Combinatorics, Probability and Computing (1998) 7, 295–305. Printed in the United Kingdom

The Size of the Giant Component of a Random Graph with a Given Degree Sequence

MICHAEL MOLLOY[1] and BRUCE REED[2]

[1] Department of Computer Science,
University of Toronto, Toronto, Canada
(e-mail: molloy@cs.toronto.edu)

[2] Equipe Combinatoire, CNRS,
Université Pierre et Marie Curie, Paris, France
(e-mail: reed@ecp6.jussieu.fr)

Received 22 November 1996; revised 29 May 1997

Given a sequence of nonnegative real numbers $\lambda_0, \lambda_1, \ldots$ that sum to 1, we consider a random graph having approximately $\lambda_i n$ vertices of degree i. In [12] the authors essentially show that if $\sum i(i-2)\lambda_i > 0$ then the graph a.s. has a giant component, while if $\sum i(i-2)\lambda_i < 0$ then a.s. all components in the graph are small. In this paper we analyse the size of the giant component in the former case, and the structure of the graph formed by deleting that component. We determine $\epsilon, \lambda'_0, \lambda'_1 \ldots$ such that a.s. the giant component, C, has $\epsilon n + \mathrm{o}(n)$ vertices, and the structure of the graph remaining after deleting C is basically that of a random graph with $n' = n - |C|$ vertices, and with $\lambda'_i n'$ of them of degree i.

1. Introduction and overview

Perhaps the most studied phenomenon in the field of random graphs is the behaviour of the size of the largest component in $G_{n,p}$† when $p = c/n$ for c near 1. For $c < 1$ the size of the largest component is almost surely‡ (a.s.) $\mathrm{O}(\log n)$, for $c = 1$ the size of the largest component is a.s. $\Theta(n^{2/3})$, and for $c > 1$ a.s. the size of the largest component is $\Theta(n)$ while the size of the second largest component is $\mathrm{O}(\log n)$ (see [8], [7] or [9]). For $c > 1$, this largest component is commonly referred to as the *giant component* and the point $p = 1/n$ is referred to as the *critical point* or the *double jump threshold*.

For $c > 1$, we can also determine the approximate size of the giant component, C, as well as the structure of the graph formed by deleting it. Its size is a.s. $\epsilon_c n + \mathrm{o}(n)$, where ϵ_c is the unique solution to $\epsilon + \mathrm{e}^{-c\epsilon} = 1$, and the graph formed by deleting C is essentially

† $G_{n,p}$ is the random graph with n vertices where each edge appears independently with probability p.
‡ We say that a random event E_n holds *almost surely* if $\lim_{n\to\infty} \mathbf{Pr}(E_n) = \infty$.

equivalent to $G_{n',p=d_c/n'}$, where $n' = n - |C| = (1-\epsilon_c)n + o(n)$, and $d_c = c(1-\epsilon_c)$ (see [1] or [10]). (Note that $d_c < 1$.) The latter property is referred to as the Discrete Duality Principle.

In [12], the authors showed that a similar phenomenon occurs among random graphs with a fixed degree sequence. Essentially, we considered random graphs on n vertices with $\lambda_i n + o(n)$ vertices of degree i, for some fixed sequence $\lambda_0, \lambda_1, \ldots$. We introduced the parameter $Q = \sum i(i-2)\lambda_i$ and showed that, if $Q < 0$, then a.s. the size of the largest component is $O(\omega^2 \log n)$, where ω is the highest degree in the graph, and if $Q > 0$, then a.s. the size of the largest component is $\Theta(n)$, and the size of the second largest component is $O(\log n)$.

In this paper we refine our arguments to determine the approximate size of the giant component in such a graph. We also find an analogue to the Discrete Duality Principle, showing that there is a sequence $\lambda'_0, \lambda'_1, \ldots$, such that the graph remaining after deleting the giant component, C, is basically equivalent to a random graph on $n' = n - |C|$ vertices, with approximately $\lambda'_i n'$ vertices of degree i for each i. Of course, $\sum i(i-2)\lambda'_i < 0$.

To be expeditious, we will state our main theorems here, momentarily postponing the definition of a well-behaved sparse asymptotic degree sequence, which was introduced in [12].

Given a sequence of nonnegative reals $\lambda_0, \lambda_1, \ldots$ summing to one, we set $K = \sum_{i \geqslant 0} i\lambda_i$, and define $\chi : [0,1] \to \mathbb{R}$ as

$$\chi(\alpha) = K - 2\alpha - \sum_{i \geqslant 1} i\lambda_i \left(1 - \frac{2\alpha}{K}\right)^{\frac{i}{2}},$$

and we denote the smallest positive solution to $\chi(\alpha) = 0$ (if such a solution exists), by $\alpha_{\mathscr{D}}$. Now setting

$$\begin{aligned} \epsilon_{\mathscr{D}} &= 1 - \sum_{i \geqslant 1} \lambda_i \left(1 - \frac{2\alpha_{\mathscr{D}}}{K}\right)^{\frac{i}{2}}, \\ \lambda'_i &= \frac{\lambda_i}{(1-\epsilon_{\mathscr{D}})} \left(1 - \frac{2\alpha_{\mathscr{D}}}{K}\right)^{\frac{i}{2}}, \end{aligned}$$

we have the following theorems.

Theorem 1. *Let $\mathscr{D} = d_0(n), d_1(n), \ldots$ be a well-behaved sparse asymptotic degree sequence where, for each $i \geqslant 0$, $\lim_{n\to\infty} d_i(n)/n = \lambda_i$ and for which there exists $\epsilon > 0$ such that, for all n and $i \geqslant n^{1/4-\epsilon}$, $d_i(n) = 0$. Suppose that $Q(\mathscr{D}) = \sum i(i-2)\lambda_i > 0$. If G is a random graph with n vertices and degree sequence $\mathscr{D}_n$, then a.s. the giant component of G has size $\epsilon_{\mathscr{D}} n + o(n)$.*

Theorem 2. *Let $\mathscr{D}$ be a degree sequence meeting the conditions of Theorem 1. Let G be a random graph with n vertices and degree sequence $\mathscr{D}_n$. Almost surely, the structure of the graph formed by deleting the largest component, C, from G is essentially the same as that of a random graph on $n' = n - |C| = (1-\epsilon_{\mathscr{D}})n + o(n)$ vertices, with degree sequence $\mathscr{D}'$, for some $\mathscr{D}' = d'_0(n), d'_1(n), \ldots$, where $d'_i(n) = \lambda'_i n + o(n)$.*

https://doi.org/10.1017/S0963548398003526 Published online by Cambridge University Press

Now we will recall the relevant definitions from [12]. Throughout this paper, all asymptotics will be taken as n tends to ∞ and we only claim things to be true for sufficiently large n. By $A \approx B$ we mean that $\lim_{n\to\infty} A/B = 1$.

Definition 1. An *asymptotic degree sequence* is a sequence of integer-valued functions $\mathscr{D} = d_0(n), d_1(n), \ldots$ such that

(a) $d_i(n) = 0$ for $i \geqslant n$;
(b) $\sum_{i\geqslant 0} d_i(n) = n$.

Given an asymptotic degree sequence $\mathscr{D}$, let $\mathscr{D}_n$ be the degree sequence $\{c_1, c_2, \ldots, c_n\}$, where $c_j \geqslant c_{j+1}$ and $|\{j : c_j = i\}| = d_i(n)$ for each $i \geqslant 0$. Define $\Omega_{\mathscr{D}_n}$ to be the set of all graphs with vertex set $[n]$ with degree sequence $\mathscr{D}_n$. A random graph on n vertices with degree sequence $\mathscr{D}$ is a uniformly random member of $\Omega_{\mathscr{D}_n}$.

Definition 2. An asymptotic degree sequence $\mathscr{D}$ is *feasible* if $\Omega_{\mathscr{D}_n} \neq \emptyset$ for all $n \geqslant 1$.

Definition 3. An asymptotic degree sequence $\mathscr{D}$ is *smooth* if there exist constants λ_i such that $\lim_{n\to\infty} d_i(n)/n = \lambda_i$.

Definition 4. An asymptotic degree sequence $\mathscr{D}$ is *sparse* if $\sum_{i\geqslant 0} i d_i(n)/n = K + o(1)$ for some constant K.

Definition 5. Given a smooth asymptotic degree sequence, $\mathscr{D}$, $Q(\mathscr{D}) = \sum_{i\geqslant 1} i(i-2)\lambda_i$.

Definition 6. An asymptotic degree sequence $\mathscr{D}$ is *well-behaved* if the following conditions hold.

(a) $\mathscr{D}$ is feasible and smooth.
(b) $i(i-2)d_i(n)/n$ tends uniformly to $i(i-2)\lambda_i$, *i.e.*, for all $\epsilon > 0$ there exists N such that, for all $n > N$ and for all $i \geqslant 0$,
$$\left|\frac{i(i-2)d_i(n)}{n} - i(i-2)\lambda_i\right| < \epsilon.$$
(c) The limit
$$L(\mathscr{D}) = \lim_{n\to\infty} \sum_{i\geqslant 1} i(i-2)d_i(n)/n$$
exists, and the sum approaches the limit uniformly, that is,
 (i) if $L(\mathscr{D})$ is finite, then for all $\epsilon > 0$ there exists i^*, N such that, for all $n > N$,
$$\left|\sum_{i=1}^{i^*} i(i-2)d_i(n)/n - L(\mathscr{D})\right| < \epsilon,$$
 (ii) if $L(\mathscr{D})$ is infinite, then for all $T > 0$ there exists i^*, N such that, for all $n > N$,
$$\sum_{i=1}^{i^*} i(i-2)d_i(n)/n > T.$$

We note that it is an easy exercise to show that if $\mathscr{D}$ is well-behaved then

$$L(\mathscr{D}) = Q(\mathscr{D}).$$

Note that, for a well-behaved asymptotic degree sequence $\mathscr{D}$, if $Q(\mathscr{D})$ is finite then $\mathscr{D}$ is sparse. Note further that if $\mathscr{D}$ is sparse and well-behaved then, since for $i^* > 1$, $\sum_{i>i^*} id_i(n) < \sum_{i>i^*} i(i-2)d_i(n)$, the sum $\lim_{n\to\infty} \sum_{i\geqslant 1} id_i(n)/n$ approaches its limit uniformly in the sense of condition (c) in Definition 6.

The main result of [12] is the following.

Theorem 3. *Let $\mathscr{D} = d_0(n), d_1(n), \ldots$ be a well-behaved sparse asymptotic degree sequence for which there exists $\epsilon > 0$ such that, for all n and $i > n^{\frac{1}{4}-\epsilon}$, $d_i(n) = 0$. Let G be a graph with n vertices, $d_i(n)$ of which have degree i, chosen uniformly at random from amongst all such graphs.*

(a) *If $Q(\mathscr{D}) > 0$ then there exist constants $\zeta_1, \zeta_2 > 0$ dependent on $\mathscr{D}$ such that G a.s. has a component with at least $\zeta_1 n$ vertices and $\zeta_2 n$ cycles. Furthermore, if $Q(\mathscr{D})$ is finite then G a.s. has exactly one component of size greater than $\gamma \log n$ for some constant γ dependent on $\mathscr{D}$.*

(b) *If $Q(\mathscr{D}) < 0$ and, for some function $0 \leqslant \omega(n) \leqslant n^{\frac{1}{8}-\epsilon}$, $d_i(n) = 0$ for all $i \geqslant \omega(n)$, then, for some constant R dependent on $Q(\mathscr{D})$, G a.s. has no component with at least $R\omega(n)^2 \log n$ vertices, and a.s. has fewer than $2R\omega(n)^2 \log n$ cycles. Also, a.s. no component of G has more than one cycle.*

To be consistent with the model $G_{n,p}$, we call the component referred to in Theorem 3(a) the *giant component*.

To prove Theorem 3, we worked with the configuration model introduced in this form by Bollobás [6] and motivated in part by the work of Bender and Canfield [4]. This model arose in a somewhat different form in the work of Békéssy, Békéssy and Komlós [3] and Wormald [13, 14]. A random configuration with n vertices and a fixed degree sequence is formed by taking a set L containing $\deg(v)$ distinct copies of each vertex v, and choosing a random matching of the elements of L. Each configuration represents an underlying multigraph whose edges are defined by the pairs in the matching. We often abuse notation by referring to a configuration as if it were a multigraph. For example, we say that a configuration has a graphical property P when we mean that its underlying multigraph does, and we discuss the components of a configuration rather than the components of its underlying multigraph.

This very useful lemma follows from the main result in [11], and allows us to prove results concerning a random graph on a particular degree sequence by analysing a random configuration.

Lemma 1. *Suppose $\mathscr{D}$ is a degree sequence meeting the conditions of Theorem 3 for which $Q(\mathscr{D}) < \infty$. If a random configuration with degree sequence $\mathscr{D}$ a.s. has a property P, then a random graph with degree sequence $\mathscr{D}$ a.s. has P.*

The key to the proof of Theorem 3 is the manner in which we exposed the configuration.

Given $\mathscr{D}$, we expose a random configuration F on n vertices with degree sequence $\mathscr{D}$ as follows.

At each step, a vertex, all of whose copies are in exposed pairs, is *entirely exposed*. A vertex, some but not all of whose copies are in exposed pairs, is *partially exposed*. All other vertices are *unexposed*. The copies of partially exposed vertices which are not in exposed pairs are *open*.

1. Form a set L consisting of i distinct copies of each of the $d_i(n)$ vertices that have degree i.
2. Repeat until L is empty.
   (a) Expose a pair of F by first choosing any member of L, and then choosing its partner at random. Remove them from L.
   (b) Repeat until there are no partially exposed vertices:
   Choose an open copy of a partially exposed vertex, and pair it with another randomly chosen member of L. Remove them both from L.

All random choices are made uniformly. Note that we have complete freedom as to which vertex copy we pick in step 2(a), but for the purposes of this paper we will choose it in the same manner in which we choose all other vertex copies, that is, we will simply pick a uniformly random member of L. It is clear that every possible matching amongst the vertex copies occurs with the same probability under this procedure, and hence this is a valid way to choose a random configuration.

Let X_i represent the number of open vertex copies after the ith pair is exposed. Initially the expected increase in X_i is approximately

$$\frac{\sum_{i\geqslant 1} i(i-2)d_i(n)}{\sum_{j\geqslant 1} jd_j(n)} = \frac{Q(\mathscr{D})}{K},$$

explaining the significance of our parameter $Q(\mathscr{D})$.

Suppose that $Q(\mathscr{D})$ is positive, and thus so is the initial expected increase in X_i. If this expected increase remained positive throughout the process then a.s. some component would keep growing in size. Of course, the expected increase does *not* remain positive: it changes as the set of unexposed vertices changes. However, we proved that it takes at least $\Theta(n)$ steps for the expected increase to change significantly, and that this was enough time for a component to become giant.

In this paper, we gain a better understanding of this process by studying the way in which the expected increase of X_i changes throughout the exposure. The key to this will be to keep track of the degrees of the unexposed vertices at each step. Recall that initially there are $d_i(n)$ unexposed vertices of degree i. We will define the random variable $d_{i,j}$ to be the number of unexposed vertices of degree i after j pairs of the configuration have been exposed. Thus $d_{i,0} = d_i(n) \approx \lambda_i n$.

In the next section, we will determine a sequence of functions $Z_0(\alpha), Z_1(\alpha), \ldots$ and prove that a.s. $d_{i,\alpha n} = Z_i(\alpha)n + o(n)$. We will do this by solving a system of differential equations with the property that

$$D_i'(\alpha) \approx \mathbf{Exp}(d_{i,j+1} - d_{i,j}), \quad \text{for} \quad j \approx \alpha n,$$

and then applying a recent theorem of Wormald, which states that, under certain con-

https://doi.org/10.1017/S0963548398003526 Published online by Cambridge University Press

ditions, random variables a.s. behave like the solution to such a system of differential equations. One of these conditions (in fact the only one that doesn't apply here) is that the number of variables is bounded. Fortunately, our differential equations are particularly well-behaved, allowing us to skirt this issue by dealing with the equations individually.

Once we have determined what the degree sequence of the set of unexposed vertices looks like throughout the exposure of the giant component, it will be a simple matter to analyse the size of that component. Furthermore, once that component is completely exposed, we will know the degree sequence of the unexposed vertices. The remainder of the graph will have the structure of a random graph on that degree sequence, and this yields the analogue to the Discrete Duality Principle.

2. A detailed analysis

Recall that a function $f(u_1,\ldots,u_j)$ satisfies a *Lipschitz condition* on $D \subseteq \mathbb{R}^j$ if a constant $L > 0$ exists with the property that

$$|f(u_1,\ldots,u_j) - f(v_1,\ldots,v_j)| \leqslant L \sum_{i=1}^{j} |u_i - v_i|$$

for all $(u_1,\ldots,u_j)$ and $(v_1,\ldots,v_j)$ in D.

The following theorem appears in a more general form in [15]. In it, 'uniformly' refers to the convergence implicit in the $o()$ terms. Hypothesis (i) ensures that Y_t does not change too quickly throughout the process, (ii) tells us what we expect the rate of change to be, and (iii) ensures that this rate does not change too quickly.

Theorem 4. *Suppose $Y_t, 0 \leqslant t \leqslant m = m(n)$ is a sequence of real-valued random variables, such that $0 \leqslant Y_t \leqslant Cn$ for some constant C, and H_j is the history of the sequence, that is, the array $(Y_0,\ldots,Y_j)$. Suppose, further, that for some function $f : \mathbb{R}^2 \to \mathbb{R}$ the following statements hold.*

(i) *There is a constant C' such that, for all $t < m$ and all l,*

$$|Y_{t+1} - Y_t| < C'$$

always.

(ii) *Uniformly over all $t < m$,*

$$\mathbf{Exp}(Y_{t+1} - Y_t | H_t) = f(t/n, Y_t/n) + o(1)$$

always.

(iii) *The function f is continuous and satisfies a Lipschitz condition on D, where D is some bounded connected open set containing the intersection of $\{(t,z) : t \geqslant 0\}$ with some neighbourhood of $\{(0,z) : \mathbf{Pr}(Y_0 = zn) \neq 0$ for some $n\}$.*

Then,

(a) *for $(0,\hat{z}) \in D$ the differential equation*

$$\frac{dz}{ds} = f(s,z)$$

https://doi.org/10.1017/S0963548398003526 Published online by Cambridge University Press

has a unique solution in D for $z : \mathbb{R} \to \mathbb{R}$ passing through

$$z(0) = \hat{z},$$

and which extends to points arbitrarily close to the boundary of D;

(b) *and*

$$Y_t = nz(t/n) + o(n)$$

with probability at least $1 - n^{-1/2}$ uniformly for $0 \leqslant t \leqslant \min\{\sigma n, m\}$ and for each l, where $z(t)$ is the solution in (a) *with $\hat{z} = Y_0/n$, and $\sigma = \sigma(n)$ is the supremum of those s to which the solution can be extended.*

Remark. The only part of Theorem 4 that does not follow directly from the statement of Theorem 1 in [15] is the bound on the probability in (b). This is implicit in its proof.

Now, suppose that we are given a well-behaved degree sequence $\mathscr{D}$, such that $Q(\mathscr{D}) > 0$. We expose a random configuration, F, with n vertices and degree sequence $\mathscr{D}$ using our branching process.

It is important to note that with high probability it will not take very many steps before we begin to expose the giant component, as demonstrated by the following lemma.

Lemma 2. *For any function $\omega(n) \to \infty$, $\omega(n) = \mathrm{o}(n/\log n)$, a.s. the largest component of F will be one of the first $\omega(n)$ components exposed.*

Proof. Let E_1 be the event that F has a cyclic component of size at least $\zeta_1 n$, and no other component of size greater than $\gamma \log n$, where ζ_1, γ are as in Theorem 3. By Theorem 3, E_1 a.s. occurs.

For any configuration with degree sequence $\mathscr{D}$, we say that $\mathscr{C}$ is the subset of the components defined as follows. We consider the components to be sorted first in non-increasing order of the sizes of their edge sets, and then by decreasing order of their highest labelled vertex. We take $\mathscr{C}$ to be the smallest initial sequence of components that contains a total of at least $\zeta_1 n$ edges. Note that, if E_1 occurs, then $\mathscr{C}$ contains only the largest component.

Let E_2 be the event that one of the first $\omega(n)$ components exposed lies in $\mathscr{C}$. Now, each time we start a new component, either we have already exposed a member of $\mathscr{C}$, *or* the probability that a uniformly selected copy of an unexposed vertex lies in $\mathscr{C}$ is at least $2\zeta_1/K$. Therefore,

$$\mathbf{Pr}(E_2) \geqslant 1 - \left(1 - \frac{2\zeta_1}{K}\right)^{\omega(n)} = 1 - \mathrm{o}(1).$$

Clearly, the probability that the largest component is one of the first $\omega(n)$ components exposed is at least the probability that E_1 and E_2 hold, thus proving the lemma. □

Corollary 1. *Almost surely, the $\lfloor \log^2 n \rfloor$th edge exposed will lie in the largest component of F.*

Proof. This follows immediately from Lemma 2 and Theorem 3(a). □

And now we can prove our main theorems.

https://doi.org/10.1017/S0963548398003526 Published online by Cambridge University Press

Proof of Theorem 1. We prove this by analysing the asymptotic value of $d_{i,j}$. Clearly $d_{i,0} = d_i(n)$ for each i. Consider any fixed $i_0 \geqslant 1$, and set $M = \sum_{i\geqslant 0} id_i(n)$. When exposing the $(j+1)$st edge, we have exactly $M - 2j - 1$ vertex copies to choose from, $i_0 d_{i_0,j}$ of which are copies of unexposed vertices of degree i_0. Therefore, if $X_j > 0$ then the expected change in $d_{i_0,j}$ is

$$\mathbf{Exp}(d_{i_0,j+1} - d_{i_0,j}) = -\frac{i_0 d_{i_0,j}}{M - 2j - 1},$$

and the distribution of this change is mutually independent of the values of $d_{i,j}$ for all $i \neq i_0$.

Thus, if it were not for the complications that arise when $X_j = 0$, it would be straightforward to apply Theorem 4 to $d_{i,j}$. To deal with these complications, we add two twists to our analysis. The first is that we begin our analysis at step $j = \lfloor \log^2 n \rfloor$. Clearly, $d_{i,\lfloor \log^2 n \rfloor} = d_{i,0} + \mathrm{o}(n)$ for each i. Furthermore, by Corollary 1, after this step, X_j will almost surely remain positive until after the giant component has been entirely exposed. However, we must still deal with the slim chance that X 'plummets' to 0 prematurely. To do this, we introduce twin random variables $\delta_{i,j}$, defined as follows.

For $j = 0$, and for each j such that $X_J > 0$ for all $\lfloor \log^2 n \rfloor \leqslant J \leqslant j + \lfloor \log^2 n \rfloor$, $\delta_{i,j} = d_{i,j+\lfloor \log^2 n \rfloor}$. For any other j, we define

$$\delta_{i,j} = \begin{cases} \delta_{i,j-1} - 1, & \text{with probability } \dfrac{i\delta_{i,j-1}}{M - 2(j-1) - 1}, \\ \delta_{i,j-1}, & \text{otherwise.} \end{cases}$$

Now, for any fixed $i_0 \geqslant 1$, by applying Theorem 4 with $Y_j = \delta_{i_0,j}$, $C' = 1, m = n$ and $f(s,z) = -iz/(K - 2s)$, we see that with probability at least $1 - \mathrm{o}(n^{-1/2})$, for every $0 < \alpha < 1$,

$$\delta_{i_0,\lceil \alpha n \rceil} = Z_{i_0}(\alpha) n + \mathrm{o}(n), \tag{2.1}$$

where

$$Z_i(\alpha) = d_{i,0}\left(1 - \frac{2\alpha}{K}\right)^{\frac{i}{2}}$$

is the unique solution to

$$\begin{aligned} Z_i(0) &= d_{i,0}/n, \\ Z_i'(\alpha) &= -\frac{iZ_i(\alpha)}{K - 2\alpha}. \end{aligned}$$

Since our degree sequence has maximum degree $\mathrm{o}(n^{1/4})$, a.s. (2.1) holds for every i.

Note that $X_j = M - 2j - \sum_{i\geqslant 1} id_{i,j}$. Thus, by applying Corollary 1 and using the fact that $\mathscr{D}$ is well-behaved, we have that, for any $0 \leqslant \alpha \leqslant \alpha_{\mathscr{D}}$ and any $I > 0$, a.s.

$$\begin{aligned} X_{\lceil \alpha n \rceil} &= M - 2\lceil \alpha n \rceil - \sum_{i=1}^{I} id_{i,j} - \sum_{i>I} id_{i,j} \\ &= Kn - 2\lceil \alpha n \rceil - \sum_{i=1}^{I} id_i(n)\left(1 - \frac{2\alpha}{K}\right)^{\frac{i}{2}} + S \\ &= \left(K - 2\alpha - \sum_{i=1}^{I} i\lambda_i \left(1 - \frac{2\alpha}{K}\right)^{\frac{i}{2}}\right) n + S, \end{aligned}$$

for each $0 \leqslant \alpha \leqslant \alpha_{\mathscr{D}}$, where $|S| < \gamma n$ for some $\gamma = \gamma(I)$ where $\lim_{I\to\infty} \gamma(I) = 0$, and so a.s.

$$X_{\lceil \alpha n \rceil} = \chi(\alpha)n + o(n), \tag{2.2}$$

and Theorem 1 now follows. □

Proof of Theorem 2. By Lemma 2, for any $\omega(n) \to \infty$ we a.s. expose less than $\omega(n)$ components prior to the exposure of the giant component. In fact, with probability $\epsilon_{\mathscr{D}}$, the giant component is the first component exposed.

Upon completion of the exposure of the giant component, the configuration induced by the unexposed vertices is a uniformly random configuration with $d_{i,j}$ vertices of degree i for each i, where j is the number of exposed pairs. By Theorem 1, this configuration a.s. has $n' = (1-\epsilon_{\mathscr{D}})n$ vertices, $\lambda'_i n' + o(n')$ of which have degree i. □

Recall that G a.s. has exactly one component of size greater than $\gamma \log n$, so it should not be surprising that $\sum_{i\geqslant 1} i(i-2)\lambda'_i < 0$, as we will now see.

For $0 \leqslant \alpha \leqslant \alpha_{\mathscr{D}}$,

$$\begin{aligned}
\chi'(\alpha) &= -2 + (K-2\alpha)^{-1}\sum_{i\geqslant 1} i^2\lambda_i\left(1-\frac{2\alpha}{K}\right)^{\frac{i}{2}} \\
&= (K-2\alpha)^{-1}\left(\sum_{i\geqslant 1} i^2\lambda_i\left(1-\frac{2\alpha}{k}\right)^{\frac{i}{2}} - 2(K-2\alpha)\right) \\
&\geqslant (K-2\alpha)^{-1}\left(\sum_{i\geqslant 1} i^2\lambda_i\left(1-\frac{2\alpha_{\mathscr{D}}}{k}\right)^{\frac{i}{2}} - 2\sum_{i\geqslant 1} i\lambda_i\left(1-\frac{2\alpha_{\mathscr{D}}}{k}\right)^{\frac{i}{2}}\right) \\
&= (K-2\alpha_{\mathscr{D}})^{-1}(1-\epsilon_{\mathscr{D}})\sum_{i\geqslant 1} i(i-2)\lambda'_i \\
&= (K-2\alpha_{\mathscr{D}})^{-1}(1-\epsilon_{\mathscr{D}})Q(\mathscr{D}').
\end{aligned}$$

Furthermore, the inequality is strict for $\alpha < \alpha_{\mathscr{D}}$, and so $Q(\mathscr{D}') < 0$, as otherwise $\alpha_{\mathscr{D}}$ could not be the smallest positive zero of $\chi(\alpha)$.

3. The model $G_{n,p}$

We close by noting that some previously known results about $G_{n,p=c/n}$ are special cases of Theorems 1 and 2.

Select $G_{n,p}$ by first exposing its degree sequence, and then choosing a random graph with that degree sequence. Note that every graph with that degree sequence occurs as $G_{n,p}$ with the same probability and so this is a valid method of selection.

It is well known that $G_{n,p=c/n}$ a.s. has $c^i/(i!)\,e^{-c}n + o(n)$ vertices of degree i, for each $i \leqslant O(\log n/\log\log n)$, and no vertices of higher degree. It is straightforward to verify that, if this property holds, then $K = c$, and so in order to apply Theorem 1 we wish to

https://doi.org/10.1017/S0963548398003526 Published online by Cambridge University Press

solve

$$c - 2\alpha - \sum_{i \geqslant 1} i \frac{c^i}{i!} e^{-c} \left(1 - \frac{2\alpha}{c}\right)^{\frac{i}{2}} = 0. \tag{3.1}$$

There are two solutions at $\alpha = 0, c/2$. We will see that there is another and so $0 < \alpha_{\mathscr{D}} < c/2$.

For $\alpha \neq 0, c/2$, (3.1) is congruent to

$$\sqrt{\frac{c - 2\alpha}{c}} = \exp\left(\sqrt{c^2 - 2c\alpha} - c\right).$$

Now set

$$\begin{aligned} \epsilon(\alpha) &= 1 - \sum_{i \geqslant 1} \frac{c^i}{i!} e^{-c} \left(1 - \frac{2\alpha}{c}\right)^{\frac{i}{2}} \\ &= 1 - \exp\left(\sqrt{c^2 - 2c\alpha} - c\right). \end{aligned}$$

By Theorem 1, a.s. the size of the giant component of $G_{n,p=c/n}$ is $\epsilon n + o(n)$, with $\epsilon = \epsilon(\alpha_{\mathscr{D}})$, where $\alpha_{\mathscr{D}}$ is the smallest positive solution to (3.1). Now, for $0 < \alpha < c/2$, $\chi(\alpha) = 0$ if and only if

$$\begin{aligned} \epsilon(\alpha) + e^{-c\epsilon(\alpha)} &= 1 - \exp\left(\sqrt{c^2 - 2c\alpha} - c\right) \\ &\quad + \exp\left(-c\left(1 - \exp\left(\sqrt{c^2 - 2c\alpha} - c\right)\right)\right) \\ &= 1 - \exp\left(\sqrt{c^2 - 2c\alpha} - c\right) \\ &\quad + \exp\left(-c\left(1 - \sqrt{\frac{c - 2\alpha}{c}}\right)\right) \\ &= 1 - \exp\left(\sqrt{c^2 - 2c\alpha} - c\right) + \exp\left(\sqrt{c^2 - 2c\alpha} - c\right) \\ &= 1, \end{aligned}$$

thus verifying that a.s. the size of the largest component of $G_{n,p=c/n}$ is $\epsilon_c n + o(n)$, where ϵ_c is the unique solution to $\epsilon + e^{-c\epsilon} = 1$.

We will now see that the Discrete Duality Principle is a special case of Theorem 1, by showing that, if

$$\lambda_i = \frac{c^i}{i!} e^{-c},$$

then

$$\lambda_i' = \frac{d_c^i}{i!} e^{-d_c}, \tag{3.2}$$

where $d_c = c(1 - \epsilon_c)$. It can easily be shown (see, for example, [1]) that $ce^{-c} = d_c e^{-d_c}$, and thus

$$\frac{e^{-c}}{1 - \epsilon_c} = e^{-d_c}.$$

Since $\epsilon_c = 1 - \exp\left(\sqrt{c^2 - 2c\alpha^*} - c\right)$, $d_c = c \exp\left(\sqrt{c^2 - 2c\alpha^*} - c\right)$. Therefore, $c \exp(-c) =$

https://doi.org/10.1017/S0963548398003526 Published online by Cambridge University Press

$d_c \exp\left(-\sqrt{c^2 - 2c\alpha^*}\right)$, and so

$$d_c = \sqrt{c^2 - 2c\alpha_{\mathscr{D}}}.$$

Therefore,

$$\begin{aligned}\lambda_i' &= \frac{\lambda_i}{1-\epsilon_c}\left(1 - \frac{2\alpha_{\mathscr{D}}}{K}\right)\\ &= \frac{c^i \mathrm{e}^{-c}}{i!(1-\epsilon_c)}\left(1 - \frac{2\alpha_{\mathscr{D}}}{c}\right)\\ &= \left(\sqrt{c^2 - 2\alpha_{\mathscr{D}} c}\right)^i \frac{\mathrm{e}^{-c}}{i!(1-\epsilon_c)}\\ &= \frac{d_c^i}{i!}\mathrm{e}^{-d_c},\end{aligned}$$

as claimed, thus verifying the Discrete Duality Principle.

Acknowledgements

The authors would like to thank Nick Wormald for an early manuscript of [15], Noel Walkington for his valuable advice concerning differential equations, Boris Pittel for his helpful comments, and a referee for some minor corrections.

References

[1] Alon, N. and Spencer, J. (1992) *The Probabilistic Method*, Wiley.
[2] Azuma, K. (1967) Weighted sums of certain dependent random variables. *Tokuku Math. Journal* **19** 357–367.
[3] Békéssy, A., Békéssy, P. and Komlós, J. (1972) Asymptotic enumeration of regular matrices. *Stud. Sci. Math. Hungar.* **7** 343–353.
[4] Bender, E. A. and Canfield, E. R. (1978) The asymptotic number of labelled graphs with given degree sequences. *J. Combin. Theory Ser. A* **24** 296–307.
[5] Billingsley, P. (1986) *Probability and Measure Theory*, Wiley.
[6] Bollobás, B. (1980) A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Europ. J. Combinatorics* **1** 311–316
[7] Bollobás, B. (1984) The evolution of random graphs. *Trans. Amer. Math. Soc.* **286** 257–274.
[8] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17–61.
[9] Jansen, S., Knuth, D., Łuczak, T. and Pittel, B. (1993) The birth of the giant component. *Random Structures and Algorithms* **4** 233–358.
[10] Karp, R. M. (1990) The transitive closure of a random digraph. *Random Structures and Algorithms* **1** 73–93.
[11] McKay, B. D. (1985) Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combinatorica* **19A** 15–25.
[12] Molloy, M. and Reed, B. (1995) A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6** 161–180.
[13] Wormald, N. C. (1981) The asymptotic connectivity of labelled regular graphs. *J. Combin. Theory Ser. B* **31** 156–167.
[14] Wormald, N. C. (1981) The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** 156–167.
[15] Wormald, N. C. (1995) Differential equations for random processes and random graphs. *Ann. Appl. Probab.* **5** 1217–1235.