INDUSTRIAL MATHEMATICS INSTITUTE

2005:07

The volume of the giant component of a random graph with given expected degrees

F. Chung and L. Lu

IMI
Preprint Series

Department of Mathematics
University of South Carolina

The volume of the giant component of a random graph with given expected degrees

Fan Chung [*]† Linyuan Lu ‡

December 12, 2005

Abstract

We consider the random graph model $G(\mathbf{w})$ for a given expected degree sequence $\mathbf{w} = (w_1, w_2, \ldots, w_n)$. If the expected average degree is strictly greater than 1, then almost surely the giant component in G of $G(\mathbf{w})$ has volume (i.e., sum of weights of vertices in the giant component) equal to $\lambda_0 \mathrm{Vol}(G) + O(\sqrt{n}\log^{3.5} n)$, where λ_0 is the unique non-zero root of the following equation;

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

and where $\mathrm{Vol}(G) = \sum_i w_i$.

1 Introduction

Among the many celebrated results of Erdős and Rényi on random graphs, one of the most well known theorems is a sharp estimate for the size of the giant component. For the random graph $G(n,p)$, as introduced by Erdős and Rényi in 1959 [17], every pair of a set of n vertices is chosen to be an edge with probability p independently. Erdős and Rényi [17] showed that the size (i.e., the number of vertices) of the giant component of $G(n,p)$ satisfies the following:

Theorem A: *If $d = np > 1$, a graph G of $G(n,p)$ almost surely contains a giant component with $(f(d)+o(1))n$ vertices, where $f(d)$ is given by*

$$f(d) = 1 - \frac{1}{d}\sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}(de^{-d})^k. \tag{1}$$

In $G(n,p)$, every vertex has the same expected degree np. Although such a random graph model is useful in some applications, most real-world networks

---

[*] University of California, San Diego, fan@ucsd.edu

† Research supported in part by NSF Grants DMS 0457215, ITR 0205061 and ITR 0426858

‡ University of South Carolina, lu@math.sc.edu

have degree distributions far from being regular. It is therefore not surprising that the random graph model $G(n, p)$ does not capture many behaviors of numerous networks.

Here we consider the random graph model $G(\mathbf{w})$ for a given expected degree sequence $\mathbf{w} = (w_1, w_2, \ldots, w_n)$, as introduced in [10]. The edges are chosen independently and randomly as follows. The probability p_{ij} that there is an edge between v_i and v_j is proportional to the product $w_i w_j$ (as well as the loop at v_i with probability proportional to w_i^2). Namely,

$$p_{ij} = \frac{w_i w_j}{\sum_k w_k} = \frac{w_i w_j}{\mathrm{Vol}(G)}. \tag{2}$$

Here the expected volume for a subset S of vertices, $\mathrm{Vol}(S)$, is defined as follows.

$$\mathrm{Vol}(S) = \sum_{v_i \in S} w_i$$

and $\mathrm{Vol}(G) = \mathrm{Vol}(V(G))$. The (actual) volume of S in a graph G is the sum of all degrees of vertices in S and is denoted by $\mathrm{vol}(S)$.

$$\mathrm{vol}(S) = \sum_{v_i \in S} d_i$$

where d_i denote the degree of vertex v_i. In order to avoid confusion when we deal with the graph G in a non-probabilistic context, we can view w_i as a weight assigned to vertex v_i.

In [10], the following theorem was given concerning the giant components for graphs in the random graph model $G(\mathbf{w})$.

Theorem B: *Suppose that G is a random graph in $G(\mathbf{w})$ with expected degree sequence $\mathbf{w}$. If the expected average degree d is strictly greater than 1, then the following holds:*
(1) Almost surely G has a unique giant component. Furthermore, the volume of the giant component is at least $(1 - \frac{2}{\sqrt{de}} + o(1))\mathrm{Vol}(G)$ if $d \geq \frac{4}{e} = 1.4715\ldots$, and is at least $(1 - \frac{1+\log d}{d} + o(1))\mathrm{Vol}(G)$ if $d < 2$.
(2) The second largest component almost surely has size at most $(1+o(1))\mu(d)\log n$, where

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

Moreover, with probability at least $1 - n^{-k}$, the second largest component has size at most $(k+1+o(1))\mu(d)\log n$, for any $k \geq 1$. [1]

In this paper, we will state a sharp asymptotic estimate for the volume of the giant component for a random graph in $G(\mathbf{w})$.

---

[1] The quantitative estimate of this probability is in the proof of Theorem 1 in [10].

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

$$\sum_{i=1}^{n} w_i e^{-w_i \lambda} = (1-\lambda)\sum_{i=1}^{n} w_i. \tag{3}$$

We remark that $\mathrm{Vol}(G)$ in the statement of Theorem 1 can be replaced by $\mathrm{vol}(G)$ since it was proved in [10] that with probability at least $1 - e^{-c}$,

$$|\mathrm{vol}(G) - \mathrm{Vol}(G)| \leq \sqrt{c\mathrm{Vol}(G)}.$$

Since the average degree is $\mathrm{vol}(G)/n$, the average degree can also be approximated by the average expected degree $\mathrm{Vol}(G)/n$.

The paper is organized as follows. Section 2 contains several facts concerning equation (3). In section 3, we show that the asymptotic formula for the volume of the giant components of a random graph in $G(\mathbf{w})$ is a generalization of Theorem A by Erdős and Rényi. Section 4 includes some improved lower bounds for the volume of the giant component of $G \in G(\mathbf{w})$ as a function of the expected average degree. In Section 5, we give the complete proof of Theorem 1. In Section 6, we derive a sharp estimate for the number of vertices in the giant components.

2 Preliminaries

Before we proceed, we examine some basic properties of the solutions to the equation in (3). The proof is quite straightforward and will be omitted here.

Let $\tilde{d}$ denote the expected second order average degree.

$$\tilde{d} = \frac{\sum_i w_i^2}{\sum_i w_i}.$$

Lemma 1 *Suppose the expected second order average degree satisfies* $\tilde{d} > 1$*. Define*

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

We have $f(0) = 0$*,* $f'(0) < 0$*, and* $f''(\lambda) > 0$*. Hence* $f(\lambda) = 0$ *has a unique positive solution* λ_0 *(see Figure 1). In particular,*

1. *If* $f(x_1) \leq 0$ *for some positive* x_1*, then* $\lambda_0 \geq x_1$*.*
2. *If* $f(x_2) \geq 0$ *for some positive* x_2*, then* $\lambda_0 \leq x_2$*.*
3. $\lambda_0 < 1$ *since* $f(1) > 0$*.*

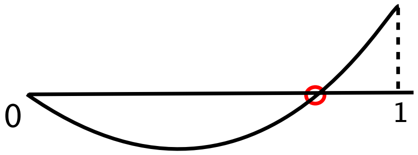

Figure 1: *When $\tilde{d} > 1$, $f(x)$ has a unique positive root.*

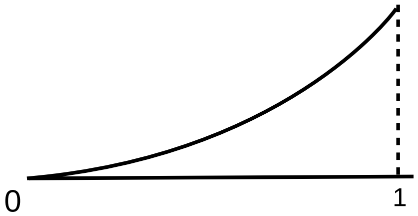

Figure 2: *When $\tilde{d} < 1$, $f(x) > 0$ for all $x > 0$.*

When $\tilde{d} < 1$, We have $f(0) = 0$, $f'(0) > 0$, and $f''(\lambda) > 0$. Zero is the only non-negative root for $f(x)$ (see Figure 2). This corresponds to the case that there is no giant component.

The following fact is useful in the proof of the main theorem.

Lemma 2 *Suppose that the expected average degree d satisfies*

$$d = \frac{1}{n}\sum_{i=1}^{n} w_i \geq 1 + \delta > 1$$

for some positive constant δ. Define $f(\lambda) = \sum_{i=1}^{n} w_i e^{-w_i\lambda} - (1-\lambda)\sum_{i=1}^{n} w_i$ and let λ_0 denote the unique non-zero root of $f(\lambda) = 0$. Then there is a positive constant $c = c(\delta)$ such that

$$f'(\lambda_0) \geq c\sum_{i=1}^{n} w_i.$$

Proof: Since $\tilde{d} \geq d > 1$, the unique root λ_0 of f exists. We have

$$f'(\lambda_0) = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 e^{-w_i\lambda_0}.$$

Case 1: $\lambda_0 \geq \frac{1}{2}$. Since $xe^{-x\lambda_0}$ attains its maximum at $x = 1/\lambda_0$, we have

$$\begin{aligned} f'(\lambda_0) &= \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 e^{-w_i\lambda_0} \\ &\geq \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i \frac{1}{e\lambda_0} \end{aligned}$$

$$
\begin{aligned}
&= (1-\frac{1}{e\lambda_0})\sum_{i=1}^{n} w_i \\
&\geq (1-\frac{2}{e})\sum_{i=1}^{n} w_i.
\end{aligned}
$$

The statement holds for this case.

Case 2: $\lambda_0 < \frac{1}{2}$. We will utilize some convexity inequalities. First we will prove the following claim.

Now we consider the function $h(x) = (x^2 + \frac{x}{\lambda_0})e^{-\lambda_0 x}$. We have

$$
\begin{aligned}
h'(x) &= (\frac{1}{\lambda_0} + x - \lambda_0 x^2)e^{-\lambda_0 x} \\
h''(x) &= -\lambda_0 x(3-\lambda_0 x)e^{-\lambda_0 x}. \qquad (4)
\end{aligned}
$$

We need the following facts whose proofs will be given at the end of this section.

Claim A:

(i) $h(x)$ is concave downward over x in $(0, \frac{3}{\lambda_0})$. The maximum value of $h(x)$ for x in $[0,\infty)$ is reached at $x_0 = \frac{\sqrt{5}+1}{2\lambda_0}$.

(ii) $d < \frac{2}{e\lambda_0} < x_0$.

(iii) $\lambda_0 > 1 - \frac{1}{d}$.

Now, we consider the following function

$$
H(x) = \begin{cases} h(x) & 0 \leq x \leq x_0 \\ h(x_0) & x \geq x_0 \end{cases}
$$

Using Claim A (i), $H(x)$ is concave downward and $H(x) \geq h(x)$ for all $x \geq 0$. We have

$$
\begin{aligned}
f'(\lambda_0) &= \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 e^{-w_i\lambda_0} \\
&= \sum_{i=1}^{n} w_i + \frac{1}{\lambda_0}\sum_{i=1}^{n} w_i e^{-w_i\lambda_0} - \sum_{i=1}^{n} h(w_i) \\
&= \sum_{i=1}^{n} w_i + \frac{1}{\lambda_0}(1-\lambda_0)\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} h(w_i) \\
&= \frac{1}{\lambda_0}\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} h(w_i) \\
&\geq \frac{1}{\lambda_0}\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} H(w_i) \\
&\geq \frac{1}{\lambda_0}\sum_{i=1}^{n} w_i - nH(\frac{1}{n}\sum_{i=1}^{n} w_i)
\end{aligned}
$$

$$= \frac{1}{\lambda_0} nd - nH(d).$$

By Claim A (ii), we have $d < \frac{2}{e\lambda_0} < x_0$. Hence, $H(d) = h(d)$.

$$\begin{aligned}
f'(\lambda_0) &\geq \frac{1}{\lambda_0} nd - nh(d) \\
&= \frac{1}{\lambda_0} nd - n(d^2 + \frac{d}{\lambda_0})e^{-\lambda_0 d} \\
&= nd\frac{1}{\lambda_0}(1 - (1 + d\lambda_0)e^{-\lambda_0 d}) \\
&\geq nd(1 - (1 + d\lambda_0)e^{-\lambda_0 d}).
\end{aligned}$$

The function $\psi(x) = 1 - (1+x)e^{-x}$ is increasing for x in $[0, \infty)$. For any $x > 0$, $\psi(x) > \psi(0) = 0$.

Hence we have

$$\begin{aligned}
f'(\lambda_0) &\geq nd\psi(\lambda_0 d) \\
&\geq nd\psi(d-1) \\
&\geq cnd
\end{aligned}$$

by choosing $c = c(\delta) = \min\{\psi(\delta), 1 - 2/e\}$.

It remains to prove Claim A.

Proof of Claim A: (i) follows from (4).

To prove (ii), we use the fact that λ_0 is a root of f and $xe^{-\lambda_0 x}$ has its maximum value $\frac{1}{e\lambda_0}$ at $x = 1/\lambda_0$. Then

$$\begin{aligned}
(1-\lambda_0)nd &= (1-\lambda_0)\sum_{i=1}^{n} w_i \\
&= \sum_{i=1}^{n} w_i e^{-\lambda_0 w_i} \\
&\leq \sum_{i=1}^{n} \frac{1}{e\lambda_0} \\
&= \frac{n}{e\lambda_0}.
\end{aligned}$$

Thus,

$$\lambda_0(1-\lambda_0) \leq \frac{1}{de}.$$

We have

$$\lambda_0 \leq \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{de}}\right) \quad \text{or} \quad \lambda_0 \geq \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{de}}\right).$$

$\lambda_0 < \frac{1}{2}$ implies

$$\begin{aligned}\lambda_0 &\leq \frac{1}{2}(1-\sqrt{1-\frac{4}{de}}) \\ &= \frac{2}{de}\frac{1}{1+\sqrt{1-\frac{4}{de}}} \\ &< \frac{2}{de}.\end{aligned}$$

Hence, we have $d < \frac{2}{e\lambda_0} < x_0$ as desired.

To prove (iii), we consider the function

$$g(x) = \begin{cases} xe^{-\lambda_0 x} & 0 \leq x \leq \frac{1}{\lambda_0} \\ \frac{1}{e\lambda_0} & x > \frac{1}{\lambda_0}. \end{cases}$$

We observe that $g(x)$ is concave downward and $g(x) \geq xe^{-\lambda_0 x}$ for all $x \geq 0$.

By the definition of λ_0, we have

$$\begin{aligned}(1-\lambda_0)nd &= (1-\lambda_0)\sum_{i=1}^n w_i \\ &= \sum_{i=1}^n w_i e^{-\lambda_0 w_i} \\ &\leq \sum_{i=1}^n g(w_i) \\ &\leq ng(d).\end{aligned}$$

By Claim A (ii), $d < \frac{2}{e\lambda_0}$. Thus, $g(d) = de^{-\lambda_0 d}$. We have

$$1-\lambda_0 \leq e^{-\lambda_0 d}.$$

Note that $\phi(\lambda) = (1-\lambda) - e^{-\lambda d}$ is concave downward over $[0,\infty)$. Since $\phi(0)=0$ and $\phi'(0) = d-1 > 0$, $\phi(x)$ has a unique positive root, which we denote by s. We have $\phi(x) > 0$, for any $0 < x < s$. Since $\phi(\lambda_0) \leq 0$ and $\lambda_0 \neq 0$, we have $\lambda_0 \geq s$.

Define $t = (1-s)d$ and then we have

$$\frac{t}{d} = 1-s = e^{-sd} = e^{-d+t}.$$

Thus t satisfies the following equation:

$$te^{-t} = de^{-d}. \tag{5}$$

The function xe^{-x} increases in $[0,1]$ and decreases in $[1,\infty]$. There is a unique $t < 1$ satisfying equation (5).

We have

$$\lambda_0 \geq s = 1 - \frac{t}{d} > 1 - \frac{1}{d}.$$

The proof of Claim A is finished and therefore the proof of Lemma 2 is complete.

□

3 Theorem 1 $\Rightarrow$ Theorem A

In this section we want to show that the formula for the size of the giant component for a random graph in $G(n,p)$ as derived by Erdős and Rényi in Theorem A is a special case of Theorem 1. In other words, if we restrict the expected degree sequence to the case when all degrees are equal, then we recover the theorem of Erdős and Rényi.

Theorem 2 *Theorem 1 implies Theorem A of Erdős and Rényi for $G(n,p)$.*

Proof: In $G(n,p)$, we have $w_1 = w_2 = \cdots = w_n = np = d$. Equation (3) becomes

$$e^{-d\lambda} = 1 - \lambda.$$

Let $\lambda = 1 - \frac{1}{d}z$. We have

$$e^{-d+z} = \frac{z}{d}.$$

Or equivalently,

$$z = de^{-d}e^{z}.$$

Here we use the following version of the well-known Lagrange inversion formula:

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.

We apply the above formula with $x = 0$, $y = de^{-d}$, and $\phi(z) = e^z$. Then we have

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

This is exactly Equation (1) in Theorem A of Erdős and Rényi. □

4 Lower bounds

Theorem 1 gives an implicit formula for the volume of the giant component for a random graph with a given expected degree sequence. It is often useful to deduce some bounds which depend only on the expected average degree d. Of particular interest is the following question:

Among all random graphs $G(\mathbf{w})$ with the same expected average degree d, which degree distributions minimize or maximize the volume of the giant component?

One obvious example comes to mind. Almost surely $G(m, p)$ with $mp = \Omega(\log m)$ is connected. By adding $n - m$ vertices to $G(m, p)$ with weights zero, we get a random graph $G(\mathbf{w})$ with the expected average degree $d = \frac{mp}{n}$, which almost surely has a giant component with volume $\mathrm{Vol}(G)$.

One might be inclined to conjecture the random graph with equal expected degrees generates the smallest giant component among all possible degree distribution with the same volume. The answer is "yes" for $1 < d \leq \frac{e}{e-1}$, and a surprising "no" if d is sufficiently large.

We will prove the following theorem.

Theorem 3 *When $d \geq \frac{4}{e}$, almost surely the giant component of $G \in G(\mathbf{w})$ has volume at least*

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

We remark that $\frac{1}{2}(1 + \sqrt{1 - \frac{4}{de}}) = 1 - \frac{1}{de} + O(\frac{1}{d^2})$ improves the bound in Theorem B. In fact, this bound is best possible as d approaches infinity as shown by the following example.

Example: Let $m = \lfloor n^{3/4} \rfloor$ and $y = 1 + \frac{n}{m}(d-1) \approx (d-1)n^{1/4}$. We choose the expected degrees

$$w_1 = w_2 = \cdots = w_m = y, \quad w_{m+1} = \cdots = w_n = 1.$$

The expected average degree of this random graph $G(\mathbf{w})$ is

$$\frac{my + (n-m)}{n} = d.$$

Let $x_0 = 1 - \frac{1}{de}$. To show the giant component of G has volume at most $(x_0 + o(1))\mathrm{Vol}(G)$, it is sufficient to verify $f(x_0) \geq 0$. Here

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

We have

$$f(x_0) \quad = \quad \sum_{i=1}^{n} w_i e^{-w_i x_0} - (1 - x_0) \sum_{i=1}^{n} w_i$$

$$\begin{aligned}
&= mye^{-yx_0} + (n-m)e^{-x_0} - (1-x_0)nd \\
&\geq \frac{n}{e}(e^{\frac{1}{de}} - 1 - O(n^{-1/4})) \\
&\geq 0
\end{aligned}$$

as desired.

We are now ready to prove Theorem 3.

Proof of Theorem 3: We note that the function $g(z) = ze^{-z\lambda}$ reaches its maximum value at $z = \frac{1}{\lambda}$. We have

$$\begin{aligned}
f(\lambda) &= \sum_{i=1}^{n} w_i e^{-w_i\lambda} - (1-\lambda)\sum_{i=1}^{n} w_i \\
&\leq \sum_{i=1}^{n} \frac{1}{\lambda} e^{-1} - (1-\lambda)\sum_{i=1}^{n} w_i \\
&= \frac{n}{e\lambda}(1 - \lambda(1-\lambda)de).
\end{aligned}$$

Since λ_0 is a solution of $f(\lambda) = 0$, we have

$$\lambda_0(1-\lambda_0) \leq \frac{1}{de}$$

which implies either $\lambda_0 \leq \frac{1}{2}(1 - \sqrt{1 - \frac{4}{de}})$ or $\lambda_0 \geq \frac{1}{2}(1 + \sqrt{1 - \frac{4}{de}})$.

We will show that $\lambda_0 \leq \frac{1}{2}(1 - \sqrt{1 - \frac{4}{de}})$ is not true by proving $f(\frac{1}{2}) \leq 0$.

We note that

$$\begin{aligned}
f(\frac{1}{2}) &= \sum_{i=1}^{n} w_i e^{-w_i/2} - \frac{1}{2}\sum_{i=1}^{n} w_i \\
&\leq 2ne^{-1} - \frac{1}{2}nd \\
&= \frac{n}{2}(\frac{4}{e} - d) \\
&\leq 0.
\end{aligned}$$

Thus we conclude that $\lambda_0 \geq \frac{1}{2}(1 + \sqrt{1 - \frac{4}{de}})$. □

When d is small and not in the range covered by Theorem 3, we can still derive the following lower bound.

Theorem 4 *When $1 < d \leq \frac{e}{e-1}$, then almost surely $G(\mathbf{w})$ has a giant component of size at least $(\lambda_1 + o(1))\mathrm{Vol}(G)$, where λ_1 is the nonzero root of the following equation:*

$$e^{-\lambda d} = 1 - \lambda. \tag{6}$$

In other words, among all random graphs $G(\mathbf{w})$ with fixed expected average degree d, the Erdős-Rényi random graph $G(n, \frac{d}{n})$ has the smallest giant component (measured in volume).

Proof: Consider the function

$$g(x) = \begin{cases} xe^{-\lambda_1 x} & 0 \leq x \leq \frac{1}{\lambda_1} \\ \frac{1}{e\lambda_1} & x > \frac{1}{\lambda_1}. \end{cases}$$

We observe that $g(x)$ is concave downward and $g(x) \geq xe^{-\lambda_1 x}$ for all $x \geq 0$. We have

$$\begin{aligned} f(\lambda_1) &= \sum_{i=1}^n w_i e^{-\lambda_1 w_i} - (1-\lambda_1)nd \\ &\leq \sum_{i=1}^n g(w_i) - (1-\lambda_1)nd \\ &\leq ng(\frac{1}{n}\sum_{i=1}^n w_i) - (1-\lambda_1)nd \\ &\leq n(g(d) - (1-\lambda_1)d). \end{aligned}$$

Since λ_1 is an increasing function of d, $d\lambda_1$ is also an increasing function of d. When $d = \frac{e}{e-1}$, it is easy to verify $\lambda = 1 - \frac{1}{e}$ is the other root of equation (6). Therefore, $d\lambda_1 \leq 1$, when $d \leq \frac{e}{e-1}$. In particular, we have

$$g(d) = de^{-\lambda_1 d}.$$

Hence

$$\begin{aligned} f(\lambda_1) &\leq n(g(d) - (1-\lambda_1)d) \\ &= nd(e^{-\lambda_1 d} - (1-\lambda_1)) \\ &= 0 \end{aligned}$$

By Remark 1, we have $\lambda_0 \geq \lambda_1$ as desired. $\square$.

5 The proof of the main theorem

A main tool that we use in the proof of the main theorem is a relaxed version of the Azuma's inequality (as seen in Theorem 1 of [12]) which can be described as follows:

Suppose that Ω is a probability space and $\mathcal{F}$ denote a σ-field on Ω (i.e., a collection of subsets of Ω which contains $\emptyset$ and Ω, and is closed under unions, intersections, and complementation.) A *filter* $\mathbf{F}$ is an increasing chain of σ-subfields

$$\{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}.$$

A martingale (obtained from) X is associated with a filter $\mathbf{F}$ and a sequence of random variables $X_0, X_1, \ldots, X_n$ satisfying $X_i = E(X \mid \mathcal{F}_i)$ and, in particular, $X_0 = E(X)$ and $X_n = X$. For undefined terminology on martingales, the reader is referred to [19].

For $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ a vector with positive entries, a martingale X is said to be c-Lipschitz if

$$|X_i - X_{i-1}| \leq c_i \tag{7}$$

for $i = 1, 2, \ldots, n$.

If the c-Lipschitz condition is not satisfied, we can still consider the following relaxed version:

A martingale X is said to be near-c-Lipschitz with an exceptional probability η if

$$\sum_i \Pr(|X_i - X_{i-1}| \geq c_i) \leq \eta. \tag{8}$$

Theorem C (Theorem 1 as in [12]) For non-negative values, $c_1, c_2, \ldots, c_n$, a martingale X is near-c-Lipschitz with an exceptional probability η. Then X satisfies

$$\Pr(|X - E(X)| < a) \leq 2e^{-\frac{a^2}{2\sum_{i=1}^n c_i^2}} + \eta.$$

The idea for the proof of Theorem 1 is to first prove that the volume of giant component concentrates on its expected value $\mathrm{E}(\mathrm{Vol}(GCC))$ and then show that $\mathrm{E}(Vol(GCC))/\mathrm{Vol}(G)$ can be approximated by the non-zero root of equation (3). To do so, we need to establish several useful facts.

Lemma 3 *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 giant component of* $G(\mathbf{w})$*.*

Proof: Consider a vertex v_i with weight $w_i \geq \max\{8k, 2(k+1+o(1))\mu(d)\} \log n$. For a random graph G in $G(\mathbf{w})$, let d_i denote the degree of v_i in G. Then, d_i is the sum of independent 0-1 random variables with $\mathrm{E}(d_i) = w_i$. For any non-negative value λ, we have

$$\Pr(d_i - \mathrm{E}(d_i) < -\lambda) \leq e^{-\frac{\lambda^2}{2\mathrm{E}(d_i)}}.$$

By choosing $\lambda = w_i/2$, we have

$$\Pr(d_i < w_i/2) \leq e^{-w_i/8} \leq n^{-k}.$$

With probability at least $1 - n^{-k}$, v_i is in a connected component of size at least $w_i/2$. If this connected component is not the giant component, then the second largest component must have size at least $w_i/2$. However, from Theorem B, this can only happen with probability at most n^{-k} because of the assumption that

$$w_i/2 \geq (k + 1 + o(1))\mu(d) \log n.$$

Hence, 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 giant component. $\square$

Lemma 4 *For any $k > 2$, with probability at least $1 - 6n^{-k+2}$, we have*

$$|\mathrm{Vol}(GCC) - \mathrm{E}(\mathrm{Vol}(GCC))| \le 2C_1(k+1)^2\sqrt{k-2}\sqrt{n}\log^{2.5} n,$$

where $C_1 = 10\mu(d) + 2\mu(d)^2$.

Proof: Let $L = L(k)$ be the set of vertices with weight greater than $\max\{8k, 2(k+1+o(1))\mu(d)\}\log n$. If $L \neq \emptyset$, we form a new graph G^* by adding a new vertex v_* to $G(\mathbf{w})$ and add edges from v_* to each vertex in L. $G(\mathbf{w})$ almost surely has a giant component, so does G^*. Let X denote the volume of the giant component in G^*. (While computing the values for Vol of the giant component in G^*, we use the convention that the weight of v_* is zero.) If $L = \emptyset$, we simply let $X = \mathrm{Vol}(GCC)$.

We wish to show the concentration of the random variable X. It is sufficient to prove the following claim.

Claim:

$$\Pr(|X - \mathrm{E}(X)| < \lambda) \le 4n^{-k+2}$$

where $\lambda = 2C_1(k+1)^2\sqrt{k-2}\sqrt{n}\log^{2.5} n$.

We observe that X does not depend on whether $\{u, v\}$ is an edge if both u and v are in L. We list all pairs of vertices with at least one vertex not in L by $\{f_1, f_2, \ldots, f_m\}$,where $m = \binom{n}{2} - \binom{|L|}{2}$. (The order of edges in the list is arbitrarily chosen.) For $i = 0, 1, 2, \ldots, m$, let $\mathcal{F}_i$ denote the σ-field generated by exposing pairs $f_1, f_2, \ldots, f_i$. We apply Theorem C on the edge-exposing martingale X with $X_i = \mathrm{E}(X|\mathcal{F}_i)$ and $X_m = X$. We wish to find a good Lipschitz or near-Lipschitz bound for $|X_i - X_{i-1}|$. By definition, X_{i-1} is the conditional expectation of X_i. Choosing the pair f_i as an edge can change X by at most the volume of a small component. Let v_i be a vertex of the pair f_i not in L. (If there is a tie, break arbitrarily.) Let G_{v_i} be the random graph obtained by deleting v_i from $G(\mathbf{w})$. The possible small component containing v before f_i is exposed can be broken into at most d_i largest connected components excluding the giant component in G_{v_i}.

First, we apply Theorem B to the random graph G_{v_i}. Note that the average degree of G_{v_i} is $(1+o(1))d$. Thus, with probability at least $1 - n^{-k}$, all small components of G_{v_i} have size at most $(k+1+o(1))\mu(d)\log n$. Also, for any positive λ', the degree d_i of v_i can be upper bounded by

$$\Pr(d_i > w_i + \lambda') < e^{-\frac{\lambda^2}{2(w_i + \lambda'/3)}}.$$

By choosing $\lambda' = w_i + 2k\log n$, we have

$$\begin{aligned}
\Pr(d_i > 2w_i + 2k\log n) &< e^{-\frac{\lambda^2}{2(w_i+\lambda/3)}} \\
&= e^{-\frac{(w_i+2k\log n)^2}{2(w_i+(w_i+2k\log n)/3)}} \\
&< n^{-k}.
\end{aligned}$$

Thus, with probability at least $1 - 2n^{-k}$, we have

$$\begin{aligned} |X_i - X_{i-1}| &\leq d_i \times (k+1+o(1))\mu(d)\log n \\ &< (2w_i + 2k\log n)(k+1+o(1))\mu(d)\log n \\ &< (10k\log n + 2(k+1+o(1))\mu(d)\log n)(k+1+o(1))\mu(d)\log n \\ &< C_1(k+1)^2\log^2 n \end{aligned}$$

where $C_1 = 10\mu(d) + 2\mu(d)^2$ is a bounded positive number.

Now we apply Theorem C on martingale X with $c_i = C_1(k+1)^2\log^2 n$ and $\eta \leq \binom{n}{2}2n^{-k}$. For any positive λ, we have

$$\begin{aligned} \Pr(|X - \mathrm{E}(X)| > \lambda) &\leq 2e^{-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}} + \eta \\ &\leq 2e^{-\frac{\lambda^2}{2C_1^2(k+1)^4 n\log^4 n}} + 2n^{-k+2}. \end{aligned}$$

For $\lambda = 2C_1(k+1)^2\sqrt{k-2}\sqrt{n}\log^{2.5} n$, we have

$$Pr(|X - \mathrm{E}(X)| > \lambda) \leq 4n^{-k+2}$$

as desired. $\square$

Proof of Theorem 1:

For any vertex v with weight w_v, the probability that v is not in the giant component of $G(\mathbf{w})$ can be estimated as follows. To simplify the notation, we write $C_k = \max\{8k, 2(k+1+o(1))\mu(d)\}$.

Case a: $w_v \geq C_k \log n$. By Lemma 3, we have

$$\Pr(v \notin GCC) \leq \frac{2}{n^k}.$$

Case b: $w_v \leq C_k \log n$. Let G_v be the random graph by removing v from G. Expose every pairs of vertices in G_v. Let H be the giant component of G_v. Apply Lemma 4 to G_v, with probability at least $1 - \frac{6}{(n-1)^{k-2}}$, we have

$$|\mathrm{Vol}(H) - \mathrm{E}(\mathrm{Vol}(H))| \leq 2C_1(k+1)^2\sqrt{k-2}\sqrt{n}\log^{2.5} n.$$

Now we expose the pairs of vertices containing v. We have

$$\begin{aligned} \Pr(v \notin GCC|H) &= \prod_{v_j \in V(H)} (1 - w_v w_j \rho) \\ &= e^{-\sum_{v_j\in V(H)} w_v w_j \rho + \sum_{v_j \in V(H)} w_v^2 w_j^2 \rho^2} \\ &= e^{-w_v \mathrm{Vol}(H)\rho(1+O(w_v \tilde{d}\rho))}. \end{aligned}$$

The probability that v is not in the giant component can be estimated as follows:

$$\begin{aligned} \Pr(v \notin GCC) &= \mathrm{E}(\Pr(v \notin GCC|H)) + O(n^{-k+2}) \\ &= \mathrm{E}(e^{-w_v \mathrm{Vol}(H)\rho}) + O(n^{-k+2}) \\ &= e^{-w_v \mathrm{E}(\mathrm{Vol}(H))\rho + O(k^2 w_i \rho\sqrt{n}\log^{2.5} n)} + O(n^{-k+2}). \qquad (9) \end{aligned}$$

Note that GCC can be formed from H by joining at most d_v's small components. Thus, we have

$$\begin{aligned}
|\mathrm{E}(GCC) - \mathrm{E}(H)| &\leq \mathrm{E}(d_v)(k+1+o(1))\mu(d)\log n + 2n^{-k} \\
&= w_v(k+1+o(1))\mu(d)\log n + 2n^{-k} \\
&= O(w_v k \log n).
\end{aligned}$$

By substituting $\mathrm{E}(H)$ by $\mathrm{E}(\mathrm{Vol}(GCC)) + O(w_v k \log n)$ in (9), we have

$$\begin{aligned}
\Pr(v \notin GCC) &= e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho + O(w_v^2 k\rho \log n) + O(k^2 w_v \rho \sqrt{n} \log^{2.5} n)} + O(n^{-k+2}) \\
&= (1 + O(k^3 \rho \sqrt{n} \log^{3.5} n)) e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} + O(n^{-k+2}).
\end{aligned}$$

Together, we have

$$\begin{aligned}
& \mathrm{Vol}(G) - \mathrm{E}(\mathrm{vol}(GCC)) \\
= & \sum_v w_v \Pr(v \notin GCC) \\
= & \sum_{w_v < C_k \log n} w_v \Pr(v \notin GCC) + \sum_{w_v \geq C_k \log n} w_v \Pr(v \notin GCC) \\
= & \sum_{w_v < C_k \log n} w_v \left[(1 + O(k^3 \rho \sqrt{n} \log^{3.5} n)) e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} + O(n^{-k+2}) \right] \\
& + \sum_{w_v \geq C_k \log n} w_v O(2n^{-k}) \\
= & \sum_{w_v < C_k \log n} w_v e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} + O(k^3 \sqrt{n} \log^{3.5} n).
\end{aligned}$$

We choose k to be a constant large enough satisfying

$$C_k \geq \begin{cases} \frac{2}{(1 - \frac{2}{\sqrt{de}})} & \text{if } d > \frac{4}{e}, \\ \frac{2}{(1 - \frac{1 + \log d}{d})} & \text{if } 1 < d < 2. \end{cases}$$

By Theorem A, we have $C_k \mathrm{E}(\mathrm{Vol}(GCC))\rho \geq 2$. In particular, for any vertex v with $w_v \geq C_k \log n$, we have

$$e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} \leq n^{-2}.$$

Thus,

$$\sum_{w_v \geq C_k \log n} w_v e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} = O(n^{-1}).$$

Therefore we have

$$\mathrm{Vol}(G) - \mathrm{E}(\mathrm{vol}(GCC)) = \sum_v w_v e^{-w_v \mathrm{E}(\mathrm{Vol}(GCC))\rho} + O(\sqrt{n} \log^{3.5} n).$$

Let $x_0 = \frac{\mathrm{Vol}(GCC)}{\mathrm{Vol}(G)}$ and $f(x) = \sum_{i=1}^n w_i e^{-w_i x} - (1-x)\sum_{i=1}^n w_i$, we have

$$f(x_0) = O(\sqrt{n}\log^{3.5} n). \qquad (10)$$

The equation $f(x) = 0$ has only two roots $x = 0$ and $x = \lambda_0$. Note that $f(x)$ is concave upward with $|f'(0)| = n(d^2 - d)$, and $|f'(\lambda_0)| > cnd$. Consider a small interval I around 0 with diameter $O(\sqrt{n}\log^{3.5} n)$. The preimage $f^{-1}(I)$ has diameter at most $O(n^{-1/2}\log^{3.5} n)$. Since x_0 is bounded away from 0 by a small constant, we have $|x_0 - \lambda_0| = O(n^{-1/2}\log^{3.5} n)$. Therefore, almost surely the giant component has volume

$$\lambda_0 \mathrm{Vol}(G) + O(\sqrt{n}\log^{3.5} n).$$

Theorem 1 is proved. $\square$

6 The complement of the giant component and its size

As we know, the giant component almost surely exists if the expected average degree $d > 1$. We consider the remaining graph G' after removing the giant component.

For a random graph G in the Erdős-Rényi model $G(n,p)$, where $p = d/n$, if $d > 1$, there is a unique $c < 1$ satisfying

$$ce^{-c} = de^{-d}.$$

We write $\lambda_0 = 1 - \frac{c}{d}$. For any vertex v, the probability that $v \in S$ is known [19] to be

$$e^{-\lambda_0 d} = e^{-d+c} = \frac{c}{d}.$$

Hence S has $(\frac{c}{d} + o(1))n$ vertices. After removing the giant component from $G(n,p)$, the remaining graph can be viewed as a random graph in $G(n', p)$, where $n' \approx \frac{c}{d}n$.

The above fact can be generalized to the random graph model $G(\mathbf{w})$. The following theorem is based on the proof of Theorem 1 and we omit the proof here.

Theorem 5 *Suppose the expected average degree d is strictly greater than 1. Let G' denote the remaining graph of a random graph G in $G(\mathbf{w})$ by removing the giant component. Then almost surely G' is an induced subgraph on a random subset S satisfying.*

1. *Any vertex v_i is contained in S with probability $e^{-\lambda_0 w_i}$ where λ_0 is as defined in (3).*
2. *For any $v_i, v_j \in S$, the probability that $v_i v_j$ is an edge of G_S is $w_i w_j/\mathrm{Vol}(G)$. The induced subgraph G_S is a random graph with given expected degrees*

$$\{(1-\lambda_0)w_i\}_{v_i \in S}.$$

3. $G' \setminus G_S$ consists of at most $O(\log n)$ components each with size $O(\log n)$.

We further analyze the size of the giant component. The proof is similar and will be omitted.

Theorem 6 *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 where λ_0 is as defined in (3).*

7 Comparing theoretical results with the data from the collaboration graph

To illustrate the effectiveness of our results, we use an example of the collaboration graph of the second kind. Based on the data of *Mathematics Review* [18], there are about 401,000 authors as vertices. Two vertices are joined by an edge if there is a paper by exactly two authors. There are about 284,000 edges. The giant component has 176,000 vertices and 248,000 edges. Suppose we model this collaboration graph as a random graph with some given expected degrees w_i. Although we do not know the exact values of w_i's, we can make the following deductions using the theorems in the previous section.

By Theorem 6, we have

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

Solving the above equation, we have $\lambda_0 \approx 0.644$.

For a fixed vertex v_i, the degree of v_i follows the Poisson distribution with expected value w_i. Namely, for a fixed k, the probability that v_i has degree k is $\frac{w_i^k}{k!} e^{-w_i}$. Let n_k denote the number of vertices of degree k. Then by the linearity of the expectation, we have

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

Theorem 6 implies that the size of the giant component satisfies:

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

$$\begin{aligned} &\approx \sum_{k\geq 0} n_k(1-(1-\lambda_0)^k) \\ &= \sum_{k\geq 1} n_k(1-(1-\lambda_0)^k) \qquad (11) \end{aligned}$$

Here we estimate n_k by

$$n_k \approx \mathrm{E}(n_k) \approx \sum_{k=1}^{n} \frac{w_i^k}{k!} e^{-w_i}.$$

Grossman [18] has computed the n_k's as follows:

n_0	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	$\cdots$
166381	145872	34227	16426	9913	6670	4643	3529	2611	2032	$\cdots$

Table 1: The degree sequence of the collaboration graph of the second kind.

By substituting the above n_k's into (11), the size of giant component is supposed to be about 177, 400. This is rather close to the actual value 176, 000, within an error bound of less than 1%.

In Figure 3 and Figure 4, we have plotted the degree distribution and the distribution of the sizes of connected components of the collaboration graph of the second kind.

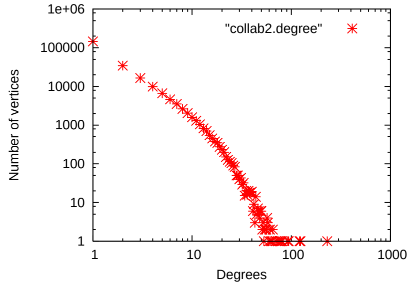

Figure 3: *Degree distribution of the collaboration graph of the second kind.*

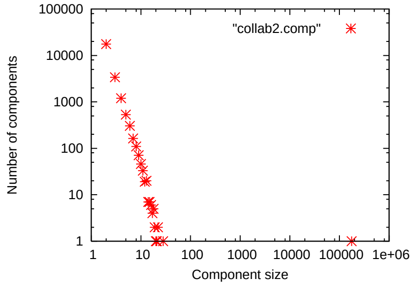

Figure 4: *Size distribution of connected components of the collaboration graph of the second kind.*

Acknowledgement
The authors are grateful to the referees for numerous valuable comments and crucial corrections on the earlier draft of this paper.

References

[1] W. Aiello, F. Chung and L. Lu, A random graph model for massive graphs, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, (2000) 171-180.

[2] W. Aiello, F. Chung and L. Lu, A random graph model for power law graphs, *Experimental Math.*, **10**, (2001), 53-66.

[3] W. Aiello, F. Chung and L. Lu, Random evolution in massive graphs, *Handbook of Massive Data Sets*, Volume 2, (Eds. J. Abello et al.), Kluwer Academic Publishers, (2002), 97-122.

[4] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286** (1999) 509-512.

[5] A.-L. Barabási, R. Albert, and H. Jeong, Scale-free characteristics of random networks: the topology of the world wide web, *Physica* A **281** (2000), 69-77.

[6] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tompkins, and J. Wiener, "Graph Structure in the Web," *Proceedings of the WWW9 Conference*, May, 2000, Amsterdam.

[7] B. Bollabás and O. Riordan, Robustness and Vulnerability of Scale-free Random Graphs. *Internet Mathematics*, **1** (2003) no. 1, 1–35.

[8] F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.

[9] F. Chung and L. Lu, The diameter of random sparse graphs, *Advances in Applied Math.* **26** (2001), 257-279.

[10] F. Chung and L. Lu, Connected components in a random graph with given degree sequences, *Annals of Combinatorics* **6**, (2002), 125–145.

[11] F. Chung and L. Lu, The average distance in random graphs with given expected degrees, *Proceeding of National Academy of Science* **99**, (2002), 15879–15882.

[12] F. Chung and L. Lu, Coupling online and offline analyses for random power law graphs, *Internet Math.* **1**, (2004), no. 4, 409–461.

[13] F. Chung, L. Lu and V. Vu, The spectra of random graphs with given expected degrees, *Proceedings of National Academy of Sciences*, **100**, no. 11, (2003), 6313-6318.

[14] C. Cooper and A. Frieze, A general model of undirected Web graphs, *Random Structures and Algorithms*, **22**, (2003), 311-335.

[15] C. Cooper, A. Frieze and J. Vera, Random deletions in a scale free random graph, *Internet Mathematics*, **1** (2004), 463-483.

[16] P. Erdős and T. Gallai, Gráfok előírt fokú pontokkal (Graphs with points of prescribed degrees, in Hungarian), *Mat. Lapok* **11** (1961), 264-274.

[17] P. Erdős and A. Rényi, On random graphs. I, *Publ. Math. Debrecen* **6** (1959), 290-297.

[18] Jerry Grossman, Patrick Ion, and Rodrigo De Castro, Facts about Erdős Numbers and the Collaboration Graph, http://www.oakland.edu/enp/trivia.html.

[19] S. Janson, T. Łuczak, and A. Rucinski, *Random Graphs*, Wiley-Interscience, 2000.

[20] H. Jeong, B. Tomber, R. Albert, Z. Oltvai and A.-L. Babárasi, The large-scale organization of metabolic networks, *Nature*, **407** (2000), 378-382.

[21] J. Kleinberg, S. R. Kumar, P. Raphavan, S. Rajagopalan and A. Tomkins, The web as a graph: Measurements, models and methods, *Proceedings of the International Conference on Combinatorics and Computing*, 1999.

[22] A. J. Lotka, The frequency distribution of scientific productivity, *The Journal of the Washington Academy of the Sciences*, **16** (1926), 317.

[23] L. Lu, The Diameter of Random Massive Graphs, *Proceedings of the Twelfth ACM-SIAM Symposium on Discrete Algorithms*, (2001) 912-921.

[24] C. McDiarmid, Concentration. *Probabilistic methods for algorithmic discrete mathematics*, 195–248, *Algorithms Combin.*, 16, Springer, Berlin, 1998.

[25] S. Milgram, The small world problem, *Psychology Today*, **2** (1967), 60-67.

[26] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions, *Internet Mathematics*, **1**, 226-251.