MAXIMIZING THE SIZE OF THE GIANT

TOM BRITTON * ** AND

PIETER TRAPMAN,* *** *Stockholm University*

Abstract

Consider a random graph where the mean degree is given and fixed. In this paper we derive the maximal size of the largest connected component in the graph. We also study the related question of the largest possible outbreak size of an epidemic occurring 'on' the random graph (the graph describing the social structure in the community). More precisely, we look at two different classes of random graphs. First, the Poissonian random graph in which each node i is given an independent and identically distributed (i.i.d.) random weight X_i with $\mathrm{E}(X_i) = \mu$, and where there is an edge between i and j with probability $1 - \mathrm{e}^{-X_i X_j/(\mu n)}$, independently of other edges. The second model is the thinned configuration model in which the n vertices of the ground graph have i.i.d. ground degrees, distributed as D, with $\mathrm{E}(D) = \mu$. The graph of interest is obtained by deleting edges independently with probability $1 - p$. In both models the fraction of vertices in the largest connected component converges in probability to a constant $1 - q$, where q depends on X or D and p. We investigate for which distributions X and D with given μ and p, $1 - q$ is maximized. We show that in the class of Poissonian random graphs, X should have all its mass at 0 and one other real, which can be explicitly determined. For the thinned configuration model, D should have all its mass at 0 and two subsequent positive integers.

1. Introduction

In this paper we study asymptotic properties of some random graphs as n, the number of nodes/vertices, tends to ∞. More specifically, we study the size of the largest connected component, within two classes of random graphs. If this largest connected component is of the same order as the number of nodes, then it is called the giant. We find the random graph that has the largest giant among all random graphs in the class having a pre-defined mean degree.

We consider two types of network in this paper. Both types of random graph are frequently used extensions of the classical Erdös–Rényi graph [3], [6].

Poissonian random graphs. Poissonian random graphs were introduced in [10] and are a main example of inhomogeneous random graphs [4]. Our model is slightly different from the model introduced in [10], but, asymptotically (for $n \to \infty$), the fraction of vertices in the largest connected component of the graph will be the same.

  We construct the random graph of n vertices as follows. First, we assign independent and identically distributed (i.i.d.) weights to the vertices distributed as the nonnegative

random variable X, with finite mean $\mu_X := E(X)$. A pair of vertices with weights x and y share an edge with probability $1 - \exp[-xy/(n\mu_X)]$, independent of other edges in the graph. In [10] this probability is $1 - \exp[-xy/L_n]$, where L_n is the sum of the n vertex weights.

We note that creating the graph and after that removing edges independently with probability $1 - p$ is asymptotically in distribution the same as immediately creating a Poissonian random graph with weight distribution pX. Later we will be interested in properties of thinned versions of Poissonian random graphs, but this observation implies that these fall under the same model and, hence, need no additional analysis.

Thinned configuration model. The configuration model [6], [8] is obtained by assigning i.i.d. numbers (distributed as the nonnegative integer-valued random variable D) of half-edges to the n vertices in the graph. We assume that $\mu_D = E(D) < \infty$. If the total number of half-edges is odd then we add one half-edge to the final vertex. Then, we pair the vertices uniformly at random. The probability that a specific graph is created is the same for all graphs with a given degree sequence. Parallel edges (edges with the same end vertices) and self-loops (an edge which connects a vertex to itself) might occur, but they will not influence the asymptotic fraction of vertices in the largest connected component (when $\mu_D < \infty$ as we have assumed) [6]. We clean the graph by removing all self-loops and merging all parallel edges. It is easy to check that this will not influence the asymptotic degree distribution for $\mu_D < \infty$. After this, edges are deleted independently with probability $1 - p$.

Within both of the two random graph models we identify the distribution, X and D, respectively, which maximizes the size of the giant among all distributions having some fixed mean μ_X in the first model and fixed mean μ_D and parameter p in the second model. The problem of identifying which (random) graph has the maximal giant among *all* graphs with a fixed mean degree μ_D is less interesting in that the giant can make up the whole population whenever $\mu_D \geq 2$, while, for $\mu_D < 2$, the maximum is obtained in any graph, which is a tree. Similarly, the problem of *minimizing* the size of the giant among Poissonian random graphs and thinned configuration models is achieved by choosing $P(X = 0)$ (respectively $P(D = 0)$) arbitrary close to 1 (and the remaining small probability at a very large value giving the correct mean), which has the effect that the relative size of the giant goes to 0.

Random graphs are interesting in their own right but also have numerous applications. One such application, which has been the inspiration to many scientists, is that of modeling the spread of an infectious disease in a socially organized human community, where the social structure of the community is described by a random graph [1], [2], [5], [9]. The degree D then signifies how many other individuals a given individual is socially connected to, a number which could be 0 or some small or large positive integer. The interpretation of thinning the random network when considering the spread of SIR (susceptible $\rightarrow$ infectious $\rightarrow$ recovered) epidemics with nonrandom infectious periods on networks is that transmission will take place (if one of the nodes gets infected) exactly along those edges that are kept, and transmission between one infected node and a susceptible neighbor is modeled to have probability p and be independent of other transmission links. In epidemic terminology our aim is hence to identify the degree distribution with given mean for which the asymptotic fraction of individuals infected, in the case of a major epidemic outbreak, is maximal. We are hence searching for the worst-case scenario if all that is known about the underlying social network is the mean degree μ.

2. Notation and some basic results

Throughout, we will use $\mathbb{N}$ for the strictly positive integers and $\mathbb{N}_0 = \mathbb{N} \cup 0$ for the nonnegative integers. Unless specified otherwise, we will use D for a random variable taking values in $\mathbb{N}_0$ and X for a nonnegative real-valued random variable. The mean of X is denoted by μ_X. A bar above a random variable denotes the size-biased variant of the random variable, i.e.

$$P(\bar{D} = k) = \frac{k\,P(D = k)}{\mu_D},$$

or in the case of a general positive random variable

$$P(\bar{X} \le x) = \frac{\int_0^x y\,dP(X \le y)}{\mu_X} = \frac{E(X\,\mathbf{1}[X \le x])}{\mu_X}.$$

The (probability) generating function of an $\mathbb{N}_0$-valued random variable is defined by $f_D(s) := E(s^D) = \sum_{k=0}^{\infty} s^k\,P(D = k)$ for $s \in [0, 1]$. We sometimes use the notation $\bar{f}_D(s) = f_{\bar{D}-1}(s)$. The smallest root of $s = \bar{f}_D(s)$ is denoted by z_D, and q_D is defined by $q_D = f_D(z_D)$. Some well-known facts about generating functions that we will use (and which are easy to check) are the following.

1. $f_D(s)$ is analytic on $(0, 1)$ and all derivatives of $f_D(s)$ are nonnegative.

2. $f_D(1) = 1$.

3. $df_D(s)/ds = \mu_D \bar{f}_D(s)$; in particular, $df_D(s)/ds|_{s=1} = \mu_D$. Or, equivalently,

$$f_D(s) = 1 - \mu_D \int_s^1 \bar{f}_D(x)\,dx \tag{1}$$

   for $s \in [0, 1]$.

4. z_D is the extinction probability of a Galton–Watson branching process [7] with offspring distribution $\bar{D} - 1$ and one ancestor; q_D is the extinction probability of a branching process for which the number of children of the ancestor is distributed as D and all other individuals have offspring distribution $\bar{D} - 1$.

5. Assuming that $P(D = 0) + P(D = 2) < 1$, q_D and z_D are strictly less than 1 if and only if $E(\bar{D} - 1) > 1$.

For a nonnegative real-valued random variable X, the distribution of a mixed Poisson(X) random variable D is given by $P(D = k) = E(X^k e^{-X}/k!)$. The generating function of this random variable D is given by $f_D(s) = E(e^{-(1-s)X})$. Furthermore, $\bar{f}_D(s) = E(e^{-(1-s)\bar{X}})$. So the generating function of $\bar{D} - 1$ is given by the generating function of a mixed Poisson distribution based on the size-biased variant of X. We note that $\mu_D = \mu_X$.

In this paper we consider undirected simple graphs. A simple graph is a graph with no parallel edges (two or more edges with the same end vertices) or self-loops (edges connecting a vertex to itself). The degree of a vertex is the number of edges a vertex is adjacent to.

A connected component in a graph is a set of vertices for which there is a path between every pair of vertices in this set. Let $\mathcal{C}^i(n)$ be the ith largest connected component (in the case of a tie, the order of the tied components is uniform at random). The number of vertices in a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$.

We consider two types of random graph (as defined in the introduction).

Poissonian random graphs. For this model, there is, for $n \to \infty$, with high probability, at most one giant component [4], i.e. for every $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P(|\mathcal{C}^2(n)| < \varepsilon n) = 1.$$

Let D be mixed Poisson(X). The fraction of vertices in the giant component is, for large n, with high probability, close to $1 - q_D$. More precisely, for every $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P(|n^{-1}|\mathcal{C}^1(n)| - (1 - q_D)| < \varepsilon) = 1.$$

Thinned configuration model. Let D be the degree distribution of the ground graph, and let p be the thinning parameter. For technical reasons (see the remark below), we exclude the (trivial) model in which both $p = 1$ and $P(D = 0) + P(D = 2) = 1$. For the thinned configuration model, it also holds that the probability that there is more than one giant component converges to 0 as $n \to \infty$. In this class of random graphs, the generating function of the degree distribution of the thinned graph is given by $g_{D,p}(s) = f_D(1 - p + ps)$ and $\bar{g}_{D,p}(s) = \bar{f}_D(1 - p + ps)$. The mean degree of a randomly chosen vertex is $p\mu_D$.

Let $z_{D,p}$ be the smallest root of $s = \bar{g}_{D,p}(s)$, and define $q_{D,p} = g_{D,p}(z_{D,p})$. The fraction of vertices in the giant component is, for large n, with high probability, close to $1 - q_{D,p}$. More precisely, for every $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P(|n^{-1}|\mathcal{C}^1(n)| - (1 - q_{D,p})| < \varepsilon) = 1.$$

Remark. If $p = 1$ and $P(D = 0) + P(D = 2) = 1$, then there exists $\varepsilon > 0$ such that $\lim_{n\to\infty} P(|\mathcal{C}^2(n)| < \varepsilon n) < 1$. In the thinned configuration model the fraction of the vertices in a component of size at least k converges to $P(D = 2)$ for every $k \in \mathbb{N}$. However, the fraction of vertices in the largest component does not converge to $q_{D,p} = P(D = 2)$.

We will show that, for the Poissonian random graphs with given μ_X, the limiting size of the giant component is maximized if all vertices have weight μ_X if μ_X is larger than $\mu_c \approx 1.756$. If $\mu_X < \mu_c$ then X should have mass only on μ_c and 0. Again, we note that thinning with a factor p is equivalent to replacing X by pX.

For the thinned configuration model with given μ_D and p, the maximal giant size is obtained if D has all mass on 0 and two subsequent positive integers k and $k + 1$. We were not able to identify a closed formula for k, and the exact mass distribution on the three possible atoms.

3. Poissonian random graph

Define μ_c as the largest root of $2x = e^{x-1/2}$. The numerical value of μ_c is approximated by $\mu_c \approx 1.756$.

Let $|\mathcal{C}_X^1(n)|$ be the size of the giant in the Poissonian random graph with weight distribution X and n vertices. Again, D is mixed Poisson(X). Furthermore, let q_D and z_D be as before. Let $\mathcal{D}_\mu$ be the collection of mixed Poisson random variables with $E(D) = \mu$.

Theorem 1. *Let D^* be mixed* Poisson(X^*) *and*

$$P(X^* = \max(\mu, \mu_c)) = 1 - P(X^* = 0) = \min\left(1, \frac{\mu}{\mu_c}\right).$$

Then

$$\min_{D \in \mathcal{D}_\mu} q_D = q_{D^*}.$$

This theorem may be interpreted as follows. If $\mu < \mu_c$ then $D \in \mathcal{D}_\mu$ defined via

$$P(X = \mu_c) = 1 - P(X = 0) = \frac{\mu}{\mu_c}$$

leads to the Poissonian graph, for which the fraction of vertices in the giant converges in probability (as $n \to \infty$) to the largest limit. If $\mu \geq \mu_c$ then $P(X = \mu) = 1$ (i.e. the Erdös–Rényi graph [3] with mean degree μ) leads to the largest giant in this class.

We first show by a series of three lemmas that the maximal giant is obtained for a weight distribution with mass only at 0 and one other real number. After that we show that in this class, X^* leads to the largest giant component, which will complete the proof.

Lemma 1. *If A and B are positive real-valued random variables, then*

$$E(A^k B) \geq E(B) \left(\frac{E(AB)}{E(B)} \right)^k.$$

Proof. Use Hölders inequality, $E(XY) \leq E((X^a))^{1/a} E((Y^b))^{1/b}$ for nonnegative random variables X and Y and $a, b > 0$ such that $a^{-1} + b^{-1} = 1$. Letting $a = k$, $X = AB^{1/k}$, and $Y = B^{(k-1)/k}$ gives the desired result.

Lemma 2. *Let X be a general nonnegative random variable, and let D be mixed* Poisson(X). *If, for some $s_* \in (0, 1)$,*

$$f_D(s_*) := E(e^{-X(1-s_*)}) = e^{-\lambda(1-s_*)},$$

then $E(e^{-X(1-s)}) \leq e^{-\lambda(1-s)}$ for $s \in [s_*, 1]$.

Proof. If $f_D(s)$ crosses $e^{-\lambda(1-s)}$ from below in s_* then we know that

$$\frac{d}{ds} f_D(s) \bigg|_{s=s_*} \geq \lambda e^{-\lambda(1-s_*)}. \tag{2}$$

Furthermore, for $k \in \mathbb{N}$, we have

$$\frac{d^k}{ds^k} f_D(s) \bigg|_{s=s_*} = E(X^k e^{-X(1-s_*)}) \geq \lambda^k e^{-\lambda(1-s_*)} = \frac{d^k}{ds^k} e^{-\lambda(1-s)} \bigg|_{s=s_*}.$$

Here the inequality follows by (2) and Lemma 1 with $A = X$ and $B = e^{-X(1-s_*)}$. Since $f_D(1) = 1$ and $f_D(s)$ is analytic on $(0, 1)$, it follows from a Taylor expansion in $s = s_*$ that $f_D(s)$ cannot cross $e^{-\lambda(1-s)}$ from below in $s \in (0, 1)$.

Let $D := D(\lambda)$ be mixed Poisson(X), where the random variable $X = X(\lambda)$ is defined by $P(X = \lambda) = 1 - P(X = 0) = \mu/\lambda$. Let

$$f(s; \lambda) := f_D(s) = 1 - \frac{\mu}{\lambda} + \frac{\mu}{\lambda} e^{-\lambda(1-s)}.$$

Note that $\bar{f}_D(s) = e^{-\lambda(1-s)}$ is the generating function of a Poisson λ distribution. Let $q(\lambda) := q_{D(\lambda)}$ and $z(\lambda) := z_{D(\lambda)}$.

Lemma 3. *Let X be a general nonnegative random variable with expectation $\mu := \mu_X$, and let D be mixed* Poisson(X). *Let $q(\lambda)$ be defined as above.*

(a) *If $z_D > z(\mu) = q(\mu)$ then $q_D > q(\mu)$.*

(b) *If $z_D \leq z(\mu) = q(\mu)$ then $q_D \geq q(\lambda)$, where $\lambda = -\log[z_D]/(1 - z_D)$.*

Proof. Part (a) follows from $q_D = \mathrm{E}(\mathrm{e}^{-X(1-z_D)})$, Jensen's inequality, and the fact that $f_D(s)$ is increasing. For (b), we use the fact that $\bar{f}_D(s)$ equals $\mathrm{e}^{-\lambda(1-s)} = \bar{f}(s, \lambda)$ in $s = z_D$ and $s = 1$. In particular, $z(\lambda) = z_D$. By (1) and Lemma 2, we obtain

$$q_D = f_D(z_D) = 1 - \mu \int_{z_D}^{1} \bar{f}_D(x)\,\mathrm{d}x \geq 1 - \mu \int_{z_D}^{1} \mathrm{e}^{-\lambda(1-x)}\,\mathrm{d}x = q(\lambda).$$

This completes the proof. □

We now show that among the distributions $(D(\lambda); \lambda > 0)$, the fraction of vertices in the giant component will converge in probability to the largest limit for D^*.

Lemma 4. *Let $\mathcal{D}'_\mu$ be the class of mixed Poisson random variables, where D is mixed* Poisson(X) *and* $\mathrm{P}(X = \lambda) = 1 - \mathrm{P}(X = 0) = \mu/\lambda$, *with $\lambda \geq \mu$. For $D \in \mathcal{D}'_\mu$, q_D is minimized for $\lambda = \max(\mu, \mu_c)$.*

Proof. We first note that we might assume that $\lambda > 1$; otherwise, $q_D = 1$ anyway. In what follows we require that $z(\lambda)$ is differentiable on $(1, \infty)$. We prove this by analyzing the derivative of its inverse $z^{-1}(x)$, and show that it is nonzero on this domain. From the definition of $z(\lambda)$, we deduce that $z^{-1}(x) = -\log[x]/(1 - x)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}x} z^{-1}(x) = -\frac{1 - x + x\log[x]}{x(1 - x)^2}.$$

For $x \in (0, 1)$, this derivative is strictly negative and finite, by

$$\frac{\mathrm{d}}{\mathrm{d}x}[-(1 - x + x\log[x])] = -\log[x] > 0$$

and $[1 - x + x\log[x]]_{x=1} = 0$. This implies that $z(\lambda)$ is differentiable on the domain where it takes values in $(0, 1)$, that is, on $(1, \infty)$.

Since $\bar{f}(z(\lambda); \lambda) - z(\lambda) = 0$, we obtain, by applying the chain rule,

$$0 = \frac{\mathrm{d}}{\mathrm{d}\lambda}[\bar{f}(z(\lambda); \lambda) - z(\lambda)] = \left[\frac{\mathrm{d}}{\mathrm{d}\lambda}\bar{f}(s; \lambda) + \frac{\mathrm{d}}{\mathrm{d}s}\bar{f}(s; \lambda)\frac{\mathrm{d}}{\mathrm{d}\lambda}z(\lambda) - \frac{\mathrm{d}}{\mathrm{d}\lambda}z(\lambda)\right]_{s=z(\lambda)}.$$

This gives

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}z(\lambda) = \left[\frac{\mathrm{d}\bar{f}(s; \lambda)/\mathrm{d}\lambda}{1 - \mathrm{d}\bar{f}(s; \lambda)/\mathrm{d}s}\right]_{s=z(\lambda)}.$$

Furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}q(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda}f(z(\lambda), \lambda) = \left[\frac{\mathrm{d}}{\mathrm{d}\lambda}f(s; \lambda) + \frac{\mathrm{d}}{\mathrm{d}s}f(s; \lambda)\frac{\mathrm{d}}{\mathrm{d}\lambda}z(\lambda)\right]_{s=z(\lambda)}.$$

Noting that

$$\frac{\mathrm{d}}{\mathrm{d}s}f(s; \lambda)\bigg|_{s=z(\lambda)} = \mu\bar{f}(z(\lambda); \lambda) = \mu z(\lambda),$$

we obtain

$$\frac{d}{d\lambda}q(\lambda) = \left[\frac{\mu z(\lambda)d\bar{f}(s;\lambda)/d\lambda}{1 - d\bar{f}(s;\lambda)/ds} + \frac{d}{d\lambda}f(s;\lambda)\right]_{s=z(\lambda)}.$$

Equating this derivative to 0 and using $f(s;\lambda) = 1 - \mu/\lambda + (\mu/\lambda)e^{-\lambda(1-s)}$ and $\bar{f}(s;\lambda) = e^{-\lambda(1-s)}$ gives

$$0 = \frac{-\mu[z(\lambda)]^2[1 - z(\lambda)]}{1 - \lambda z(\lambda)} + \frac{\mu}{\lambda^2}[1 - z(\lambda)] - \frac{\mu}{\lambda}[1 - z(\lambda)]z(\lambda).$$

The solutions of this equation are $z(\lambda) = 1$ and $z(\lambda) = (2\lambda)^{-1}$. The first solution is of no use because if $z(\lambda) = 1$ then $q(\lambda) = 1$ as well. Substituting the second solution into $z(\lambda) = e^{-\lambda(1-z(\lambda))}$ gives $2\lambda = e^{\lambda-1/2}$. Because the root of this equality is strictly larger than 1 and because $\lim_{\lambda\to\infty} q(\lambda) = 1$, $q(\lambda)$ takes its minimum on $(1,\infty)$ in this largest root. The lemma follows by observing that $\lambda \geq \mu$. $\qquad\square$

Proof of Theorem 1. From Lemma 3, it follows that, for any mixed Poisson distribution $D \in \mathcal{D}_\mu$, there is a distribution $D' \in \mathcal{D}'_\mu$ such that $q_{D'} \leq q_D$. The theorem now follows from Lemma 4. $\qquad\square$

4. Thinned configuration model

Let $\mathcal{B}_\mu$ be the collection of all $\mathbb{N}_0$-valued random variables D with $E(D) = \mu$. Let $z_{D,p}$ be the smallest root of the equation $s = \bar{g}_{D,p}(s)$, and let $q_{D,p} = g_{D,p}(z_{D,p})$.

Theorem 2. *Let $q^* = \inf_{D\in\mathcal{B}_\mu} q_{D,p}$. Then there exist $k \in \mathbb{N}$ and a degree distribution $D^* \in \mathcal{B}(\mu)$, which satisfies*

$$P(D^* = 0) + P(D^* = k) + P(D^* = k+1) = 1,$$

such that $q_{D^*,p} = q^*$.

Proof. Let $q := q_{D,p}$ and $z := z_{D,p}$. First we show that we can always find a distribution D' with mass only at 0 and two subsequent integers, such that the associated $q' = q_{D',p}$ satisfies $q \geq q'$.

Let $q'' = q_{D'',p}$ and $z'' = z_{D'',p}$, where the degree distribution D'' is defined by

$$P(D'' = \lfloor\mu\rfloor) = 1 - \mu + \lfloor\mu\rfloor = 1 - P(D'' = \lfloor\mu\rfloor + 1).$$

Here we use the notation $\lfloor x\rfloor := \max\{y \in \mathbb{N}_0 : y \leq x\}$ for the floor of x. The generating functions associated with the model before thinning are $f''(s)$ and $\bar{f}''(s)$. We now distinguish two cases.

Case 1: $z > z''$. Let $k = \lfloor\mu\rfloor$ and $\beta = \mu - k$. We rewrite D as a mixture of two distributions D_1 and D_2, with $E(D_1) = k$ and $E(D_2) = k+1$. So D is D_1 with probability $1 - \beta$ and D_2 with probability β. Then

$$\begin{aligned}
q &= f(1 - p + pz) \\
&= E((1 - p + pz)^D) \\
&= (1 - \beta)E((1 - p + pz)^{D_1}) + \beta E((1 - p + pz)^{D_2}) \\
&\geq (1 - \beta)(1 - p + pz)^k + \beta(1 - p + pz)^{k+1} \\
&= f_{D''}(1 - p + pz),
\end{aligned}$$

where the inequality is obtained by Jensen's inequality. As $z > z''$, and by the observation that $f_{D''}(s)$ is increasing, we note that

$$q \geq f_{D''}(1 - p + pz'') = q'',$$

and the statement holds for $z > z''$.

Case 2: $z \leq z''$. In this case we may choose an integer k and a random variable $D^{\dagger}$ with mass only at 0, k, and $k + 1$, such that $E(D) = \mu$ and $z_{D^{\dagger}, p} =: z^{\dagger} = z$. Let $f^{\dagger}(s) := f_{D^{\dagger}}(s)$ be its generating function, and let $\bar{f}^{\dagger}(s) := \bar{f}_{D^{\dagger}}(s)$. Let $\alpha := P(\bar{D}^{\dagger} - 1 = k)$. As in the previous section, we use the fact that, by (1),

$$f(1 - p + pz) = 1 - \mu \int_{1-p+pz}^{1} \bar{f}(s) \, ds. \tag{3}$$

We note that

$$\frac{d^2}{ds^2} \frac{\bar{f}(s) - \bar{f}^{\dagger}(s)}{s^{k-1}} = \frac{d^2}{ds^2} \left[\left(\sum_{m=0}^{\infty} P(\bar{D} - 1 = m) s^{m-k+1} \right) - (1 - \alpha) + \alpha s \right]$$

$$= \sum_{m=0}^{\infty} P(\bar{D} - 1 = m)(m - k + 1)(m - k) s^{m-k-1}.$$

Since $m - k + 1$ and $m - k$ cannot have opposite signs, all summands are nonnegative. Now if, for some $s \in (0, 1)$, it holds that

$$\bar{f}(s) - \bar{f}^{\dagger}(s) = 0 \quad \text{and} \quad \frac{d}{ds}((\bar{f}(s) - \bar{f}^{\dagger}(s))s^{-k+1}) \geq 0,$$

then it is impossible that

$$(\bar{f}(1) - \bar{f}^{\dagger}(1))s^{-k+1} = 0,$$

which leads to a contradiction. So $\bar{f}^{\dagger}(s) \geq \bar{f}(s)$ on $s \in [z, 1]$ and, by (3), the statement of the theorem follows for $z \leq z''$.

An additional question to consider for the thinned configuration model is the following. Let $\bar{\mathcal{B}}_c$ be the class of all $p \in (0, 1]$ and distributions D with $E[D] = c/p$. For which p and distribution of D in $\bar{\mathcal{B}}_c$ is the (in probability as $n \to \infty$) limit of $n^{-1}|\mathcal{C}^1(n)|$ maximized? Note that in these models the expected number of edges in the thinned graph is kept constant.

If $c > 2$ then $p = 1$ and $P(D = \lfloor c \rfloor + 1) = 1 - P(D = \lfloor c \rfloor) = c - \lfloor c \rfloor$ give $q_{D,p} = 0$, so the (in probability as $n \to \infty$) limit of $n^{-1}|\mathcal{C}^1(n)|$ is maximized.

For $c \leq 2$, we first consider the assymptotic branching process and minimize $q_{D,p}$ over $\bar{\mathcal{B}}_c$. The following heuristic argument shows that $q_{D,p}$ is maximized if $p = 1$ and $P(D = 0) + P(D = 2) = 1$. Assume that either $p = 1$ or $P(D = 0) + P(D = 2) = 1$ does not hold. Then $n^{-1}|\mathcal{C}^1(n)|$ converges in probability to $1 - q_{D,p}$. The limit of $n^{-1}|\mathcal{C}^1(n)|$ is at most $c/2$, because the number of edges in the thinned graph is roughly $cn/2$ and the number of vertices in the giant component of a graph is at most 1 higher than the number of edges. Note that, for $p = 1$ and $P(D = 0) + P(D = 2) = 1$, $q_{D,p} = 1 - c/2$. However, if $p = 1$ and $P(D = 0) + P(D = 2) = 1$, $n^{-1}|\mathcal{C}^1(n)|$ does not converge to $c/2$ in probability. Still, the (in probability) limit of $n^{-1}|\mathcal{C}^1(n)|$ can be made arbitrary close to $c/2$ by chosing $\varepsilon > 0$ arbitrarily small and taking $P(D = 2) = c/2 - 3\varepsilon$, $P(D = 3) = 2\varepsilon$, and $P(D = 0) = 1 - c/2 + \varepsilon$.

Formally, we can minimize $q_{D,p}$ by applying the change of variable $s = 1 - pt$ to the right-hand side of (3) to obtain

$$1 - q_{D,p} = c \int_0^{1-z_{D,p}} \bar{f}_D(1 - pt) \, dt.$$

Assume that $z_{D,p} < 1$; otherwise, there will be no giant component. Furthermore, assume that $P(\bar{D} = 1) < 1$. We note that $\bar{f}_D(1 - pt) = 1 - t$ has roots in $t = 0$ and $t = z_{D,p}$. By the convexity of $\bar{f}_D(1 - pt)$ in t, those are the only roots and $\bar{f}_D(1 - pt) < 1 - t$ for $0 < t < z_{D,p}$. This implies that

$$1 - q_{D,p} \leq c \int_0^{1-z_{D,p}} (1 - t) \, dt \leq c \int_0^1 (1 - t) \, dt \leq \frac{c}{2}.$$

So, if $c \leq 2$, $1 - q_{D,p} < c/2$. On the other hand, the (in probability) limit of the size of the giant, $n^{-1}|\mathcal{C}^1(n)|$, can be taken arbitrarily close to $c/2$ by taking $p = 1$ and $P(D = 2) = c/2 - 3\varepsilon$, $P(D = 3) = 2\varepsilon$, and $P(D = 0) = 1 - c/2 + \varepsilon$ for arbitrarily small $\varepsilon > 0$. This answers the question posed.

5. Discussion

The main conclusion of the current paper is that, for both random graph models, the degree should be 'homogenized' in order to maximize the size of the giant component if the given mean degree μ is larger than some critical value. On the other hand, if μ is smaller than this critical value, certain nodes should be given degree 0 and the remaining nodes should be homogenized, with the effect that these later nodes have large enough degree to result in a large enough giant component.

In terms of epidemics, we found the maximal possible outbreak size for a given mean degree μ and transmission parameter p. If the epidemic is initiated by one randomly selected index case, the outbreak size is equivalent to the *probability* of a major outbreak. As a consequence, the construction mentioned above not only maximizes the size of an outbreak, but it also maximizes the probability of having a major outbreak.

It might at a first glance seem unnatural to allow for individuals to have degree 0. However, in many applications, e.g. sexual networks, the fraction of individuals having degree 0 is distinctly different from 0. Besides, it is not obvious that giving a fraction of the community degree 0 (and the rest being 'homogenized') may in some cases (i.e. when μ is small) make the size of the giant maximal.

It was mentioned in the introduction that the relative size of the *minimal* giant can be made arbitrarily close to 0. An interesting problem would be to investigate if there is a natural definition of the size of a *random* giant, where the randomness is taken over some set of random graphs. Similarly, it would be interesting to derive maximal and minimal sizes of the giant for different classes of random graphs and/or if other local properties of the network, beside the mean degree, are given and fixed. This could, for example, be the degree distribution, the clustering index, and/or the degree correlation.

References

[1] ANDERSSON, H. (1999). Epidemic models and social networks. *Math. Sci.* **24,** 128–147.

[2] BALL, F. G., SIRL, D. AND TRAPMAN, P. (2009). Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv. Appl. Prob.* **41,** 765–796.

[3] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd edn. Cambridge University Press.

[4] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31,** 3–122.

[5] BRITTON, T., DEIJFEN, M., LAGERÅS, A. N. AND LINDHOLM, M. (2008). Epidemics on random graphs with tunable clustering. *J. Appl. Prob.* **45,** 743—756.

[6] DURRETT, R. (2006). *Random Graph Dynamics*. Cambridge University Press.

[7] JAGERS, P. (1975). *Branching Processes with Biological Applications*. John Wiley, London.

[8] MOLLOY, M. AND REED, B. (1998). The size of the giant component of a random graph with a given degree sequence. *Combinatorics Prob. Comput.* **7,** 295–305.

[9] NEWMAN, M. E. J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E* **66,** 016128, 11pp.

[10] NORROS, I. AND REITTU, H. (2006). On a conditionally Poissonian graph process. *Adv. Appl. Prob.* **38,** 59–75.