Annals of Combinatorics 6 (2002) 125-145
0218-0006/02/020125-21

Annals of Combinatorics

Connected Components in Random Graphs with Given Expected Degree Sequences

Fan Chung* and Linyuan Lu

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA
{fan, llu}@euclid.ucsd.edu

Received August 11, 2002

AMS Subject Classification: 05C80

Abstract. We consider a family of random graphs with a given expected degree sequence. Each edge is chosen independently with probability proportional to the product of the expected degrees of its endpoints. We examine the distribution of the sizes/volumes of the connected components which turns out depending primarily on the average degree d and the second-order average degree $\tilde{d}$. Here $\tilde{d}$ denotes the weighted average of squares of the expected degrees. For example, we prove that the giant component exists if the expected average degree d is at least 1, and there is no giant component if the expected second-order average degree $\tilde{d}$ is at most 1. Examples are given to illustrate that both bounds are best possible.

Keywords: random graphs, connected components, expected degree sequence, power law, power law graphs

1. Introduction

The primary subject in the study of random graph theory is the classical random graph $G(n, p)$, as introduced by Erdős and Rényi in 1959 [19]. In $G(n, p)$, every pair of a set of n vertices is chosen to be an edge with probability p. Such random graphs are fundamental and useful for modeling problems in many applications. However, a random graph in $G(n, p)$ has the same expected degree at every vertex and therefore does not capture some of the main behaviors of numerous graphs arising from the real world. It is imperative to consider a versatile and generalized version of random graphs. In this paper, we consider random graphs with given expected degree sequences which include as special cases both the classical random graphs and the random graphs with "power-law" degree distributions. Many realistic graphs satisfy the power-law [1–3,7,8, 12,13,20,21,25,26,36]. Namely, the fraction of vertices with degree d is proportional to $1/d^{\beta}$ for some constant $\beta > 1$. Although here we consider random graphs with general expected degree distributions, special emphasis will be given to sparse graphs (with average degree a small constant) and to power law graphs (see Section 9). The methods

* Research supported in part by NSF Grant DMS 0100472.

and results that we derive in dealing with random graphs with given expected degree distribution are useful not only for modeling and analyzing realistic graphs but also leading to improvements for some problems on classical random graphs as well [14,29].

We consider the following class of random graphs with a given expected degree sequence $\mathbf{w} = (w_1, w_2, \ldots, w_n)$. The vertex v_i is assigned vertex weight w_i. The edges are chosen independently and randomly according to the vertex weights as follows. The probability p_{ij} that there is an edge between v_i and v_j is proportional to the product $w_i w_j$ where i and j are not required to be distinct. There are possible loops at v_i with probability proportional to w_i^2, i.e.,

$$p_{ij} = \frac{w_i w_j}{\sum_k w_k} \text{ and we assume } \max_i w_i^2 < \sum_k w_k. \tag{1.1}$$

This assumption ensures that $p_{ij} \leq 1$ for all i and j. In addition, (1.1) implies that the sequence w_i is graphic (in the sense that the necessary and sufficient conditions for a sequence to be realized by a graph [18] are satisfied) except that we do not require the w_i's to be integers.

We denote a random graph with a given expected degree sequence $\mathbf{w}$ by $G(\mathbf{w})$. For example, a typical random graph $G(n, p)$ (see [19]) on n vertices and edge density p is just a random graph with expected degree sequence $(pn, pn, \ldots, pn)$. The random graph $G(\mathbf{w})$ is different from the random graphs with an exact degree sequence such as the configuration model (more discussion in Section 8). In [31, 32], Molloy and Reed obtained results on the sizes of connected components for random graphs with exact degree sequences which satisfy certain "smoothing" conditions. There are also a number of evolution models for generating a power-law degree random graphs as in Bollobás, Spencer et al. [11], Cooper and Freeze [17] and Aiello, Chung and Lu [2]. In Section 8, some of these models will be discussed.

Here we give some definitions. The expected average degree d of a random graph G in $G(\mathbf{w})$ is defined to be

$$d = \frac{1}{n}\sum_{i=1}^{n} w_i.$$

For a subset S of vertices, the volume of S, denoted by $\mathrm{Vol}(S)$, is the sum of expected degrees in S.

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

In particular, the volume $\mathrm{Vol}(G)$ of $G(\mathbf{w})$ is just $\sum_i w_i$. The edge probability p_{ij} in (1.1) can be written as:

$$p_{ij} = \frac{w_i w_j}{\mathrm{Vol}(G)} = w_i w_j \rho,$$

where

$$\rho := \frac{1}{\mathrm{Vol}(G)} = \frac{1}{nd}.$$

A connected component C is said to be ε-*small* for an $\varepsilon < 1/2$ if the volume of C is at most $\varepsilon \mathrm{Vol}(G)$. We say that a component is *c-giant* if its volume is at least $c\mathrm{Vol}(G)$, for some constant $c > 0$. A giant component, if exists, is almost surely unique (to be proved later in Section 6).

For a subset S of vertices, a typical measure is the number of vertices in S that we call the size of S. In the classical random graph $G(n, p)$, a giant component is a connected component having at least cn vertices for some constant c. Our definition of the giant component involves the volume instead of the size of the connected component. In fact, the definition for the giant component using the size of the component simply does not work for random graphs with general degree distributions, as illustrated in the following example.

Example 1.1. We consider the degree sequence $\mathbf{w}$ consisting of n^α vertices with weight 2 and the other vertices with weight 0. Here α is a constant satisfying $\frac{1}{2} < \alpha < 1$. The random graph $G(\mathbf{w})$ is a union of a random graph $G(n^\alpha, \frac{2}{n^\alpha})$ and some isolated vertices. Therefore, the largest connected component does not have $\Theta(n)$ vertices.

If the average degree d satisfies $d \geq 1+\delta$, where δ is a positive constant, we will show that almost surely any ε-small connected component has size at most $O(\log n)$ (detailed in Theorems 1.1–1.2). Such components will be called small components. An upper bound for the sizes of small components will be given in terms of the average degree d.

Here we state the main results which will be proved in subsequent sections.

Theorem 1.1. *For any positive $\varepsilon < 1$ and $d > \frac{4}{e(1-\varepsilon)^2} \approx (1+2\varepsilon)1.4715\ldots$, in a random graph in $G(\mathbf{w})$ with average degree d, almost surely every connected component has volume either at least εn or has size at most $\frac{\log n}{1+\log d-\log 4+2\log(1-\varepsilon)}$. The upper bound $\frac{\log n}{1+\log d-\log 4}$ for small components is asymptotically best possible for large d.*

Theorem 1.2. *For any positive $\varepsilon < 1$ and d satisfying $\frac{1}{1-\varepsilon} < d < \frac{2}{1-\varepsilon}$, in a random graph in $G(\mathbf{w})$ with average degree d, every connected component almost surely has volume either at least εn or has at most $\frac{\log n}{d-1-\log d-\varepsilon d}$ vertices. The upper bound $\frac{\log n}{d-1-\log d}$ is asymptotically best possible.*

We consider the second-order average degree $\tilde{d}$ which is the weighted average of the squares of the vertex weights. Namely,

$$\tilde{d} = \sum w_i^2 \rho.$$

Clearly,

$$\tilde{d} = \frac{\sum w_i^2}{\sum w_i} \geq \frac{\sum w_i}{n} = d.$$

For the classical random graphs $G(n, p)$, we have $\tilde{d} = d = np$. In the seminal paper of Erdős and Rényi [19], it was shown that for any $\varepsilon > 0$, there is a giant component if $np \geq 1+\varepsilon$, and there is no giant component if $np \leq 1-\varepsilon$. Furthermore, a double jump occurs near $np = 1$, with the largest component of size $\Theta(n^{2/3})$ if $|np-1| = o(n^{-1/3})$. For random graphs $G(\mathbf{w})$ of general degree distribution, the evolution is more complicated.

Theorem 1.3. *For a random graph G with a given expected degree sequence having average degree $d > 1+\delta > 1$, almost surely G has a unique giant component.*

(i) *If $d \geq e$, the volume of the unique giant component is almost surely at least*

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

where e denotes the base of the natural logarithm.

(ii) *If $1+\delta \leq d \leq e$, the volume of the unique giant component is almost surely at least*

$$\left(1 - \frac{1+\log d}{d} + o(1)\right) \mathrm{Vol}(G).$$

If the second-order average degree $\tilde{d} \leq 1-\delta$, then almost surely, there is no giant component.

The proof of Theorem 1.3 is given in Section 7. A natural question arises concerning the relationship of the degrees to the emergence of the giant component for the range of $\tilde{d} > 1 > d$. The examples in Section 3 illustrate both the existence and non-existence for some degree distributions satisfying $\tilde{d} > 1 > d$. Numerous questions arise. For example, it would be of interest to characterize degree sequences for which the phase transition occurs at $\tilde{d} = 1$. For what degree sequences, are there double jumps at such phase transition?

2. Basic Facts and Examples

We will use the following inequality which is a generalization of the Chernoff inequalities for binomial distribution:

Lemma 2.1. *Let $X_1, \ldots, X_n$ be independent random variables with*

$$Pr(X_i = 1) = p_i, \qquad Pr(X_i = 0) = 1 - p_i.$$

For $X = \sum_{i=1}^n a_i X_i$, we have $E(X) = \sum_{i=1}^n a_i p_i$ and we define $\nu = \sum_{i=1}^n a_i^2 p_i$. Then we have

$$Pr(X < E(X) - \lambda) \leq e^{-\lambda^2/2\nu}, \tag{2.1}$$

$$Pr(X > E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\nu + a\lambda/3)}}, \tag{2.2}$$

where $a = \max\{a_1, a_2, \ldots, a_n\}$.

Inequality (2.2) is a corollary of a general concentration inequality (see Theorem 2.7 in the survey paper by McDiarmid [30]). Inequality (2.1) which is a slight improvement of the inequality in [30] can be proved as follows.

Proof. For any $0 \leq p \leq 1$, and $x \geq 0$, we define $f(x) = px + \ln(1 - p + pe^{-x})$ and

$g(x) = px^2/2$. Then we have $f(0) = g(0) = 0$, and $f'(0) = g'(0) = 0$. Also,

$$\begin{aligned} f''(x) &= \frac{p(1-p)e^{-x}}{(1-p+pe^{-x})^2} \\ &\leq \frac{p(1-p)e^{-x}}{(2\sqrt{(1-p)e^{-x}} - (1-p)e^{-x})^2} \\ &\leq \frac{p(1-p)e^{-x}}{(\sqrt{(1-p)e^{-x}})^2} \\ &= p = g''(x). \end{aligned}$$

Hence we have $px + \ln(1-p+pe^{-x}) \leq px^2/2$ for any $x \geq 0$. For any $t > 0$, we have

$$E(e^{-a_i t(X_i - p_i)}) = p_i e^{-ta_i(1-p_i)} + (1-p_i)e^{p_i t a_i} = e^{p_i t a_i + \ln(1-p_i+p_i e^{-ta_i})} \leq e^{\frac{p_i (ta_i)^2}{2}}.$$

Hence

$$\begin{aligned} E(e^{-t(X - \sum_{i=1}^n a_i p_i)}) &= \prod_{i=1}^n e^{-t(X_i - p_i a_i)} \\ &\leq \prod_{i=1}^n e^{\frac{p_i (ta_i)^2}{2}} \\ &= e^{\sum_{i=1}^n \frac{p_i (ta_i)^2}{2}} \\ &= e^{\frac{t^2 \mathrm{v}}{2}}. \end{aligned}$$

We have

$$\begin{aligned} Pr(X - \sum_{i=1}^n a_i p_i < -\lambda) &= Pr(e^{-t(X - \sum_{i=1}^n a_i p_i)} > e^{t\lambda}) \\ &\leq E(e^{-t(X - \sum_{i=1}^n a_i p_i)}) e^{-t\lambda} \\ &\leq e^{\frac{t^2 \mathrm{v}}{2} - t\lambda} \\ &= e^{-\frac{\lambda^2}{2\mathrm{v}}} \end{aligned}$$

by choosing $t = \frac{\lambda}{\mathrm{v}}$. This completes the proof of Lemma 2.1. ∎

As immediate consequences of Lemma 2.1, the following facts then follow.

Fact 1. For a graph G in $G(\mathbf{w})$, with probability $1 - e^{-c^2/2}$, the number d_i of edges incident to a vertex v_i satisfies

$$d_i > w_i - c\sqrt{w_i},$$

and

$$Prob(d_i < (1+\varepsilon)w_i) > 1 - e^{-\varepsilon^2 w_i/(2+2\varepsilon/3)}.$$

Fact 2. With probability $1-2e^{-c^2/2}$, the number $e(G)$ of edges in G, satisfies

$$2e(G) > \mathrm{Vol}(G) - c\sqrt{\mathrm{Vol}(G)}.$$

In the other direction,

$$Prob(2e(G) < (1+\varepsilon)\mathrm{Vol}(G)) > 1 - e^{-\varepsilon^2 \mathrm{Vol}(G)/(2+2\varepsilon/3)}.$$

With probability $1-\frac{2}{n}$, all vertices v_i satisfy

$$2\sqrt{w_i \log n} \le d_{v_i} - w_i \le \frac{2}{3}\log n + \sqrt{\left(\frac{2}{3}\log n\right)^2 + 4w_i \log n}.$$

Fact 3. With probability at least $1-e^{-c}$, the number of edges $e(S)$ between pairs of vertices in S is at least $\frac{1}{2}\mathrm{Vol}(S)^2\rho - \mathrm{Vol}(S)\sqrt{\rho c}$.

In the remainder of this section, we will give several examples with proofs which illustrate the sharpness of the main results. These examples are also instrumental for developing methods later on for dealing with random graphs with given expected degree distributions.

Example 2.1. For the following choices of $\mathbf{w}$ with $d \le 1$ and $\tilde{d} > 1$, a random graph in $G(\mathbf{w})$ almost surely has no giant component.

Let ε be a constant satisfying $1 > \varepsilon > 0$. For each of the first $n-m$ vertices, the weight is set to be $1-\varepsilon$. For each of the remaining m vertices, the weight is set to be x satisfying

$$mx = o(\frac{n}{\log n}) \quad \text{and} \quad mx^2 > Cn > n.$$

(For example, we can choose $m = \lceil \log n \rceil$, $x = \sqrt{(1-\varepsilon)n/2}$ and $C = 10$.) We have

$$\mathrm{Vol}(G) = (n-m)(1-\varepsilon) + mx \approx (1-\varepsilon)n.$$

$$d = \frac{\mathrm{Vol}(G)}{n} \approx (1-\varepsilon).$$

$$\tilde{d} = \frac{\mathrm{Vol}_2(G)}{\mathrm{Vol}(G)} = \frac{(n-m)(1-\varepsilon)^2 + mx^2}{(1-\varepsilon)n} > 1 - \varepsilon + \frac{C}{1-\varepsilon} > 1.$$

Let G_1 denote the induced graph on the set S_1 of vertices with weight $1-\varepsilon$, and G_2 denote the induced subgraph on S_2, the set of vertices with weight x.

From [19], if $Np > 1+\varepsilon$, almost surely $G(N,p)$ has a giant component. If $Np < 1-\varepsilon$, then almost surely $G(N,p)$ does not have a giant component and all components have sizes at most $O(\log N)$.

To apply the above result to G_1, we select $N = n-m \approx n$, and $p = (1-\varepsilon)^2\rho \approx (1-\varepsilon)\frac{1}{n}$. Thus, we have $Np \approx (1-\varepsilon) < 1$ and consequently almost surely all components of G_1 have size at most $O(\log N) = O(\log n)$.

We will next show that there is no giant component in G. We first construct an auxiliary graph G' from G as follows. A new vertex v is added to G, and is connected to all vertices in S_2 but to no vertex in S_1. The following facts are immediate.

1. Every connected component of G must be contained in some component in G'.
2. G' has a special component C containing v and all vertices in S_2.
3. Components of G' other than C are components of G_1, which almost surely have at most $O(\log n)$ vertices and volume $O(\log n)$.

Now we will use a branching process starting from v to reveal the component C. For a subset S, we define the i-boundary $\Gamma_i(S) = \{u\colon d(u,S) = i\}$. We have

$$\Gamma_1(v) = S_2.$$

For each $u \in S_1$, the probability that $u \in \Gamma(S_2)$ is

$$1-(1-(1-\varepsilon)x\rho)^m \approx (1-\varepsilon)mx\rho, \quad \text{since } mx\rho = o(1).$$

The size of $\Gamma(S_2)$ can be upper bounded by a sum of $n-m$ independent 0-1 variables. The probability of each random variable X_i having value one is about $(1-\varepsilon)mx\rho$. These random variables are mutually independent. Using Lemma 2.1, $X = \sum_i X_i$ satisfies

$$Pr(X - E(X) > \lambda) < e^{-\frac{\lambda^2}{2(E(X)+\lambda/3)}},$$

where we choose $\lambda = E(X) \approx (n-m)(1-\varepsilon)mx\rho \approx mx$.

With probability at least $1-e^{-3mx/8} = 1-o(1)$, the size of $\Gamma(S_2)$ is at most $2mx$. Note that $\Gamma_2(v) = \Gamma(S_2)$ is completely contained in S_1, and so are the i-boundary $\Gamma_i(v)$ for all $i \geq 2$. Since in G_1, almost surely any branching process can expand at most $O(\log n)$ vertices, the total size of C is almost surely at most $2mx\ O(\log n) + m + 1 = O(mx\log n)$. The volume of $C \setminus \{v\}$ is almost surely at most $2mx\ O(\log n)(1-\varepsilon) + mx = O(mx\log n)$. Hence each component in G almost surely can have volume at most $O(mx\log n) = o(n)$ and consequently there is no giant component in G.

Example 2.2. For the following choice of degree sequence $\mathbf{w}$ with $d < 1$ and $\tilde{d} > 1$, a random graph in $G(\mathbf{w})$ almost surely has a giant component.

Let M be a very large but fixed constant. For each of the first $\lceil n - \frac{n}{M} \rceil$ vertices, the weight is set to be $x = o(1)$. For the remaining $\frac{n}{M}$ vertices, each weight is set to $1+\varepsilon$. In this example, we have

$$\begin{aligned}
\mathrm{Vol}(G) &\approx \frac{(M-1)n}{M}x + \frac{1+\varepsilon}{M}n = \frac{1+\varepsilon+o(1)}{M}n,\\
d &= \frac{\mathrm{Vol}(G)}{n} = \frac{1+\varepsilon+o(1)}{M} \ll 1,\\
\tilde{d} &= \frac{\mathrm{Vol}_2(G)}{\mathrm{Vol}(G)} = 1+\varepsilon-o(1) > 1.
\end{aligned}$$

Note that $G(\mathbf{w})$ contains a classical random graph $G(N,p)$, where $N = \frac{n}{M}$, and $p = \frac{M(1+\varepsilon+o(1))}{n}$. Since $Np = \frac{n}{M}\frac{M(1+\varepsilon+o(1))}{n} = 1+\varepsilon+o(1) > 1$, almost surely $G(N,p)$ has a giant component of size $\Theta(N) = \Theta(n)$. The component of G containing this connected subset has at least $\Theta(n)$ vertices and at least $\Theta(\mathrm{Vol}(G))$ edges.

3. The Expected Number of Components of Size k

In this section, we consider the probability of having a connected component of size k.

Suppose that we have a subset of vertices $S = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$ with weights w_{i_1}, $w_{i_2}, \ldots, w_{i_k}$. The probability that there is no edge leaving S is

$$\prod_{v_i \in S, v_j \notin S} (1 - w_i w_j \rho) \approx e^{-\rho \sum_{v_i \in S, v_j \notin S} w_i w_j}$$

$$= e^{-\rho \mathrm{Vol}(S)(\mathrm{Vol}(G) - \mathrm{Vol}(S))}. \tag{3.1}$$

If S is a connected component, the induced subgraph on S contains at least one spanning tree T. The probability of containing a spanning tree T is

$$Pr(T) = \prod_{(v_{i_j} v_{i_l}) \in E(T)} w_{i_j} w_{i_l} \rho.$$

Hence the probability of having a connected spanning graph on S is at most

$$\sum_T Pr(T) = \sum_T \prod_{(v_{i_j} v_{i_l}) \in E(T)} w_{i_j} w_{i_l} \rho,$$

where T ranges over all spanning trees on S.

By a generalized version of the matrix-tree Theorem [34], the above sum equals the determinant of any $k-1$ by $k-1$ principal sub-matrix of the matrix $D - A$, where A is the matrix

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

and D is the diagonal matrix $\mathrm{diag}(w_{i_1}(\mathrm{Vol}(S) - w_{i_1})\rho, \ldots, w_{i_k}(\mathrm{Vol}(S) w_{i_k} - w_{i_k})\rho)$. By evaluating the determinant, we conclude that

$$\sum_T P(T) = w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{Vol}(S)^{k-2} \rho^{k-1}. \tag{3.2}$$

By combining (3.1) and (3.2), we have proved the following:

Lemma 3.1. *The expected value $E(X_k)$ of the number of connected components of size k is at most*

$$E(X_k) \leq \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{Vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{Vol}(S)(1 - \mathrm{Vol}(S)/\mathrm{Vol}(G))}, \tag{3.3}$$

where the sum ranges over all sets S of k vertices.

Lemma 3.2. *For a positive $\varepsilon < 1$, The expected value $E(Y_k)$ of the number of ε-small connected components of size k is at most*

$$E(Y_k) \leq \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{Vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{Vol}(S)(1-\varepsilon)}, \tag{3.4}$$

where the sum ranges over all sets S of k vertices with $\mathrm{Vol}(S) < \varepsilon \mathrm{Vol}(G)$.

4. Proof of Theorem 1.1

Suppose that G is a random graph in $G(\mathbf{w})$ with expected average degree $d > 1+\delta$. We want to show that the expected number $E(Y_k)$ of ε-small components of size k is small.

From Lemma 3.2, it suffices to upper bound

$$f(k) = \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{Vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{Vol}(S)(1-\varepsilon)}.$$

By using the fact that the function $x^{2k-2}e^{-x(1-\varepsilon)}$ achieves its maximum value at $x = (2k-2)/(1-\varepsilon)$, we have

$$\begin{aligned}
f(k) &= \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} \mathrm{Vol}(S)^{k-2} \rho^{k-1} e^{-\mathrm{Vol}(S)(1-\varepsilon)} \\
&\le \sum_S \frac{\rho^{k-1}}{k^k} \mathrm{Vol}(S)^{2k-2} e^{-\mathrm{Vol}(S)(1-\varepsilon)} \\
&\le \sum_S \frac{\rho^{k-1}}{k^k} \left(\frac{2k-2}{1-\varepsilon}\right)^{2k-2} e^{-(2k-2)} \\
&\le \frac{n^k}{k!} \frac{\rho^{k-1}}{k^k} \left(\frac{2k-2}{1-\varepsilon}\right)^{2k-2} e^{-(2k-2)} \\
&\le \frac{1}{4\rho(k-1)^2} (n\rho)^k \left(\frac{2}{1-\varepsilon}\right)^{2k} e^{-k} \\
&\le \frac{1}{4\rho(k-1)^2} \left(\frac{4}{de(1-\varepsilon)^2}\right)^k.
\end{aligned}$$

The above inequality is useful when $d > \frac{4}{e(1-\varepsilon)^2}$ which is an assumption for Theorem 1.1. If k satisfies $\frac{\log n}{1+\log d-\log(4)-2\varepsilon} < k < \frac{2\log n}{1+\log d-\log(4)-2\varepsilon}$, then

$$f(k) \le \frac{1}{4n\rho(k-1)^2} = O\left(\frac{1}{\log n}\right).$$

When k satisfies $\frac{2\log n}{1+\log d-\log(4)-2\varepsilon} \le k \le n$, we have

$$f(k) \le \frac{1}{4n^2\rho(k-1)^2} = O\left(\frac{1}{n\log n}\right).$$

For $k_0 = \frac{\log n}{1+\log d-\log(4)-2\varepsilon}$, the probability that a small component has size $k > k_0$ is at most

$$\sum_{k>k_0} f(k) \le \frac{\log n}{1+\log d-\log(4)-2\varepsilon} \times o\left(\frac{1}{\log n}\right) + n \times o\left(\frac{1}{n\log n}\right) = o(1).$$

Therefore, almost surely the size of a ε-small component is at most $k_0 = \frac{\log n}{1+\log d-\log 4-2\varepsilon}$. We have proved the first part of Theorem 1.1.

To show that the above upper bound is asymptotically best possible for large d, we consider the following example.

Example 4.1. We consider a random graph with the following weights as the expected degree sequence. Here we assume that $d > 10$.

There are $n^{2/3}$ vertices with weights $(d-1)n^{1/3}+1$. Each of the remaining $n-n^{2/3}$ vertices has weight 1. The average (weight) degree is exactly d.

Let S_1 denote the set of vertices with weight 1, and S_2 denote the set of vertices with weight $(d-1)n^{1/3}+1$. Let G_i be the induced graph of G on S_i, for $i = 1, 2$. The graph G_2 is a classical random graph $G(N, p)$ with $N = n^{2/3}$ and $Np = n^{2/3}((d-1)n^{1/3}+1)^2/(nd) = \Theta(\sqrt{N})$. Almost surely G_2 is connected. In fact, G_2 is contained in the giant component of G. Let c denote the fraction of vertices, which is not in the giant component. We claim that c is bounded away from 0.

To prove the claim, we consider a special branching process. We first reveal all edges in G_2. Then we examine the boundary of S_2 in S_1, the 2-boundary of S_2, and so on, which eventually exposing all vertices in the giant component of G. For any vertex $u \in S_1$, the probability of u in $\Gamma(S_2)$ is

$$1-\left(1-\frac{(d-1)n^{1/3}+1}{nd}\right)^{n^{2/3}} \approx 1-e^{-1+\frac{1}{d}}.$$

The size of $\Gamma(S_2)$ can be well approximated by the binomial distribution with $N = n-n^{2/3}$ and $p = 1-e^{-1+\frac{1}{d}}$. Thus with high probability, the size is about $(1-e^{-1+\frac{1}{d}})n$. We will estimate the size of $\Gamma_i(S_2)$ for $i > 1$ by induction. Suppose $|\Gamma_i(S_i)|$ is highly concentrated on $a_i n$ for some constant a_i, for $i \geq 2$. Let $c_i = 1-\sum_{k=1}^{i} a_i$. For any vertex u not in $\cup_{j\leq i}\Gamma_i(S_i)$, the probability of $u \in \Gamma_{i+1}(S_2)$ is

$$1-\left(1-\frac{1}{nd}\right)^{a_i n} \approx 1-e^{-\frac{a_i}{d}}.$$

The size of $\Gamma_{i+1}(S_2)$ can be approximated by the binomial distribution with $N = c_i n$ and $p = 1-e^{-\frac{a_i}{d}}$. By the definition of a_i. We have

$$a_{i+1} = c_i(1-e^{-\frac{a_i}{d}}),$$

$$c_{i+1} = c_i - a_{i+1} = c_i e^{-\frac{a_i}{d}}.$$

Hence

$$c_{i+1} = c_1 \prod_{k=1}^{i} e^{-\frac{a_k}{d}} = (1-e^{-1+\frac{1}{d}})e^{-\frac{1-c_i}{d}}.$$

From the above recurrence for c_i, we see that the limit $c = \lim_{i\to\infty} c_i$ exists and satisfies

$$c = (1-e^{-1+\frac{1}{d}})e^{-\frac{1-c}{d}}.$$

It is easy to see that the above equation has a unique solution of c in $[0, 1]$ for $d > 1$ and the solution for c increases as a function of d. Since we choose $d > 10$, c is bounded away from zero. The claim is proved.

The size of the second largest component can be estimated as follows. After removing the giant component from G, the remaining graph is a classical random graph $G(t, p)$ with $t = cn$ and $p = \frac{1}{nd} = \frac{c}{dt}$. By [19], the largest component of $G(t, \frac{c}{dt})$ with $d < 1$ has size at most

$$\frac{\log n - 5/2 \log\log n}{\frac{c}{d} - 1 - \log\frac{c}{d}} = \frac{(1+o(1))\log n}{\log d - \log c - 1 + \frac{c}{d}}.$$

The constant $\frac{1}{\log d - \log c - 1 + \frac{c}{d}}$ is asymptotically close to $\frac{1}{1+\log d - \log 4}$ when d is large and ε is arbitrarily small. This completes the proof for Theorem 1.1.

Remark. For the classical random graph $G(n, \frac{d}{n})$ with $d > 1+\varepsilon$, it was shown [19] that the size of the second largest connected components is approximately the same as the size of the largest connected component of $G(m, \frac{c}{m})$, where c is the unique solution of $ce^{-c} = de^{-d}$ for c in $(0, 1)$, and $m = \frac{c}{d}n$. From [19], the largest component of $G(m, \frac{c}{m})$ has size about

$$\frac{\log m - 5/2 \log\log m}{c - 1 - \log c} = \frac{(1+o(1))\log n}{d - 1 - \log d},$$

which is consistent with Theorem 1.1.

5. Proof of Theorem 1.2

In this section, we consider $\frac{1}{1-\varepsilon} < d < \frac{2}{1-\varepsilon}$. The methods for proving Theorem 1.1 no longer work and a different estimate for $f(k)$ is needed here. We will derive an upper bound for the expected number $E(Y_k)$ of connected components of size k by using inequality (3.1). First, we split $f(k)$ into two parts as follows:

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

where

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

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

To bound $f_1(k)$, we note that $x^{2k-2}e^{-x(1-\varepsilon)}$ is an increasing function when $x < (2k-2)/(1-\varepsilon)$. Thus we have

$$\mathrm{Vol}(S)^{2k-2} e^{-\mathrm{Vol}(S)} \leq (dk)^{2k-2} e^{-dk(1-\varepsilon)},$$

since $\mathrm{Vol}(S) < dk < (2k-2)/(1-\varepsilon)$. This implies

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

Next, we consider $f_2(k)$. Since $x^{k-2}e^{-x(1-\varepsilon)}$ is a decreasing function when $x > (k-2)/(1-\varepsilon)$, we have

$$\mathrm{Vol}(S)^{k-2}e^{-\mathrm{Vol}(S)(1-\varepsilon)} \le (dk)^{k-2}e^{-dk(1-\varepsilon)}$$

by using $\mathrm{Vol}(S) \ge dk > \frac{k-2}{1-\varepsilon}$. Therefore, we have

$$\begin{aligned}
f_2(k) &= \sum_{\mathrm{Vol}(S)\ge dk} w_{i_1}w_{i_2}\cdots w_{i_k}\mathrm{Vol}(S)^{k-2}\rho^{k-1}e^{-\mathrm{Vol}(S)(1-\varepsilon)} \\
&\le \sum_{\mathrm{Vol}(S)\ge dk} w_{i_1}\cdots w_{i_k}\rho^{k-1}(dk)^{k-2}e^{-dk(1-\varepsilon)} \\
&\le \sum_{S} w_{i_1}w_{i_2}\cdots w_{i_k}\rho^{k-1}(dk)^{k-2}e^{-dk(1-\varepsilon)} \\
&< \frac{\mathrm{Vol}(G)^k}{k!}\rho^{k-1}(dk)^{k-2}e^{-dk(1-\varepsilon)} \\
&\le \frac{1}{d^2k^2\rho}d^k e^{-(d(1-\varepsilon)-1)k} \\
&\le \frac{n}{dk^2}\left(\frac{d}{e^{(d(1-\varepsilon)-1)}}\right)^k.
\end{aligned}$$

Together we have

$$f(k) = f_1(k) + f_2(k) \le \frac{2n}{dk^2}\left(\frac{de}{e^{d(1-\varepsilon)-1}}\right)^k.$$

If $\frac{\log n}{d(1-\varepsilon)-1-\log d} < k < \frac{2\log n}{d(1-\varepsilon)-1-\log d}$, we have

$$f(k) \leq \frac{2}{dk^2} = O\left(\frac{1}{\log^2 n}\right).$$

If $\frac{2\log n}{d(1-\varepsilon)-1-\log d} \leq k \leq n$, then

$$f(k) \leq \frac{2}{ndk^2} = O\left(\frac{1}{n\log^2 n}\right).$$

By setting $k_1 = \frac{\log n}{d(1-\varepsilon)-1-\log d}$, the probability of having a small component of size $k > k_1$ is at most

$$\sum_{k>k_1} f(k) \leq \frac{\log n}{d(1-\varepsilon)-1-\log d} \times O\left(\frac{1}{\log^2 n}\right) + n \times O\left(\frac{1}{n\log^2 n}\right) = o(1).$$

Therefore, almost surely the size of a small component is at most $k_1 = \frac{\log n}{d(1-\varepsilon)-1-\log d}$.

To see that this upper bound is best possible, we consider the following example. In the random graph $G(n, \frac{d}{n})$ with $d < 1$, the largest component has size about $\frac{\log n - 5/2 \log\log n}{d-1-\log d}$ (see [19]), as desired.

6. Proof of Theorem 1.3

Before proving Theorem 1.3, we first prove several reductions.

Fact 4. Suppose that a random graph G in $G(\mathbf{w})$ has average degree $d > 1+\delta$, and contains a connected subset having more than $C\log n$ vertices, where $C = max\{\frac{2}{\delta-\log\delta}, 10\}$. Then almost surely there is a giant component in G.

Proof. Fact 4 is an immediate consequence of Theorems 1.1 and 1.2, subject to verifying the required assumptions which follow from the definition of C as follows:

$$C \geq \frac{2}{\delta-\log\delta} > \frac{1}{d-1-\log d-\varepsilon_1 d}$$

for some $\varepsilon_1 > 0$ when $1+\delta < d \leq 2$, and

$$C \geq 10 > \frac{1}{1+\log d-\log 4+2\log(1-\varepsilon_2)}$$

for some $\varepsilon_2 > 0$ when $2 < d$. ∎

Fact 5. An induced subgraph H in $G \in G(\mathbf{w})$ is a random graph with given expected sequence $\mathbf{w}'$ which consists of $w_i' = w_i \mathrm{Vol}(H)\rho$ for v_i in H.

The proof follows from the fact that the expected degree of v_i in H is just

$$\sum_{j\in V(H)} w_i w_j \rho = w_i \mathrm{Vol}(H)\rho.$$

Lemma 6.1. *Suppose that in a random graph $G \in G(\mathbf{w})$, there is a value M (independent of n) so that $w_i \le M$ for all i, and the average expected degree $d \ge 1+\delta$, where δ is a positive constant. Then almost surely G has a unique giant component.*

Proof. We use a branching process as follows: First choose any vertex u with weight greater than 1 and carry out a breadth first search of its connected component. A vertex is called *unexamined* if it has been discovered to be in the component, but we have not yet exposed its neighbors. Let X_k be the sum of the weights of all unexamined vertices at depth k. For any vertex v_j not yet been exposed, the probability that v_j is to be discovered in the component at depth $k+1$ is well approximated by $X_k w_j \rho$. Hence, the expected value of X_{k+1} is $\sum_j X_k w_j^2 \rho \approx X_k \tilde{d}$. By Lemma 2.1, we have

$$Pr(X_{k+1} < X_k \tilde{d} - \lambda) \le e^{-\lambda^2/2\nu},$$

where $\nu = \sum_j X_k w_j^3 \rho \le M X_k \tilde{d}$. By choosing $\lambda = X_k(\tilde{d}-1)/2$, we have

$$Pr(X_{k+1} < \frac{1}{2}(\tilde{d}+1)X_k) \le e^{-\frac{(\tilde{d}-1)^2}{8M\tilde{d}}X_k}.$$

For each k, X_k increases by a factor of $\frac{1}{2}(\tilde{d}+1) > 1$ with failure probability at most $(1-c)^X$, where c is a positive constant satisfying $1-c > e^{-\frac{(\tilde{d}-1)^2}{8M\tilde{d}}}$. Since $\sum_{j=1}^{\infty}(1-c)^j$ converges, there exists a constant t_0 satisfying $\sum_{j\ge t_0}(1-c)^j < 1-\varepsilon$ for a positive constant ε. With positive constant probability, X will increase at least by a factor of $\frac{1}{2}(\tilde{d}+1) > 1$ if $X_1 > t_0$.

Since t_0 is an absolute constant, the event $X_1 > t_0$ occurs with some positive constant probability. If the branching process dies early (i.e., the connected component is small), then we just start another branching process from a new vertex with weight greater than 1. (There are enough such vertices since the number of vertices with weight greater than 1 is at least $\frac{d-1}{M}n$.) After at most $\Theta(\log n)$ tries, almost surely the giant component will be revealed. ∎

The proof of Theorem 1.3. Let $y < 1$ be a constant satisfying $(1-y)^2 d > 1+\delta/2$ (for example, choose $y = 1-\delta/4$). We sort the vertices so that $w_1 \le w_2 \le \cdots \le w_n$. Let i_0 denote the largest integer satisfying

$$\sum_{i \ge i_0} w_i \approx y\mathrm{Vol}(G).$$

If $w_{i_0} > \frac{2}{y}$, we use Fact 5 which implies there is an induced subgraph on $n-i_0$ vertices having expected degrees > 2. It contains the Erdős-Rényi graph $G(n-i_0, 2/(n-i_0))$ and therefore it contains a component of $c_1(n-i_0)$ vertices for some constant c_1.

If $w_{i_0} \le \frac{2}{y}$, we consider the induced subgraph on the first i_0 vertices. By Fact 5, it has volume $(1-y)^2 dn$ and therefore has average degree at least $1+\delta/2$. Furthermore, all weights are bounded by the constant $2/y$. By Lemma 6.1, it contains a component of volume $c_2 i_0$.

From (1.1), the maximum weight is no more than $\sqrt{\mathrm{Vol}(G)}$. Both i_0 and $n-i_0$ have at least $\sqrt{n}$ vertices. By Fact 4, the giant component almost surely exists.

The above arguments can be used to show the uniqueness of the giant component as well. For any two vertices u and v, we begin a branching process starting at u but stop at the moment when the volume of the set S_1 of exposed vertices reaches $\sqrt{(2+\varepsilon)\mathrm{Vol}(G)\log n}$. Then we begin a new branching process starting at v and stop at the moment when the volume of the exposed vertices S_2 reaches $\sqrt{(2+\varepsilon)\mathrm{Vol}(G)\log n}$. Then the probability of S_1 and S_2 being not connected by any edge is at most

$$\begin{aligned}\prod_{u\in S_1, v\in S_2}(1-w_u w_v\rho) &\le \prod_{u\in S_1, v\in S_2} e^{-w_u w_v\rho}\\ &= e^{-\sum_{u\in S_1, v\in S_2} w_u w_v\rho}\\ &= e^{-\mathrm{Vol}(S_1)\mathrm{Vol}(S_2)\rho}\\ &\le e^{-(2+\varepsilon)\log n}\\ &= n^{-2-\varepsilon}.\end{aligned}$$

The probability that every pair of vertices each in a giant component are connected is at least $1-n^{-\varepsilon}$. Thus, the giant component is almost surely unique.

Next, we consider the volume of the giant component. We want to show the following:

(i) If $d \ge e$, the volume of the giant component is at least $\left(1-\frac{2}{\sqrt{de}}+o(1)\right)\mathrm{Vol}(G)$.

(ii) If $1+\delta \le d \le e$, the volume of the giant component is at least

$$\left(1-\frac{1+\log d}{d}+o(1)\right)\mathrm{Vol}(G).$$

We first consider the case of $d \ge e$. If (i) does not hold, then the giant component is ε-small for some ε satisfying $\varepsilon < 1-\frac{2}{\sqrt{de}}$. By Theorem 1.1, the size of the giant component is at most $\frac{\log n}{1+\log d-\log 4+2\log(1-\varepsilon)}$. Hence there is one vertex with weight w greater than or equal to the average:

$$w \ge \frac{\varepsilon\mathrm{Vol}(G)}{\frac{\log n}{1+\log d-\log 4+2\log(1-\varepsilon)}} \ge c_\varepsilon\frac{\mathrm{Vol}(G)}{\log n}.$$

It is easy to check that

$$w^2\rho \gg 1,$$

which contradicts our assumption (1.1). Hence the volume of the giant component is at least $\left(1-\frac{2}{\sqrt{de}}+o(1)\right)\mathrm{Vol}(G)$ if $d \ge e$.

For the case of $d_0 \le d \le e$, we again prove by contradiction. Suppose that all connected components are ε-small for some ε satisfying $\varepsilon < 1-\frac{1+\log n}{d}$. By Theorem 1.2, the size of the giant component is at most $\frac{\log n}{d-1-\log d-\varepsilon d}$. Hence there is one vertex with weight w greater than or equal to the average:

$$w \ge \frac{\varepsilon\mathrm{Vol}(G)}{\frac{\log n}{d-1-\log d-\varepsilon d}} \ge c'_\varepsilon\frac{\mathrm{Vol}(G)}{\log n}.$$

This contradicts the assumption (1.1) and (ii) is proved.

It remains to show that for $\tilde{d}$ smaller than 1, almost surely all components have volumes at most $\sqrt{n}\log n$ and therefore there is no giant component in this case.

Claim. If $\tilde{d} < 1-\delta$, with probability at least $1 - \frac{d\tilde{d}^2}{C^2(1-\tilde{d})}$, all components have volume at most $C\sqrt{n}$.

Proof. Let x be the probability that there is a component having volume greater than $C\sqrt{n}$. Now we choose two random vertices with the probability of being chosen proportional to their weights. Under the condition that there is a component with volume greater than $C\sqrt{n}$, the probability of each vertex in this component is at least $C\sqrt{n}\rho$. Therefore, the probability that the random pair of vertices are in the same component is at least

$$x(C\sqrt{n}\rho)^2 = C^2 x n \rho^2. \tag{6.1}$$

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

$$P_k(u,v) \le \sum_{i_1 i_2 \ldots i_k} (w_u w_{i_1}\rho)\,(w_{i_1} w_{i_2}\rho)\cdots(w_{i_k} w_v \rho) \le w_u w_v \rho \tilde{d}^k.$$

The probability that u and v belong to the same component is at most

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

Since the probabilities of u and v being selected are $w_u\rho$ and $w_v\rho$ respectively, the probability that the random pair of vertices are in the same connected component is at most

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

Combining with (6.1), we have

$$C^2 x n \rho^2 \le \frac{\tilde{d}^2}{1-\tilde{d}}\rho,$$

which implies

$$x \le \frac{d\tilde{d}^2}{C^2(1-\tilde{d})}.$$

Therefore with probability at most $1 - \frac{d\tilde{d}^2}{C^2(1-\tilde{d})}$, all components have size at most $C\sqrt{n}$ as desired. This completes the proof for the claim. By choosing C to be $\log n$, we have shown that with probability at least $1-o(1)$, all components are small. We have completed the proof for Theorem 1.3. ∎

7. Several Random Graph Models

In the literature, the following model, so called *the configuration model*, is often used to construct a random graph with a prescribed degree sequence. It was first introduced by Bender and Canfield [9], refined by Bollobás [10] and also Wormald [35]. A random graph G with given degrees d_v is associated with a random matching in a set N of $\sum d_v$ nodes. Each vertex v corresponds to a set S_v of d_v nodes in N. The number of edges between two vertices u and v is the number of edges in the associated matching with one node in S_u and one node in S_v. It is easy to see that the resulting graph (as a multi-graph) has degrees exactly as required.

Molloy and Reed [31, 32] used the configuration model to show that if there are $d_i(n) \approx \lambda_i n$ vertices of degree i, where $\sum_i \lambda_i = 1$ and $\sum i(i-2)\lambda_i > 0$, then the graph almost surely has a giant component if the following conditions are satisfied.

1. The maximum degree is at most $n^{1/4-\varepsilon}$.
2. $i(i-2)d_i(n)/n$ tends uniformly to $i(i-2)\lambda_i$.
3. The limit
$$L(\mathcal{D}) = \lim_{n\to\infty} \sum_{i\geq 1} i(i-2)d_i(n)/n$$
exists, and the sum approaches the limit uniformly.
4. The degree sequence is graphic.

The advantage of the configuration model is to generate graphs exactly with the prescribed degrees and it is the primary model for examining regular graphs with constant degrees. There are several disadvantages of the configuration model. The analysis of the configuration model is much more complicated due to the dependency of the edges. A random graph from the configuration model is in fact a multigraph instead of a simple graph. The probability of having multiple edges increases rapidly when the degrees increase. In the papers of Molloy and Reed, the condition on maximum degree with an upper bound of $n^{1/4-\varepsilon}$ is required because of the occurrence of multiple edges in the configuration model. Consequently, this model is restrictive for power-law graphs, where the largest degree can be quite large. Furthermore, additional conditions (e.g., Conditions 2 and 3 as in [31, 32]) are often required for the configuration models. In the same way, the classical random graph model $G(n, p)$ is often preferred to the configuration models of random graphs with $p\binom{n}{2}$ edges.

The advantage of the generalized model that we use here is the simplicity without any condition on the degree sequence except for the only assumption (1.1). Our model does not produce the graph with exact given degree sequence. Instead, it yields a random graph with given expected degree sequence.

Another line of approach which simulates realistic graphs is to generate a vertex/edge at a time, starting from one node or a small graph. Although we will not deal with such models in this paper, we will briefly mention several evolution models. Barabási and Albert [7] describe the following graph evolution process. Starting with a small initial graph, at each time step they add a new node and an edge between the new node and each of m random nodes in the existing graph, where m is a parameter of the model. The random nodes are not chosen uniformly. Instead, the probability of

picking a node is weighted according to its existing degree (the edges are assumed to be undirected). Using heuristic analysis with the assumption that the discrete degree distribution is differentiable, they derive a power law for the degree distribution with a power of 3, regardless of m. A power law with power 3 for the degree distribution of this model was independently derived and proved by Bollobás et al. [11].

Kumar et al. [28] proposed three evolution models — "linear growth copying", "exponential growth copying", and "linear growth variants." The *Linear growth coping* model adds one new vertex with d out-links at a time. The destination of i-th out-link of the new vertex is either copied from the corresponding out-link of a "prototype" vertex (chosen randomly) or a random vertex. They showed that the in-degree sequence follows the power law. These models were designed explicitly to model the World Wide Web. Indeed, they show that their model has a large number of complete bipartite subgraphs, as has been observed in the WWW graph, whereas several other models do not. This (and the linear growth variants model) has the similar drawback as the first model in [27]. The out-degree of every vertex is always a constant. Edges and vertices in the *exponential growth copying* model increase exponentially.

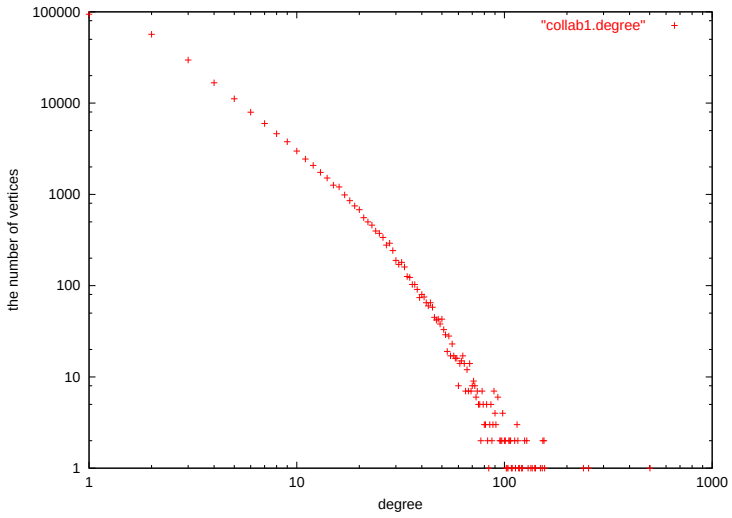

Figure 1: Degree distribution of the collaboration graph.

Aiello et al. described a general random graph evolution process in [3] for generating directed power law graphs with given expected in-degrees and out-degrees. At each time t, a new node is generated and certain edges are added as follows. The end points of new edges can be either the new node or one of the existing nodes. An existing node is selected as the destination (or the origin) with probability proportional to its in-degree (or out-degree). There are four types of edges according to their destinations and origins. A probability space P_t controls the number and the type of edges to be added at time t. Under the assumption that the number of edges added at each time is bounded and P_t has a limiting distribution, Aiello et al. [3] proved this general process

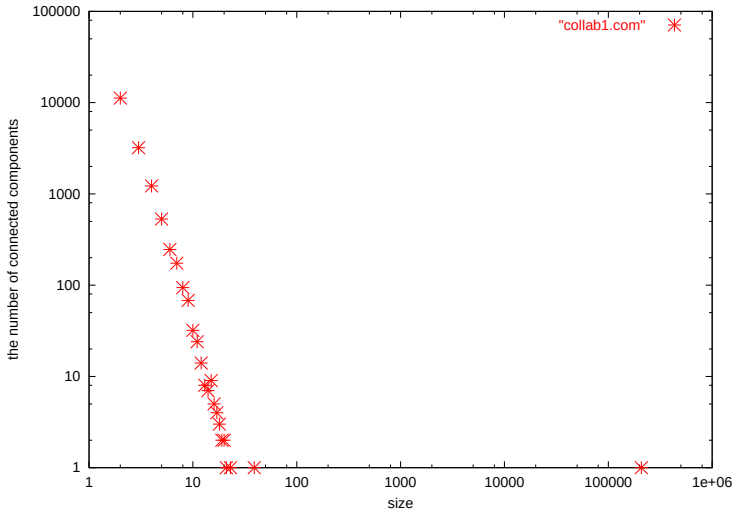

Figure 2: Connected component distribution.

generates power law graphs. The power of the power law of out-degree (or in-degree) equals to $2+\frac{A}{B}$, where A is the expected number of edges per step with the new node as the origin (or the destination) and B is the expected number of edges per step with an existing node as the origin (or the destination). Recently, Cooper and Frieze [17] independently analyzed the above evolution of adding either new vertices or new edges and derived power law degree distribution for vertices of small degrees.

8. Remarks on Power Law Graphs

In this paper, we examine the sizes of connected components of a random graph with given degree sequences. The results and methods here can be useful to examine power law graphs that arise in various context. A power law graph with power α has the number of vertices of degree k proportional to k^{α}. For example, the collaboration based on the data from *Mathematics Review* [22] has about 337,000 vertices (as authors) and about 496,000 edges (as joint publications). So, the average degree d is 2.94. To model the collaboration graph as a random power law graph, the exponent is approximately $\beta = 2.97$ as shown in Figure 1. We also include the actual data on the sizes of connected components in Figure 2.

References

1. L.A. Adamic and B.A. Huberman, Growth dynamics of the World Wide Web, Nature **401** September 9 (1999) 131.
2. 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, pp. 171–180.

3. W. Aiello, F. Chung, and L. Lu, Random evolution in massive graphs, Extended abstract appeared in: The 42th Annual Symposium on Foundation of Computer Sciences, October, 2001, Handbook on Massive Data Sets, Vol. 2, J. Abello, et al., Eds., Kluwer Academic Publishers, 2002, pp. 97–122.
4. L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, Classes of small-world networks, Proc. Natl. Acad. Sci. USA **97** (21) (2000) 11149–11152.
5. N. Alon and J. H. Spencer, The Probabilistic Method, Wiley and Sons, New York, 1992.
6. R.B.R. Azevedo and A.M. Leroi, A power law for cells, Proc. Natl. Acad. Sci. USA **98** (10) (2001) 5699–5704.
7. A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286** (1999) 509–512.
8. A. Barabási, R. Albert, and H. Jeong, Scale-free characteristics of random networks: the topology of the world wide web, Physica A **272** (1999) 173–187.
9. E.A. Bender and E.R. Canfield, The asymptotic number of labelled graphs with given degree sequences, J. Combin. Theory, Ser. A **24** (1978) 296–307.
10. B. Bollobás, Random Graphs, Academic, New York, 1985.
11. B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, The degree sequence of a scale-free random graph process, Random Structures Algorithms **18** (3) (2001) 279–290.
12. 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, Comput. Networks **33** (1-6) (2000) 309–321.
13. K. Calvert, M. Doar, and E. Zegura, Modeling internet topology, IEEE Commun. Mag. **35** (6) (1997) 160–163.
14. F. Chung and L. Lu, The diameter of random sparse graphs, Adv. Appl. Math. **26** (2001) 257–279.
15. F. Chung and L. Lu, Average distances in random graphs with given expected degree sequences, Proc. Natl. Acad. Sci. USA, to appear.
16. F. Chung, L. Lu, and V. Vu, Eigenvalues of random power law graphs, preprint.
17. C. Cooper and A. Frieze, A general model of web graphs, Proceedings of ESA 2001, pp. 500–511.
18. P. Erdős and T. Gallai, Gráfok előírt fokú pontokkal (Graphs with points of prescribed degrees, in Hungarian), Mat. Lapok **11** (1961) 264–274.
19. P. Erdős and A. Rényi, On random graphs, I, Publ. Math. Debrecen **6** (1959) 290–291.
20. M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the Internet topology, Proceedings SIGCOMM, 1999, pp. 251–262.
21. N. Gilbert, A simulation of the structure of academic science, Socialogical Research Online, **2** (2) 1997.
22. J. Grossman, The Erdős Number Project, `http://www.oakland.edu/~grossman/erdoshp.html`.
23. S. Jain and S. Krishna, A model for the emergence of cooperation, interdependence, and structure in evolving networks, Proc. Natl. Acad. Sci. USA **98** (2) (2001) 543–547.
24. S. Janson, T. Łuczak, and A. Ruciński, Random Graphs, Wiley-Interscience, New York, 2000.
25. J. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, The web as a graph: measurements, models and methods, Proceedings of the International Conference on Combinatorics and Computing, 1999, pp. 26–28.

26. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, Trawling the web for emerging cyber communities, Proceedings of the 8th World Wide Web Conference, Toronto, 1999, Extended version in Computer Networks **31** (11–16) (1999) 1481–1493.
27. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, Extracting large-scale knowledge bases from the web, Proceedings of the 25th VLDB Conference, Edinburgh, Scotland, 1999, pp. 639–650.
28. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, Stochastic models for the web graph, Proceedings of the 41st Annual Symposium on Foundations of Computer Science, 2000, pp. 57–65.
29. L. Lu, The diameter of random massive graphs, Proceedings of the Twelfth ACM-SIAM Symposium on Discrete Algorithms, 2001, pp. 912–921.
30. C. McDiarmid, Concentration, Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithms Combin., **16**, Springer, Berlin, 1998, pp. 195–248.
31. M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Structures Algorithms **6** (2–3) (1995) 161–179.
32. M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, Combin. Probab. Comput. **7** (3) (1998) 295–305.
33. M.E.J. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. USA **98** (2) (2001) 404–409.
34. D. West, Introduction to Graph Theory, Prentice Hall, 1996.
35. N. Wormald, Some problems in the enumeration of labelled graphs, Ph.D. Thesis, Newcastle University, 1978.
36. E. Zegura, K. Calvert, and M. Donahoo, A quantitative comparison of graph-based models for Internet topology, IEEE/ACM Transactions on Networking **5** (6) (1997) 770–783.