Complex Graphs and Networks

Lecture 4: The rise of the giant component

Linyuan Lu

lu@math.sc.edu

University of South Carolina

BASICS2008 SUMMER SCHOOL
July 27 – August 2, 2008

Overview of talks

- Lecture 1: Overview and outlines
- Lecture 2: Generative models - preferential attachment schemes
- Lecture 3: Duplication models for biological networks
- Lecture 4: The rise of the giant component
- Lecture 5: The small world phenomenon: average distance and diameter
- Lecture 6: Spectrum of random graphs with given degrees

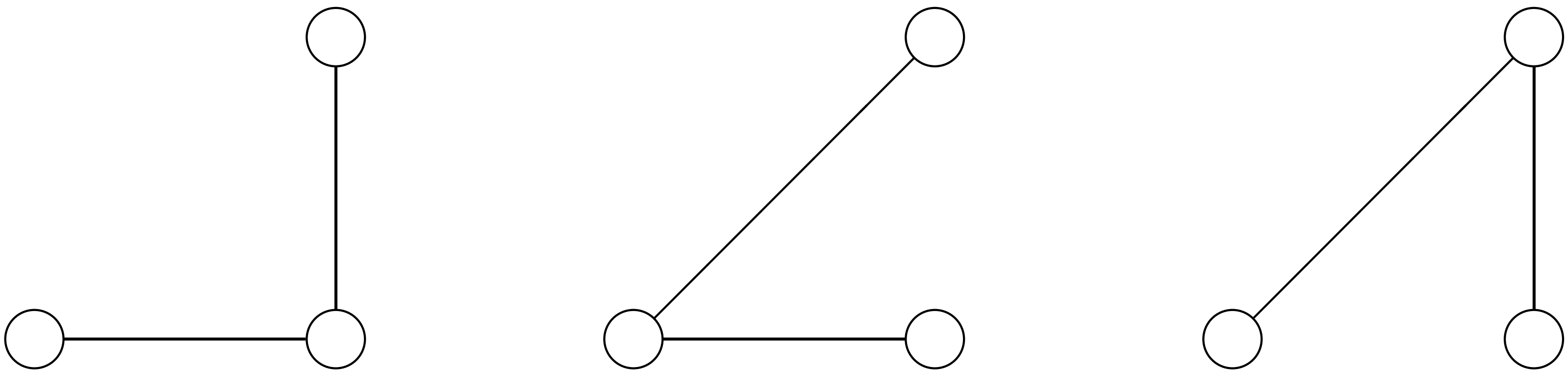
Random graphs

A random graph is a set of graphs together with a probability distribution on that set.

Random graphs

A random graph is a set of graphs together with a probability distribution on that set.

Example: A random graph on 3 vertices and 2 edges with the uniform distribution on it.

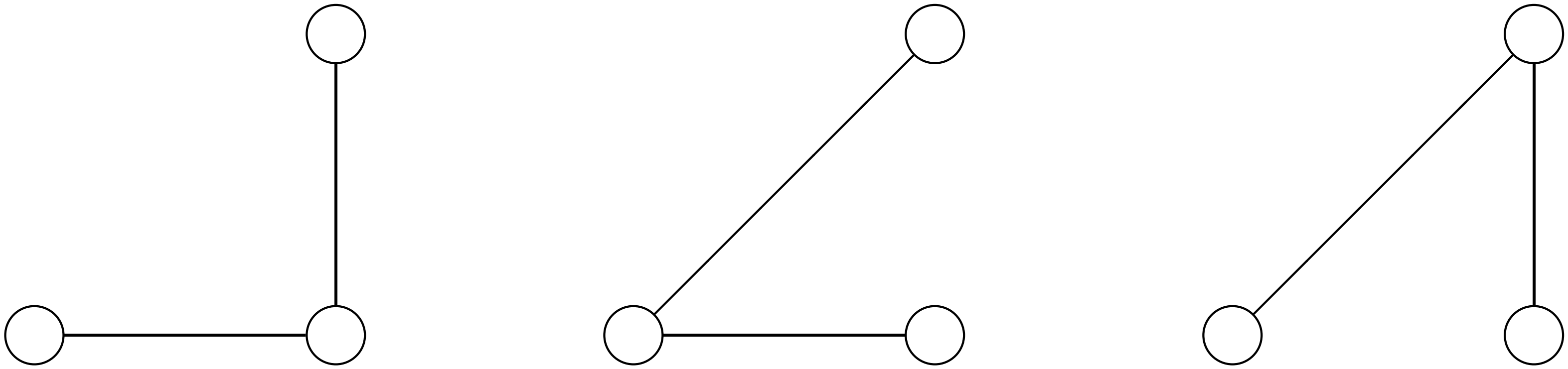


Probability $\frac{1}{3}$ Probability $\frac{1}{3}$ Probability $\frac{1}{3}$

Random graphs

A random graph is a set of graphs together with a probability distribution on that set.

Example: A random graph on 3 vertices and 2 edges with the uniform distribution on it.



Probability $\frac{1}{3}$ Probability $\frac{1}{3}$ Probability $\frac{1}{3}$

A random graph G *almost surely* satisfies a property P, if

$$Pr(G \text{ satisfies } P) = 1 - o_n(1).$$

Erdős-Rényi model $G(n, p)$

- n nodes

Erdős-Rényi model $G(n,p)$

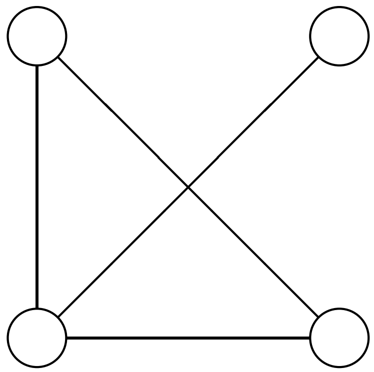
- n nodes
- For each pair of vertices, create an edge independently with probability p.

Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.

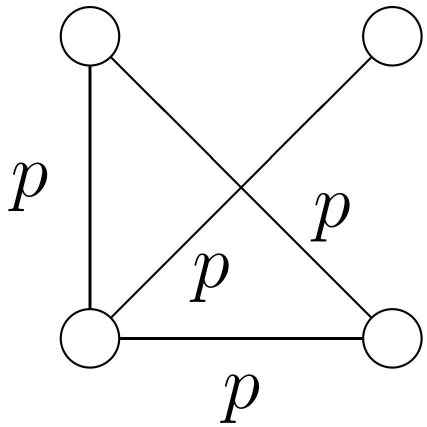
Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



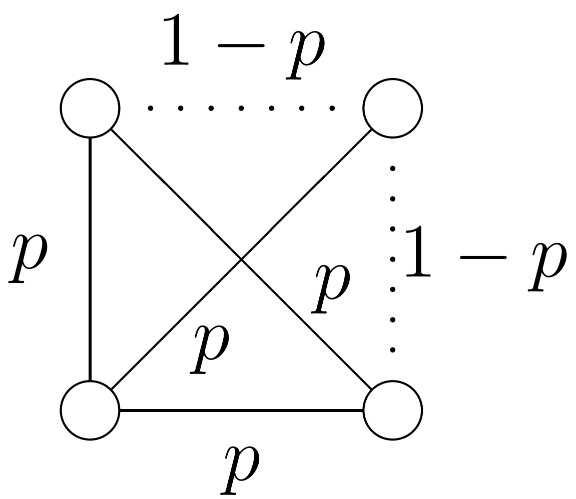
Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



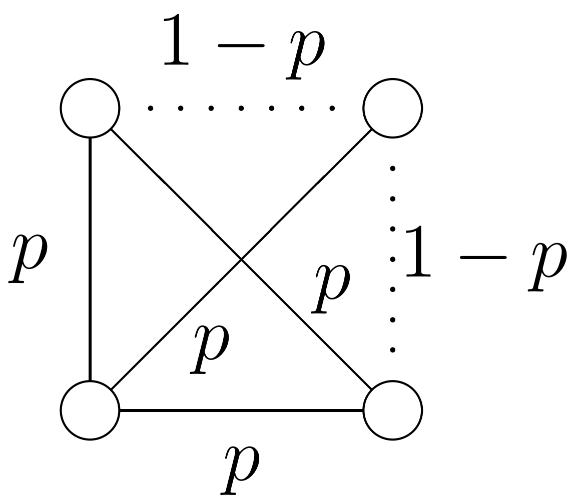
Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



Erdős-Rényi model $G(n,p)$

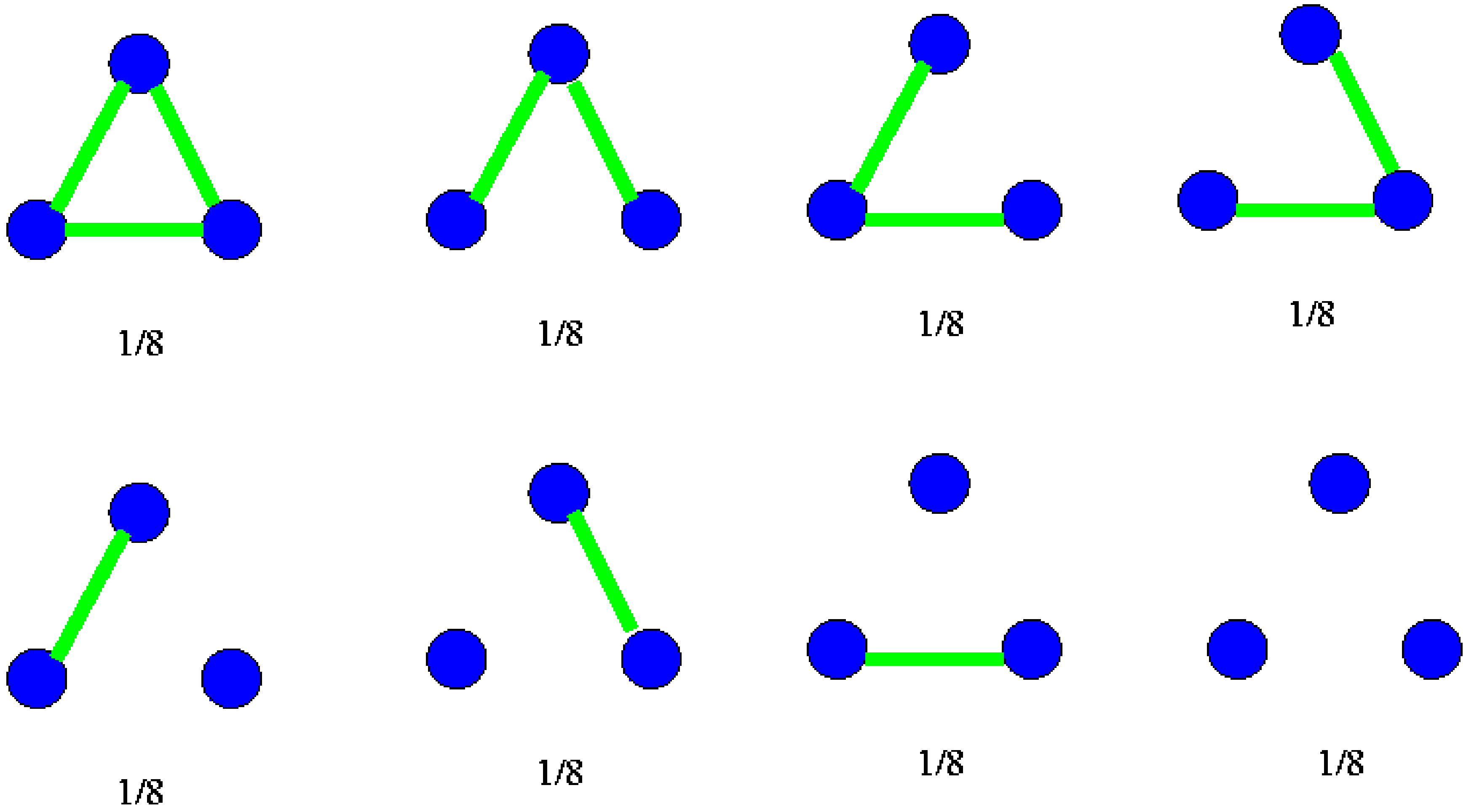
- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.



The probability of this graph is

$$p^4(1-p)^2.$$

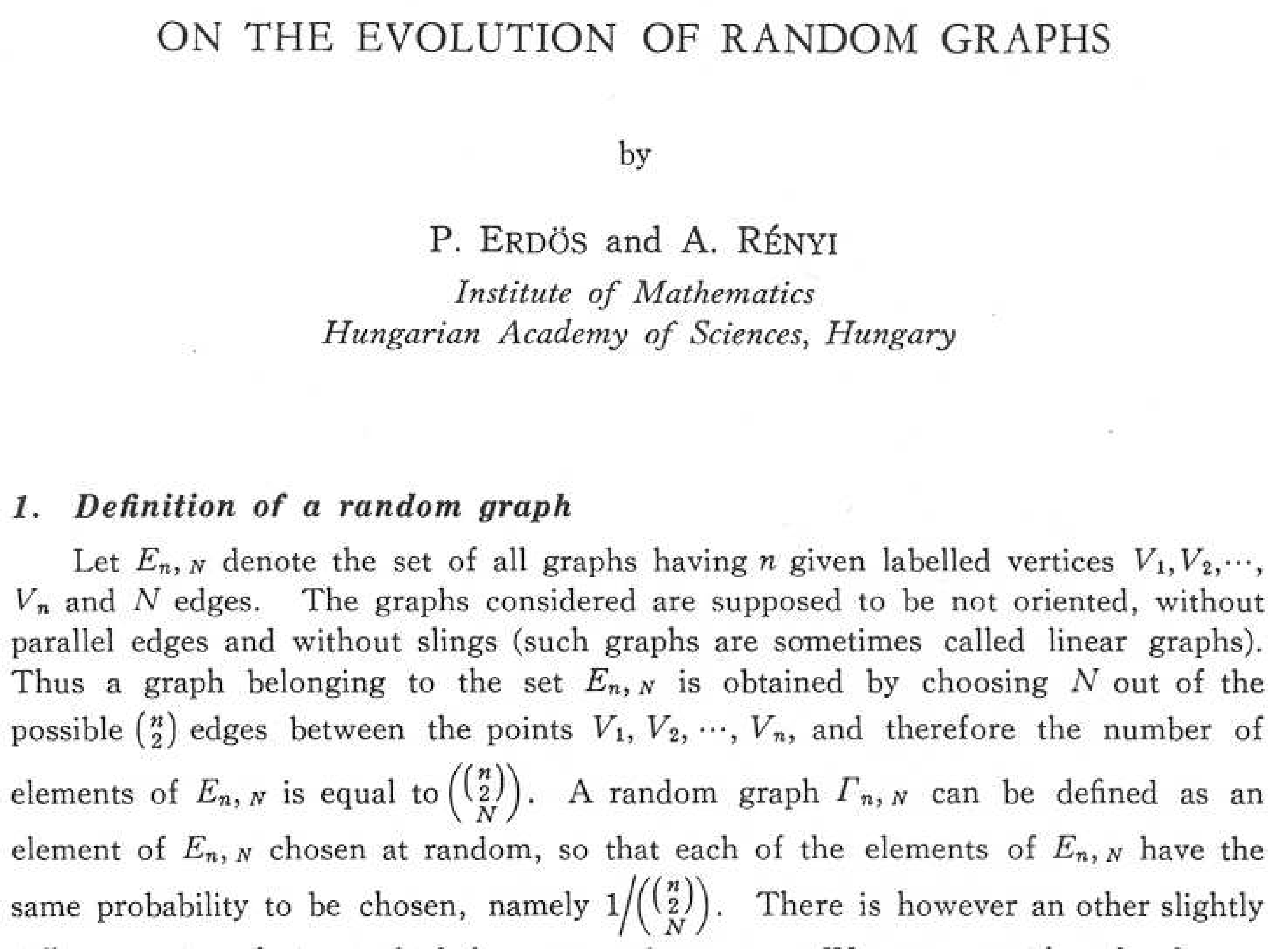
A example: $G(3, \frac{1}{2})$



The birth of random graph theory

Paul Erdős and A. Rényi, On the evolution of random graphs *Magyar Tud. Akad. Mat. Kut. Int. Kozl.* **5** (1960) 17-61.

The birth of random graph theory



ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDŐS and A. RÉNYI

Institute of Mathematics
Hungarian Academy of Sciences, Hungary

1. Definition of a random graph

Let $E_{n,N}$ denote the set of all graphs having n given labelled vertices $V_1, V_2, \cdots$, V_n and N edges. The graphs considered are supposed to be not oriented, without parallel edges and without slings (such graphs are sometimes called linear graphs). Thus a graph belonging to the set $E_{n,N}$ is obtained by choosing N out of the possible $\binom{n}{2}$ edges between the points $V_1, V_2, \cdots, V_n$, and therefore the number of elements of $E_{n,N}$ is equal to $\binom{\binom{n}{2}}{N}$. A random graph $\Gamma_{n,N}$ can be defined as an element of $E_{n,N}$ chosen at random, so that each of the elements of $E_{n,N}$ have the same probability to be chosen, namely $1/\binom{\binom{n}{2}}{N}$. There is however an other slightly

Evolution of $G(n, p)$

p	
0	the empty graph.
	disjoint union of trees.
$\frac{c}{n}$	cycles of any size.
$\frac{1}{n}$	the double jumps.
$\frac{c'}{n}$	one giant component, others are trees.
$\frac{\log n}{n}$	$G(n, p)$ is connected.
$\Omega(\frac{\log n}{n})$	
	connected and almost regular.
$\Omega(n^{\epsilon-1})$	finite diameter.
$\Theta(1)$	dense graphs, diameter is 2.
1	the complete graph.

Evolution of $G(n,p)$

Range I $p = o(1/n)$

The random graph $G_{n,p}$ is the disjoint union of trees. In fact, trees on k vertices, for $k = 3, 4, \ldots$ only appear when p is of the order $n^{-k/(k-1)}$.

Evolution of $G(n,p)$

Range I $p = o(1/n)$

The random graph $G_{n,p}$ is the disjoint union of trees. In fact, trees on k vertices, for $k = 3, 4, \ldots$ only appear when p is of the order $n^{-k/(k-1)}$.

Furthermore, for $p = cn^{-k/(k-1)}$ and $c > 0$, let $\tau_k(G)$ denote the number of connected components of G formed by trees on k vertices and $\lambda = (2c)^{k-1}k^{k-2}/k!$. Then,

$$\Pr(\tau_k(G_{n,p}) = j) \to \frac{\lambda^j e^{-\lambda}}{j!}$$

for $j = 0, 1, \ldots$ as $n \to \infty$.

Evolution of $G(n, p)$

Range II $p \sim c/n$ for $0 < c < 1$

- In this range of p, $G_{n,p}$ contains cycles of any given size with probability tending to a positive limit.

Evolution of $G(n, p)$

Range II $p \sim c/n$ for $0 < c < 1$

- In this range of p, $G_{n,p}$ contains cycles of any given size with probability tending to a positive limit.
- All connected components of $G_{n,p}$ are either trees or unicyclic components. Almost all (i.e., $n - o(n)$) vertices are in components which are trees.

Evolution of $G(n,p)$

Range II $p \sim c/n$ for $0 < c < 1$

- In this range of p, $G_{n,p}$ contains cycles of any given size with probability tending to a positive limit.
- All connected components of $G_{n,p}$ are either trees or unicyclic components. Almost all (i.e., $n - o(n)$) vertices are in components which are trees.
- The largest connected component of $G_{n,p}$ is a tree and has about $\frac{1}{\alpha}(\log n - \frac{5}{2}\log\log n)$ vertices, where $\alpha = c - 1 - \log c$.

Evolution of $G(n, p)$

Range III $p \sim 1/n + \mu/n$, the double jump

- If $\mu < 0$, the largest component has size $(\mu - \log(1 + \mu))^{-1} \log n + O(\log \log n)$.

Evolution of $G(n, p)$

Range III $p \sim 1/n + \mu/n$, the double jump

- If $\mu < 0$, the largest component has size $(\mu - \log(1+\mu))^{-1} \log n + O(\log \log n)$.
- If $\mu = 0$, the largest component has size of order $n^{2/3}$.

Evolution of $G(n,p)$

Range III $p \sim 1/n + \mu/n$, the double jump

- If $\mu < 0$, the largest component has size $(\mu - \log(1+\mu))^{-1}\log n + O(\log\log n)$.
- If $\mu = 0$, the largest component has size of order $n^{2/3}$.
- If $\mu > 0$, there is a unique giant component of size αn where $\mu = -\alpha^{-1}\log(1-\alpha) - 1$.

Evolution of $G(n,p)$

Range III $p \sim 1/n + \mu/n$, the double jump

- If $\mu < 0$, the largest component has size $(\mu - \log(1+\mu))^{-1}\log n + O(\log\log n)$.
- If $\mu = 0$, the largest component has size of order $n^{2/3}$.
- If $\mu > 0$, there is a unique giant component of size αn where $\mu = -\alpha^{-1}\log(1-\alpha) - 1$.
- Bollobás showed that a component of size at least $n^{2/3}$ in $G_{n,p}$ is almost always unique if p exceeds $1/n + 4(\log n)^{1/2} n^{-4/3}$.

Evolution of $G(n, p)$

Range IV $p \sim c/n$ for $c > 1$

- Except for one "giant" component, all the other components are relatively small, and most of them are trees.

Evolution of $G(n,p)$

Range IV $p \sim c/n$ for $c > 1$

- Except for one “giant” component, all the other components are relatively small, and most of them are trees.
- The total number of vertices in components which are trees is approximately $n - f(c)n + o(n)$.

Evolution of $G(n,p)$

Range IV $p \sim c/n$ for $c > 1$

- Except for one "giant" component, all the other components are relatively small, and most of them are trees.
- The total number of vertices in components which are trees is approximately $n - f(c)n + o(n)$.
- The largest connected component of $G_{n,p}$ has approximately $f(c)n$ vertices, where

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

Evolution of $G(n, p)$

Range V $p = c \log n / n$ with $c \geq 1$

- The graph $G_{n,p}$ almost surely becomes connected.

Evolution of $G(n,p)$

Range V $p = c\log n/n$ with $c \geq 1$

- The graph $G_{n,p}$ almost surely becomes connected.
- If

$$p = \frac{\log n}{kn} + \frac{(k-1)\log\log n}{kn} + \frac{y}{n} + o(\frac{1}{n}),$$

then there are only trees of size at most k except for the giant component. The distribution of the number of trees of k vertices again has a Poisson distribution with mean value $\frac{e^{-ky}}{k \cdot k!}$.

Evolution of $G(n, p)$

Range VI $p \sim \omega(n) \log n / n$ where $\omega(n) \to \infty$.

In this range, $G_{n,p}$ is not only almost surely connected, but the degrees of almost all vertices are asymptotically equal.

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.
- The graph H has probability

$$\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij}).$$

Model $G(w_1, w_2, \ldots, w_n)$

Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.
- The graph H has probability

$$\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij}).$$

- The expected degree of vertex i is w_i.

Model $G(w_1, w_2, \ldots, w_n)$

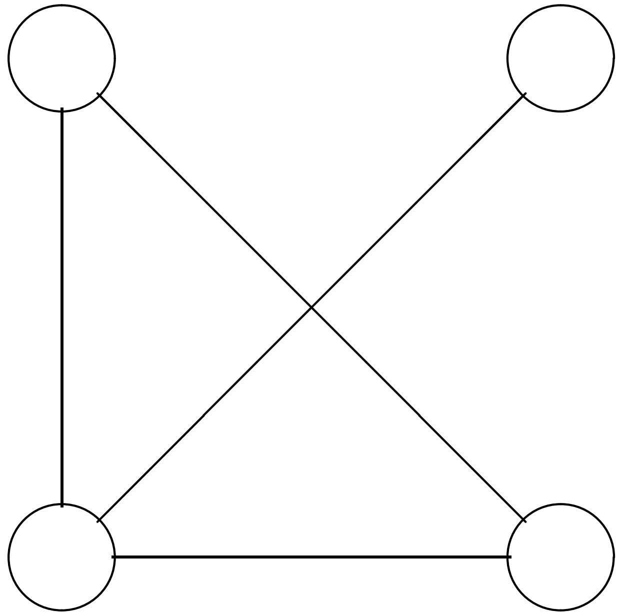
Random graph model with given expected degree sequence

- n nodes with weights $w_1, w_2, \ldots, w_n$.
- For each pair (i, j), create an edge independently with probability $p_{ij} = w_i w_j \rho$, where $\rho = \frac{1}{\sum_{i=1}^n w_i}$.
- The graph H has probability

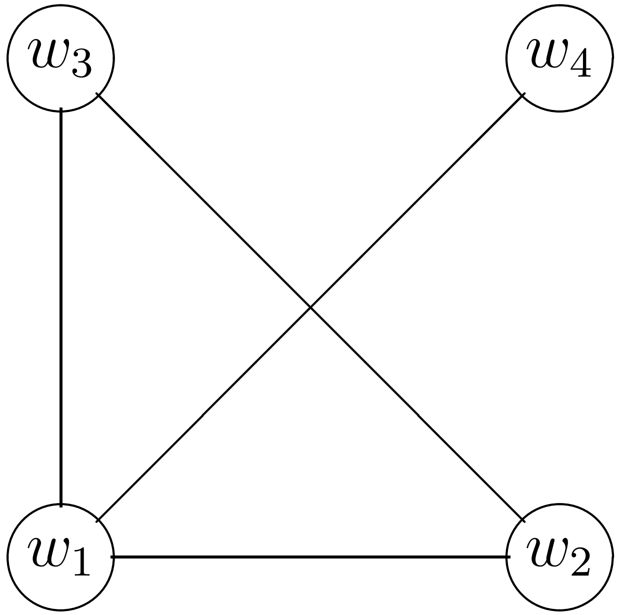
$$\prod_{ij \in E(H)} p_{ij} \prod_{ij \notin E(H)} (1 - p_{ij}).$$

- The expected degree of vertex i is w_i.

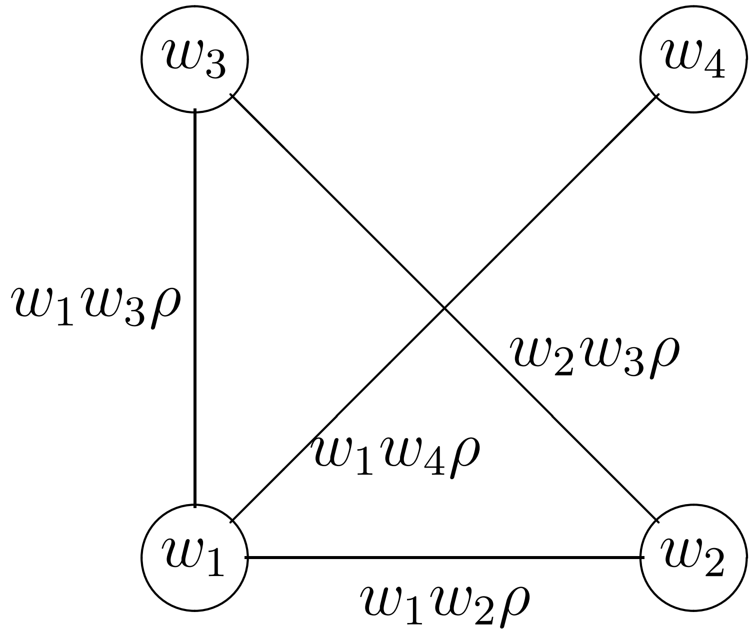
An example: $G(w_1, w_2, w_3, w_4)$



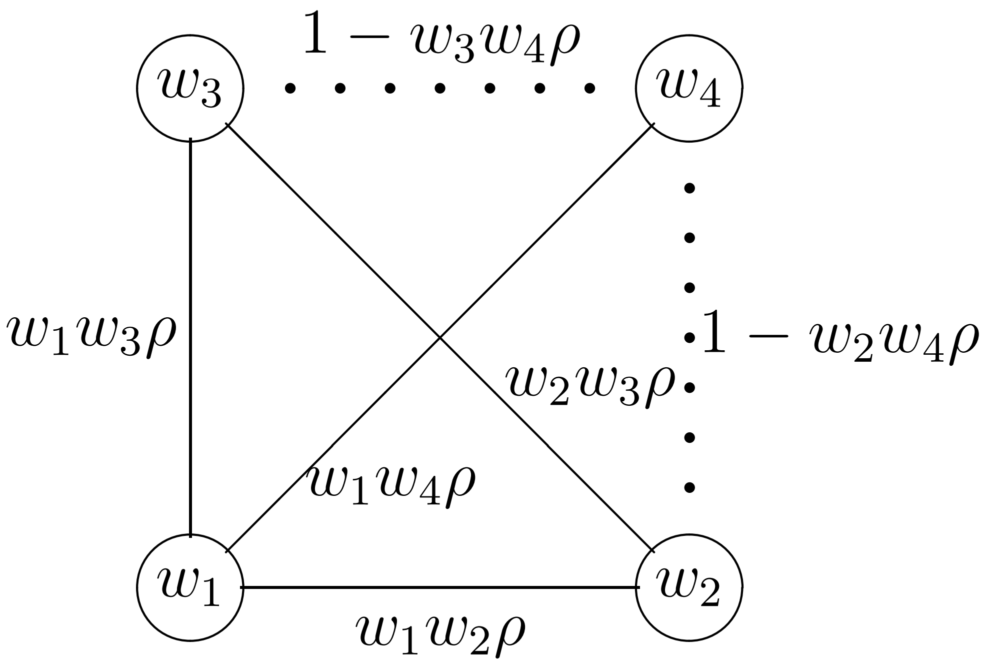
An example: $G(w_1, w_2, w_3, w_4)$



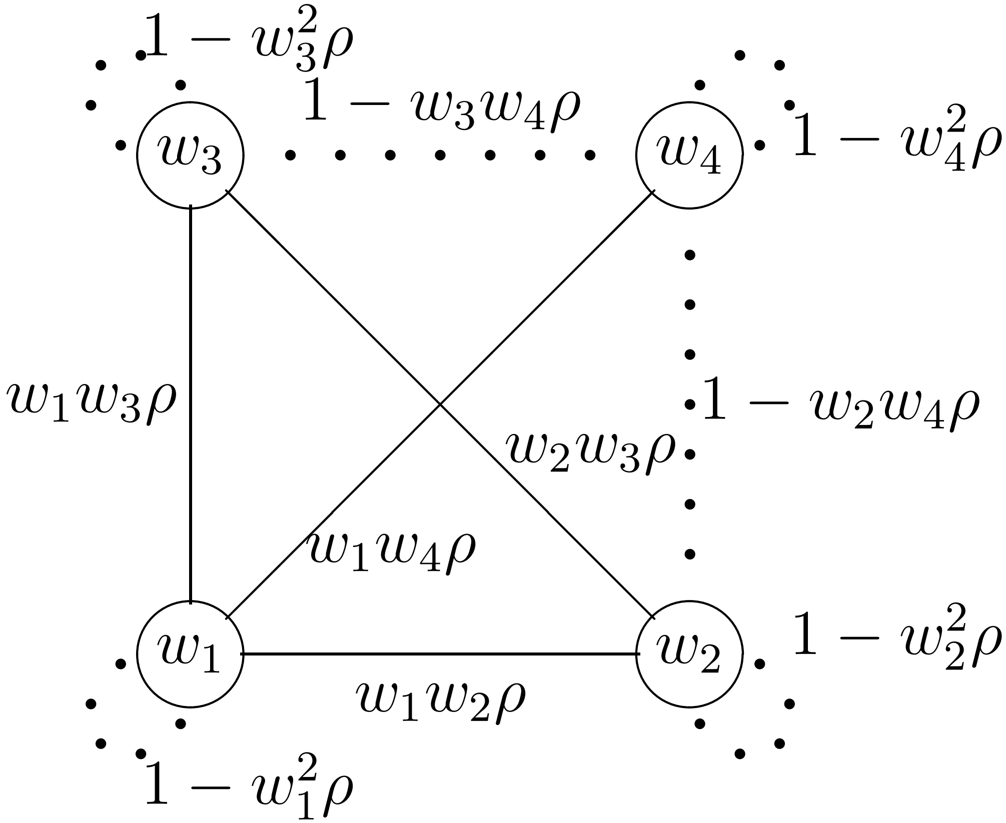
An example: $G(w_1, w_2, w_3, w_4)$



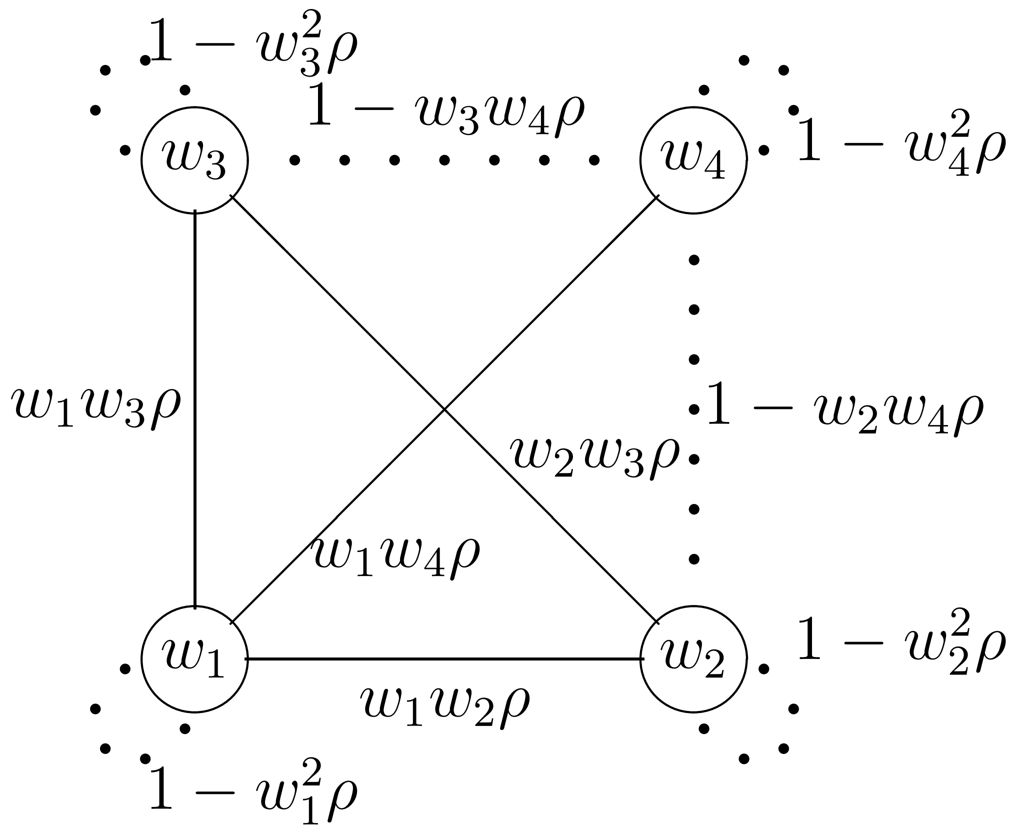
An example: $G(w_1, w_2, w_3, w_4)$



An example: $G(w_1, w_2, w_3, w_4)$



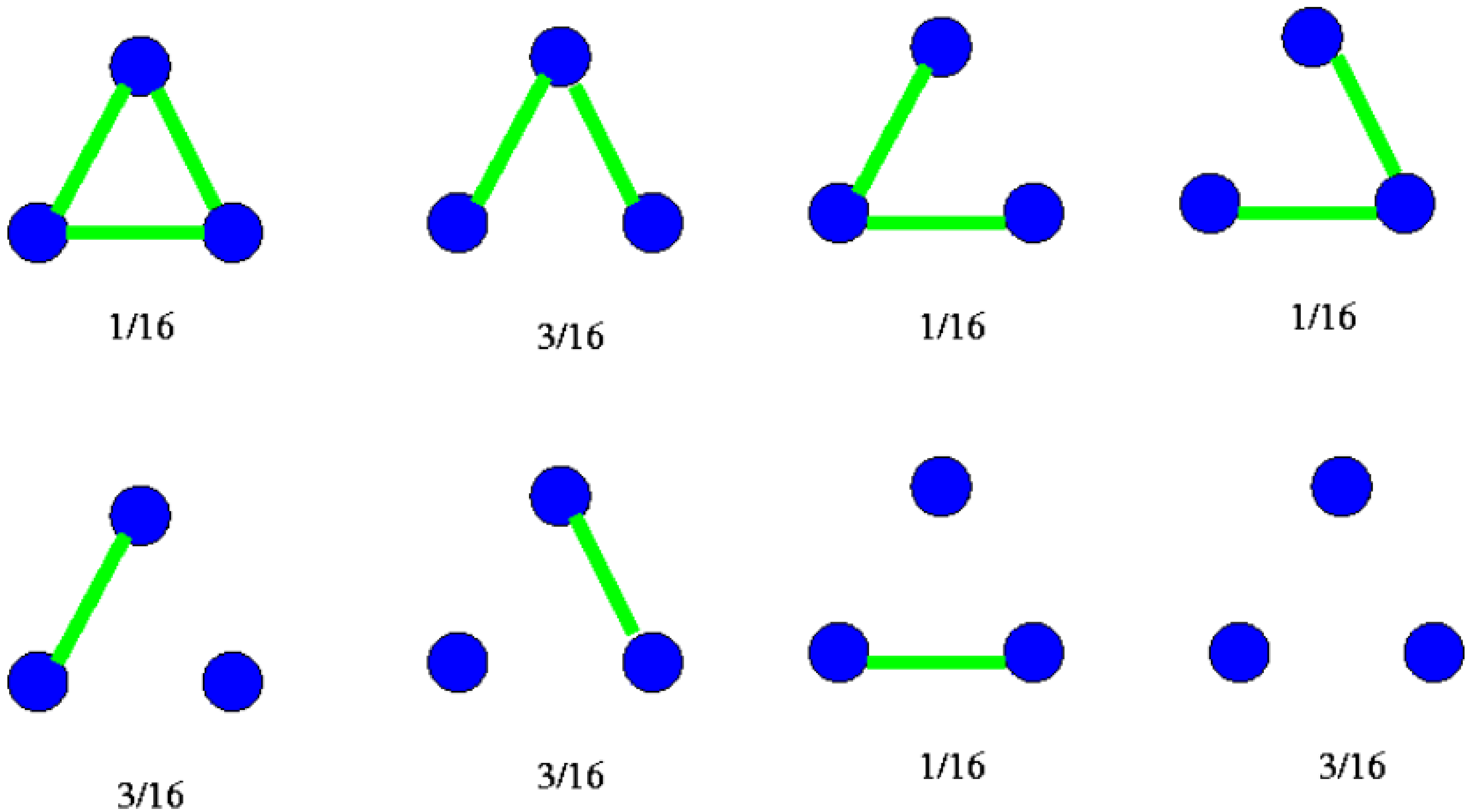
An example: $G(w_1, w_2, w_3, w_4)$



The probability of the graph is

$$w_1^3w_2^2w_3^2w_4\rho^4(1 - w_2w_4\rho) \times (1 - w_3w_4\rho)\prod_{i=1}^{4}(1 - w_i^2\rho).$$

A example: $G(1,2,1)$



Loops are omitted here.

Notations

For $G = G(w_1, \ldots, w_n)$, let

- $d = \frac{1}{n}\sum_{i=1}^n w_i$
- $\tilde{d} = \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i}$.
- The volume of S: $\mathrm{Vol}(S) = \sum_{i \in S} w_i$.

Notations

For $G = G(w_1, \ldots, w_n)$, let

- $d = \frac{1}{n}\sum_{i=1}^n w_i$
- $\tilde{d} = \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i}$.
- The volume of S: $\mathrm{Vol}(S) = \sum_{i \in S} w_i$.

We have

$$\tilde{d} \geq d$$

"$=$" holds if and only if $w_1 = \cdots = w_n$.

Notations

For $G = G(w_1, \ldots, w_n)$, let

- $d = \frac{1}{n}\sum_{i=1}^n w_i$
- $\tilde{d} = \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i}$.
- The volume of S: $\mathrm{Vol}(S) = \sum_{i\in S} w_i$.

We have

$$\tilde{d} \geq d$$

"$=$" holds if and only if $w_1 = \cdots = w_n$.

A connected component S is called a giant component if

$$\mathrm{vol}(S) = \Theta(\mathrm{vol}(G)).$$

Classical result on $G(n,p)$

- If $np < 1$, almost surely there is no giant component.
- If $np > 1$, almost surely there is a unique giant component.

$$\tilde{d} = d = np.$$

Four questions

- Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?

Four questions

- Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?

- Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?

Four questions

- Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?
- Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?
- Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $\tilde{d} < 1$?

Four questions

- Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?
- Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?
- Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $\tilde{d} < 1$?
- Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $d > 1$?

Case $d < 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?

Case $d < 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?

No.

Case $d < 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $d < 1$?

No. A counter-example: $G(\frac{n}{2}, 0) + G(\frac{n}{2}, \frac{3}{n})$.
Since $G(\frac{n}{2}, \frac{3}{n})$ has

$$n'p' = \frac{n}{2}\frac{3}{n} = \frac{3}{2} > 1.$$

It has a giant component. But as the whole graph, the average degree is $d = \frac{3}{4} < 1$.

Case $\tilde{d} > 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?

Case $\tilde{d} > 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?

No.

Case $\tilde{d} > 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?

No. A counter-example: $V = S \cup T$ (with $|S| = \log n$), weights are defined as follows.

$$w_i = \begin{cases} \sqrt{n} & \text{if } v_i \in S \\ 1 - \epsilon & \text{otherwise.} \end{cases}$$

Case $\tilde{d} > 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?

No. A counter-example: $V = S \cup T$ (with $|S| = \log n$), weights are defined as follows.

$$w_i = \begin{cases} \sqrt{n} & \text{if } v_i \in S \\ 1 - \epsilon & \text{otherwise.} \end{cases}$$

Every component in $G|_T$ has size at most $O(\log n)$. Adding S can join at most $O(\sqrt{n}\log^2 n)$ vertices in T. The volume of maximum component is at most $O(\sqrt{n}\log^2 n)$.

Case $\tilde{d} > 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has a giant component if $\tilde{d} > 1$?

No. A counter-example: $V = S \cup T$ (with $|S| = \log n$), weights are defined as follows.

$$w_i = \begin{cases} \sqrt{n} & \text{if } v_i \in S \\ 1 - \epsilon & \text{otherwise.} \end{cases}$$

Every component in $G|_T$ has size at most $O(\log n)$. Adding S can join at most $O(\sqrt{n}\log^2 n)$ vertices in T. The volume of maximum component is at most $O(\sqrt{n}\log^2 n)$.

$$\tilde{d} = \frac{n\log n + (1-\epsilon)(n - \log n)}{\sqrt{n}\log n + \sqrt{(1-\epsilon)}(n - \log n)} > \log n.$$

Case $\tilde{d} < 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $\tilde{d} < 1$?

Case $\tilde{d} < 1$

Is it true that $G(w_1, \ldots, w_n)$ almost surely has no giant component if $\tilde{d} < 1$?

Yes. Chung and Lu (2001) Suppose that $\tilde{d} < 1 - \delta$. For any $\alpha > 0$, with probability at least $1 - \frac{d\tilde{d}^2}{\alpha^2(1-\tilde{d})}$, a random graph G in $G(w_1, \ldots, w_n)$ has all connected components with volume at most $\alpha\sqrt{n}$.

Proof

Let $x = \Pr(\exists$ a component S, $\mathrm{vol}(S) \geq \alpha\sqrt{n})$.

Proof

Let $x = \Pr(\exists$ a component S, $\mathrm{vol}(S) \geq \alpha\sqrt{n})$.

Choose two vertices u and v randomly with probability proportional to their weights.

Proof

Let $x = \Pr(\exists$ a component S, $\mathrm{vol}(S) \geq \alpha\sqrt{n})$.

Choose two vertices u and v randomly with probability proportional to their weights.

Two ways to estimate $z = \Pr(u \sim v)$ the probability that u and v are connected by a path.

Proof

Let $x = \Pr(\exists$ a component S, $\mathrm{vol}(S) \geq \alpha\sqrt{n})$.

Choose two vertices u and v randomly with probability proportional to their weights.

Two ways to estimate $z = \Pr(u \sim v)$ the probability that u and v are connected by a path.

One the one hand,

$$\begin{aligned} z &\geq \Pr(u \sim v, \exists \text{ a component } S \ \mathrm{vol}(S) \geq \alpha\sqrt{n}) \\ &= \Pr(u \sim v \mid \exists \text{ a component } S \ \mathrm{vol}(S) \geq \alpha\sqrt{n})x \\ &\geq \Pr(u, v \in S \mid \exists \text{ a component } S \ \mathrm{vol}(S) \geq \alpha\sqrt{n})x \\ &\geq \alpha^2 n \rho^2 x. \end{aligned}$$

Proof

On the other hand, the probability $P_k(u,v)$ of u and v being connected by a path of length $k+1$ is at most

$$\begin{aligned} P_k(u,v) &\leq \sum_{i_1,i_2,\ldots,i_k} (w_u w_{i_1}\rho)\,(w_{i_1} w_{i_2}\rho)\cdots(w_{i_k} w_v \rho) \\ &= w_u w_v \rho \tilde{d}^k. \end{aligned}$$

Proof

On the other hand, the probability $P_k(u,v)$ of u and v being connected by a path of length $k+1$ is at most

$$\begin{aligned} P_k(u,v) &\leq \sum_{i_1,i_2,\ldots,i_k} (w_u w_{i_1}\rho)\,(w_{i_1}w_{i_2}\rho)\cdots(w_{i_k}w_v\rho) \\ &= w_u w_v \rho \tilde{d}^k. \end{aligned}$$

The probability that u and v are connected is at most

$$\sum_{k=0}^{n} P_k(u,v) \leq \sum_{k\geq 0} w_u w_v \rho \tilde{d}^k = \frac{1}{1-\tilde{d}} w_u w_v \rho.$$

Proof

$$z \leq \sum_{u,v} w_u \rho \; w_v \rho \; \frac{1}{1-\tilde{w}} w_u w_v \rho = \frac{\tilde{d}^2}{1-\tilde{d}} \rho.$$

Combining this with $z \geq x\alpha^2 n\rho^2$
we have

$$\alpha^2 x n \rho^2 \leq \frac{\tilde{d}^2}{1-\tilde{d}} \rho$$

which implies that

$$x \leq \frac{d\tilde{d}^2}{\alpha^2(1-\tilde{d})}.$$

The proof is finished. □

Case $d > 1$

Gap theorem:

- Almost surely G has a unique giant component (GCC).

$$\mathrm{vol}(GCC) \geq \begin{cases} (1 - \frac{2}{\sqrt{de}} + o(1))\mathrm{Vol}(G) & \text{if } d \geq \frac{4}{e}. \\ (1 - \frac{1+\log d}{d} + o(1))\mathrm{Vol}(G) & \text{if } d < 2. \end{cases}$$

- The second largest component almost surely has size at most $(1 + o(1))\mu(d)\log n$, where

$$\mu(d) = \begin{cases} \frac{1}{1+\log d - \log 4} & \text{if } d > 4/e; \\ \frac{1}{d-1-\log d} & \text{if } 1 < d < 2. \end{cases}$$

Matrix-tree theorem

Kirchhofff (1847) The number of spanning trees in a graph G is equal to any cofactor of $L = D - A$, where $D = diag(d_1, \ldots, d_n)$ is the diagonal degree matrix and A is the adjacency matrix.

Matrix-tree theorem

Kirchhofff (1847) The number of spanning trees in a graph G is equal to any cofactor of $L = D - A$, where $D = diag(d_1, \ldots, d_n)$ is the diagonal degree matrix and A is the adjacency matrix.

The matrix-tree theorem holds for weighted graphs.

$$\sum_{T} \prod_{f \in E(T)} w_e = |\det M|.$$

Here M is obtained by deleting one row and one column from $D - A$.

A set S as component

Let $S = \{v_{i_1}, v_{i_2} \ldots, v_{i_k}\}$ with weights $w_{i_1}, w_{i_2}, \ldots, w_{i_k}$. The probability that there is no edge leaving S is

$$\begin{aligned} \prod_{v_i \in S, v_j \notin S} & (1 - w_i w_j \rho) \\ \approx \quad & e^{-\rho \sum_{v_i \in S, v_j \notin S} w_i w_j} \\ = \quad & e^{-\rho \mathrm{vol}(S)(\mathrm{vol}(G) - \mathrm{vol}(S))}. \end{aligned}$$

A set S as component

Let $S = \{v_{i_1}, v_{i_2} \ldots, v_{i_k}\}$ with weights $w_{i_1}, w_{i_2}, \ldots, w_{i_k}$. The probability that there is no edge leaving S is

$$\begin{aligned} \prod_{v_i \in S, v_j \notin S} & (1 - w_i w_j \rho) \\ \approx \quad & e^{-\rho \sum_{v_i \in S, v_j \notin S} w_i w_j} \\ = \quad & e^{-\rho \mathrm{vol}(S)(\mathrm{vol}(G) - \mathrm{vol}(S))}. \end{aligned}$$

The probability $G\mid_S$ is connected is at most

$$\sum_T \prod_{(v_{i_j} v_{i_l}) \in E(T)} w_{i_j} w_{i_l} \rho = w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1}.$$

Computation is done by matrix-tree theorem.

Detail computation

Let

$$A = \begin{pmatrix} 0 & w_{i_1} w_{i_2} \rho & \cdots & w_{i_1} w_{i_k} \rho \\ w_{i_2} w_{i_1} \rho & 0 & \cdots & w_{i_2} w_{i_k} \rho \\ \vdots & \vdots & \ddots & \vdots \\ w_{i_k} w_{i_1} \rho & w_{i_k} w_{i_2} \rho & \cdots & 0 \end{pmatrix}$$

and D is the diagonal matrix $diag(w_{i_1}(\mathrm{vol}(S) - w_{i_1})\rho, \ldots, w_{i_k}(\mathrm{vol}(S) w_{i_k} - w_{i_k})\rho)$. Then compute the determinant of any $k-1 \times k-1$ sub-matrix.

A set S as component

The probability that S is a component is at most

$$\sum_{S} w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{vol}(S)(1-\mathrm{vol}(S)/\mathrm{vol}(G))}.$$

The probability that there exists a connected component on size k with volume less than $\epsilon \mathrm{vol}(G)$ is at most

$$f(k,\epsilon) = \sum_{|S|=k} w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{vol}(S)(1-\epsilon)}.$$

Case $d > \frac{4}{e(1-\epsilon)^2}$

$$
\begin{aligned}
f(k,\epsilon) &= \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{vol}(S)(1-\epsilon)} \\
&\leq \sum_S \frac{\rho^{k-1}}{k^k} \mathrm{vol}(S)^{2k-2} e^{-\mathrm{vol}(S)(1-\epsilon)} \\
&\leq \sum_S \frac{\rho^{k-1}}{k^k} (\frac{2k-2}{1-\epsilon})^{2k-2} e^{-(2k-2)} \\
&\leq \frac{n^k}{k!} \frac{\rho^{k-1}}{k^k} (\frac{2k-2}{1-\epsilon})^{2k-2} e^{-(2k-2)} \\
&\leq \frac{1}{4\rho(k-1)^2} (\frac{4}{de(1-\epsilon)^2})^k
\end{aligned}
$$

Case $\frac{1}{1-\epsilon} < d < \frac{2}{1-\epsilon}$

First, we split $f(k, \epsilon)$ into two parts as follows:

$$f(k, \epsilon) = f_1(k, \epsilon) + f_2(k, \epsilon)$$

where

$$f_1(k, \epsilon) = \sum_{\mathrm{vol}(S) < dk} w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{vol}(S)(1-\epsilon)}$$

$$f_2(k, \epsilon) = \sum_{\mathrm{vol}(S) \geq dk} w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{vol}(S)(1-\epsilon)}$$

Bounding $f_1(k,\epsilon)$

$$
\begin{aligned}
f_1(k,\epsilon) &= \sum_{\mathrm{vol}(S)<dk} w_{i_1}\cdots w_{i_k}\mathrm{vol}(S)^{k-2}\rho^{k-1}e^{-\mathrm{vol}(S)(1-\epsilon)} \\
&\le \sum_{\mathrm{vol}(S)<dk} \frac{\rho^{k-1}}{k^k}\mathrm{vol}(S)^{2k-2}e^{-\mathrm{vol}(S)(1-\epsilon)} \\
&\le \sum_{\mathrm{vol}(S)<dk} \frac{\rho^{k-1}}{k^k}(dk)^{2k-2}e^{-dk(1-\epsilon)} \\
&\le \binom{n}{k}\frac{\rho^{k-1}}{k^k}(dk)^{2k-2}e^{-dk(1-\epsilon)} \\
&\le \frac{n}{dk^2}\left(\frac{d}{e^{d(1-\epsilon)-1}}\right)^k
\end{aligned}
$$

Bounding $f_2(k, \epsilon)$

$$
\begin{aligned}
f_2(k,\epsilon) &= \sum_{\mathrm{vol}(S)\geq dk} w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{vol}(S)(1-\epsilon)} \\
&\leq \sum_{\mathrm{vol}(S)\geq dk} w_{i_1} \cdots w_{i_k} \rho^{k-1} (dk)^{k-2} e^{-dk(1-\epsilon)} \\
&\leq \sum_{S} w_{i_1} w_{i_2} \cdots w_{i_k} \rho^{k-1} (dk)^{k-2} e^{-dk(1-\epsilon)} \\
&< \frac{\mathrm{vol}(G)^k}{k!} \rho^{k-1} (dk)^{k-2} e^{-dk(1-\epsilon)} \\
&\leq \frac{n}{dk^2} \left(\frac{d}{e^{(d(1-\epsilon)-1)}}\right)^k
\end{aligned}
$$

Put together

If $d > \frac{4}{e(1-\epsilon)^2}$, then

$$f(k,\epsilon) \leq \frac{1}{4\rho(k-1)^2}\left(\frac{4}{de(1-\epsilon)^2}\right)^k.$$

If $\frac{1}{1-\epsilon} < d < \frac{2}{1-\epsilon}$, then

$$f(k,\epsilon) \leq 2\frac{n}{dk^2}\left(\frac{d}{e^{(d(1-\epsilon)-1)}}\right)^k.$$

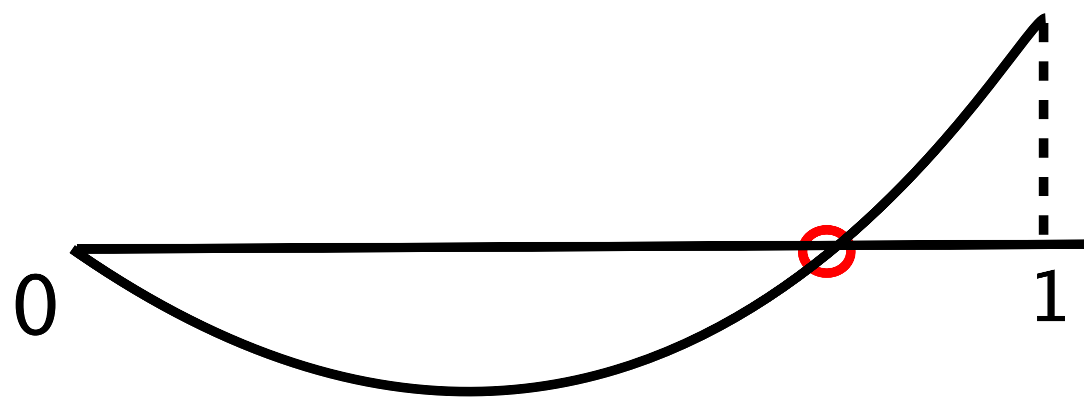
Choose $k = \mu(d)\log n$, then $f(k,\epsilon) = o(1)$. The gap theorem is proved.

Volume of Giant Component

Chung and Lu (2004)

If the average degree is strictly greater than 1, then almost surely the giant component in a graph G in $G(\mathbf{w})$ has volume $(\lambda_0 + O(\sqrt{n}\frac{\log^{3.5} n}{\mathrm{Vol}(G)}))\mathrm{Vol}(G)$, where λ_0 is the unique positive root of the following equation:

$$\sum_{i=1}^{n} w_i e^{-w_i\lambda} = (1-\lambda)\sum_{i=1}^{n} w_i.$$



Sketch proof

- With probability at least $1 - 2n^{-k}$, a vertex with weight greater than $\max\{8k, 2(k+1+o(1))\mu(d)\} \log n$ is in the GCC.

Sketch proof

- With probability at least $1 - 2n^{-k}$, a vertex with weight greater than $\max\{8k, 2(k+1+o(1))\mu(d)\} \log n$ is in the GCC.
- For any $k > 2$, with probability at least $1 - 6n^{-k+2}$, we have $|\mathrm{Vol}(GCC) - \mathrm{E}(\mathrm{Vol}(GCC))| \leq 2C_1(k+1)^2\sqrt{k-2}\sqrt{n}\log^{2.5} n$, where $C_1 = 10\mu(d) + 2\mu(d)^2$.

Sketch proof

- With probability at least $1 - 2n^{-k}$, a vertex with weight greater than $\max\{8k, 2(k+1+o(1))\mu(d)\}\log n$ is in the GCC.
- For any $k > 2$, with probability at least $1 - 6n^{-k+2}$, we have $|\mathrm{Vol}(GCC) - \mathrm{E}(\mathrm{Vol}(GCC))| \leq 2C_1(k+1)^2\sqrt{k-2}\sqrt{n}\log^{2.5} n$, where $C_1 = 10\mu(d) + 2\mu(d)^2$.
- $\mathrm{Vol}(G) - \mathrm{E}(\mathrm{vol}(GCC)) = \sum_{w_v < C_k \log n} w_v e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} + O(k^3\sqrt{n}\log^{3.5} n)$.

Lagrange inversion formula

Suppose that z is a function of x and y in terms of another analytic function ϕ as follows:

$$z = x + y\phi(z).$$

Then z can be written as a power series in y as follows:

$$z = x + \sum_{k=1}^{\infty} \frac{y^k}{k!} D^{(k-1)} \phi^k(x)$$

where $D^{(t)}$ denotes the t-th derivative.

Apply it to $G(n,p)$

For the $G(n,p)$, the equation is simply $e^{-d\lambda} = (1-\lambda)$. Let $\lambda = 1 - \frac{z}{d}$. We have $z = de^{-d}e^{z}$.

Apply it to $G(n,p)$

For the $G(n,p)$, the equation is simply $e^{-d\lambda} = (1-\lambda)$. Let $\lambda = 1 - \frac{z}{d}$. We have $z = de^{-d}e^{z}$. We apply Lagrange inversion formula with $x = 0$, $y = de^{-d}$, and $\phi(z) = e^{z}$. Then we have

$$\begin{aligned} z &= \sum_{k=1}^{\infty} \frac{y^k}{k!} D^{(k-1)} e^{kx} \mid_{x=0} \\ &= \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (de^{-d})^k \end{aligned}$$

This is exactly Erdős and Rényi's result on $G(n,p)$.

$G(n,p)$ verse $G(w_1,\ldots,w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?

$G(n,p)$ verse $G(w_1,\ldots,w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?

Chung Lu (2004)

- Yes, for $1 < d \leq \frac{e}{e-1}$.

$G(n,p)$ verse $G(w_1,\ldots,w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?
Chung Lu (2004)

- Yes, for $1 < d \leq \frac{e}{e-1}$.
- No, for sufficiently large d.

$G(n,p)$ **verse** $G(w_1, \ldots, w_n)$

Question: Does the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume?

Chung Lu (2004)

- Yes, for $1 < d \leq \frac{e}{e-1}$.
- No, for sufficiently large d.
- When $d \geq \frac{4}{e}$, almost surely the giant component of $G(w_1, \ldots, w_n)$ has volume at least

$$\left(\frac{1}{2}\left(1+\sqrt{1-\frac{4}{de}}\right)+o(1)\right)\mathrm{Vol}(G).$$

This is asymptotically best possible.

Sizes and edges in GCC

Chung, Lu (2004) *If the expected average degree is strictly greater than* 1*, then almost surely the giant component in a random graph of given expected degrees* w_i*,* $i = 1, \ldots, n$*, has* $n - \sum_{i=1}^{n} e^{-w_i \lambda_0} + O(\sqrt{n} \log^{3.5} n)$ *vertices and* $(\lambda_0 - \frac{1}{2}\lambda_0^2)\mathrm{Vol}(G) + O(\sqrt{\mathrm{Vol}(G)} \log^{3.5} n)$ *edges.*

In the collaboration graph

$$\lambda_0(2-\lambda_0) \approx \frac{\mathrm{Vol}(GCC)}{\mathrm{Vol}(G)} \approx \frac{248000}{284000}.$$

We have $\lambda_0 \approx 0.644$.
Let n_k denote the number of vertices of degree k. We have

$$n_k \approx \mathrm{E}(n_k) \approx \sum_{i\geq 0} \frac{w_i^k}{k!} e^{-w_i}.$$

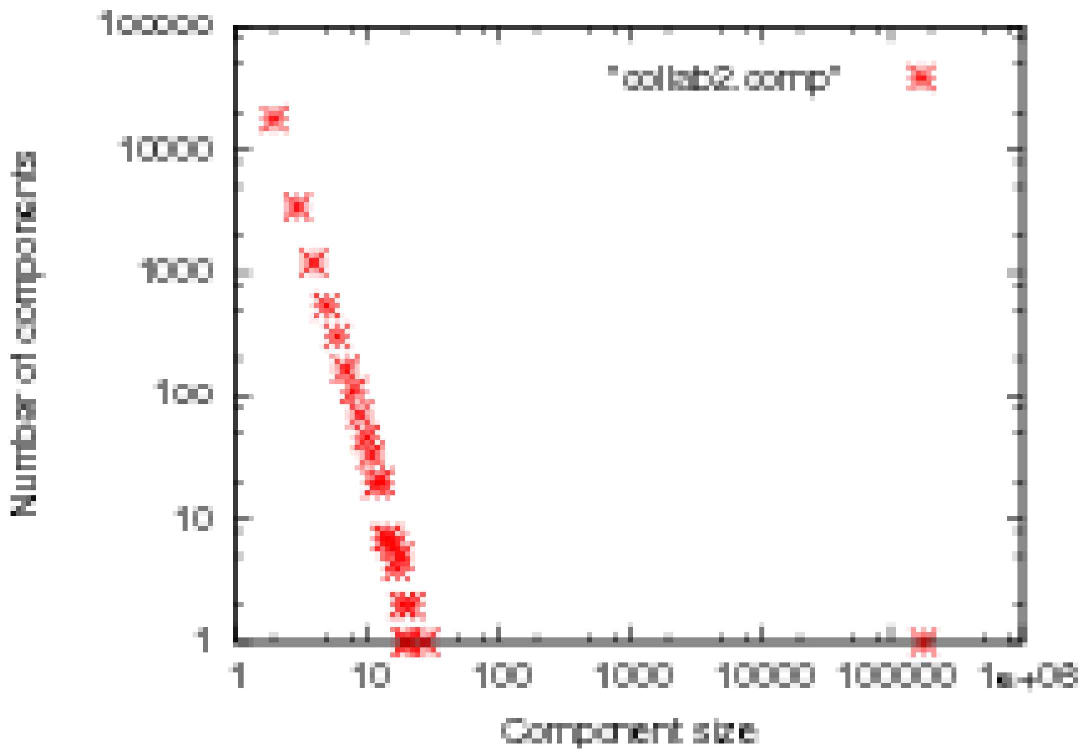
n_0	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
166381	145872	34227	16426	9913	6670	4643	3529	2611	203

Compute $|GCC|$

$$
\begin{aligned}
|GCC| &\approx n - \sum_{i=1}^{n} e^{-\lambda_0 w_i} \\
&= n - \sum_{i=1}^{n} e^{(1-\lambda_0)w_i} e^{-w_i} \\
&= \sum_{k\geq 0} n_k - \sum_{i=1}^{n} \sum_{k=0}^{\infty} \frac{(1-\lambda_0)^k}{k!} w_i^k e^{-w_i} \\
&\approx \sum_{k\geq 0} n_k (1 - (1-\lambda_0)^k) \\
&= \sum_{k\geq 1} n_k (1 - (1-\lambda_0)^k).
\end{aligned}
$$

Conclusion

The size of giant component is predicted to be about $177,400$ by our theory. This is rather close to the actual value $176,000$, within an error bound of less than 1%.



References

- Fan Chung and Linyuan Lu. Connected components in a random graph with given degree sequences, *Annals of Combinatorics*, **6** (2002), 125–145.
- Fan Chung and Linyuan Lu, The volume of the giant component for a random graph with given expected degrees , *SIAM J. Discrete Math.*, **20** (2006), No. 2, 395–411.

Overview of talks

- Lecture 1: Overview and outlines
- Lecture 2: Generative models - preferential attachment schemes
- Lecture 3: Duplication models for biological networks
- Lecture 4: The rise of the giant component
- Lecture 5: The small world phenomenon: average distance and diameter
- Lecture 6: Spectrum of random graphs with given degrees