GRAPHS WITH SPECIFIED DEGREE DISTRIBUTIONS, SIMPLE EPIDEMICS AND LOCAL VACCINATION STRATEGIES

TOM BRITTON, SVANTE JANSON AND ANDERS MARTIN-LÖF

Abstract. Consider a random graph, having a pre-specified degree distribution F but other than that being uniformly distributed, describing the social structure (friendship) in a large community. Suppose one individual in the community is externally infected by an infectious disease and that the disease has its course by assuming that infected individuals infect their not yet infected friends independently with probability p. For this situation the paper determines R_0 and τ_0, the basic reproduction number and the asymptotic final size in case of a major outbreak. Further, the paper looks at some different local vaccination strategies where individuals are chosen randomly and vaccinated, or friends of the selected individuals are vaccinated, prior to the introduction of the disease. For the studied vaccination strategies the paper determines R_v: the reproduction number, and τ_v: the asymptotic final proportion infected in case of a major outbreak, after vaccinating a fraction v.

1. Introduction

Simple undirected random graphs can be used to describe the social network in a large community (e.g. [20]), vertices corresponding to individuals and edges to some type of social relation, from now on denoted friendship. Given such a graph, a model for the spread of the disease may be defined, where individuals at first are susceptible but may then become infected by a friend. An infected individual has the potential to spread the disease to its not yet infected friends before it recovers and becomes immune. The final outbreak, both its size and who gets infected, depends on properties of the social graph as well as on properties of disease transmission. In order to prevent an outbreak it is possible to vaccinate, or immunize in some other way, individuals prior to arrival of the disease. Who and how many that are to be vaccinated specifies the vaccination strategy.

The present paper studies questions arising from such modeling. In particular, we consider random graphs where the degree distribution (the number of friends) follows some pre-specified distribution F, typically having heavy tails, but where the random graph G is otherwise uniformly distributed. The

Date: January 25, 2007; revised October 26, 2007.

2000 *Mathematics Subject Classification.* 60C05; 05C80; 92D30.

Key words and phrases. Degree distribution, epidemic model, final size, limit theorem, networks, random graphs, vaccination.

epidemic model is the simplest possible model for a susceptible-infectious-removed (SIR) disease (e.g. [2]). One randomly selected individual is initially externally infected. Any individual who becomes infected infects each of his/her not yet infected friends independently with probability p, and after that the individual recovers and becomes immune, a state called removed. For this graph and epidemic model we study different vaccination strategies: the uniform strategy and the acquaintance strategy [7]. In both strategies individuals are chosen randomly from the community. In the uniform strategy the selected individuals are vaccinated and in the acquaintance strategy a randomly chosen friend of the selected individual is vaccinated. Both vaccination strategies are local in the sense that the global social network need not be known in order to perform the strategy. We also study a vaccination strategy where, instead of selecting individuals at random, friendships are selected and one or two of the corresponding friends get vaccinated.

As the population size n tends to infinity, we prove that the initial phase of the epidemic may be approximated by a suitable branching process. The largest eigenvalue of the branching process, often denoted R_0 and called the basic reproduction number when applied to epidemics [2], determines whether a major outbreak can occur or not: if $R_0 \leq 1$ only minor outbreaks can occur whereas if $R_0 > 1$ outbreaks of order $O(n)$ can also occur with positive probability. In case of a major outbreak the total number of individuals infected during the outbreak, the final size, is shown to satisfy a law of large number. The corresponding (random) proportion is shown to converge in probability to a deterministic limit τ_0. Similar results are obtained when a vaccination strategy with vaccination coverage v has been performed prior to disease introduction. In this situation the strategy-specific reproduction number R_v, and the major outbreak size τ_v, are determined. From this it is possible to determine the (strategy-specific) critical vaccination coverage v_c which determines the necessary proportion to vaccinate in order to surely prevent a major outbreak, so $v_c = \inf_v\{v; R_v \leq 1\}$.

Stochastic epidemic models on networks with pre-specified degree distributios have mainly been studied in the physics literature (e.g. [17], [19], [7]), Andersson [1] being one exception. Some of the problems studied in the present paper have been analysed before whereas others have not, in particular the final size proportion τ_v as a function of v. Beside contributing with some new results another aim of the paper is to give formal proofs to results which have previously only been obtained heuristically.

The rest of the paper is structured as follows. In Section 2 we define the models for the random graph, the epidemic and the vaccination strategies. In Section 3 we present the main results, motivate them with some heuristics and give some examples and illustrations. The proofs are given in Sections 4 and 5.

2. Models

2.1. **Graphs.** Let G denote a random *multigraph*, allowing for multiple edges and loops, and let $n = |G|$ denote the number of vertices of G, i.e. the population size. Later we shall consider limits as $n \to \infty$. We define our random multigraph as follows. Let $n \in \mathbb{N}$ and let $(d_i)_1^n = (d_i^{(n)})_1^n$ be a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even. We define a *random multigraph with given degree sequence* $(d_i)_1^n$, denoted by $G^*(n, (d_i)_1^n)$, by the configuration model (see e.g. [4]): take a set of d_i half-edges for each vertex i, and combine the half-edges into pairs by a uniformly random matching of the set of all half-edges.

Note that $G^*(n, (d_i)_1^n)$ does not have exactly the uniform distribution over all multigraphs with the given degree sequence; there is a weight with a factor $1/j!$ for every edge of multiplicity j, and a factor $1/2$ for every loop, see [11, §1]. However, conditioned on the multigraph being a (simple) graph, we obtain a uniformly distributed random graph with the given degree sequence, which we denote by $G(n, (d_i)_1^n)$. It is also worth mentioning that the distribution of $G^*(n, (d_i)_1^n)$ is the same as the one obtained by sampling the edges as ordered pairs of vertices uniformly with replacement, and then conditioning on the vertex degrees being correct.

Let us write $2m := \sum_{i=1}^n d_i$, so that $m = m(n)$ is the number of edges in the multigraph $G^*(n, (d_i)_1^n)$. We assume that we are given $(d_i)_1^n$ satisfying the following regularity conditions, cf. Molloy and Reed [15, 16].

Condition 2.1. *For each n, $(d_i)_1^n = (d_i^{(n)})_1^n$ is a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even and, for some probability distribution $(p_j)_{j=0}^\infty$ independent of n, and with $n_j := \#\{i : d_i = j\}$,*

(i) *$n_j/n \to p_j$ for every $j \geq 0$ as $n \to \infty$;*
(ii) *$\mu := \sum_j j p_j \in (0, \infty)$;*
(iii) *$2m/n \to \mu$ as $n \to \infty$.*
(iv) *$p_2 < 1$.*

Remark 2.2. Note that $2m = \sum_i d_i = \sum_j j n_j$. Thus, Condition 2.1 implies that the sum $\sum_j j n_j/n$ converges uniformly for $n \geq 1$, i.e.

$$\lim_{J \to \infty} \sup_n \sum_{j > J} j n_j/n = 0. \tag{2.1}$$

Conversely, (2.1) together with (i) and (ii) implies (iii). (This follows from, e.g., [8, Theorem 5.5.4], taking X_n to be the degree of a random vertex.)

Note that our condition is slightly weaker than the one in Molloy and Reed [15, 16]; they also assume (in an equivalent formulation) that if $\sum_j j^2 p_j < \infty$, then the sums $\sum_j j^2 n_j/n$ converge uniformly; moreover they assume that $j^2 n_j/n \to j^2 p_j$ uniformly.

Condition 2.1 is all we need to study the random multigraph $G^*(n, (d_i)_1^n)$. In order to treat the random simple graph $G(n, (d_i)_1^n)$, which is our main model, we need an additional assumption.

Condition 2.3. $\sum_i d_i^2 = O(n)$.

Note that $\sum_i d_i^2 = \sum_j j^2 n_j$, so Conditions 2.1 and 2.3 imply, by Fatou's lemma, that $\sum_j j^2 p_j < \infty$; in other words, the asymptotic degree distribution has finite variance.

When Conditions 2.1 and 2.3 hold, the probability that $G^*(n, (d_i)_1^n)$ is a simple graph is bounded away from 0, see Subsection 5.2 for details, and thus all results that can be stated in terms of convergence in probability for $G^*(n, (d_i)_1^n)$ transfer to the random simple graph $G(n, (d_i)_1^n)$ too.

2.2. **Alternative graph models.** We will in the remainder of the paper consider $G(n, (d_i)_1^n)$ as our underlying graph model, but we believe that similar results hold for other random graph models too, and that they could be proved by suitable modifications of the branching process arguments below. Good candidates are the classical random graphs $G(n, p)$ and $G(n, m)$, with $p = \mu/n$ and $m = n\mu/2$ (rounded to an integer), respectively, and random graphs of the general type $G(n, \kappa)$ defined in [5]. We will not pursue this here, and leave such attempts to modify the proofs to the interested reader, but we will discuss one interesting case (including $G(n, p)$) where the result easily follow from the results proved below for $G(n, (d_i)_1^n)$.

This example is a random graph defined by Britton, Deijfen and Martin-Löf [6, Section 3], see also [5, Subsection 16.4], as follows. Let W be a non-negative random variable with finite expectation $\mu_W := \mathbb{E}\, W$. We first assign random weights W_i, $i = 1, \dots, n$ to the vertices; these weights are i.i.d. with the same distribution as W. Secondly, given $\{W_i\}_1^n$, we draw an edge between vertices i and j with probability

$$p_{ij} := \frac{W_i W_j}{n + W_i W_j}; \tag{2.2}$$

this is done independently (conditioned on $\{W_i\}$) for all pairs $\{i, j\}$ with $1 \le i < j \le n$. We denote this random graph by $G_W(n)$. It is easily seen [6] that (2.2) implies that all graphs with a given degree sequence $(d_i)_1^n$ have the same probability; in fact, if G is any graph with degree sequence $(d_i)_1^n$, then

$$\mathbb{P}\big(G_W(n) = G \mid (W_i)_{i=1}^n\big) = \frac{n^{\binom{n}{2} - \frac{1}{2}\sum_i d_i} \prod_i W_i^{d_i}}{\prod_{i<j}(n + W_i W_j)}.$$

Hence, if we denote the (random) vertex degrees by $D_1, \dots, D_n$, then conditioned on $D_i = d_i$, $i = 1, \dots, n$, we have a random graph $G(n, (d_i)_1^n)$. Moreover, it is not difficult to verify that Condition 2.1 holds in probability, with $(p_j)_0^\infty$ the mixed Poisson distribution $\mathrm{Po}(\mu_W W)$ and $\mu = \mu_W^2$, see [6, Theorem 3.1] and [5, Theorem 3.13]; in other words, $n_j/n \xrightarrow{\mathrm{p}} p_j$ and $2m/n = n^{-1}\sum_i d_i \xrightarrow{\mathrm{p}} \mu$. Assume from now on that $\mathbb{E}\, W^2 < \infty$; it may then be shown by similar arguments that $n^{-1}\sum_i d_i^2 \xrightarrow{\mathrm{p}} \mu_w^2(\mathbb{E}\, W^2 + 1)$. Using the Skorohod coupling theorem, see e.g. [13, Theorem 4.30]), we can assume

that these limits hold a.s.; hence Conditions 2.1 and 2.3 hold a.s. Consequently, by conditioning on $(D_1, \dots, D_n)$, we can apply the results proved in the present paper for $G(n, (d_i)_1^n)$, and it follows that the theorems below hold for the random graph $G_W(n)$ too, with (p_j) and μ as above.

Further, it is easy to see that this remains true if (2.2) is modified to

$$p_{ij} := \min\Big(\frac{W_i W_j}{n}, 1\Big); \tag{2.3}$$

we may use suitable couplings and compare the random graph defined by (2.3) by the ones defined by (2.2) for the same W_i (giving a lower bound), or by (2.2) with W_i replaced by $(1+\varepsilon)W_i$ (giving an upper bound, assuming as we may that $W_i \le \sqrt{\varepsilon n}$), and then letting $\varepsilon \searrow 0$; we omit the details. More precisely, it can be shown [10] that under the assumptions above on (W_i), the random graphs defined by (2.2) and (2.3) are asymptotically equivalent in a strong sense (the total variation distance tends to 0). Random graphs defined by (2.3) and minor variations of it have been studied by several authors, see [5, Subsection 16.4] and the references given there. Note that the special (deterministic) case $W = \sqrt{\mu}$ for a constant $\mu > 0$ gives the classical random graph $G(n, \mu/n)$. The results in this paper thus holds for $G(n, \mu/n)$ too, with (p_j) a $\mathrm{Po}(\mu)$ distribution; in other words, with D defined in Section 3, $D \sim \mathrm{Po}(\mu)$.

2.3. **Epidemic model.** We consider an infectious disease that spreads along the edges of a graph G. We will in this paper assume that $G = G(n, (d_i)_1^n)$ is the random graph defined above, where we condition the graph $G^*(n, (d_i)_1^n)$ on being simple. The vertices of G are the individuals in the population, and the edges represent friendships through which infection might spread.

The disease has its course in the following way. Initially, one randomly chosen individual (vertex) is infected from the outside. This individual then spreads the disease to each of its friends independently and with the same probability p. Those who get infected make out the first generation infected in the epidemic. These individuals then do the same thing to their not yet infected friends thus infecting a second generation, and so forth. Note that an individual can only get infected once – we then consider such an individual either recovered and immune (or dead). This epidemic continues until there are no new infections in a generation, when it stops. Since the population is finite this happens after a finite number of generations ($\le n$, where $n = |G|$ is the size of the population). The individuals who get infected during the course of the epidemic make up the total outbreak, and the number of such individuals is called the final size of the epidemic.

Note that each edge is a possible path of infection at most once, namely when the first of its endpoints has been infected. Hence we may just as well determine in advance for every edge in G whether it will spread the disease or not, provided that one of the endpoints gets infected. Equivalently, we may consider the graph G_p obtained by randomly deleting edges from G, with each edge kept with probability p, independently of the others. The

final size of the epidemic is thus the size of the component of G_p containing the initially infected individual.

2.4. **Vaccination strategies.** Assume now that a perfect vaccine is available. By this we mean that an individual who is vaccinated is completely protected from (i.e., immune to) the disease and is not able to spread the disease further. We assume that a part of the population is vaccinated before the epidemic starts, or as soon as the first individual is infected. The epidemic progresses as defined above, with the only difference that infected individuals can only infect unvaccinated friends.

Note that for the study of the epidemic in the vaccinated population, we may simply remove all vaccinated individuals from G (and edges connected to these individuals). If we let G_{v} denote the remaining graph, and we assume that the initially infected individual x is not vaccinated, the final size of the epidemic is thus the size of the component of $G_{\mathsf{v};p} := (G_{\mathsf{v}})_p$ that contains x. We thus have to study the combined effect on G of vertex deletion by the vaccination and edge deletion by the randomness of infection.

The goal is to contain the disease, so that the final size of the epidemic is small, and it is preferable to do this with a rather small number of vaccinations. For this we look at different local vaccination strategies. The first two strategies are local in the sense that they require no global knowledge of the social network G (which is rarely available in applications, [18, Section 8.2]) and the latter two selects friendhsips rather than individuals at random which may also be thought of needing only local information. We let V denote the (usually random) number of vaccinations.

Uniform vaccination. Let us assume that we sample a fraction $c \in [0,1]$ chosen uniformly in the population without replacement and that this fraction is immunized, so the fraction v being immunized satisfies $v = c$. This vaccination strategy is the most commonly studied vaccination strategy due to its simplicity [18, Section 8.2].

More precisely, for convenience, we assume that each individual is vaccinated with a given probability v, independently of each other. The number V of vaccinations is thus $\mathrm{Bi}(n, v)$, and $V/n \xrightarrow{\mathrm{p}} v$ as $n \to \infty$ (with v fixed). We denote the remaining graph of unvaccinated individuals by G_v^{U}; this is thus obtained from G by random vertex deletions. Remember that our main concern is with the graph $G_{v;p}^{\mathsf{U}} = (G_v^{\mathsf{U}})_p$; this is obtained from G by random vertex and edge deletions, independently for all vertices and edges. (In this case, it does not matter whether we delete edges or vertices first.)

Acquaintance vaccination. It is intuitively clear that a better vaccination strategy would be to vaccinate the individuals with highest degrees (most friends) since this would reduce potential spread the most. However, for this targeted vaccination strategy to be achievable the whole social graph (or at least the degrees of all individuals) would have to be known, and this

is rarely the case [18, Section 8.2]. A different strategy aiming at vaccinating individuals with high degree, but still only using local graph-knowledge from selected individuals, proposed by Cohen et al. [7], goes under the name *acquaintance vaccination*. In this vaccination strategy a fraction c of individuals are sampled, and for each sampled individual one of its friends, chosen randomly among all friends, is vaccinated. Of course it may happen that some individuals are chosen more than once for immunization (being selected as friends of more than one individual) so the fraction $v = v(c)$ actually immunized is smaller than c. This vaccination strategy has two slightly different variants depending on whether the "fraction" c is chosen with or without replacement. We will use the version with replacement. For this case the "fraction" c may in fact exceed 1 without having everyone vaccinated (individuals who are selected more than once are asked for friends independently each time and friends not yet immunized are vaccinated). To be precise, we let the number of individuals sampled be Poisson distributed $\mathrm{Po}(cn)$, with $c \in [0, \infty)$. Equivalently, each individual is sampled $\mathrm{Po}(c)$ times, and each time reports a randomly chosen friend. Again, for simplicity, we assume that each individual does this with replacement. Consequently, an individual with degree d will report each of its friends $\mathrm{Po}(c/d)$ times, and these random numbers are all independent. (An individual that is sampled but has no friends is ignored. An individual is only vaccinated once, even if he or she is reported several times.)

For any initial graph G and $0 \le c < \infty$, we denote the remaining graph of unvaccinated individuals by G_c^{A}. We further write $G_{c;p}^{\mathsf{A}} = (G_c^{\mathsf{A}})_p$ for the graph obtained by additional edge deletions. (For acquaintance vaccination, the order of the deletions is important, since the vaccination strategy uses all edges, without knowing whether they may be selected to transmit the disease or not.)

Edgewise vaccination. In some situations it may be possible to observe, or at least sample, the edges representing friendships. If this is the case, another reasonable vaccination strategy is to sample a number of the edges and then either vaccinate both endpoints or one (randomly selected) endpoint; we denote these two versions by $\mathsf{E1}$ and $\mathsf{E2}$.

For $\mathsf{E2}$, we assume that we sample each edge with probability $1 - \alpha$, where $\alpha \in (0, 1]$ is a fixed number. (Equivalently, we sample $\mathrm{Po}(cm)$ edges with replacement, with $\alpha = e^{-c}$.) For $\mathsf{E1}$, we assume for simplicity that we sample $\mathrm{Po}(2cm)$ edges with replacement; thus each end of each edge is sampled with probability $1 - \alpha = 1 - e^{-c}$, independently of all other edge ends. Hence, for both versions, a vertex with degree d is unvaccinated with probability α^d, and for $\mathsf{E1}$, this is independent of all other vertices.

For an initial graph G and $0 < \alpha \le 1$, we denote the remaining graph of unvaccinated individuals by G_α^{E1} and G_α^{E2}, for the two versions. We further write, for $j = 1, 2$, $G_{\alpha;p}^{\mathsf{E}j} = (G_\alpha^{\mathsf{E}j})_p$ for the graph obtained by additional edge deletions.

3. Main results

We now state our main results together with heuristic motivations. We assume that the underlying graph is the random graph $G(n,(d_i)_1^n)$ and that Conditions 2.1 and 2.3 hold. Complete proofs are given in Section 5.

3.1. **Original epidemic model.** Assume that n, the number of nodes, is large. The regularity assumption on the degrees of the graph (Condition 2.1) implies that no separate node will contain a large fraction of all edges, see (2.1). This in turn implies that self loops, multiple edges and short cycles will be rare.

The epidemic starts by a randomly selected individual being infected from outside, so this individual has (approximately) the degree distribution $(p_j)_{j=0}^\infty$. The friends of this individual, or friends of any given individual, have the size biased degree distribution $(\tilde{p}_j)_{j=0}^\infty$, where

$$\tilde{p}_j = jp_j/\sum_k kp_k. \tag{3.1}$$

Let D and $\tilde{D}$ be random variables having these degree distributions respectively. Thus (asymptotically), we can interpret D is the number of friends of a random person, while $\tilde{D}$ is the number of friends of a random friend of a given person. Then, given that $D=d$, the number of individuals that the initially infected infects is $\mathrm{Bi}(d,p)$, and the unconditional distribution is hence mixed binomial $\mathrm{MixBi}(D,p)$. Those then infected, as well as infecteds in the following generations, have degree distribution $(\tilde{p}_j)_{j=0}^\infty$. Given that $\tilde{D}=\tilde{d}$, the number of individuals an infected individual infects in the next generation has distribution $\mathrm{Bi}(\tilde{d}-1,p)$. This follows because the infected was infected by one of his friends (which cannot get reinfected) and, since short cycles are rare, it is very unlikely that any of the remaining $\tilde{d}-1$ friends have already been infected. Unconditionally, the number infected in the next generation is hence $\mathrm{MixBi}(\tilde{D}-1,p)$. Further, the property that short cycles are unlikely implies that the number of infections caused by different individuals are (approximately) independent random variables.

The above paragraph motivates why the early stages of the epidemic may be approximated by a branching process (e.g. [3]), as is common for epidemic models (e.g. [2]), and where "giving birth" corresponds to infecting someone. The branching process is a simple Galton–Watson process starting with one ancestor having off-spring distribution $X\sim\mathrm{MixBi}(D,p)$ and the following generations have off-spring distribution $\tilde{X}\sim\mathrm{MixBi}(\tilde{D}-1,p)$. The mean of this latter off-spring distribution plays an important role in branching process theory and also in in epidemic theory where it is denoted R_0 and

denoted the basic reproduction number. We get the following, using (3.1),

$$R_0 = \mathbb{E}(\tilde{X}) = p\,\mathbb{E}(\tilde{D}-1) = p\left(\frac{\sum_j j^2 p_j}{\mu} - 1\right) = p\left(\mu + \frac{\mathrm{Var}(D)-\mu}{\mu}\right), \tag{3.2}$$

where $\mu = \mathbb{E}(D) = \sum_k k p_k$ and $\mathrm{Var}(D) = \sum_j j^2 p_j - \mu^2$ (a very related expression is obtained in [1]). The branching process is subcritical, critical or supercritical depending on whether $R_0 < 1$, $R_0 = 1$ or $R_0 > 1$. For the epidemic, this means a major outbreak infecting a non-negligible fraction of the community, is possible if and only if $R_0 > 1$. Note that, for fixed μ, R_0 is increasing in $\mathrm{Var}(D)$, so the more variance in the degree distribution, the higher R_0, and if the degree distribution has infinite variance then $R_0 = \infty$ (a case not treated in the present manuscript due to Condition 2.3).

The probability π that the branching process dies out is derived in the standard way as follows. First, we derive the probability $\tilde{\pi}$ that a branching process with all individuals having off-spring distribution $\tilde{X}$ dies out. This is obtained by conditioning on the number of individuals born in the first generation: for the branching process to die out, all branching processes initiated by the individuals of the first generation must die out, i.e.

$$\tilde{\pi} = \sum_{k=0}^{\infty} \tilde{\pi}^k\, \mathbb{P}(\tilde{X} = k).$$

Let $f_{\tilde{X}}(\cdot)$ denote the probability generating function for $\tilde{X}$, and $f_D(\cdot)$ the probability generating function of the original degree distribution D. Then we see that $\tilde{\pi}$ is a solution to the equation $f_{\tilde{X}}(t) = t$, and it is known from branching process theory (e.g. [3, Theorem I.5.1]) that it is the smallest non-negative such solution. The fact that $\tilde{X}$ is MixBi$(\tilde{D}-1, p)$ implies that

$$f_{\tilde{X}}(t) = \mathbb{E}(t^{\tilde{X}}) = \mathbb{E}(\mathbb{E}(t^{\tilde{X}}|\tilde{D})) = \mathbb{E}((pt+1-p)^{\tilde{D}-1}) = \mathbb{E}((1-p(1-t))^{\tilde{D}-1}).$$

Further,

$$\mathbb{E}(a^{\tilde{D}-1}) = \sum_k a^{k-1}\frac{k p_k}{\mu} = \frac{d}{da}\sum_k a^k \frac{p_k}{\mu} = \frac{d}{da}\frac{f_D(a)}{\mu} = \frac{f_D'(a)}{\mu} = \frac{f_D'(a)}{f_D'(1)}.$$

In terms of $f_D(\cdot)$ the probability $\tilde{\pi}$ that the branching process dies out is hence the smallest non-negative solution to

$$\frac{f_D'(1-p(1-\tilde{\pi}))}{f_D'(1)} = \tilde{\pi}. \tag{3.3}$$

The probability π that the branching process, in which the ancestor has different off-spring distribution X, dies out, is obtained from $\tilde{\pi}$ by conditioning on the number of off-spring of the ancestor:

$$\begin{aligned}\pi &= \sum_k \tilde{\pi}^k\, \mathbb{P}(X=k) = \mathbb{E}(\tilde{\pi}^X) = \mathbb{E}(\mathbb{E}(\tilde{\pi}^X|D)) = \mathbb{E}((p\tilde{\pi}+1-p)^D) \\ &= f_D(1-p(1-\tilde{\pi})).\end{aligned} \tag{3.4}$$

We now look at the final size of the epidemic in case it takes off, corresponding to the case that the branching process grows beyond all limits. We do this by considering the epidemic from a graph representation. The social structure was represented by a random graph G. If this graph is thinned by removing each edge independently with probability $1-p$ we get a thinned graph denoted G_p. Edges in G_p represent potential spread of infection: if one of the nodes get infected from elsewhere, its neighbour will get infected. As a consequence, the final outbreak of the epidemic will consist of all nodes in G_p that are connected to the initially infected. From random graph theory it is known that if $R_0 > 1$ there will be exactly one connected component of order n, the giant component, and all remaining connected components will be of smaller order. If $R_0 \leq 1$ there will be no giant component. The initially infected was chosen uniformly in the community so it will belong to the giant component with a probability that equals the relative size of the giant component. On the other hand, the initially infected belongs to the giant component if and only if its branching process of new infections grows beyond all limits, and we know from before that this happens with probability $1-\pi$ defined in equation (3.4). From this it follows that the asymptotic final proportion infected, τ, equals $1-\pi$. So, τ is both the probability of a major outbreak, and the relative size of the outbreak in case a major outbreak occurs.

The above arguments motivate the following theorem, which is proven in Section 5, and where Z_n denotes the final number infected in the epidemic.

Theorem 3.1. *If $R_0 \leq 1$ then $Z_n/n \xrightarrow{\mathrm{p}} 0$. If $R_0 > 1$, then Z_n/n converges to a two-point distribution Z for which $\mathbb{P}(Z=0) = \pi$ and $\mathbb{P}(Z=\tau) = \tau$, where π is defined by* (3.3) *and* (3.4) *and $\tau = 1-\pi$.*

3.2. **Uniform vaccination.** Prior to arrival of the infectious disease, each individual is vaccinated independently and with the same probability v which implies that the total number of vaccinated V is $\mathrm{Bi}(n,v)$, and from the law of large number the random proportion vaccinated $V/n \xrightarrow{\mathrm{p}} v$.

Vaccinated individuals, and edges connecting to them, can be removed from the graph since there will be no spreading between these individuals and their friends in either direction. As a consequence, an individual who originally had d friends now has $\mathrm{Bi}(d, 1-v)$ unvaccinated friends. If an individual gets infected during the early stages of the epidemic he will infect each of his unvaccinated friends independently with probability p. Given that the initially infected has degree d he will hence infect $\mathrm{Bi}(d, p(1-v))$ friends, so without the conditioning he will infect a mixed binomial number $X_v \sim \mathrm{MixBi}(D, p(1-v))$. Similarly, during the early stages an infected individual with degree d will infect $\mathrm{Bi}(d-1, p(1-v))$, and unconditionally an individual has degree distribution $\{\tilde{p}_k\}$, so the unconditional number he will infect $\tilde{X}_v$ will be $\mathrm{MixBi}(\tilde{D}-1, p(1-v))$.

It is seen that we have the same type of distributions as in the case without vaccination. As a consequence, all results for the case with uniform

vaccination can be obtained from the case without vaccination simply by replacing p by $p(1-v)$. We hence have that the reproduction number $R^{\mathsf{U}}_{v;p}$ after vaccinating a fraction v chosen uniformly satisfies

$$R^{\mathsf{U}}_{v;p} = \mathbb{E}(\tilde{X}_v) = (1-v)R_0 = p(1-v)\left(\mu + \frac{\mathrm{Var}(D)-\mu}{\mu}\right). \tag{3.5}$$

The probability $\tilde{\pi}^{\mathsf{U}}_{v;p}$ that the epidemic never takes off, assuming the initially infected has $\tilde{X}_v$ unvaccinated friends, is the smallest solution to

$$\frac{f'_D\big(1-p(1-v)(1-\tilde{\pi}^{\mathsf{U}}_{v;p})\big)}{f'_D(1)} = \tilde{\pi}^{\mathsf{U}}_{v;p}. \tag{3.6}$$

The probability $\pi^{\mathsf{U}}_{v;p}$ that the epidemic never takes off if the initially infected is selected randomly among the unvaccinated is given by

$$\pi^{\mathsf{U}}_{v;p} = f_D\big(1-p(1-v)(1-\tilde{\pi}^{\mathsf{U}}_{v;p})\big)\big), \tag{3.7}$$

where $\tilde{\pi}^{\mathsf{U}}_{v;p}$ is the smallest solution to (3.6). Finally, the final size is determined from the probability of a major outbreak as before. This means that the final proportion infected (among the unvaccinated!) will converge to $1-\pi^{\mathsf{U}}_{v;p}$ in case of a major outbreak. We have the following corollary, where $Z^{\mathsf{U}}_n(v)$ denotes the final number infected in the epidemic where each individual was vaccinated independently with probability v ($0 \le v < 1$) prior to the outbreak, and where the initially infected was chosen randomly among the unvaccinated.

Theorem 3.2. *If $R^{\mathsf{U}}_{v;p} \le 1$, then $Z^{\mathsf{U}}_n(v)/((1-v)n) \xrightarrow{\mathrm{p}} 0$. If $R^{\mathsf{U}}_{v;p} > 1$, then $Z^{\mathsf{U}}_n(v)/((1-v)n)$ converges to a two-point distribution $Z^{\mathsf{U}}_{v;p}$ for which $\mathbb{P}(Z^{\mathsf{U}}_{v;p}=0) = \pi^{\mathsf{U}}_{v;p}$ and $\mathbb{P}(Z^{\mathsf{U}}_{v;p} = \tau^{\mathsf{U}}_{v;p}) = \tau^{\mathsf{U}}_{v;p}$, where $\pi^{\mathsf{U}}_{v;p}$ is defined by* (3.6) *and* (3.7) *and $\tau^{\mathsf{U}}_{v;p} = 1-\pi^{\mathsf{U}}_{v;p}$.*

3.3. **Acquaintance vaccination.** Recall that each individual is sampled, independently, a $\mathrm{Po}(c)$ number of times, where $0 \le c < \infty$, so in total $\mathrm{Po}(nc)$ individuals are sampled. Each time an individual is sampled, a randomly chosen friend of the individual is selected and vaccinated (unless it already was vaccinated). The effect of this strategy is that vaccinated individuals, being selected as somebody's friend, have the size biased degree distribution $(\tilde{p}_j)_{j=0}^{\infty}$, where $\tilde{p}_j = jp_j/\sum_k kp_k$, rather than the original degree distribution $\{p_k\}$ for uniformly selected individuals. The proportion vaccinated $v = v(c)$ is obtained as follows. An individual avoids being vaccinated if he is not vaccinated "through" any of its friends. The friends of the individual have independent degree distributions $(\tilde{p}_j)_{j=0}^{\infty}$, and the probability of not being vaccinated "through" an individual with degree k is $e^{-c/k}$ (the number of vaccination attempts on a specific friend is $\mathrm{Po}(c/k)$). It follows that the probability to avoid being vaccinated from one friend equals

$$\alpha = \alpha(c) = \sum_{k=1}^{\infty} e^{-c/k}\tilde{p}_k = \sum_{k=1}^{\infty} e^{-c/k}\frac{kp_k}{\mu}. \tag{3.8}$$

(Note that α has the same interpretation as for α introduced for the edgewise strategies, but it is a different function of c.) If the individual in question has j friends it hence avoids being vaccinated with probability α^j. The proportion $1-v(c)$ not being vaccinated equals the probability that a randomly selected individual is *not* vaccinated, which hence equals

$$1 - v(c) = \sum_{j=0}^{\infty} \alpha^j p_j = f_D(\alpha), \tag{3.9}$$

where as before $f_D(\cdot)$ is the probability generating function of a random variable D having distribution $(p_j)_{j=0}^{\infty}$.

Note that in this model, given the graph, individuals are vaccinated independently of each other (although with different probabilities). It follows easily that the actual (random) number V of vaccinated persons satisfy

$$V/n \xrightarrow{\mathrm{p}} v(c) \qquad \text{as } n \to \infty. \tag{3.10}$$

Hence we will ignore the randomness in V and regard $v(c)$ given by (3.9) as the proportion of vaccinated persons.

We now approximate the initial stages of an epidemic, occurring in a community having been vaccinated according to the acquaintance strategy, with a suitable branching process. To find "the right" branching process approximation is harder for the acquaintance strategy because the vaccination status of an individual depends on the degrees of its friends. We therefore introduce some convenient terminology.

We say that *transmission may take place* through an edge, and through its two half-edges, if it is one of the edges in G_p, i.e., one of the randomly selected edges which will spread the disease if one of its endpoints is infected. (Recall that we may assume this random selection to take place before the start of the infection.) Further, there is a natural correspondence between half-edges and *directed* edges, with a half-edge corresponding to the edge it is part of, directed so that the it begins with this half-edge. We say that a directed edge, or the corresponding half-edge, is *used for vaccination*, if the person at the start of the edge is selected and names the person at the end of the edge, who thus gets vaccinated.

It turns out that a suitable "individual" in the branching process is a pair (x, ϵ) consisting of an unvaccinated person x *together* with a directed edge ϵ from this person satisfying the conditions that transmission may take place through the edge ϵ and that ϵ is not used for vaccination. It is worth noting that a person may be part of several "individuals" in the branching process (if the person was not vaccinated and has several friends such that the connecting edges satisfy the conditions above). See Figure 1 for an illustration of an individual (a) and situations where the individual "gives birth" to 2 (b) and 0 (c) individuals. In (b) the "individual" to the left gives birth to two individuals: the person in the centre together with the upgoing edge, and the person in the middle together with the edge to the right. These two edges are parts of individuals since the edges are open for

transmission *and* the middle person did not name the friends at the other end for vaccination. The down-going edge is not part of an individual since the person in the middle named the friend below for vaccination. In (c) no individual is born since the middle person is vaccinated (being named for vaccination by one of his friends excluding the person to the left).

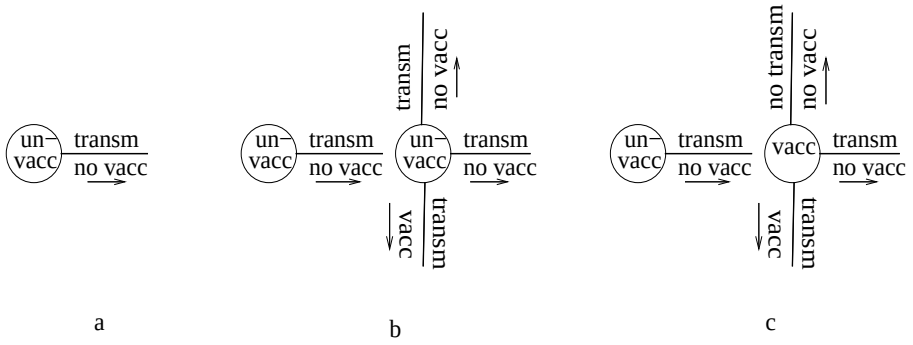

FIGURE 1. (a) An illustration of an "individual" in the branching process. In (b) the left "individual" has two offspring: the person in the middle with the edge to its right, and the person in the middle with the up-going edge. In (c) no individual is born.

In order to analyse the corresponding branching process we have to determine the distribution of how many new "individuals" one "individual" will infect during the early stages of the epidemic assuming a large population (large n). First we determine the distribution of the degree K of the friend z at the other end of the edge ϵ of our "individual" (x, ϵ). We know that the person x of our "individual" is unvaccinated, so the edge ϵ has not been used for vaccination backwards, i.e. in the opposite direction. As a consequence, we have to condition on this, and then the friend z at the other end of ϵ has degree $K = k$ with probability

$$\mathbb{P}(K = k) = \frac{\tilde{p}_k e^{-c/k}}{\sum_{j=1}^{\infty} \tilde{p}_j e^{-c/j}} = \frac{\tilde{p}_k e^{-c/k}}{\alpha}, \quad k = 1, 2, \ldots, \tag{3.11}$$

i.e. the size biased degree distribution conditional on not having vaccinated backwards. In order for this friend z to create new "individuals", it must not have been vaccinated by any of its other $K - 1$ friends (by assumption it was not vaccinated from our original individual x). Conditioned on $K = k$, this happens with probability α^{k-1}. Each of the friend's remaining $k-1$ edges then will be *open* (i.e., transmission may take place but it is not used for vaccination) *independently*, each open with probability $pe^{-c/k}$. The number of open edges (equal to the number of new "individuals") is hence $\mathrm{Bi}(k-1, pe^{-c/k})$. If the friend is vaccinated (probability $1 - \alpha^{k-1}$) no new individuals are born. The unconditional number Y of new "individuals" an

individual "gives birth" to, i.e. the off-spring distribution of the approximating branching process, can thus be obtained by conditioning on the number K of friends our friend z has, using (3.11) and recalling that 0 individuals are born whenever the friend is vaccinated or if the binomial variable equals 0:

$$\begin{aligned}\mathbb{P}(Y=0) &= \sum_{k=1}^{\infty}\Big((1-\alpha^{k-1})+\alpha^{k-1}(1-pe^{-c/k})^{k-1}\Big)\frac{\tilde{p}_k e^{-c/k}}{\alpha},\\ \mathbb{P}(Y=j) &= \sum_{k=j+1}^{\infty}\alpha^{k-1}\binom{k-1}{j}(pe^{-c/k})^j(1-pe^{-c/k})^{k-1-j}\frac{\tilde{p}_k e^{-c/k}}{\alpha},\quad j\geq 1.\end{aligned} \tag{3.12}$$

This off-spring distribution determines both $R^{\mathsf{A}}_{c;p}$, the probability of a major outbreak, and the final size in case of a major outbreak. For instance, the reproduction number is the mean of this distribution, and this mean is obtained by first conditioning on the degree of the node in question. Given that the degree equals k, the average number of off-spring equals $\alpha^{k-1}(k-1)pe^{-c/k}$, which gives the following reproduction number:

$$R^{\mathsf{A}}_{c;p} = \mathbb{E}(Y) = \sum_{k\geq 1}\alpha^{k-1}(k-1)pe^{-c/k}\frac{\tilde{p}_k e^{-c/k}}{\alpha} = p\sum_{k\geq 1}(k-1)\alpha^{k-2}e^{-2c/k}\tilde{p}_k \tag{3.13}$$

(cf. [7]). Let $f_Y(a)=\mathbb{E}(a^Y)$ be the probability generating function of this off-spring distribution. If the epidemic starts by one "individual", i.e. one person with one open directed edge, then the probability $\tilde{\pi}^{\mathsf{A}}_{c;p}$ that the epidemic never takes off is the smallest solution to the equation

$$\tilde{\pi}^{\mathsf{A}}_{c;p} = f_Y(\tilde{\pi}^{\mathsf{A}}_{c;p}). \tag{3.14}$$

If we start with one infected person that is unvaccinated and has degree j, then each of its j half-edges is open with probability $pe^{-c/j}$, and the probability that a given half-edge does not start a large epidemic is $1-pe^{-c/j}+pe^{-c/j}\tilde{\pi}^{\mathsf{A}}_{c;p}$, so the probability that the epidemic never takes off equals $(1-pe^{-c/j}(1-\tilde{\pi}^{\mathsf{A}}_{c;p}))^j$, for $j\geq 1$, and 1 for $j=0$.

If instead the initially infected is chosen randomly among the unvaccinated as we assume, then the probability that it has degree j is $p_j\alpha^j/\sum_j p_j\alpha^j$, cf. (3.9), and thus the probability that the epidemic never takes off equals

$$\pi^{\mathsf{A}}_{c;p} = \frac{p_0+\sum_{j\geq 1}p_j\alpha^j\big(1-pe^{-c/j}(1-\tilde{\pi}^{\mathsf{A}}_{c;p})\big)^j}{\sum_j p_j\alpha^j}. \tag{3.15}$$

Finally, using the same reasoning as before, the limiting proportion infected in case of a major outbreak equals $\tau^{\mathsf{A}}_{c;p} = 1-\pi^{\mathsf{A}}_{c;p}$. We summarize our results in the following theorem, proved in Section 5, where $Z^{\mathsf{A}}_n(c)$ denotes the final number infected in the epidemic where vaccination is done prior to the outbreak according to the acquaintance vaccination strategy. Recall

that $0 \le c < \infty$ and that $v(c)$, the proportion of the population vaccinated, is given by (3.9) with $\alpha = \alpha(c)$ given by (3.8).

Theorem 3.3. *$Z_n^{\mathsf{A}}(c)/\big((1-v(c))n\big) \xrightarrow{\mathrm{p}} 0$ if $R_{c;p}^{\mathsf{A}} \le 1$, where $R_{c;p}^{\mathsf{A}}$ is defined by (3.13). If $R_{c;p}^{\mathsf{A}} > 1$, then $Z_n^{\mathsf{A}}(c)/\big((1-v(c))n\big)$ converges to a two-point distribution $Z_{c;p}^{\mathsf{A}}$ for which $\mathbb{P}\big(Z_{c;p}^{\mathsf{A}} = 0\big) = \pi_{c;p}^{\mathsf{A}}$ and $\mathbb{P}\big(Z_{c;p}^{\mathsf{A}} = \tau_{c;p}^{\mathsf{A}}\big) = \tau_{c;p}^{\mathsf{A}}$, where $\pi_{c;p}^{\mathsf{A}}$ is defined by (3.14) and (3.15), and $\tau_{c;p}^{\mathsf{A}} = 1 - \pi_{c;p}^{\mathsf{A}}$.*

3.4. **Edgewise vaccination.** Recall that, for both E1 and E2, a person with d friends is unvaccinated with probability α^d (here α has the same meaning in the previous subsection, but it can be treated as a free parameter). Thus,

$$\mathbb{E}\, V = n \sum_d p_d(1-\alpha^d) + o(n)$$

and a simple variance estimate shows that the vaccinated proportion

$$V/n \xrightarrow{\mathrm{p}} v(\alpha) := \sum_d p_d(1-\alpha^d), \tag{3.16}$$

just as for acquaintance vaccination, see (3.9) and (3.10).

We define open (directed) edges as for acquaintance vaccination, and argue as there with the following modifications. The other endpoint of an open edge has just the size-biased distribution $(\tilde{p}_k)$. If this vertex, z say, has degree k, it is unvaccinated with probability α^{k-1}, and in that case, the number of new open edges originating at z is $\mathrm{Bi}(k-1, p\alpha)$ for E1 and $\mathrm{Bi}(k-1,p)$ for E2. The difference between the two versions is because we already know that these edges do not vaccinate z, and for E2, this implies that they do not vaccinate their other endpoint either, while for E1 that is an independent event with probability α.

We thus have the offspring distributions for E1 and E2, cf. (3.12),

$$\mathbb{P}(Y_1 = j) = \sum_{k=j+1}^{\infty} \tilde{p}_k \alpha^{k-1} \binom{k-1}{j} (p\alpha)^j (1-p\alpha)^{k-1-j}, \qquad j \ge 1,$$

$$\mathbb{P}(Y_2 = j) = \sum_{k=j+1}^{\infty} \tilde{p}_k \alpha^{k-1} \binom{k-1}{j} p^j (1-p)^{k-1-j}, \qquad j \ge 1;$$

we leave the formulas for $\mathbb{P}(Y_1 = 0)$ and $\mathbb{P}(Y_2 = 0)$ to the reader.

This gives the reproduction numbers

$$\begin{aligned} R_{\alpha;p}^{\mathsf{E1}} &= \mathbb{E}(Y_1) = \sum_{k\ge 1} \tilde{p}_k \alpha^{k-1}(k-1)p\alpha = p\sum_k (k-1)\tilde{p}_k \alpha^k, \\ R_{\alpha;p}^{\mathsf{E2}} &= \mathbb{E}(Y_2) = \sum_{k\ge 1} \tilde{p}_k \alpha^{k-1}(k-1)p = p\sum_k (k-1)\tilde{p}_k \alpha^{k-1}. \end{aligned} \tag{3.17}$$

Note that $R_{\alpha;p}^{\mathsf{E1}} = \alpha R_{\alpha;p}^{\mathsf{E2}} < R_{\alpha;p}^{\mathsf{E2}}$, which shows that, with the same number of vaccinations, E1 is a better strategy than E2. In particular, the critical critical vaccination coverage v_c is smaller for E1 than for E2. An intuitive

explanation to why E2 is not as efficient as E1 is that in E2 both individuals of selected friendships are vaccinated, and since an individual is partly protected by friends getting vaccinated the second vaccination is less "efficient".

We let $\tilde\pi_{\alpha;p}^{\mathsf{E1}}$ and $\tilde\pi_{\alpha;p}^{\mathsf{E2}}$ be the probabilities that the Galton–Watson processes with offspring distributions Y_1 and Y_2, respectively, starting with one individual, die out; they are thus the smallest positive solutions to $t = f_{Y_1}(t)$ and $t = f_{Y_2}(t)$, where f_{Y_1} and f_{Y_2} are the corresponding probability generating functions.

If we start with one unvaccinated person x with degree d, the number of open edges from x is $\mathrm{Bi}(d, p\alpha)$ for E1 and $\mathrm{Bi}(d, p)$ for E2, for the same reason as for the number of new edges above. The probability that the epidemic never takes off is thus $(1 - p\alpha + p\alpha\tilde\pi_{\alpha;p}^{\mathsf{E1}})^d$ for E1 and $(1 - p + p\tilde\pi_{\alpha;p}^{\mathsf{E2}})^d$ for E2.

If the initially infected is chosen randomly among the unvaccinated, we thus find the probabilities that the epidemic never takes off

$$\begin{aligned}
\pi_{\alpha;p}^{\mathsf{E1}} &= \frac{\sum_j p_j\alpha^j\bigl(1 - p\alpha(1 - \tilde\pi_{\alpha;p}^{\mathsf{E1}})\bigr)^j}{\sum_j p_j\alpha^j} = \frac{f_D\bigl(\alpha\bigl(1 - p\alpha(1 - \tilde\pi_{\alpha;p}^{\mathsf{E1}})\bigr)\bigr)}{f_D(\alpha)},\\
\pi_{\alpha;p}^{\mathsf{E2}} &= \frac{\sum_j p_j\alpha^j\bigl(1 - p(1 - \tilde\pi_{\alpha;p}^{\mathsf{E2}})\bigr)^j}{\sum_j p_j\alpha^j} = \frac{f_D\bigl(\alpha\bigl(1 - p(1 - \tilde\pi_{\alpha;p}^{\mathsf{E2}})\bigr)\bigr)}{f_D(\alpha)}.
\end{aligned} \tag{3.18}$$

We summarize our results as before, letting $Z_n^{\mathsf{E1}}(\alpha)$ and $Z_n^{\mathsf{E2}}(\alpha)$ denote the final numbers infected in the epidemic for the two strategies. Recall that $v(\alpha)$ is given by (3.16).

Theorem 3.4. *For $j = 1, 2$, $Z_n^{\mathsf{E}j}(\alpha)/\bigl((1 - v(\alpha))n\bigr) \xrightarrow{\mathrm{p}} 0$ if $R_{\alpha;p}^{\mathsf{E}j} \le 1$, where $R_{\alpha;p}^{\mathsf{E}j}$ is defined by (3.17). If $R_{\alpha;p}^{\mathsf{E}j} > 1$, then $Z_n^{\mathsf{E}j}(\alpha)/\bigl((1 - v(\alpha))n\bigr)$ converges to a two-point distribution $Z_{\alpha;p}^{\mathsf{E}j}$ for which $\mathbb{P}\bigl(Z_{\alpha;p}^{\mathsf{E}j} = 0\bigr) = \pi_{\alpha;p}^{\mathsf{E}j}$ and $\mathbb{P}\bigl(Z_{\alpha;p}^{\mathsf{E}j} = \tau_{\alpha;p}^{\mathsf{E}j}\bigr) = \tau_{\alpha;p}^{\mathsf{E}j}$, where $\pi_{\alpha;p}^{\mathsf{E}j}$ is defined by (3.18), and $\tau_{\alpha;p}^{\mathsf{E}j} = 1 - \pi_{\alpha;p}^{\mathsf{E}j}$.*

3.5. **Examples.** We now compare the performance of the different vaccination strategies on two examples. In the first example we have chosen the degree distribution to be Poisson distributed with mean $\lambda = 6$, and the transmission probability to equal $p = 0.5$. Using (3.2) we conclude that this implies that $R_0 = 3$. The assumption of Poisson distributed degree means that this applies to the simple $G(n, p = 6/n)$ graph with transmission probability $p = 0.5$; in the epidemic literature this model is knowns as the Reed-Frost model (e.g. [2]). In Figure 2 we show τ, the final proportion infected among unvaccinated in case of a major outbreak, as a function of the vaccination coverage v, for the 4 different vaccination strategies treated. It is seen that the acquaintance and edgewise E1 strategies perform best in the sense that, for a fixed proportion vaccinated, the proportion τ getting infected in case of a major outbreak is smallest for these two strategies. As a consequence, the critical vaccination coverage, $v_c = \inf_v\{v; R_v \le 1\}$, is also smallest for these two strategies. There is no unique ordering of the two

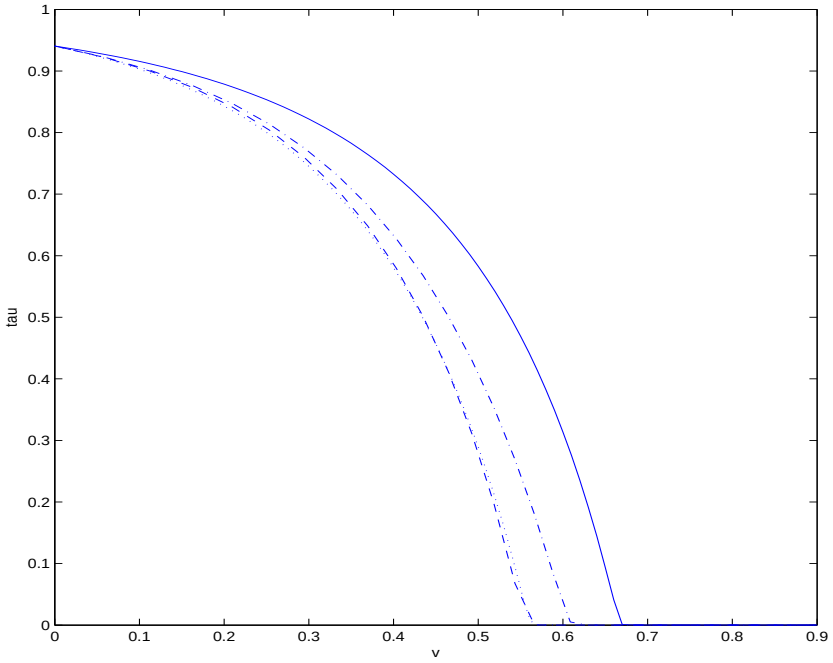

FIGURE 2. Final proportion infected τ as a function of the vaccination coverage v for four vaccination strategies: uniform (—), acquaintance ($\cdots$), E1 (- - -) and E2 ($- \cdot - \cdot -$). The degree distribution is Po(6) and transmission probability $p = 0.5$.

strategies – the acquaintance strategy is slightly better for small vaccination coverages and E1 is slightly better for higher vaccination coverages and hence also has slightly smaller v_c. The edgewise strategy E2 is not as good as these two strategies but still better than the uniform vaccination coverage. (Indeed, E2 is always less efficient than E1, see above.) Acquaintance, E1 and E2 all perform better than the uniform strategy, the reason being that they tend to find individuals with high degrees. For the parameter choices of this example, the critical vaccination coverages equal $v_c \approx 0.56$ for the acquaintence and E1 strategies, $v_c \approx 0.61$ for E2 and $v_c \approx 0.67$ for the uniform vaccination strategy.

In the second example (illustrated in Figure 3) we chose a more heavy tailed degree distribution having $p_d \propto d^{-3.5}$ (in the computations it was truncated at $d = 200$). The initial values were modified such that $E(D) \approx 6$ to make it more comparable to the previous example, with a resulting variance equal to 18.9. The transmission parameters was set $p = 0.5$ as before.

Using (3.2) we hence see that $R_0 \approx 4.1$. In the figure we see the same

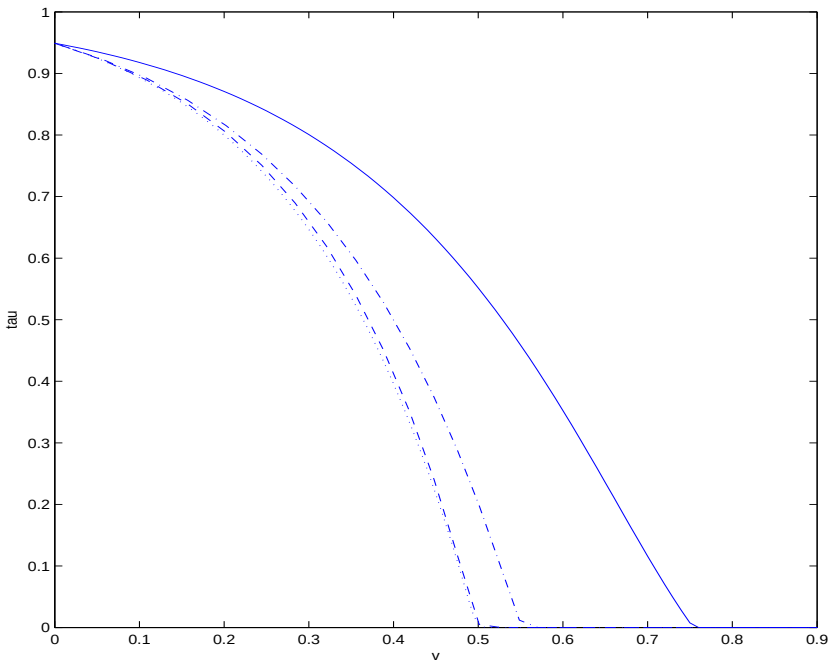

FIGURE 3. Final proportion infected as a function of the vaccination coverage for four vaccination strategies: uniform (—), acquaintance ($\cdots$), E1 (- - -) and E2 ($- \cdot - \cdot -$.) The degree distribution is heavy-tailed ($p_d \propto d^{-3.5}$) with mean $E(D) \approx 6$ and $p = 0.5$.

type of pattern as in the previous example. However, the difference between the strategies is more pronounced with $v_c \approx 0.50$ for the acquaintence and E1 strategies, $v_c \approx 0.55$ for E2 and $v_c \approx 0.75$ for the uniform vaccination strategy. In other words, if the uniform strategy is applied in these two examples we have to vaccinate *more* individuals if the degree distribution is heavy-tailed, but if any of the other strategies is performed, the heavy-tailed degree distribution require *less* vaccinations to surely prevent an outbreak. Another minor difference from the previous example is that, for the present heavy-tailed distribution, the acquaintance strategy is (slightly) better than E1 for all vaccination coverages and hence also has a smaller critical vaccination coverage. However, the difference between the two strategies is negligible.

Note that all τ's in both examples denote the proportion of infected among the unvaccinated (in case of an outbreak) and can hence be thought of as an

indirect protection from those getting vaccinated. Of course, by assumption, all vaccinated are also protected from getting infected.

4. Preliminaries on branching processes

As said above, our method is based on comparison with branching processes, more precisely Galton–Watson processes, see e.g. [3] for definitions and basic facts. If $\mathfrak{X}$ is a Galton–Watson process started with 1 initial particle, we let $\mathfrak{X}^d$ denote the same branching process with d initial particles, i.e. the union of d independent copies of $\mathfrak{X}$. Further, for any Galton–Watson process $\mathfrak{X}$, we let $|\mathfrak{X}|$ denote its total progeny, i.e. the total number of particles in all generations, and we let $\rho(\mathfrak{X})$ be the survival probability of $\mathfrak{X}$, i.e. $\rho(\mathfrak{X}) := \mathbb{P}(|\mathfrak{X}| = \infty)$. Note that if $\mathfrak{X}$ starts with 1 particle, then

$$\rho(\mathfrak{X}^d) = 1 - \bigl(1 - \rho(\mathfrak{X})\bigr)^d, \tag{4.1}$$

since $\mathfrak{X}^d$ dies out if and only if all d copies of $\mathfrak{X}$ in it do.

We will need the following simple continuity result, which presumably is well known although we have failed to find a reference.

Lemma 4.1. *Let X_ν and X be non-negative integer-valued random variables, and let $\mathfrak{X}_\nu^d$ and $\mathfrak{X}^d$ be the corresponding Galton–Watson processes with offspring distributions X_ν and X, starting with d particles. If $X_\nu \xrightarrow{\mathrm{d}} X$ as $n \to \infty$, and $\mathbb{P}(X = 1) < 1$, then $\rho(\mathfrak{X}_\nu^d) \to \rho(\mathfrak{X}^d)$, for every fixed $d \geq 0$.*

Proof. By (4.1), it suffices to show this for $d = 1$, and we then drop the superscript 1.

Consider the probability generating functions $f_X(t) := \mathbb{E}\, t^X$ and $f_{X_\nu}(t) := \mathbb{E}\, t^{X_\nu}$ for $0 \leq t \leq 1$. It is well-known, see e.g. [3, Theorem I.5.1], that the extinction probability $q := 1-\rho(\mathfrak{X})$ is the smallest root in $[0,1]$ of $f_X(q) = q$. It follows easily, since we have excluded the possibility $f_X(t) \equiv t$, that if $0 \leq t < q$, then $f_X(t) > t$, and if $q < t < 1$, then $f_X(t) < t$.

Since $X_\nu \xrightarrow{\mathrm{d}} X$, we have $f_{X_\nu}(t) \to f_X(t)$ for every $t \in [0,1]$. Hence, if $0 \leq t < q$, then $f_{X_\nu}(t) > t$ for large n, and thus $q_\nu := 1 - \rho(\mathfrak{X}_\nu) > t$. Similarly, if $q < t < 1$, then, for large n, $f_{X_\nu}(t) < t$ and thus $q_\nu < t$. It follows that $q_\nu \to q$ as $n \to \infty$. □

Remark 4.2. The case $\mathbb{P}(X = 1) = 1$, i.e. $X = 1$ a.s., really is an exception. If we let $X_\nu \sim \mathrm{Be}(1 - \nu^{-1})$, we have $X_\nu \xrightarrow{\mathrm{d}} X = 1$, but $\rho(\mathfrak{X}_\nu) = 0$ for every ν while $\rho(\mathfrak{X}) = 1$.

5. The giant component

Our ultimate goal is to describe the large component(s) of $G^*(n, (d_i)_1^n)_{\mathsf{v};p}$ and $G(n, (d_i)_1^n)_{\mathsf{v};p}$, where v is one of the vaccination strategies defined above. The basic strategy will be to relate the neighbourhoods of a vertex to a branching process. We do this for $G^*(n, (d_i)_1^n)$, which is technically easier to handle; as explained in Subsection 5.2, the results then transfer to

$G(n,(d_i)_1^n)$ too, provided Condition 2.3 holds. We first do the argument in detail in the simplest case, viz. $G^*(n,(d_i)_1^n)$ without edge deletion (i.e. $p=1$) or vaccination and prove our main results concerning the existence, size and uniqueness of the giant component. We use and adapt the method in Bollobás, Janson and Riordan [5] (for a different random graph model). This will provide a new proof of the results by Molloy and Reed [15, 16] (under our slightly weaker condition). We will then describe the modifications needed to make the results valid also when there is edge deletion or vaccination.

We say that an event holds *with high probability* (whp), if it holds with probability tending to 1 as $n\to\infty$. We shall use o_p in the standard way (see e.g. Janson, Łuczak and Ruciński [12]); for example, if (X_n) is a sequence of random variables, then $X_n=o_p(1)$ means that $X_n\xrightarrow{\mathrm{p}}0$. We shall often use the basic fact that, if $a\in\mathbb{R}$, then $X_n\xrightarrow{\mathrm{p}}a$ if and only if, for every $\varepsilon>0$, the relations $X_n>a-\varepsilon$ and $X_n<a+\varepsilon$ hold whp. All unspecified limits are taken as $n\to\infty$, while p and the vaccination parameters v or c are kept fixed.

We denote the orders of the components of a graph G by $C_1(G)\ge C_2(G)\ge\dots$, with $C_j(G)=0$ if G has fewer than j components. We let $N_k(G)$ denote the total number of vertices in components of order k, and write $N_{\ge k}(G)$ for $\sum_{j\ge k}N_j(G)$, the number of vertices in components of order at least k. Similarly, we let $N_{k,d}(G)$ and $N_{\ge k,d}(G)$ denote the number of such vertices that have degree d.

Remark 5.1. Our results are typically of the form $C_1(G_n)=\tau n+o_p(n)$ and $C_2(G_n)=o_p(n)$ for some number $\tau\ge0$ (or, equivalently, $C_1(G_n)/n\xrightarrow{\mathrm{p}}\tau$ and $C_2(G_n)/n\xrightarrow{\mathrm{p}}0$). Hence, if $\tau>0$, then there is exactly one "giant" component, and all other components are much smaller. In our epidemic setting, this means that if $\tau=0$, then every epidemic will be "small", i.e. $o(n)$, while if $\tau>0$, then the epidemic is large with probability τ (allowing the case that the initially infected person is vaccinated and thus never becomes ill), and in that case, a fraction τ of the population will be infected. (τ thus has a double role.)

5.1. $G^*(n,(d_i)_1^n)$**, with** $p=1$ **and no vaccination.** As said above, we will use a branching process approximation. The particles in the branching process correspond to free (not yet paired) half-edges. Note that there are jn_j half-edges belonging to vertices of degree j. Hence, a random half-edge shares a vertex with $j-1$ other half-edges with probability $jn_j/\sum_k kn_k$. By Condition 2.1, $jn_j/\sum_k kn_k\to jp_j/\mu$, and recall the definition of $\tilde p_j=jp_j/\mu$ defined in (3.1). Let $\mathfrak{X}$ be the Galton–Watson branching process starting with one particle and with the offspring distribution $(\tilde p_{j+1})_{j=0}^\infty$. (This is the distribution $(p_j)_j$ size-biased and shifted one step.) In other words, the offspring distribution is $\tilde D-1$, with $\tilde D$ as in Section 3.

We let $\rho = \rho(\mathfrak{X})$ denote the survival probability of $\mathfrak{X}$, and define

$$\tau := \sum_{d=1}^{\infty} p_d\bigl(1-(1-\rho)^d\bigr); \tag{5.1}$$

this is the survival probability for the branching process $\mathfrak{X}$ started with a random number of particles having the distribution $(p_d)_{d=0}^{\infty}$.

Consider a vertex x of degree d in $G^*(n,(d_i)_1^n)$. We explore the component containing x by a breadth-first search. We concentrate on the half-edges, so we begin by taking the d half-edges at x, and label them as *active*. We then process the active half-edges one by one as follows. We take an active half-edge, relabel it as *used*, and find the half-edge that it connects to and the corresponding vertex; this partner is chosen uniformly among all half-edges that are not yet used. We then label the partner as used and all other half-edges at the same vertex as active, provided that they are not already used (which would mean that we have found a cycle or a multiple edge). The active half-edges will behave essentially as a Galton–Watson process (where we reveal the children of the particles one by one), but the probability distribution of the children will vary slightly; it will depend on the numbers of vertices of different degrees that we already have found. Nevertheless, it is obvious that at each step in the beginning, the probability of $j-1$ new half-edges is close to $jn_j/\sum_k kn_k \approx \tilde{p}_j$.

To be more precise, first, let k be a fixed number, and consider the event that x belongs to a component with at least k vertices. This is almost the same as the probability that we will find at least $k-1$ active half-edges in the process just described. (This is not exact, because if we stop when we have found $k-1$ half-edges, some of these may connect back to vertices already found; the probability of this tends to 0, however, as $n\to\infty$.) The complementary event, that the process finds less than $k-1$ active half-edges, consists of a finite number of cases, where each case describes the sequence of new active half-edges found at each step. It is obvious that the probability of each of these cases converges, as $n\to\infty$, to the corresponding probability in $\mathfrak{X}^d$, and thus we find, for a vertex x of degree d, with $\mathcal{C}(x)$ denoting the corresponding component of $G^*(n,(d_i)_1^n)$,

$$\mathbb{P}(|\mathcal{C}(x)|\ge k) = \mathbb{P}(|\mathfrak{X}^d|\ge k-1)+o(1). \tag{5.2}$$

Recall that $N_{\ge k}$ is the number of vertices of degree d belonging to a component of size at least k. The expectation $\mathbb{E}\,N_{\ge k,d}$ equals n_d times the probability that a given vertex x of degree d satisfies $|\mathcal{C}(x)|\ge k$, and thus, by (5.2) and Condition 2.1(i), for every fixed $d\ge 0$ and $k\ge 1$,

$$\mathbb{E}(N_{\ge k,d}/n) \to p_d\,\mathbb{P}(|\mathfrak{X}^d|\ge k-1). \tag{5.3}$$

We next want to let $k\to\infty$ here. We thus, for the remainder of this section, assume that $\omega(n)$ is a function such that $\omega(n)\to\infty$ but $\omega(n)/n\to 0$ as $n\to\infty$. We regard components as *big* if they contain at least $\omega(n)$ vertices, and *small* otherwise. (The flexibility in the choice of $\omega(n)$ is useful,

but we will see that it does not matter much; the asymptotics we find do not depend on ω.)

Lemma 5.2. *If $\omega(n) \to \infty$ and $\omega(n)/n \to 0$, then,*

$$\mathbb{E}(N_{\ge\omega(n)}/n) \to \tau \tag{5.4}$$

and, for every fixed $d \ge 0$,

$$\mathbb{E}(N_{\ge\omega(n),d}/n) \to p_d\,\mathbb{P}(|\mathfrak{X}^d| = \infty) = p_d\bigl(1-(1-\rho)^d\bigr). \tag{5.5}$$

Proof. We begin with an upper bound in (5.5). For any fixed k, we have $\omega(n) > k$ for large n, and thus $N_{\ge\omega(n),d} \le N_{\ge k,d}$. Consequently, (5.3) yields

$$\limsup_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \le \limsup_{n\to\infty} \mathbb{E}(N_{\ge k,d}/n) = p_d\,\mathbb{P}(|\mathfrak{X}^d| \ge k-1). \tag{5.6}$$

As $k \to \infty$, the right hand side converges to $p_d\,\mathbb{P}(|\mathfrak{X}^d| = \infty)$, and we find

$$\limsup_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \le p_d\,\mathbb{P}(|\mathfrak{X}^d| = \infty). \tag{5.7}$$

For a lower bound, let $\nu \ge 1$ be fixed and let X_ν be a random variable taking values in $\{0,1,\dots,\nu\}$ with $\mathbb{P}(X_\nu = j) = (1-\nu^{-1})\tilde p_{j+1}$ for $1 \le j \le \nu$ (and a suitable value for $\mathbb{P}(X_\nu = 0)$ so that the sum becomes 1). Consider the breadth-first exploration process described above. As long as we have found less than $\omega(n)$ vertices, the number of new active half-edges at each step stochastically dominates X_ν, provided n is large enough, since the remaining number of vertices of degree $j+1$ is $n_{j+1} - o(n) = p_{j+1}n - o(n) \ge (1-\nu^{-1})p_{j+1}n$ for n large. (If $p_{j+1} = 0$, the result is trivial.) Consequently, letting $\mathfrak{X}^d_\nu$ be the Galton–Watson process with d initial particles and the number of children distributed as X_ν, if n is large enough, we can couple the exploration process and $\mathfrak{X}^d_\nu$ such that as long as we have found less than $\omega(n)$ vertices, the number of active half-edges is at least the number of active particles in $\mathfrak{X}^d_\nu$ (i.e., the particles whose children have not yet been revealed.) In particular, if the exploration process stops before $\omega(n)$ vertices are found, then $\mathfrak{X}^d_\nu$ stops, and thus the probability that a vertex x of degree d satisfies $|\mathcal{C}(x)| < \omega(n)$ is at most $\mathbb{P}(|\mathfrak{X}^d_\nu| < \infty)$. Consequently, for large n,

$$\mathbb{E}\, N_{\ge\omega(n),d} \ge n_d\,\mathbb{P}(|\mathfrak{X}^d_\nu| = \infty) \tag{5.8}$$

and thus

$$\liminf_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \ge p_d\,\mathbb{P}(|\mathfrak{X}^d_\nu| = \infty). \tag{5.9}$$

Now let $\nu \to \infty$. Then $X_\nu \xrightarrow{\mathrm{d}} X$, where X has the distribution $\mathbb{P}(X = j) = \tilde p_{j+1}$, and thus, by Lemma 4.1, $\mathbb{P}(|\mathfrak{X}^d_\nu| = \infty) \to \mathbb{P}(|\mathfrak{X}^d| = \infty)$. Consequently,

$$\liminf_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \ge p_d\,\mathbb{P}(|\mathfrak{X}^d| = \infty),$$

which together with (5.7) and (4.1) yields (5.5).

Finally, noting that $N_{\ge\omega(n),d} \le n_d$, it follows easily from the uniform summability in (2.1) that we can sum (5.5) over d and take the limit outside the sum, i.e.

$$\mathbb{E}(N_{\ge\omega(n)}/n) = \sum_d \mathbb{E}(N_{\ge\omega(n),d}/n) \to \sum_d p_d\bigl(1-(1-\rho)^d\bigr) = \tau.$$

□

Note that the limits do not depend on the choice of $\omega(n)$. Hence, it follows that the expected number of vertices belonging to components of size between, say, $\log n$ and $n^{0.99}$ is $o(n)$.

We next show that we have convergence not only of the expectations but also of the random variables in (5.4) and (5.5), i.e. that these random variables are concentrated close to their expectations.

Lemma 5.3. *If $\omega(n) \to \infty$ and $\omega(n)/n \to 0$, then,*

$$N_{\ge\omega(n)}/n \xrightarrow{\mathrm{p}} \tau \tag{5.10}$$

and, for every fixed $d \ge 0$,

$$N_{\ge\omega(n),d}/n \xrightarrow{\mathrm{p}} p_d\bigl(1-(1-\rho)^d\bigr). \tag{5.11}$$

Proof. Start with two distinct vertices x and z of the same degree d and explore their components as above. We can repeat the arguments above, and find

$$\mathbb{P}\bigl(|\mathcal{C}(x)|<k,\,|\mathcal{C}(y)|<k\bigr) = \mathbb{P}(|\mathfrak{X}^d|<k-1)^2 + o(1)$$

and thus, using (5.2),

$$\mathbb{P}\bigl(|\mathcal{C}(x)|\ge k,\,|\mathcal{C}(y)|\ge k\bigr) = \mathbb{P}(|\mathfrak{X}^d|\ge k-1)^2 + o(1).$$

Multiplying with the number $n_d(n_d-1)$ of pairs (x,y) of the same degree d, and noting that the number of such pairs where both x and z belong to components of size $\ge k$ (the same or not) is $N_{\ge k,d}(N_{\ge k,d}-1)$, we find

$$\mathbb{E}(N_{\ge k,d}^2/n^2) = \mathbb{E}(N_{\ge k,d}(N_{\ge k,d}-1)/n^2) + O(1/n) \to p_d^2\,\mathbb{P}(|\mathfrak{X}^d|\ge k-1)^2.$$

Hence, $\limsup_{n\to\infty}\mathbb{E}(N_{\ge\omega(n),d}^2/n^2) \le p_d^2\,\mathbb{P}(|\mathfrak{X}^d|\ge k-1)^2$ for every k, and thus

$$\limsup_{n\to\infty}\mathbb{E}(N_{\ge\omega(n),d}^2/n^2) \le p_d^2\,\mathbb{P}(|\mathfrak{X}^d|=\infty)^2.$$

Since, by the Cauchy–Schwarz inequality and (5.5), further

$$\mathbb{E}(N_{\ge\omega(n),d}^2/n^2) \ge (\mathbb{E}(N_{\ge\omega(n),d}/n))^2 \to p_d^2\,\mathbb{P}(|\mathfrak{X}^d|=\infty)^2,$$

it follows that

$$\mathbb{E}(N_{\ge\omega(n),d}^2/n^2) \to p_d^2\,\mathbb{P}(|\mathfrak{X}_\nu^d|=\infty)^2.$$

This and (5.5) show that

$$\mathrm{Var}\bigl(N_{\ge\omega(n),d}/n\bigr) \to 0,$$

and thus

$$\bigl(N_{\ge\omega(n),d} - \mathbb{E}(N_{\ge\omega(n),d})\bigr)/n \xrightarrow{\mathrm{p}} 0,$$

which by (5.5) implies (5.11).

Finally, again we can sum over d because of (2.1); this yields (5.10). □

Theorem 5.4. *Assume that Condition 2.1 holds. Then*

$$C_1(G^*(n,(d_i)_1^n)) = \tau n + o_p(n),$$
$$C_2(G^*(n,(d_i)_1^n)) = o_p(n).$$

Proof. We have already shown that roughly τn vertices lie in big components. It remains to show that most of them belong to the same component. We write $G_n = G^*(n,(d_i)_1^n)$.

First, if $C_1(G_n) \ge \omega(n)$, then $N_{\ge\omega(n)}(G_n) \ge C_1(G_n)$. Thus, for every $\varepsilon > 0$ and n so large that $\omega(n) < \varepsilon n$, we have by Lemma 5.3

$$\mathbb{P}\big(C_1(G_n) > \tau n + \varepsilon n\big) \le \mathbb{P}\big(N_{\ge\omega(n)}(G_n) > \tau n + \varepsilon n\big) \to 0. \tag{5.12}$$

This completes the proof if $\tau = 0$.

In the sequel we assume $\tau > 0$ and show a corresponding estimate from below. First, if $p_d = 0$ for every $d \ge 2$, then $\tilde{p}_{j+1} = 0$ for all $j \ge 1$, so $\mathfrak{X}$ dies immediately and $\rho = 0$ and $\tau = 0$. Hence $p_d > 0$ for some $d \ge 2$. We fix such a d for the remainder of the proof, and fix δ with $0 < \delta < 1/2$. Further, take (rather arbitrarily) $\omega(n) = n^{0.9}$.

We assume in the sequel that n is so large that $n_d > n^{1-\delta}$. We then split the $n^{1-\delta}$ first of the vertices of degree d in G_n into d vertices of degree 1 each; we colour these $dn^{1-\delta}$ new vertices *red*. (To be precise, we should round $n^{1-\delta}$ to an integer.) We denote the resulting graph by G_n'; note that G_n' is a random multigraph $G^*(n',(d_i'))$ where n_j', the number of vertices of degree j, is given by $n_d' = n_d - n^{1-\delta}$, $n_1' = n_1 + dn^{1-\delta}$, and $n_j' = n_j$ for $j \ne 1, d$. Note that the total number of vertices in G_n' is $n' := n + (d-1)n^{1-\delta} = n + o(n)$, and that (d_i') satisfies Condition 2.1 with the same (p_j) (except that n is replaced by n', which only makes a notational difference). Consequently, our results above apply to G_n' too.

By symmetry, we may assume that the $dn^{1-\delta}$ red vertices in G_n' are chosen at random among all vertices of degree 1, and that G_n is obtained by partitioning the red vertices at random into groups with d vertices and then coalescing each group into one vertex.

During the exploration of the component $\mathcal{C}'(x)$ in G_n' containing a vertex x, in each step, the active half-edge is paired with the single half-edge leading to a red vertex with probability at least $c_1 n^{-\delta}$, for some $c_1 > 0$, unless at least $n^{1-\delta}$ red vertices already have been found. Consequently, if the component $\mathcal{C}'(x)$ has at least $\omega(n)$ vertices, the number of red vertices stochastically dominates $\min\big(n^{1-\delta}, \mathrm{Bi}(\omega(n)-1, c_1 n^{-\delta})\big)$. A Chernoff bound, see e.g. [12, Corollary 2.3], shows that the probability that $\mathcal{C}'(x)$ has at least $\omega(n)$ vertices but less than $c_2 n^{-\delta}\omega(n) = c_2 n^{0.9-\delta}$ red vertices is at most $\exp(-c_3 n^{0.9-\delta}) = o(n^{-1})$, for $c_2 = c_1/2$ and some $c_3 > 0$. Summing over all x, we see that whp, *every* big component of G_n' contains at least $c_2 n^{0.9-\delta}$ red vertices.

Assume that this holds, and consider two big components K_1 and K_2 in G'_n. We can construct the random partition of the red vertices by taking first the red vertices in K_1 one by one, unless already used, and randomly selecting $d-1$ partners. We thus do this at least $m := c_2 n^{0.9-\delta}/d$ times, and each time the probability of not including a red vertex in K_2 is at most $1 - c_2 n^{0.9-\delta}/(dn^{1-\delta}) = 1 - c_4 n^{-0.1}$, with $c_4 = c_2/d$. Consequently, the probability of not joining K_1 and K_2 in the coalescing phase is at most

$$\exp(-mc_2 n^{-0.1}) = \exp(-c_4^2 n^{0.8-\delta}) = o(n^{-2}).$$

Since there are at most $(n')^2 = O(n^2)$ such pairs K_1 and K_2, we see that whp all big components in G'_n are connected in G_n. Hence, if B' is the union of all big components in G'_n, and B is the corresponding set of vertices in G_n, we see that whp B is connected in G_n, and, using Lemma 5.3 for G'_n,

$$C_1(G_n) \geq |B| \geq |B'| - (d-1)n^{1-\delta} = \tau n' + o_p(n) = \tau n + o_p(n). \quad (5.13)$$

Combining (5.13) and (5.12) we obtain $C_1(G_n) = \tau n + o_p(n)$.

Finally, we observe that if $C_2(G_n) \geq \omega(n)$, then $N_{\geq\omega(n)}(G_n) \geq C_1(G_n) + C_2(G_n)$, and thus, by (5.10) and (5.13),

$$C_2(G_n) \leq \max\bigl(\omega(n), N_{\geq\omega(n)}(G_n) - C_1(G_n)\bigr) = o_p(n). \qquad \square$$

5.2. **The simple random graph $G(n,(d_i)_1^n)$.** We transfer the results to the simple random graph $G(n,(d_i)_1^n)$ by the following result proved in [9]; see also e.g. Bollobás [4] and McKay [14] for earlier versions.

Lemma 5.5. *If Conditions 2.1 and 2.3 hold, then*

$$\liminf_{n\to\infty} \mathbb{P}\bigl(G^*(n,(d_i)_1^n) \text{ is a simple graph}\bigr) > 0.$$

All results for $G^*(n,(d_i)_1^n)$ that can be stated in terms of convergence in probability, as our results in this section, thus hold also if we condition on the graph being simple. In other words, the results proved for $G^*(n,(d_i)_1^n)$ hold for $G(n,(d_i)_1^n)$ too. Thus, Theorem 5.4 has the following version for $G(n,(d_i)_1^n)$.

Theorem 5.6. *Assume that Conditions 2.1 and 2.3 hold. Then*

$$C_1(G(n,(d_i)_1^n)) = \tau n + o_p(n),$$
$$C_2(G(n,(d_i)_1^n)) = o_p(n).$$

5.3. **Uniform vaccination.** We now extend Theorem 5.4 to the graph $G^*(n,(d_i)_1^n)_{v;p}^{\mathsf{U}}$ where $0 \leq v < 1$ and $0 < p \leq 1$, see Section 2. Recall that we obtain this graph from $G^*(n,(d_i)_1^n)$ by randomly and independently deleting edges with probability $1-p$ (non-transmission) and vertices with probability v (vaccination). The branching process approximation arguments above still work, with the difference that each new individual found is kept with probability $p(1-v)$, and otherwise discarded. Hence the offspring distribution is changed from $\tilde{D}-1$ to $\tilde{X}_v \sim \mathrm{MixBi}(\tilde{D}-1, p(1-v))$, and the branching process corresponding to an unvaccinated person with d friends starts

with $\mathrm{Bi}(d, p(1-v))$ individuals. Let now $\mathfrak{X}^d$ denote the branching process with this offspring distribution, starting with d individuals. The probability generating function of $\tilde{X}_v$ is, as shown in Subsections 3.1 and 3.2, given by

$$\mathbb{E}\, t^{\tilde{X}_v} = \frac{f_D'\big(1-p(1-v)(1-t)\big)}{f_D'(1)}.$$

Hence, the extinction probability of $\mathfrak{X}^1$ is $\tilde{\pi}_{v;p}^{\mathsf{U}}$ given by (3.6). If we start the branching process with $D' \sim \mathrm{Bi}(d, p(1-v))$ individuals, the extinction probability is thus, writing $\bar{p} = p(1-v)$,

$$\pi^{(d)} := \sum_k \binom{d}{k} \bar{p}^k (1-\bar{p})^{d-k} (\tilde{\pi}_{v;p}^{\mathsf{U}})^k = \big(1-\bar{p}+\bar{p}\tilde{\pi}_{v;p}^{\mathsf{U}}\big)^d.$$

The arguments in the proofs of Lemmas 5.2 and 5.3 show, recalling that each vertex has probability $1-v$ of being unvaccinated, that (5.11) holds in the form

$$N_{\ge\omega(n),d}/n \xrightarrow{\mathrm{p}} p_d(1-v)\big(1-\pi^{(d)}\big),$$

for every fixed $d \ge 0$, assuming $\omega(n) \to \infty$ and $\omega(n)/n \to 0$. Hence,

$$N_{\ge\omega(n)}/(n(1-v)) \xrightarrow{\mathrm{p}} \sum_d p_d(1-\pi^{(d)}) = 1-\sum_d p_d\pi^{(d)} = 1-f_D(1-\bar{p}+\bar{p}\tilde{\pi}_{v;p}^{\mathsf{U}}).$$

This limit equals $\tau_{v;p}^{\mathsf{U}} := 1-\pi_{v;p}^{\mathsf{U}}$ with $\pi_{v;p}^{\mathsf{U}}$ given by (3.7).

To extend Theorem 5.4, it remains to show that there is only one very large component. More precisely, we show again that, with $\omega(n) = n^{0.9}$, there is whp only one big component. We argue as for Theorem 5.4, splitting some vertices of degree d in $G_n = G^*(n, (d_i)_1^n)$ into d red vertices of degree 1, calling the resulting graph G_n'.

We vaccinate the vertices in G_n' with probability v each, independently; we then recombine the red vertices to vertices of degree d in G_n and consider each such vertex as vaccinated if at least one of its red parts in G_n' is. This means that some vertices in G_n are vaccinated with probability larger than v, but this does not hurt since the aim of the argument is to provide a lower bound for C_1, the size of the largest component, and any extra vaccinations can only decrease C_1.

By a Chernoff bound, there are whp at least $(1-v)n^{1-\delta}$ unvaccinated red vertices, and it follows as before that whp every big component of $(G_n')_{v;p}^{\mathsf{U}}$ contains at least $c_2 n^{0.9-\delta}$ red vertices (although the value of c_2 may change). Given two big components K_1 and K_2 it follows similarly as before that with probability $1-o(n^{-2})$ there exists a vertex in G_n that is split into d red vertices, of which at least one is in K_1, at least one in K_2, and all are unvaccinated. The proof is completed as before.

Consequently, using also Lemma 5.5, we have the following theorem. Theorems 3.2 and 3.1 (the special case $v = 0$) are immediate consequences.

Theorem 5.7. *Assume Condition 2.1, and let* $0 < p \le 1$, $0 \le v < 1$. *Then,*

$$C_1\big(G^*(n,(d_i)_1^n)^{\mathsf{U}}_{v;p}\big) = \tau^{\mathsf{U}}_{v;p} n(1-v) + o_p(n),$$
$$C_2\big(G^*(n,(d_i)_1^n)^{\mathsf{U}}_{v;p}\big) = o_p(n),$$

where $\tau^{\mathsf{U}}_{v;p} = 1 - \pi^{\mathsf{U}}_{v;p}$ *with* $\pi^{\mathsf{U}}_{v;p}$ *given by* (3.7). *If also Condition 2.3 holds, then the same results hold for* $G(n,(d_i)_1^n)^{\mathsf{U}}_{v;p}$ *too.*

5.4. **Acquaintance vaccination.** As explaind in Subsection 3.3, in order to obtain (asymptotically) a Galton–Watson branching process, with the right independence properties, we consider directed edges, or equivalently half-edges, that are *open*, i.e. transmission may take place but the edge is not used for vaccination. Moreover, we consider only open edges originating at an unvaccinated person.

Let x be a given vertex with degree d in $G^*(n,(d_i)_1^n)$, and let us explore the component of x in $G^*(n,(d_i)_1^n)^{\mathsf{A}}_{c;p}$, conditioned on x being unvaccinated (otherwise x does not belong to $G^*(n,(d_i)_1^n)^{\mathsf{A}}_{c;p}$). In order to be kept in $G^*(n,(d_i)_1^n)^{\mathsf{A}}_{c;p}$, an edge has to be open, but not all edges are kept since some may lead to vertices that are vaccinated, see Figure 1c). Nevertheless, we consider all open edges found during the exploration. We declare the open edges starting at x to be *active*. We then investigate the active edges. If an active edge leads to a person that is unvaccinated, we declare the open edges going from that person, except the one going back to where we just came from, to be new active edges. We continue until no more active edges are found; we then have found the component containg x (plus some extra open edges leading to vaccinated persons).

We investigate this process probabilistically, revealing the structure of $G^*(n,(d_i)_1^n)$ by combining half-edges at random as we proceed the exploration. We consider asymptotics as $n \to \infty$, and some of the statements below are only approximatively correct for finite n.

Note first that each of the d edges leading from x is open with probability $pe^{-c/d}$, independently of each other, so we start with $\mathrm{Bi}(d, pe^{-c/d})$ open edges.

The vertex x has d friends; in $G^*(n,(d_i)_1^n)$ they are chosen by randomly choosing d half-edges and their degrees have the size-biased distribution $(\tilde{p}_j)$, independently of each other. Conditioning on x being unvaccinated means that we condition on none of the d edges being used for vaccination in the opposite direction. Since the probability that a friend with degree j does not name x is $e^{-c/j}$, this preserves the independence of the degrees of the friends, but shifts their distribution to, as asserted in (3.11), $(\tilde{p}_j e^{-c/j}/\alpha)_j$, where $\alpha = \alpha(c) = \sum_j \tilde{p}_j e^{-c/j}$ as in (3.8) is the probability of not being named by a random friend.

Now suppose that an open edge goes from x to a friend z of degree k. In order for this to define an edge in $G^*(n,(d_i)_1^n)^{\mathsf{A}}_{c;p}$, z must not be vaccinated through another of its friends; this has the probability α^{k-1}. In this case, z

has $k-1$ further edges, and each of them is open with probability $pe^{-c/k}$. It follows that the number of new open edges at z has a distribution that is the mixture $(1-\alpha^{k-1})\delta_0 + \alpha^{k-1}\operatorname{Bi}(k-1, pe^{-c/k})$. Using the distribution (3.11) for the degree of z, we finally see that the distribution of the number Y of new active edges found when exploring a single active edge is given by (3.12).

Hence, observing obvious independence properties, the process of active edges is (asymptotically) a Galton-Watson branching process with offspring distribution Y, starting with $\operatorname{Bi}(d, pe^{-c/d})$ active edges. Denote his branching process by $\mathfrak{X}^{(d)}$. Let, as in Subsection 3.3, $\tilde{\pi}^{\mathsf{A}}_{c;p}$ by the probability that a branching proess with this offspring distribution Y and starting with a single individual dies out. Then, the extinction probability of $\mathfrak{X}^{(d)}$ is

$$\begin{aligned}\pi^{(d)} &:= \mathbb{P}(|\mathfrak{X}^{(d)}| < \infty) = \sum_{j=0}^{d} \binom{d}{j} (pe^{-c/d})^j (1-pe^{-c/d})^{d-j} (\tilde{\pi}^{\mathsf{A}}_{c;p})^j \\ &= (1 - pe^{-c/d} + pe^{-c/d}\tilde{\pi}^{\mathsf{A}}_{c;p})^d.\end{aligned}$$

A minor complication is that the branching process approximation counts open edges and, as remarked above, not all open edges lead to vertices in $G^*(n, (d_i)_1^n)^{\mathsf{A}}_{c;p}$. Thus (5.2) does not extend directly. However, we still have the inequality

$$\mathbb{P}(|\mathcal{C}(x)| \ge k) \le \mathbb{P}(|\mathfrak{X}^{(d)}| \ge k-1) + o(1).$$

Furthermore, a vertex of degree d in $G^*(n, (d_i)_1^n)$ is unvaccinated with probabilty α^d, and thus

$$\mathbb{E}(N_{\ge k,d}) \le n_d \alpha^p \big(\mathbb{P}(|\mathfrak{X}^{(d)}| \ge k-1) + o(1)\big),$$

which arguing as in (5.6) and (5.7) leads to

$$\limsup_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \le p_d \alpha^p \,\mathbb{P}(|\mathfrak{X}^{(d)}| = \infty) = p_d \alpha^p (1-\pi^{(d)}). \tag{5.14}$$

For a lower bound, we note that an open edge creates new open edges in the exploration process only if it leads to an unvaccinated person. Hence, if $f(\mathfrak{X}^{(d)})$ denotes the number of individuals in the branching process $\mathfrak{X}^{(d)}$ with at least one child, we have, for every $k \ge 1$,

$$\mathbb{P}(|\mathcal{C}(x)| \ge k) \ge \mathbb{P}(f(\mathfrak{X}^{(d)}) \ge k-1) + o(1).$$

In order to replace the fixed k by $\omega(n)$, we do as in the proof of Lemma 5.2 and define a Galton–Watson process $\mathfrak{X}^{(d)}_\nu$, now starting with $\operatorname{Bi}(d, pe^{-c/d}(1-\nu^{-1}))$ individuals and with an offspring distribution Y_ν on $\{0,\dots,\nu\}$ with $\mathbb{P}(Y_\nu = j) = (1-\nu^{-1})\,\mathbb{P}(Y=j)$ for $j = 1,\dots,\nu$.

For each ν and each fixed $A < \infty$, we can for large n couple the exploration process and $\mathfrak{X}^{(d)}_\nu$ as in the proof of Lemma 5.2 as long as we have found at most $A\omega(n)$ open edges. Hence, if $|\mathcal{C}(x)| < \omega(n)$, then either $f(\mathfrak{X}^{(d)}_\nu) < \omega(n)$ or the process $\mathfrak{X}^{(d)}_\nu$ reaches more than $A\omega(n)$ individuals while less than

$\omega(n)$ of them, plus the root, have had children. The probability of the latter event is at most, since the root has at most d children,

$$\mathbb{P}\Big(1+d+\sum_{i=1}^{\omega(n)} Y^*_{\nu,i} > A\omega(n)\Big),$$

where $Y^*_{\nu,i}$ are independent random variables with the distribution $\mathcal{L}(Y \mid Y>0)$, and thus this probability tends to 0 by the law of large numbers provided we have chosen $A > \mathbb{E}(Y \mid Y>0)$.

Consequently,

$$\mathbb{P}\big(|\mathcal{C}(x)| < \omega(n)\big) \le \mathbb{P}\big(f(\mathfrak{X}^{(d)}_\nu) < \omega(n)\big) + o(1) \le \mathbb{P}\big(f(\mathfrak{X}^{(d)}_\nu) < \infty\big) + o(1).$$

Using again that a person with degree d is unvaccinated with probability α^d, it follows that

$$\mathbb{E}\, N_{\ge\omega(n),d} \ge n_d\alpha^p\big(1-\mathbb{P}(|\mathfrak{X}^{(d)}_\nu|<\infty)+o(1)\big)$$

and thus

$$\liminf_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \ge p_d\alpha^p\,\mathbb{P}(|\mathfrak{X}^{(d)}_\nu|=\infty).$$

We let $\nu\to\infty$ and obtain by Lemma 4.1

$$\liminf_{n\to\infty} \mathbb{E}(N_{\ge\omega(n),d}/n) \ge p_d\alpha^p\,\mathbb{P}(|\mathfrak{X}^{(d)}|=\infty) = p_d\alpha^p(1-\pi^{(d)}),$$

which together with (5.14) yields

$$\mathbb{E}(N_{\ge\omega(n),d}/n) \to p_d\alpha^p(1-\pi^{(d)}).$$

Arguing as in the proof of Lemma 5.3, we find also

$$N_{\ge\omega(n),d}/n \xrightarrow{\mathrm{p}} p_d\alpha^p(1-\pi^{(d)})$$

and, recalling (3.15) and (3.9),

$$N_{\ge\omega(n)}/n \xrightarrow{\mathrm{p}} \sum_d p_d\alpha^p(1-\pi^{(d)}) = \sum_d p_d\alpha^p(1-\pi^{\mathsf{A}}_{c;p}) = (1-v(c))\tau^{\mathsf{A}}_{c;p},$$

with $\tau^{\mathsf{A}}_{c;p} = 1-\pi^{\mathsf{A}}_{c;p}$. In particular,

$$C_1(G^*(n,(d_i)_1^n)^{\mathsf{A}}_{c;p}) \le \omega(n) + N_{\ge\omega(n)} \le (1-v(c))\tau^{\mathsf{A}}_{c;p}n + o_p(n).$$

Finally, we argue again as in the proof of Theorem 5.4 to show that most vertices in large components belong to a single component. We split some of the vertices in $G_n = G^*(n,(d_i)_1^n)$ as above and perform acquaintance vaccination on the resulting graph G'_n. This corresponds to acquaintance vaccination on G_n, except that the vertices that are split now are asked to name a friend $\mathrm{Po}(dc)$ times instead of $\mathrm{Po}(c)$. We perform thus some extra vaccinations, but this can only decrease C_1 and we obtain as in (5.13) the lower bound

$$C_1(G^*(n,(d_i)_1^n)^{\mathsf{A}}_{c;p}) \ge (1-v(c))\tau^{\mathsf{A}}_{c;p}n + o_p(n).$$

Summing up, and using Lemma 5.5, we have shown he following theorem. Theorem 3.3 is an immediate consequence.

Theorem 5.8. *Assume Condition 2.1, and let* $0 < p \le 1$, $0 \le c < \infty$. *Then,*

$$C_1\bigl(G^*(n,(d_i)_1^n)_{c;p}^{\mathsf{A}}\bigr) = \tau_{c;p}^{\mathsf{A}} n(1-v(c)) + o_p(n),$$
$$C_2\bigl(G^*(n,(d_i)_1^n)_{c;p}^{\mathsf{A}}\bigr) = o_p(n),$$

where $\tau_{c;p}^{\mathsf{A}} = 1 - \pi_{c;p}^{\mathsf{A}}$ *with* $\pi_{c;p}^{\mathsf{A}}$ *given by* (3.15). *If also Condition 2.3 holds, then the same results hold for* $G(n,(d_i)_1^n)_{c;p}^{\mathsf{A}}$ *too.*

5.5. **Edgewise vaccination.** We argue as for acquaintance vaccination with the modifications (simplifications) explained in Subsection 3.4. There are no new complications, and we obtain the following. Theorem 3.4 is an immediate consequence.

Theorem 5.9. *Assume Condition 2.1, and let* $0 < p \le 1$, $0 < \alpha \le 1$. *Then, for* $j = 1, 2$,

$$C_1\bigl(G^*(n,(d_i)_1^n)_{\alpha;p}^{\mathsf{E}j}\bigr) = \tau_{\alpha;p}^{\mathsf{E}j} n(1-v(\alpha)) + o_p(n),$$
$$C_2\bigl(G^*(n,(d_i)_1^n)_{\alpha;p}^{\mathsf{E}j}\bigr) = o_p(n),$$

where $\tau_{\alpha;p}^{\mathsf{E}j} = 1 - \pi_{\alpha;p}^{\mathsf{E}j}$ *with* $\pi_{\alpha;p}^{\mathsf{E}j}$ *given by* (3.18). *If also Condition 2.3 holds, then the same results hold for* $G(n,(d_i)_1^n)_{\alpha;p}^{\mathsf{E}j}$ *too.*

Acknowledgements

We thank Mathias Lindholm for help in producing the figures. T.B. gratefully acknowledges financial support from the Swedish Research Council.

References

[1] H. Andersson, Epidemic models and social networks. *Math. Scientist* **24** (1999), 128–147.
[2] H. Andersson and T. Britton, *Stochastic Epidemic Models and their Statistical Analysis*. Springer Lecture Notes in Statistics, 151, Springer, New York, 2000.
[3] K.B. Athreya and P.E. Ney, *Branching Processes*. Springer, Berlin, 1972.
[4] B. Bollobás, *Random Graphs*. 2nd ed., Cambridge Univ. Press, Cambridge, 2001.
[5] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs. *Random Struct. Alg.* **31** (2007), 3–122.
[6] T. Britton, M. Deijfen and A. Martin-Löf, Generating simple random graphs with prescribed degree distribution, *J. Statist. Phys.*, to appear.
[7] R. Cohen, S. Havlin and D. ben-Avrahan, Efficient immunization strategies for computer networks and populations. *Phys. Rev. Let.* **91** (2003), 247901.
[8] A. Gut, *Probability: A Graduate Course*. Springer, New York, 2005.
[9] S. Janson, The probability that a random multigraph is simple. `http://arxiv.org/math.CO/0609802`
[10] S. Janson, Asymptotic equivalence and contiguity of some random graphs. In preparation.
[11] S. Janson, D. Knuth, T. Łuczak and B. Pittel, The birth of the giant component. *Random Struct. Alg.* **4** (1994), 231–358.
[12] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*. Wiley, New York, 2000.
[13] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[14] B. D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combin.* **19** A (1985), 15–25.
[15] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence. *Random Struct. Alg.* **6** (1995), 161–179.
[16] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.* **7** (1998), 295–305.
[17] C. Moore and M.E.J. Newman, Epidemics and percolation in small world networks. *Phys. Rev. E* **61** (2000), 5678–5682.
[18] M.E.J. Newman, The structure and function of complex networks. *SIAM Rev.* **45** (2003), 167–256.
[19] M.E.J. Newman, S.H. Strogatz and J. Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64** (2001), 026118.
[20] J. Scott, *Social Network Analysis, A Handbook*. 2nd ed., Sage, London, 2000.

Tom Britton, Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden.
E-mail address: tom.britton@math.su.se

Svante Janson, Department of Mathematics, Uppsala University, P.O. Box 480, SE-75106 Uppsala, Sweden.
E-mail address: svante.janson@math.uu.se

Anders Martin-Löf, Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden.
E-mail address: andersml@math.su.se