for large d, Frieze (1990).

Suppose $k \sim \frac{2\log d}{d} n$ and X_k is the number of independent k-sets in $G_{n,d/n}$

$$\mathbf{Pr}(X_k \neq 0) \geq \frac{\mathbf{E}(X_k)^2}{\mathbf{E}(X_k)^2} \geq e^{-a_1 n}.$$

But Azuma-Hoeffding gives

$$\mathbf{Pr}(|\alpha(G_{n,d/n}) - \mathbf{E}(\alpha)| \geq \epsilon_1 n) \leq e^{-a_2 n}.$$

Here $a_2 > a_1$ and so $\mathbf{E}(\alpha) \geq \frac{(2-\epsilon_2)\log d}{d} n$ and ...

Taking a similar (but much more computationally challenging) approach Łuczak (1991) showed that

$$\chi(G_{n,d/n}) \sim \frac{d}{2\log d}.$$

Taking a similar (but much more computationally challenging) approach Łuczak (1991) showed that

$$\chi(G_{n,d/n}) \sim \frac{d}{2\log d}.$$

Then Łuczak (1991) proved that **whp** there was a two point concentration for $\chi(G_{n,d/n})$ i.e. $\exists k_d$ such that **whp**

$$\chi(G_{n,d/n}) \in \{k_d, k_d+1\}.$$

Achlioptas and Naor (2005) showed that k_d is the smallest integer ≥ 2 such that $d < d_k = 2k \log k$.

If $d > d_k$ and X_k is the number of k-colorings of $G_{n,d/n}$ then $\mathbf{E}(X_k) \to 0$.

If $d \leq d_{k-1}$ then
$\mathbf{Pr}(G_{n,d/n} \text{ is } k - \text{colorable}) \geq \mathbf{E}(X_k)^2/\mathbf{E}(X_k^2) \geq \xi > 0$.

Using the results of Friedgut (1999) and Achlioptas and Friedgut (1999) we see that this implies $G_{n,d/n}$ *is k − colorable* **whp** for $d \leq d_{k-1}$.

Some Open Problems

Some Open Problems

Is it the case that there exist $d_3 < d_4 < \cdots < d_k < \cdots$ such that $d_k < d < d_{k+1}$ implies that **whp** $\chi(G_{n,d/n}) = k$?

The results of Friedgut (1999) and Achlioptas and Friedgut (1999) suggests strongly that this is true.

Some Open Problems

What is the Chromatic number of a random r-regular graph $G_{n,r}$?

Achlioptas and Moore (2005) show that provided $r = O(1)$ the chromatic number is 3 point concentrated around the smallest integer k such that $r < 2k \log k$.

Shi and Wormald (2005) show that **whp** a random 4-regular graph has chromatic number 3 and a random 6-regular graph has chromatic number 4.

Cooper, Frieze, Reed and Riordan (2002) show that if $r \to \infty$ then **whp**

$$\chi(G_{n,r}) \sim \frac{r}{2 \log r}.$$

Some Open Problems

Is there a polynomial time algorithm that **whp** can color $G_{n,1/2}$ with $\frac{(1-\epsilon)n}{\log_2 n}$ colors?

Randomly generated k-colorable graphs, $k = O(1)$, with $O(n)$ edges can be colored quickly, Alon and Kahale (1994).

Some Open Problems

What is the game chromatic number χ_g of the random graph $G_{n,1/2}$?

There are two players: A and B who alternately *properly* color the vertices of G. A tries to color the whole graph and B tries to force a situation where some vertex cannot be colored. χ_g is the minimum number of colors which guarantees a win for A.

Bohman, Frieze and Sudakov (2005) show that **whp**

$$(1-\epsilon)\frac{n}{\log_2 n} \leq \chi_g(G_{n,1/2}) \leq (2+\epsilon)\frac{n}{\log_2 n}.$$

The diameter of random graphs

The diameter of random graphs

Suppose $d \geq 2$ is a positive integer and $p^d n^{d-1} = \log(n^2/c)$ so that average degree is $\tilde{\Theta}(n^{1/d})$. Then

$$\lim_{n\to\infty} \mathbf{Pr}(\textit{diameter } G_{n,p} = d + \delta) = \begin{cases} e^{-c/2} & \delta = 0 \\ 1 - e^{-c/2} & \delta = 1 \end{cases}$$

Bollobás (1981).

Basically, there are $\tilde{\Theta}(n^{k/d})$ vertices at distance $\leq k$ from a fixed vertex v.

The diameter of random graphs

Diameter of the Giant Component of $G_{n,c/n}$: Fernholz and Ramanchandran (2005).
One would expect this to be $\sim A(c)\log n$ **whp**. They show that

$$A(c) = \frac{2}{-\log W} + \frac{1}{\log c}$$

where W is the solution in $(0,1)$ of $We^{-W} = ce^{-c}$.

Here $W \to 0$ as $c \to \infty$, so the diameter is “like” $\log_c n$ for large c, as one would expect.

Algorithms and Differential Equations

Karp and Sipser (1981) described a simple greedy matching algorithm for finding a large matching in the random graph $G_{n,c/n}$.

If there is a vertex v of degree one, choose a random degree one vertex and the edge incident to it; otherwise choose a random edge.

Algorithms and Differential Equations

Karp and Sipser (1981) described a simple greedy matching algorithm for finding a large matching in the random graph $G_{n,c/n}$.

If there is a vertex v of degree one, choose a random degree one vertex and the edge incident to it; otherwise choose a random edge.

They show that the algorithm is asymptotically optimal i.e. the matching it produces is within $1 - o(1)$ of optimal.

Aronson, Frieze and Pittel (1998) showed that **whp** this algorithm only makes $\tilde{\Theta}(n^{1/5})$ "mistakes".

The proof of the above results rests on the fact that the progress of the algorithm can **whp** be tracked by the solution of a differential equation.

The proof of the above results rests on the fact that the progress of the algorithm can **whp** be tracked by the solution of a differential equation.

Karp and Sipser introduced this approach (via Kurtz theorem) to the “CS/Probabilistic Combinatorics” community and Wormald has “championed” its applications.

Toy Example: Number of isolated vertices in G_m.

Toy Example: Number of isolated vertices in G_m.

Let $X_0(m)$ be the number of isolated vertices in G_m. Then

$$\mathbf{E}(X_0(m+1) - X_0(m) \mid G_m) = -2\frac{X_0(m)}{n}. \tag{1}$$

Toy Example: Number of isolated vertices in G_m.

Let $X_0(m)$ be the number of isolated vertices in G_m. Then

$$\mathbf{E}(X_0(m+1) - X_0(m) \mid G_m) = -2\frac{X_0(m)}{n}. \tag{1}$$

Let $x_0(t) = X_0(tn)/n$ for $t > 0$. Then (1) suggests the equation

$$x_0' = -2x_0$$

which has the solution

$$x_0 = e^{-2t}$$

or

$$X_0(m) \sim ne^{-2m/n}.$$

More typical example: From “Hamilton Cycles in 3-Out” – Bohman and Frieze (2006).

More typical example: From “Hamilton Cycles in 3-Out” – Bohman and Frieze (2006).

$$
\begin{aligned}
\mathbf{E}(y'_{i,j,0} - y_{i,j,0}) &= -\frac{jy_{i,j,0}}{\mu} - \sum_{a,b} \frac{by_{a,b,1}}{\mu} \left((b-1)\frac{iy_{i,j,0}}{\mu-1} + \hat{a}\frac{jy_{i,j,0}}{\mu-1} \right) \\
&\quad + \sum_{a,b} \frac{by_{a,b,1}}{\mu} \left((b-1)\frac{(i+1)y_{i+1,j,0}}{\mu-1} + \hat{a}\frac{(j+1)y_{i,j+1,0}}{\mu-1} \right) + \tilde{O}(\mu^{-1}) \\
\mathbf{E}(y'_{i,j,1} - y_{i,j,1}) &= -\frac{jy_{i,j,1}}{\mu} + \frac{(j+1)y_{i,j+1,0}}{\mu} - \sum_{a,b} \frac{by_{a,b,1}}{\mu} \left((b-1)\frac{iy_{i,j,1}}{\mu-1} + \hat{a}\frac{jy_{i,j,1}}{\mu-1} \right) \\
&\quad + \sum_{a,b} \frac{by_{a,b,1}}{\mu} \left((b-1)\frac{(i+1)y_{i+1,j,1}}{\mu-1} + \hat{a}\frac{(j+1)y_{i,j+1,1}}{\mu-1} \right) + \tilde{O}(\mu^{-1}) \\
\mathbf{E}(y'_{L,j,0} - y_{L,j,0}) &= -\frac{jy_{L,j,0}}{\mu} - \sum_{a,b} \frac{by_{a,b,1}}{\mu} \left((b-1)\frac{3y_{3,j,0}}{\mu-1} + \hat{a}\frac{jy_{L,j,0}}{\mu-1} \right) \\
&\quad + \sum_{a,b} \frac{by_{a,b,1}}{\mu} \cdot \hat{a}\frac{(j+1)y_{L,j+1,0}}{\mu-1} + \tilde{O}(\mu^{-1}). \\
&\overset{d}{=} \phi^{in}_{L,j,0}(y) + \tilde{O}(\mu^{-1}) \\
\mathbf{E}(y'_{L,j,1} - y_{L,j,1}) &= -\frac{jy_{L,j,1}}{\mu} + \frac{(j+1)y_{L,j+1,0}}{\mu} - \sum_{a,b} \frac{by_{a,b,1}}{\mu} \left((b-1)\frac{3y_{3,j,1}}{\mu-1} + \hat{a}\frac{jy_{L,j,1}}{\mu-1} \right) \\
&\quad + \sum_{a,b} \frac{by_{a,b,1}}{\mu} \cdot \hat{a}\frac{(j+1)y_{L,j+1,1}}{\mu-1} + \tilde{O}(\mu^{-1})
\end{aligned}
$$

Open Problem

Is there a linear expected time algorithm for finding a maximum size matching in $G_{n,c/n}$?

The Karp-Sipser algorithm is only asymptotically optimal and algorithms based on augmenting paths have an extra $\log n$ factor.

Typical Graphs

Typical Graphs

Unstructured, randomly generated(?) real world graphs like the **WWW** seem to have a different distribution to $G_{n,p}$, e.g. the number of vertices of degree k drops off like $k^{-\alpha}$ instead of $e^{-\alpha k}$.

Typical Graphs

Unstructured, randomly generated(?) real world graphs like the **WWW** seem to have a different distribution to $G_{n,p}$, e.g. the number of vertices of degree k drops off like $k^{-\alpha}$ instead of $e^{-\alpha k}$.

Modelling Choices:

Typical Graphs

Unstructured, randomly generated(?) real world graphs like the **WWW** seem to have a different distribution to $G_{n,p}$, e.g. the number of vertices of degree k drops off like $k^{-\alpha}$ instead of $e^{-\alpha k}$.

Modelling Choices:
Fix a degree sequence and make each graph with this degree sequence equally likely: Bender and Canfield (1978), Bollobás (1980), Molly and Reed (1995) and Cooper and Frieze(digraphs) (2004).

Typical Graphs

Unstructured, randomly generated(?) real world graphs like the **WWW** seem to have a different distribution to $G_{n,p}$, e.g. the number of vertices of degree k drops off like $k^{-\alpha}$ instead of $e^{-\alpha k}$.

Modelling Choices:
Fix a degree sequence $d_1, d_2, \ldots, d_n$ and make edge (i, j) occur independently with probability proportional to $d_i d_j$: Chung and Lu (2002), Mihail and Papadimitriou (2002)

Typical Graphs

Unstructured, randomly generated(?) real world graphs like the **WWW** seem to have a different distribution to $G_{n,p}$, e.g. the number of vertices of degree k drops off like $k^{-\alpha}$ instead of $e^{-\alpha k}$.

Modelling Choices:
Preferential Attachment Model: Vertex set $v_1, v_2, \ldots, v_n, \ldots$;
Vertex v_{n+1} chooses m random neighbours in $v_1, \ldots, v_n$ with probability proportional to their degree.

Introduced as a model of the web by Barabási and Albert (1999).

Properties of the Preferential Attachment Model PAM

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).
- Cover Time $\sim \frac{2m}{m-1} n \log n$: Cooper and Frieze (2005).

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).
- Cover Time $\sim \frac{2m}{m-1} n \log n$: Cooper and Frieze (2005).
- Conductance: Gkantsidis, Mihail and Saberi (2003)

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log\log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).
- Cover Time $\sim \frac{2m}{m-1} n \log n$: Cooper and Frieze (2005).
- Conductance: Gkantsidis, Mihail and Saberi (2003)
- Randomly deleting vertices preserves a giant component: Bollobás and Riordan.

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).
- Cover Time $\sim \frac{2m}{m-1} n \log n$: Cooper and Frieze (2005).
- Conductance: Gkantsidis, Mihail and Saberi (2003)
- Randomly deleting vertices preserves a giant component: Bollobás and Riordan.
- Adversarially deleting vertices preserves a giant component: Flaxman, Frieze and Vera (2005).

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).
- Cover Time $\sim \frac{2m}{m-1} n \log n$: Cooper and Frieze (2005).
- Conductance: Gkantsidis, Mihail and Saberi (2003)
- Randomly deleting vertices preserves a giant component: Bollobás and Riordan.
- Adversarially deleting vertices preserves a giant component: Flaxman, Frieze and Vera (2005).
- Spread of viruses: Berger, Borgs, Chates and Saberi (2005).

Properties of the Preferential Attachment Model PAM

- Power Law Degree Distribution: Bollobás, Riordan, Spencer and Tusanády (2001).
- Diameter $\sim \log n / \log \log n$: Bollobás and Riordan (2004).
- Spectral Properties: Flaxman, Frieze and Fenner (2005).
- Cover Time $\sim \frac{2m}{m-1} n \log n$: Cooper and Frieze (2005).
- Conductance: Gkantsidis, Mihail and Saberi (2003)
- Randomly deleting vertices preserves a giant component: Bollobás and Riordan.
- Adversarially deleting vertices preserves a giant component: Flaxman, Frieze and Vera (2005).
- Spread of viruses: Berger, Borgs, Chates and Saberi (2005).
- Classifying special interest groups in web graphs: Cooper (2002)

Power Law:

Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

Power Law:
Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

$$d_k(t+1) = d_k(t) + m\frac{(k-1)d_{k-1}(t)}{2mt} - m\frac{kd_k(t)}{2mt} + 1_{k=m} + \textit{error terms}.$$

Power Law:

Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

$$d_k(t+1) = d_k(t) + m\frac{(k-1)d_{k-1}(t)}{2mt} - m\frac{kd_k(t)}{2mt} + 1_{k=m} + \textit{error terms}.$$

Assume that $d_k(t) \sim d_k t$. Then

$$d_k\left(\frac{k}{2}+1\right) \sim d_{k-1}\frac{k-1}{2} + 1_{k=m}$$

Power Law:
Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

$$d_k(t+1) = d_k(t) + m\frac{(k-1)d_{k-1}(t)}{2mt} - m\frac{kd_k(t)}{2mt} + 1_{k=m} + \textit{error terms}.$$

Assume that $d_k(t) \sim d_k t$. Then

$$d_k\left(\frac{k}{2}+1\right) \sim d_{k-1}\frac{k-1}{2} + 1_{k=m}$$

$$d_k \sim \frac{2m(m+1)}{(k+2)(k+1)k}t \qquad \textit{for } k \geq m.$$

Some Open Problems

What is the second eigenvalue of the transition matrix of a random walk on PAM?

It should be $O(1/m)$.

Some Open Problems

What is the size of the smallest dominating set in PAM?

Achlioptas Problem

Suppose that $e_1, f_1, e_2, f_2, \ldots,$ is a random sequence of pairs of edges e_i, f_i. You have to choose, on-line, one of e_i, f_i for $i = 1, 2, \ldots$. Can you avoid creating a giant component for significantly beyond $n/2$ choices?

Achlioptas Problem

Suppose that $e_1, f_1, e_2, f_2, \dots,$ is a random sequence of pairs of edges e_i, f_i. You have to choose, on-line, one of e_i, f_i for $i = 1, 2, \dots$. Can you avoid creating a giant component for significantly beyond $n/2$ choices?

Bohman and Frieze (2001): If one of e_i, f_i is disjoint from $e_1, f_1, \dots, e_{i-1}, f_{i-1}$ then choose this edge, otherwise just take e_i.

Whp one can choose $.544n$ edges before creating a giant.

Achlioptas Problem

Suppose that $e_1, f_1, e_2, f_2, \ldots,$ is a random sequence of pairs of edges e_i, f_i. You have to choose, on-line, one of e_i, f_i for $i = 1, 2, \ldots$. Can you avoid creating a giant component for significantly beyond $n/2$ choices?

Subsequently several authors: Bohman and Kravitz (2005), Spencer and Wormald (2005) and Flaxman, Gamarnik and Sorkin (2004) studied algorithms for delaying and/or speeding up the emergence of a giant component.

Achlioptas Problem

Suppose that $e_1, f_1, e_2, f_2, \ldots,$ is a random sequence of pairs of edges e_i, f_i. You have to choose, on-line, one of e_i, f_i for $i = 1, 2, \ldots$. Can you avoid creating a giant component for significantly beyond $n/2$ choices?

Subsequently several authors: Bohman and Kravitz (2005), Spencer and Wormald (2005) and Flaxman, Gamarnik and Sorkin (2004) studied algorithms for delaying and/or speeding up the emergence of a giant component.

Related off-line problems were considered in Bohman, Frieze and Wormald, Bohman and Kim.

In particular, the BK and SW papers show that for a restricted class of algorithm, differential equations can be used to accurately predict the emergence of a giant, by tracking the parameter

$$Z = \frac{1}{n} \sum_i |C_i|^2.$$

Where $C_1, C_2, \ldots$ are the components of the graph induced by the edges selected so far.

The giant should appear when this parameter becomes unbounded.

Open Questions

Open Questions

Analyze the algorithm that always chooses the edge which produces the smallest increase in Z. When does a giant component appear?

The differential equations method has problems here, because the natural system of equations is infinite.

Open Questions

Consider speeding up or delaying the occurrence of other graph properties e.g. avoid 3-colorability.

Game Version

Suppose there are two players, Creator and Destroyer. Creator plays on odd rounds and Destroyer plays on even rounds. Creator wants to construct a giant component as soon as possible and Destroyer wants to delay the occurrence for as long as possible.

Game Version

Suppose there are two players, Creator and Destroyer. Creator plays on odd rounds and Destroyer plays on even rounds. Creator wants to construct a giant component as soon as possible and Destroyer wants to delay the occurrence for as long as possible.

Beveridge, Bohman, Frieze and Pikhurko (2006) show that the best strategy for Creator is to try to maximize the increase in Z and the best strategy for Destroyer is to try to minimize the increase in Z.

If they both play optimally, then it takes roughly $n/2$ rounds to create a giant, since they tend to cancel each others advantage over just choosing randomly.

THANK YOU