THRESHOLDS FOR VIRUS SPREAD ON NETWORKS

By Moez Draief, Ayalvadi Ganesh and Laurent Massoulié

Imperial College London, *Microsoft Research and Thomson Research*

We study how the spread of computer viruses, worms and other self-replicating malware is affected by the logical topology of the network over which they propagate. We consider a model in which each host can be in one of 3 possible states—susceptible, infected or removed (cured and no longer susceptible to infection). We characterize how the size of the population that eventually becomes infected depends on the network topology. Specifically, we show that if the ratio of cure to infection rates is larger than the spectral radius of the graph, and the initial infected population is small, then the final infected population is also small in a sense that can be made precise. Conversely, if this ratio is smaller than the spectral radius, then we show in some graph models of practical interest (including power law random graphs) that the average size of the final infected population is large. These results yield insights into what the critical parameters are in determining virus spread in networks.

1. Introduction. Computer viruses and worms are self-replicating pieces of code that propagate in a network. The essential difference between them is that a virus typically needs some form of human intervention, such as opening an email attachment or executing some software, to cause them to be replicated, whereas worms do not require such intervention. They use a number of different methods to identify new targets for infection; for example, many worms scan randomly generated IP addresses to locate vulnerable hosts, while email viruses send copies of themselves to all addresses in the address book of the victim. A survey of techniques for target location can be found in [25].

The particular mechanism chosen by a worm or virus to propagate itself defines a topology over which the infection can potentially spread. What impact does the topology have on the speed of spread of the epidemic and, moreover, what are the key features of the topology that determine how virulent the epidemic is? These are the questions that we address in this paper.

In this paper we use a susceptible-infected-removed (SIR) model to describe the spread of the epidemic. Here, each susceptible node can be infected by its infected neighbors at a rate proportional to their number, and remains infected for a deterministic or random time until it is removed. While it is infected, it has

the potential to infect its neighbors. Removal can correspond to either (i) patching the computer represented by the node, or (ii) its disconnection from the network by some quarantining mechanism, or (iii) the exhaustion of its infectious period either by a time-out mechanism or because it has tried all its neighbors. Once a node is removed, it cannot become susceptible or infected again. Our model ignores the possibility that susceptible nodes can also be removed, for example, because they have received a patch or virus signature conferring immunity. This is justified if the timescale for patching of susceptible hosts is much larger (happens much more slowly) than that of epidemic spread.

In the context of worms, there has recently been considerable interest in automatic mechanisms for detecting whether hosts are infected, and throttling or quarantining them; see, for example, [26]. There has also been work on automatic generation of self-certifying alerts [13] which are equivalent to patches. Thus, it is possible to view removal as happening on the same time scale as infection. In the case of viruses, it takes longer to generate virus signatures and update antivirus software, but their spread is also slower. Hence, again, the model is not unrealistic.

There is a substantial literature on the SIR model in epidemiology, starting with the work of Kermack and McKendrick [18]. A commonly used approach in early work was to approximate a stochastic model by a deterministic one in a large population (law of large numbers) limit. More recent work has considered stochastic aspects, such as obtaining Poisson or normal limiting distributions for the number of survivors; see, for example, [5, 19]. A key concept in these studies is the basic reproductive number R_0, which denotes the expected number of secondary infectives caused by a single primary infective. If $R_0 > 1$, the infection spreads to some sizeable fraction of the entire population; if $R_0 < 1$, then the fraction eventually infected is close to zero. The concept of basic reproductive number is easy to define with uniform mixing (i.e., when any infective can infect any susceptible equally easily), but it is not clear how to apply it to general networks, where this number could be different for every node. One approach is to consider networks with special structure, where either nodes or links belong to one of a small number of types. This is the approach taken, for example, by Ball, Mollison and Scalia-Tomba [3], who consider two-level models where network links can belong to one of two types—(i) local, for example, within a household or (ii) global, between households.

In this paper we obtain conditions for the number eventually infected to be small, in arbitrary networks. Conversely, we obtain conditions for the average number of infected nodes to be large in some specific network models of practical interest, including Erdős–Rényi and power law random graphs. The rest of the paper is structured as follows. We introduce the epidemic spreading model in Section 2. Sufficient conditions for small epidemic size (where the size is defined as the number that ever become infected) are obtained in Section 2. Applications of these results to the star, clique, Erdős–Rényi graph and power law graph are found in Section 3. Furthermore, in this section we took advantage of results on

the giant component for various families of graphs to give a lower bound to the number of nodes ultimately removed.

2. Model. We consider a closed population of n individuals, connected by a neighborhood structure which is represented by an undirected, labelled graph $\mathcal{G} = (V, E)$ with node set V and edge set E. Each node can be in one of three possibly states, susceptible (S), infective (I) or removed (R). The initial set of infectives at time 0 is assumed to be nonempty, and all other nodes are assumed to be susceptible at time 0. The evolution of the epidemic is described by the following discrete-time model. Let $X_v(t)$ denote the indicator that node v is infected at the beginning of time slot t and $Y_v(t)$ the indicator that it is removed. Each node that is infected at the beginning of a time slot attempts to infect each of its neighbors; each infection attempt is successful with probability β independent of other infection attempts. Each infected node is removed at the end of the time slot. Thus, the probability that a susceptible node u becomes infected at the end of time slot t is given by $1 - \prod_{v \sim u}(1 - \beta X_v(t))$, where we write $v \sim u$ to mean that $(u, v) \in E$. Note that the evolution stops when there are no more infectives in the population. At this time, we want to know how many nodes are removed.

The above model is known as the Reed–Frost model. It corresponds to a deterministic infectious period which is the same at every node. It is one of the earliest stochastic SIR models to be studied in depth, because of its analytical tractability. Note that the evolution can be described by a Markov chain in this case. Another commonly used model assumes that infectious periods are i.i.d. and exponentially distributed, so that the system evolves as a continuous time Markov process. General infectious periods give rise to non-Markovian systems. These are outside the scope of this work.

The object of interest is the number of nodes that eventually become infected (and removed) compared to the number initially infected. As noted earlier, in mean field models of SIR epidemics, the number of nodes removed exhibits a sharp threshold; as β is increased, it suddenly jumps from a constant (which doesn't depend on n) to a nonzero fraction of n, the number of nodes in the system. We wish to ask if a similar threshold is exhibited on general graphs and, if so, how the critical value of β is related to properties of the graph.

We now state general conditions for the number of nodes removed to be small. Let A denote the adjacency matrix of the undirected graph G, that is, $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. Since A is a symmetric, nonnegative matrix, all its eigenvalues are real, the eigenvalue with the largest absolute value $\lambda_1(A)$ is positive and its associated eigenvector has nonnegative entries (by the Perron–Frobenius theorem). If the graph is connected, as we shall assume, then this eigenvalue has multiplicity one, the corresponding eigenvector is strictly positive and is the only one with all elements being nonnegative.

THEOREM 2.1. *Suppose $\beta\lambda_1(A) < 1$. Then, the total number of nodes removed, $|Y(\infty)|$, satisfies, for any $\varepsilon > 0$ and for some constant $C > 0$ depending*

on the graph,

$$\mathbb{P}\big(|Y(\infty)| > \sqrt{|X(0)|n^{\varepsilon+1/2}}\big) \le Cn^{-\varepsilon},$$

where $|X(0)|$ is the number of initial infectives.

If the graph G is regular (i.e., each node has the same number of neighbors) with node degree d, then for any $\varepsilon > 0$ and for some constant $C' > 0$ depending on the graph,

$$\mathbb{P}\big(|Y(\infty)| > |X(0)|n^{\varepsilon}\big) \le C'n^{-\varepsilon}.$$

PROOF.   Let $v(i) = \mathbb{P}(\text{node } i \text{ is ever infected})$. Then, $v(i) = 1$ if $i \in I$, where I denotes the set of initial infectives, and otherwise $v(i) \le \sum_{j \sim i} \beta v(j)$, where we write $j \sim i$ to mean that (i, j) is an edge. Thus,

$$(2.1) \qquad\qquad\qquad (I - \beta A)v \le \mathbf{1}_I,$$

where $\mathbf{1}_I$ denotes the vector with components 1 for $i \in I$ and 0 for $i \notin I$, and the inequality holds in the usual partial order, namely, componentwise. Now, if $\beta\lambda_1(A) < 1$, then we have the power series expansion

$$(I - \beta A)^{-1} = \sum_{k=0}^{\infty} \beta^k A^k.$$

Therefore, we can multiply both sides of the inequality in (2.1) by $(I - \beta A)^{-1}$ to obtain

$$(2.2) \qquad\qquad v = \mathbb{E}[Y(\infty)] \le (I - \beta A)^{-1} X(0),$$

and so

$$(2.3) \qquad \begin{aligned} \mathbb{E}[|Y(\infty)|] &\le \mathbf{1}^T (I - \beta A)^{-1} X(0) \\ &\le \|\mathbf{1}\| \, \|(I - \beta A)^{-1}\| \, \|X(0)\|, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm in the case of a vector, and the matrix or operator norm in the case of a matrix. Now the operator norm of a symmetric matrix is its spectral radius, the largest of its eigenvalues in absolute value. Hence, $\|(I - \beta A)^{-1}\| = (1 - \beta\lambda_1(A))^{-1}$. Moreover, $\|X(0)\| = \sqrt{\sum_{v \in V} X_v^2(0)} = \sqrt{|X(0)|}$. Likewise, $\|\mathbf{1}\| = \sqrt{n}$. Substituting these in (2.3) yields

$$\mathbb{E}[|Y(\infty)|] \le \frac{1}{1 - \beta\lambda_1(A)} \sqrt{n|X(0)|}.$$

The first claim of the theorem follows using Markov's inequality.

Next, let $(\lambda_1(A), \lambda_2(A), \dots, \lambda_n(A))$ be the eigenvalues of A associated with the eigenvectors $(x^1, x^2, \dots, x^n)$. Using the spectral decomposition

$$(I - \beta A)^{-1} = \sum_{i=1}^{n} \frac{1}{1 - \beta\lambda_i(A)} x^i (x^i)^T,$$

where x^i denotes the eigenvector corresponding to the eigenvalue λ_i of A and $(x^i)^T$ its transpose, we can rewrite (2.3) as

$$(2.4) \qquad \mathbb{E}[|Y(\infty)|] \leq \sum_{i=1}^{n} \frac{1}{1 - \beta\lambda_i(A)} \mathbf{1}^T x^i (x^i)^T X(0).$$

Now, if G is a regular graph and each node has degree d (i.e., has exactly d neighbors), then each row sum of its adjacency matrix A is equal to d. Hence, it is clear that the positive vector $\frac{1}{\sqrt{n}}\mathbf{1}$ is an eigenvector of A corresponding to the eigenvalue d. By the Perron–Frobenius theorem, this is therefore the largest eigenvalue. Hence, $\lambda_1(A) = d$, $x_1 = \frac{1}{\sqrt{n}}\mathbf{1}$, and all other eigenvectors $x_2, \ldots, x_n$ are orthogonal to $\mathbf{1}$. Hence, by (2.4),

$$\mathbb{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta\lambda_1} \mathbf{1}^T x_1 x_1^T X(0)$$

$$= \frac{1}{n(1 - \beta\lambda_1)} \mathbf{1}^T \mathbf{1}\mathbf{1}^T X(0)$$

$$= \frac{1}{1 - \beta\lambda_1} |X(0)|.$$

A direct application of Markov's inequality yields the second claim of the theorem. $\qquad \square$

REMARKS. The theorem says that, if $\beta\lambda_1(A) < 1$, then starting from a "small" population of initial infectives, the final size of the epidemic is small with high probability, that is, with probability going to 1 as n tends to infinity. For example, if $|X(0)| = 1$, then the final size of the epidemic is bounded by a constant in the case of regular graphs, and by a multiple of $\sqrt{n}$ in general. Thus, the fraction of nodes infected goes to zero as n tends to infinity.

The upper bounds in the theorem are close to the best possible in general, as the example of the star-shaped network in Section 3.1 demonstrates. Nevertheless, the factor of $\sqrt{n}$ "gap" between the bound on arbitrary graphs and on regular graphs leads us to ask whether we can obtain tighter bounds for some other families of graphs besides regular graphs. We now present two such bounds.

Let G be an arbitrary connected graph with adjacency matrix A, and let x^1 denote the eigenvector corresponding to the largest eigenvalue $\lambda_1(A)$. Note that A is an irreducible matrix since G is connected. Hence, by the Perron–Frobenius theorem, all elements of the vector x^1 are positive. We now have the following:

THEOREM 2.2. *Suppose $\beta\lambda_1(A) < 1$. Then,*

$$\mathbb{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta\lambda_1(A)} \frac{x_{\max}^1}{x_{\min}^1} |X(0)|,$$

where x^1_{max} and x^1_{min} denote the largest and smallest elements of the eigenvector x^1.

PROOF. We obtain from (2.2) and the spectral expansion of $(I - \beta A)^{-1}$ that

$$(x^1)^T v = \sum_{j=1}^{n} x^1_j \mathbb{E}[Y_j(\infty)]$$

$$\leq (x^1)^T (I - \beta A)^{-1} X(0)$$

$$= \sum_{i=1}^{n} \frac{1}{1 - \beta \lambda_i(A)} (x^1)^T x^i (x^i)^T X(0)$$

$$= \frac{1}{1 - \beta \lambda_1(A)} (x^1)^T X(0).$$

The last equality holds because the eigenvectors are orthogonal, since A is symmetric.

Recall that all elements of x^1 are positive. Hence, we obtain from the above that

$$\mathbb{E}[|Y(\infty)|] = \sum_{j=1}^{n} \mathbb{E}[Y_j(\infty)] \leq \frac{1}{1 - \beta \lambda_1(A)} \frac{x^1_{max}}{x^1_{min}} X(0),$$

as claimed. □

The bound above coincides with the first bound in Theorem 2.1 when applied to a star network (see Section 3.1) and with the second bound when applied to regular graphs. Intuitively, we might expect that graphs with bounded node degree should exhibit behavior very similar to regular graphs. However, we do not know of any general upper bounds on the ratio x^1_{max}/x^1_{min} based only on node degrees, and it is not even clear that such a bound exists. But, as the next lemma demonstrates, it is indeed possible to bound the final epidemic size via the maximal node degree. While this bound does not capture the correct epidemic threshold, it may nevertheless may be quite good away from the threshold.

THEOREM 2.3. *Let G be an arbitrary graph with maximal node degree denoted Δ. If $\beta \Delta < 1$, then*

$$\mathbb{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta \Delta} |X(0)|.$$

PROOF. We use a well-known monotonicity property of epidemics on graphs. Let G be a subgraph of G', that is, G and G' have the same vertex set and the edge set of G is a subset of the edge set of G'. Let $(X(t), Y(t))$ and $(X'(t), Y'(t))$ denote the indicators of the set of infected and removed nodes corresponding to

Reed–Frost epidemics on G and G', respectively, starting with the same initial condition $(X(0), Y(0))$. Then, the epidemics can be coupled in such a way that $Y(t) \leq Y'(t)$ for all t.

This result extends, with suitable modifications described below, to multigraphs, namely, graphs which may have self-loops as well as multiple edges between pairs of nodes. The definition of the epidemic is extended to multigraphs as follows. Self-loops can cause an infected node to become re-infected in the next time slot, so that each infected node remains infected for a geometric number of time slots (with parameter depending on the number of self-loops) prior to removal. Also, in each time slot, each infected node draws an independent sample of its edges (counting multiple edges separately) along which to propagate infection. Let $N_i(t)$ and $N_i'(t)$ denote the number of time slots for which node i has been infectious up to time t, for the epidemic on the multigraphs G and G', respectively. If these epidemics start from the same initial condition, there is a coupling such that $N(t) \leq N'(t)$ for all $t \geq 0$.

Moreover, for the epidemic on G, we have the inequality $\mathbb{E}[N_i(t)] \leq \sum_{j=1}^{n} \beta \times A_{ij}\mathbb{E}[N_j(t)] + \mathbf{1}_I(i)$, for each node i, where $\mathbf{1}_I(i)$ is the indicator that node i is initially infected. Thus, analogous to (2.2), we obtain

$$(I - \beta A)N \leq \mathbf{1}_I,$$

and a similar equation for N' with A replaced by A'; here A and A' are the "adjacency" matrices of G and G', whose entries count the number of edges between any pair of nodes. Thus, Theorem 2.1 remains true if we replace $Y(\infty)$ with $N(\infty)$ in its statement. (In the case of simple graphs, they are identical.)

Now observe that a graph G with maximal node degree $d_{\max}$ can be extended, by the addition of multiple edges and self-loops, to a *regular* multi-graph G' with constant node degree $d_{\max}$. Thus, along the lines of Theorem 2.1, we have

$$\mathbb{E}\left[\sum_{i=1}^{n} N_i'(\infty)\right] \leq \frac{1}{1 - \beta\Delta}|X(0)|.$$

But $N \leq N'$ by coupling, and $Y_i(t) = 1(N_i(t) \geq 1) \leq N_i(t)$ for all i and t. Hence,

$$\mathbb{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta\Delta}|X(0)|,$$

and the second claim follows by direct application of Markov's inequality. $\quad\square$

REMARK. Using Markov's inequality, it is clear that both Theorem 2.2 and Theorem 2.3 imply that, starting from a small number of infectives, the proportion of infected individuals is, with high probability, a zero fraction of the population.

Next, we note that the proof of Theorem 2.1 above doesn't require us to assume that the epidemic be of Reed–Frost type. It works for general infectious period. More precisely, the model is as follows and evolves in continuous time. We as-

sociate i.i.d. nonnegative random variables $\mathcal{T}_u$ with arbitrary distribution with the nodes $u \in V$, and independent i.i.d. Exponential(μ) random variables $\mathcal{E}_{u,v}$ with the edges $(u, v) \in E$. If node u becomes infected, it stays infected for a random time $\mathcal{T}_u$. If $\mathcal{E}_{u,v} \leq \mathcal{T}_u$, then $\mathcal{E}_{u,v}$ time units after infection, it attempts to infect node v.

For our purposes, the key difference from the Reed–Frost model is that edges are not independent. Conditional on the existence of an edge (u, v), the distribution of $\mathcal{T}_u$ is biased toward larger values, which makes the existence of adjacent edges (u, w) more likely. Nevertheless, this lack of independence does not affect the union bound used to derive equation (2.1). Therefore, following the steps of the above proof and replacing β by the probability that a node gets infected by an infected neighbor, given by $p = 1 - \mathbb{E}[e^{-\mu \mathcal{T}_u}]$, we have the following:

THEOREM 2.4. *If $p\lambda_1(A) < 1$, then the total number of nodes removed,* $|Y(\infty)|$, *satisfies*

$$\mathbb{E}[|Y(\infty)|] \leq \frac{1}{1 - p\lambda_1(A)} \sqrt{n|X(0)|},$$

where $|X(0)|$ is the number of initial infectives. Moreover, if the graph G is regular (i.e., each node has the same number of neighbors), then

$$\mathbb{E}[|Y(\infty)|] \leq \frac{1}{1 - p\lambda_1(A)} |X(0)|.$$

The converse of Theorem 2.1 is not true in general. Consider the ring on n nodes. As a regular graph with degree 2, its adjacency matrix has maximum eigenvalue 2. Nevertheless, for any $\beta < 1$, the size of the epidemic starting from a single initial infective is bounded by the sum of the sizes of two independent branching processes, where each branching process has Bernoulli offspring distribution with parameter β. (The branching processes decide if the epidemic will spread one node left or right from its current position before dying out.) Since each branching process is subcritical, its final size is finite almost surely and in expectation. Thus, the expected size of the epidemic is a constant that does not depend on n. In other words, the epidemic is small even if $\beta\lambda_1 = 2\beta > 1$.

In particular, the SIR epidemic on the ring does not exhibit a sharp threshold on the open interval $(0, 1)$. On this interval, the final size of the epidemic is a smooth function of the infectiousness parameter β, even in the limit as n tends to infinity. It is shown in Section 3.1 that a similar result holds for star-shaped networks as well; in fact, there is no threshold even on the closed interval $[0, 1]$ in this case.

However, while there isn't always a threshold, it turns out that there is one in many networks of practical interest: there is a lower bound on β, above which the epidemic infects a positive fraction of the population on average. We now illustrate this through several examples.

3. Examples.

3.1. *Star-shaped networks.* The star-shaped network is of interest because it illustrates that the bounds in Theorems 2.1 and 2.2 are close to the best possible for general networks. It also exhibits a smooth dependence of the final size of the epidemic on the infectiousness parameter β, thereby demonstrating that threshold behavior doesn't always occur. Finally, understanding the star is important to understanding certain power-law networks.

Consider the star network, consisting of a hub and $n-1$ leaves, each of which is attached only to the hub. Its adjacency matrix A has ones along the first row and column, except for the $(1,1)$ element, which is zero; all other elements are zero. Thus, A is a rank-two matrix and can have only two nonzero eigenvalues. It is readily verified that $(\sqrt{n-1}, 1, \ldots, 1)^T$ and $(-\sqrt{n-1}, 1, \ldots, 1)^T$ are eigenvectors corresponding to the eigenvalues $\sqrt{n-1}$ and $-\sqrt{n-1}$ respectively, and so the spectral radius of A is $\sqrt{n-1}$. The principal eigenvector is $x^1 = (\sqrt{n-1}, 1, \ldots, 1)^T$, so that $x^1_{\max}/x^1_{\min} = \sqrt{n-1}$.

Now suppose $\beta\sqrt{n-1} = c < 1$. Consider the initial condition where only the hub is infected, so that $|X(0)| = 1$. The number of leaves infected before the hub is cured is binomial with parameters $n-1$ and β. No other leaves can be infected subsequently. Hence,

$$\mathbb{E}[|Y(\infty)|] = 1 + \beta(n-1) = 1 + c\sqrt{n-1},$$

which is comparable to the upper bound, $\sqrt{n-1}/(1-c)$, given by Theorem 2.1 or, equivalently, by Theorem 2.2. In addition, using large deviations properties of the binomial distribution ([1], Corollary A.1.14), we can show that, for $c' < c$, $\mathbb{P}(|Y(\infty)| < c'\sqrt{n})$ goes to 0, exponentially in n, when n goes to infinity.

We observe in this case that $\mathbb{E}[|Y(\infty)|]$ is a smooth (almost linear) function of β and does not exhibit any threshold behavior.

Suppose next that the hub is initially uninfected but k leaves are infected. The hub becomes infected in the next time step with probability $1 - (1-\beta)^k$. It subsequently infects a number of leaves which is binomial with parameters $n-1-k$ and β. The epidemic dies out at $t = 3$. So, in this case,

$$\mathbb{E}[|Y(\infty)|] = k + [1 - (1-\beta)^k][1 + \beta(n-1-k)]$$
$$\leq k + \beta k[1 + \beta(n-1-k)] \leq |X(0)|(1 + c^2/2).$$

Thus, when the hub is initially uninfected, the expected final size of the epidemic is only a constant multiple of the initial number of infectives. This illustrates that the initial condition can have a big impact in general.

3.2. *Complete graph.* A complete graph is one in which an edge is present between every pair of nodes. Much of the early work on SIR epidemics was based on mean field models. These are rigorously justifiable only in the case of complete

graphs, and motivates our interest in them. We shall recover the classical result that the epidemic has a threshold at $R_0 = 1$, where the basic reproduction number $R_0 = \beta(n-1)$ is defined as the mean number of secondary infections caused by a single primary infective, when the entire population is susceptible.

The complete graph is a regular graph with common node degree $n - 1$. Therefore, its spectral radius is $\lambda_1 = n - 1$, and we have by Theorem 2.1 that, if $\beta(n-1) < 1$, then the final size of the epidemic is bounded by $1/(1 - \beta(n-1))$ times the initial number of infectives. We now establish a converse.

Suppose $\beta(n-1) = c > 1$ is held constant. (We don't need to assume this, but the results will need to be restated in terms of the limits superior and inferior of the sequence c_n; it should be clear to the reader how to do so based on the discussion below.) Let $|X_0| = 1$ and let u be the initial infected node. Consider the random subgraph of the complete graph obtained by retaining each edge with probability β, independent of all other edges, and let C_u denote the connected component containing u in this random graph (possibly just the singleton $\{u\}$). It is clear that C_u can be interpreted as the set of infected nodes in the epidemic; each neighbor of u is infected with probability β, and is hence a neighbor of u in the random graph described above, and so on iteratively. Thus, the number of infected nodes in the epidemic has the same probability law as the size of the component C_u. The connection between Reed–Frost epidemics and random graphs has been previously noted in [2].

The above random graph model was introduced by Erdős and Rényi [14]; we denote it by $\mathcal{G}(n, \beta)$, where n denotes the number of nodes, and β the probability that the edge between each pair of nodes is present. It is also called a Bernoulli random graph because the indicators of edges are i.i.d. Bernoulli random variables.

We now use the following fact, which was established by Erdős and Rényi [14]; see [17], Theorem 5.4, for a more recent reference. Here, we assume that $c = \beta(n-1)$ is held constant while $n \to \infty$, and that $c > 1$.

THEOREM 3.1. *Let γ be the unique positive solution of $\gamma + e^{-\gamma c} = 1$. Then, as $n \to \infty$, the size of the largest connected component in the random graph $\mathcal{G}(n, \beta)$ is $(1 + o(1))\gamma n$, with high probability (with probability going to 1 as n tends to infinity).*

The uniqueness of γ follows from the convexity of the function $f(x) = x + e^{-cx}$ and the fact that $f(0) = 1$, while its existence follows from the continuity of f and the fact that $f'(0) = 1 - c < 0$, but that $f(x) \to \infty$ as $x \to \infty$.

We now estimate the size of C_u, the connected component containing the initial infective. If u belongs to the "giant component," then $|C_u| = (1 + o(1))\gamma n$. Since a fraction γ of nodes belong to the giant component, the probability that node u does so is γ. Hence, $\mathbb{E}[|C_u|] = (1 + o(1))\gamma^2 n$. Moreover, it is proved in [17], Theorem 5.4, that the size of the second largest component in the supercritical random graph is $O(\log(n))$. Hence, $|Y(\infty)| = O(\log(n))$ with probability $1 - \gamma$, which is

the probability that the initial node does not belong to the giant component, and $\mathbb{P}(|Y(\infty)| > (1 + o(1))\gamma n) = \gamma$. We have thus shown the following:

THEOREM 3.2. *Let* $G = (V, E)$ *be the complete graph on* n *nodes, and let* $\beta = \frac{c}{n-1}$ *for an arbitrary constant* $c > 1$. *Then, the final size of the epidemic satisfies*

$$\mathbb{E}[|Y(\infty)|] \geq (1 + o(1))\gamma^2 n,$$

for any $|X(0)| \geq 1$, *where* $\gamma > 0$ *solves* $\gamma + e^{-\gamma c} = 1$. *Moreover,* $|Y(\infty)| = (1 + o(1))\gamma n$ *with probability tending to* γ *as* n *tends to infinity, while* $|Y(\infty)| = O(\log n)$ *with probability tending to* $1 - \gamma$.

There is thus a threshold at $c = 1$ for the final size of the epidemic; starting with a constant number of initial infectives, the final size is a constant independent of n with high probability if $c < 1$, and a fraction of n with positive probability (depending only on c and not on n) if $c > 1$.

3.3. *Erdős–Rényi random graphs.* The Erdős–Rényi graph $\mathcal{G}(n, p)$ with parameters n and p is defined as a random graph on n nodes, where the edge between each pair of nodes is present with probability p, independent of all other edges. If $p = 1$, then this is the complete graph.

The spreading behavior of an epidemic on an Erdős–Rényi graph is of interest for a number of reasons. First, it is a graph that has received considerable attention in the past [6]. Second, it is an important component of the class of power-law random graphs that model the emails graph and number of other social networks.

We shall consider a sequence of such graphs indexed by n. Denote by d the corresponding average degree, that is, $d = (n-1)p$. Note that p and d may depend on n, but this is suppressed in the notation. We consider the regime $\log(n) \ll d$, that is, $\log(n)/d \to 0$ as $n \to \infty$. In this case, it is known that the graphs are connected with high probability [6].

Define $c_n = \beta d = (n-1)\beta p$; we have omitted the dependence of β and p on n in the notation, but make it explicit in the case of c. Consider an SIR epidemic on such a graph starting with one node initially infected. We have the following:

THEOREM 3.3. *If* $\limsup_{n\to\infty} c_n \leq c < 1$, *then for all* n *sufficiently large,* $\mathbb{E}[|Y(\infty)|]$ *is bounded by a constant that does not depend on* n.

On the other hand, if $\liminf_{n\to\infty} c_n \geq c > 1$, *then* $|Y(\infty)| = (1 + o(1))\gamma n$ *with probability tending to* γ *as* n *tends to infinity, while* $|Y(\infty)| = O(\log n)$ *with probability tending to* $1 - \gamma$, *where* $\gamma > 0$ *solves* $\gamma + e^{-\gamma c} = 1$.

Here, expectations and probabilities are taken both with respect to the random realization of a graph and the evolution of the epidemic conditional on the graph.

PROOF.   Suppose first that $\liminf_{n\to\infty} c_n \geq c > 1$. As in the case of the complete graph, we identify the infected individuals in the epidemic with the connected component containing the initial infective u in an Erdős–Rényi random graph with parameters n and βp. [If edge (u, v) is present in the original Erdős–Rényi graph, which happens with probability p, then u succeeds in infecting v with probability β. This yields the new graph with edge probability βp; the independence of the edges is obvious.] Thus, the second claim of the theorem follows in the same way as Theorem 3.2.

The first claim is stronger than what the upper bound of Theorem 2.1 yields. Note that, by the Perron–Frobenius theorem, the spectral radius $\lambda_1(A)$ of the adjacency matrix lies between the smallest and largest node degree. For the random graph $\mathcal{G}(n, p)$, the node degrees are binomial random variables with parameters $n - 1$ and p. If the average node degree $d = (n - 1)p$ satisfies $d \gg \log(n)$, that is, $\log(n)/d \to 0$ as $n \to \infty$, then it can be shown using Chernoff's bound that both the minimal and maximal node degree are $(1 + o(1))d$ with high probability; hence, so is the spectral radius. In this case, Theorem 2.1 yields that, if $\beta\lambda_1(A) \sim (n - 1)\beta p \leq c < 1$, then the expected final size of the epidemic is bounded by a constant times $\sqrt{n}$. To show that it is in fact bounded by a constant, and that this holds even without the assumption that $d \gg \log(n)$, we use a branching process construction.

Rather than fixing the random graph $\mathcal{G}(n, p)$ in advance, we use the principle of deferred decisions to generate it dynamically as the epidemic progresses. Thus, starting with the initial infective u, we put down all edges from it to other nodes. Then, we decide whether u succeeds in infecting its neighbors along each of those edges. For each neighbor v so infected, we repeat the process. Thus, the number of nodes infected by u is binomial with parameters $n - 1$ and βp; the number of nodes infected by each subsequent infective is stochastically dominated by such a binomial random variable. Thus, the size of the epidemic is bounded above by the size of a branching process whose offspring distribution is binomial, $B(n - 1, \beta p)$. The branching process is subcritical by the assumption that $(n - 1)\beta p \leq c < 1$, and so it becomes extinct with probability 1, that is, its final population size is finite almost surely, and in expectation. It can be shown directly, using generating functions, that it is bounded uniformly in n. Alternatively, note that if $(n - 1)\beta p = c$ for all n, then the binomial offspring distributions converge in distribution to a Poisson with parameter c as $n \to \infty$; the population sizes of the corresponding branching processes also converge, both in distribution and in expectation. Since $c < 1$, the branching process with Poisson(c) offspring distribution is subcritical, and so it has a finite mean population size.   □

3.4. *Power law random graphs.*   There has been considerable interest in power law graphs since it was first noticed that the Internet AS-level graph exhibits a power law degree distribution [15]. Briefly, a power law graph is one where the number of nodes with degree k is proportional to $k^{-\gamma}$ for some $\gamma > 1$. For the

mean degree to be finite, we need $\gamma > 2$ and this is the range we shall consider. It is suggested in [15] that the Internet Autonomous System level graph is characterized by $\gamma \approx 2.1$.

There have been several different models proposed for graphs with power law degree distributions; see, for example, [4, 8]. In this paper we consider the following model of random graphs on n vertices, introduced in [11]. Let $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ be a sequence of positive weights assigned to the nodes of the graph; we assume without loss of generality that $w_1 \geq w_2 \geq \cdots \geq w_n$. The edge between the pair of vertices (i, j) is present with probability

$$p_{ij} = \frac{w_i w_j}{\sum_{k=1}^n w_k},$$

independent of all other edges; we assume that $w_1^2 \leq \sum_{k=1}^n w_k$. The resulting random graph is denoted $G(\mathbf{w})$. For example, taking $w_i = np$ for all i yields the Erdős–Rényi model with parameters n and p.

It is easy to see that w_i is the expected value of the degree of node i; hence, this model is referred to as the expected degree model. (Reed and Molloy [21, 22] have studied a model where one conditions on actual rather than expected degrees. The expected degree model has the advantage that edges are independent, which makes it much easier to analyze.) We do not assume that the w_i are integer-valued. Note that the resulting graph may have self-loops, but it does not have multiple edges. The self-loops do not affect the spread of the epidemic and are not important to our analysis.

Let d denote the average and m the maximum expected degree. (Thus, $m = w_1$, but it is convenient to distinguish it in the notation as the model is parametrized by d, m and the exponent of the power law degree distribution.) Chung and Lu [11] propose the following explicit power law model for the expected degree sequence:

$$(3.1) \qquad w_i = c(i_0 + i)^{-1/(\gamma-1)}, \qquad 1 \leq i \leq n,$$

where

$$(3.2) \qquad c = \frac{\gamma - 2}{\gamma - 1} d n^{1/(\gamma-1)}, \qquad i_0 = n \left(\frac{d(\gamma - 2)}{m(\gamma - 1)} \right)^{\gamma - 1}.$$

The number of nodes with weight bigger than k (equal to the largest i such that $w_i \geq k$) scales like $k^{1-\gamma}$. Thus, γ is the exponent of the power law distribution of expected node degrees. The weights w_i are the order statistics of this distribution. The distribution is shifted by i_0 and scaled by c, where these constants are chosen so as to achieve the specified average d and maximum m for the expected degrees.

The eigenvalues of the adjacency matrix for this random graph model have been studied by Chung, Lu and Vu. They show ([12], Theorem 4) that, with high probability, the spectral radius of the graph is

$$\lambda_1(A) = \begin{cases} (1 + o(1))\sqrt{m}, & \gamma > 2.5, \\ (1 + o(1))\dfrac{d(\gamma - 2)^2}{(\gamma - 1)(3 - \gamma)} \left(\dfrac{(\gamma - 1)m}{(\gamma - 2)d} \right)^{3-\gamma}, & 2 < \gamma < 2.5. \end{cases}$$

By Theorem 2.1, if $\beta\lambda_1(A) < 1$, then the size of the epidemic is bounded by $\sqrt{n}$ times the size of the initial infective population.

We show that, if $\gamma \in (2, 3)$, then the expected size of the final epidemic is large (a nonzero fraction of n) for any nonempty set of initial infectives, and any $\beta > 0$. We now establish a partial converse. We show that the graph has a core such that, if $\gamma \in (2, 3)$, then the expected size of the final epidemic is large (a nonzero fraction of n) for any nonempty set of initial infectives, and any $\beta > 0$. It is easy to see from the description above that the expected degree of node i is precisely w_i, and the expected average degree of the graph is given by $d = \frac{1}{n}\sum_{i=1}^n w_i$. We can now write $p_{ij} = \frac{w_i w_j}{nd}$.

It is straightforward to describe the evolution of a Reed–Frost epidemic on the expected degree random graph model. Consider a single initial infective, say, node i. Node j becomes infected at time 1 if edge (i, j) is present in the random graph and if i infects j in the first time slot; this has probability βp_{ij}, and is independent of whether node i infects some other node k. Moreover, node i cannot infect node j in any subsequent time step since it is removed at time 1. Using the principle of deferred decisions, we can construct a realization of the random graph as the epidemic spreads. It is clear from this construction that the set of nodes that eventually become infected can be identified with the connected components containing the initial infectives in the random graph with weight sequence $\beta\mathbf{w}$, namely, $G(\beta\mathbf{w})$. Suppose there is a single initial infective. The question of whether there is a large epidemic is equivalent to that of whether the random graph $G(\beta\mathbf{w})$ possesses a giant component, and whether the initial infective belongs to this giant component. If there is more than one initial infective, the final set of removed nodes is the union of the connected components containing the initial infectives, in the random graph $G(\beta\mathbf{w})$.

A sufficient condition for the existence of a giant component is derived in [10], Theorem 3. The condition can be stated in terms of the average expected degree d, as follows:

THEOREM 3.4. *For a random graph $G(\mathbf{w})$ with expected degree sequence having average expected degree $d > 1 + \delta > 1$, with positive probability, there is a unique giant component C such that $\sum_{i\in C} w_i \geq (1 - c_\delta)\sum_{i\in V} w_i$, where $c_\delta \in (0, 1)$ is a constant that depends only on δ.*

In words, the giant component contains a nonzero fraction of the total weight of all nodes. Later, we will show that this implies that it contains a nonzero fraction of the nodes.

We use this result to obtain estimates on the final size of an epidemic on a power law random graph. Fix k (as a function of n) and consider the subgraph induced by the k nodes with the largest weight in the random graph $G(\beta\mathbf{w})$. The average

expected degree of this subgraph is easily seen to be

$$(3.3) \qquad d_k = \frac{(\sum_{i=1}^{k} \beta w_i)^2}{k \sum_{i=1}^{n} \beta w_i} = \frac{\beta(\sum_{i=1}^{k} w_i)^2}{knd}.$$

If this is strictly larger than 1, then by Theorem 3.4 above, this subgraph has a giant component. We now find conditions on k such that $d_k > 1$.

We have from (3.1) and (3.2) that

$$w_i = m\left(1 + \frac{i}{i_0}\right)^{-1/(\gamma-1)},$$

and so,

$$\sum_{i=1}^{k} w_i \sim \int_0^k m\left(1 + \frac{x}{i_0}\right)^{-1/(\gamma-1)} dx = m i_0 \int_1^{1+(k/i_0)} y^{-1/(\gamma-1)} dy$$

$$= \frac{\gamma-1}{\gamma-2} m i_0 \left[\left(\frac{k}{i_0} + 1\right)^{(\gamma-2)/(\gamma-1)} - 1\right].$$

Now, if $k \gg i_0$, it follows that

$$(3.4) \qquad \sum_{i=1}^{k} w_i \sim \frac{\gamma-1}{\gamma-2} m k^{(\gamma-2)/(\gamma-1)} i_0^{1/(\gamma-1)} = dn^{1/(\gamma-1)} k^{(\gamma-2)/(\gamma-1)}.$$

Substituting this in (3.3) yields

$$(3.5) \qquad d_k = \beta d n^{(3-\gamma)/(\gamma-1)} k^{(\gamma-3)/(\gamma-1)}.$$

We now distinguish two cases. Suppose first that $\gamma > 3$. Then d_k is a nondecreasing function of k, and its maximum value, attained at $k = n$, is βd. This only yields the weak result that there is a large epidemic if $\beta d > 1$.

Suppose next that $2 < \gamma < 3$. Then d_k is a decreasing function of k. Fix $\delta > 0$. Defining N_δ to be the largest value of k for which $d_k > 1 + \delta$, we see that

$$(3.6) \qquad N_\delta = \left\lfloor \left(\frac{\beta d}{1+\delta}\right)^{(\gamma-1)/(3-\gamma)} n \right\rfloor + 1,$$

where $\lfloor x \rfloor$ denotes the integer part of x. The following result is now an easy consequence.

THEOREM 3.5. *For $2 < \gamma < 3$, and $\beta > 0$ fixed arbitrarily small, the expected size of the epidemic, starting from an arbitrary initial infective, is bounded below by a constant multiple of n, where the constant may depend on β. Here the expectation is taken both with respect to the random realization of a graph, and the evolution of the epidemic conditional on the graph.*

PROOF.   Consider the N_δ nodes of largest weight, where N_δ is given by (3.6). Since $\beta, d > 0$ are constants, N_δ is a constant multiple of n. By Theorem 3.4, the random graph $G(\beta\mathbf{w})$ restricted to these nodes contains, with positive probability, a giant component C such that

$$(3.7) \qquad \sum_{i \in C} w_i \geq (1 - c_\delta) \sum_{i=1}^{N_\delta} w_i \sim dn(1 - c_\delta)\left(\frac{\beta d}{1 + \delta}\right)^{(\gamma-2)/(3-\gamma)},$$

where we have used (3.4) and (3.6) to obtain the last asymptotic equivalence. Recall that $c_\delta \in (0, 1)$ is a constant that depends on δ. Since β, d and δ are constants, while $\sum_{i=1}^n w_i = nd$, equation (3.7) tells us that the giant component C contains a constant fraction of the total weight of the graph. We now deduce that it must also contain a constant fraction of the total number of nodes. Indeed, for a given weight, the size (in number of nodes) would be minimized if C contained the highest weight nodes. Thus, we ask what is the smallest value of k such that $\sum_{i=1}^k w_i$ exceeds the weight of C. It follows from (3.4) and (3.7) that we require

$$dn^{1/(\gamma-1)}k^{(\gamma-2)/(\gamma-1)} \sim dn(1 - c_\delta)\left(\frac{\beta d}{1 + \delta}\right)^{(\gamma-2)/(3-\gamma)},$$

and so,

$$(3.8) \qquad k \sim n(1 - c_\delta)^{(\gamma-1)/(\gamma-2)}\left(\frac{\beta d}{1 + \delta}\right)^{(\gamma-1)/(3-\gamma)};$$

in particular, k is equivalent to a constant multiple of n.

Now, if any of the initially infected nodes belongs to the giant component C, then C is a subset of the set of nodes ever infected; hence, the final size of the epidemic is proportional to n. On the other hand, suppose none of the initial infectives belongs to C. Let i be an initially infected node. Now, the probability that there is an edge between i and C in the random graph $G(\beta\mathbf{w})$ is given by

$$p(i, C) = \frac{w_i \sum_{j \in C} w_j}{nd} = (1 + o(1))(1 - c_\delta)\left(\frac{\beta d}{1 + \delta}\right)^{(\gamma-2)/(3-\gamma)} w_i,$$

which is a positive constant bounded away from zero. Conditional on this edge being present, C is a subset of the set of eventually infected nodes. Thus, in this case too, the expected final size of the epidemic is proportional to n. This completes the proof of the theorem. $\square$

This theorem shows that, at least in the regime $2 < \gamma < 3$, the epidemic threshold tends to zero in power law graphs. In other words, even arbitrarily small infection rates give rise to large epidemics (with positive probability). Similar results were obtained by Pasto–Satorras and Vespignani using mean-field methods [23, 24]. Analogous results for the SIS model were obtained by Ganesh, Massoulie and Towsley [16] and by Berger, Borgs, Chayes and Saberi [9].

3.5. *Inhomogeneous W-graphs.* We are interested in the following family of graphs $G(W)$ with vertex set $\{1, \ldots, n\}$ where i and j are connected by an edge with probability $p_{i,j}$ defined as follows: let $\{X_i, i = 1, \ldots, n\}$ be a sequence of i.i.d. random variables uniformly distributed on $[0, 1]$. Then $p_{ij} = W(X_i, X_j)$, where $W : [0, 1]^2 \to [0, 1]$ is measurable and symmetric. The edges are conditionally independent given the X_is. This model has been introduced in [20] and generalized in [7] to a larger family of inhomogeneous networks. As a matter of example, $W(x, y) = p$ for all x and y gives the classical Erdős–Rényi random graph. For more details on this family of graphs, we refer the reader to [7]. Our main interest in what follows is to understand the emergence of the giant component.

Let T_W be the integral operator with kernel nW defined by

$$(T_W f)(x) = \int_0^1 nW(x, y) f(y) \, dy,$$

and define

$$\|T_W\| = \sup\{\|T_W f\|_2 : f \geq 0, \|f\|_2 \leq 1\}.$$

As previously, we check that starting with a $G(W)$ graph, the epidemic graph is described by $G(\beta W)$.

The next theorem ([7], Corollary 3.2) characterizes the emergence of the giant component in this family of graphs.

THEOREM 3.6. *Consider the graph $G(W)$, then $\|T_W\| = 1$ is the threshold for the existence of the giant component. More precisely*:

- *if $\|T_W\| \leq 1$, then the size of the largest component is negligible with respect to n, with high probability;*
- *if $\|T_W\| > 1$ and W is irreducible [roughly speaking, T_W is reducible if the vertex set of $G(W)$ can be split into two parts so that the probability of an edge from one part to the other is zero, and irreducible otherwise], then with positive probability (uniformly in n), there is a unique giant component consisting of a nonzero fraction of all n nodes.*

Hence, $\|T_W\|^{-1}$ is the threshold for the final size of the epidemic on $G(W)$, that is, if $\beta \leq \|T_W\|^{-1}$, there is a negligible fraction of removed nodes, whereas if $\beta > \|T_W\|^{-1}$, there is a nonzero fraction of removed nodes. We are now going to give explicit computations in the case of scale-free networks.

Let us examine the case where

$$W(x, y) = \frac{w(x)w(y)}{\int_0^1 w(u) \, du},$$

and such that for X uniformly distributed on $[0, 1]$, the variable $w(X)$ follows a
Pareto (power-law) distribution, that is,

$$(3.9) \qquad \mathbb{P}\big(w(X) \in (t, t + dt)\big) = \frac{\gamma - 1}{(1 + t)^{\gamma}} \, dt, \qquad t \in [0, \infty).$$

This yields $w(x) = (1 - x)^{-1/(\gamma - 1)} - 1$ and if we assume that $\gamma > 3$, we have

$$W(x, y) = \frac{\gamma - 2}{n}\big((1 - x)^{-1/(\gamma - 1)} - 1\big)\big((1 - y)^{-1/(\gamma - 1)} - 1\big).$$

Let us now go back to the operator T_W, our goal is to compute the L_2-module of

$$(T_W f)(x) = \int_0^1 n \frac{\gamma - 2}{n}\big((1 - x)^{-1/(\gamma - 1)} - 1\big)\big((1 - y)^{-1/(\gamma - 1)} - 1\big) f(y) \, dy.$$

It is not difficult to see that the maximum is reached for $f(y) = \frac{((1-y)^{-1/(\gamma-1)}-1)}{\int_0^1 w(u)\,du}$
and, thus,

$$\|T_W\| = (\gamma - 2) \int_0^1 w(y)^2 \, dy = \frac{2}{\gamma - 3}.$$

Applying the above results for this specific W-graph we see that there will be a
large outbreak if we have a giant component in the corresponding graph, that is, if
$\frac{2\beta}{\gamma - 3} > 1$. Hence, if $\beta > \frac{\gamma - 3}{2}$, then the epidemic will eventually reach a proportion
$\tau > 0$ of the total population. As a by product of [7], Theorem 6.2, it turns out that
τ is related to the solution of the following functional equation:

$$(3.10) \qquad\qquad\qquad f = 1 - e^{T_W f}.$$

We omit this last computation as it is a bit intricate and does not bring any fur-
ther insight into the model. As previously, we can conclude that if $\beta > \frac{\gamma - 3}{2}$, then
the final size of the epidemic is bounded below by $(1 + o(1))\tau^2 n$, where τ is
related to the solution of the functional equation (3.10). Moreover, in [7], The-
orem 3.12, it is proven that, in the supercritical case $\|T_W\| > 1$, the size of the
remaining components in the graph is $O(\log(n))$. Therefore, if $\beta > \frac{\gamma - 3}{2}$, the final
size of the epidemic is a nonzero fraction of the population with probability tend-
ing to τ when n, and its size is negligeable ($O(\log n)$) with probability tending to
$1 - \tau$.

Finally, note that if T is a random variable with probability distribution given
by (3.9), then the epidemic threshold $\frac{\gamma - 3}{2}$ corresponds to the ratio $\mathbb{E}(T)/\mathbb{E}(T^2)$.
Moreover, this gives a stronger result for the case $\gamma > 3$ than the one derived
in Section 3.4 since $\frac{\gamma - 3}{2} < \gamma - 2$, where $\frac{1}{\gamma - 2}$ is the average node degree of the
graph.

REFERENCES

[1] ALON, N. and SPENCER, J. H. (2000). *The Probabilistic Method*. Wiley, New York. MR1885388

[2] BALL, F. (1983). A threshold theorem for the Reed–Frost chain-binomial epidemic. *J. Appl. Probab.* **20** 153–157. MR0688089

[3] BALL, F., MOLLISON, D. and SCALIA-TOMBA, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7** 46–89. MR1428749

[4] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

[5] BARBOUR, A. D. and UTEV, S. (2004). Approximating the Reed–Frost epidemic process. *Stoch. Proc. Appl.* **113** 173–197. MR2087957

[6] BOLLOBÁS, B. (2001). *Random Graphs*. Cambridge Univ. Press. MR1864966

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

[8] BOLLOBÁS, B. and RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **4** 5–34. MR2057681

[9] BERGER, N., BORGS, C., CHAYES, J. and SABERI, A. (2005). On the spread of viruses on the internet. In *Proc. Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 301–310. SIAM, Philadelphia.

[10] CHUNG, F. and LU, L. (2002). Connected components in random graphs with given degree sequences. *Ann. Combinatorics* **6** 125–145. MR1955514

[11] CHUNG, F. and LU, L. (2003). The average distances in random graphs with given expected degrees. *Internet Math.* **1** 91–114. MR2076728

[12] CHUNG, F., LU, L. and VU, V. (2003). Eigenvalues of random power law graphs. *Ann. Combinatorics* **7** 21–33. MR1984643

[13] COSTA, M., CASTRO, M., CROWCROFT, J., ROWSTRON, A., ZHOU, L., ZHANG, L. and BARHAM, P. (2005). Vigilante: End-to-end containment of Internet worms. In *Proc. Twentieth ACM Symp. on Operating Systems Principles* 133–147. ACM Press, New York.

[14] ERDŐS, P. and RENYI, A. (1960). On the evolution of random graphs. *Mat Kutato Int. Közl* **5** 17–60. MR0125031

[15] FALOUTSOS, M., FALOUTSOS, P. and FALOUTSOS, C. (2003). Power laws and the AS-level Internet topology. *IEEE/ACM Trans. Netw.* **11** 514–524.

[16] GANESH, A., MASSOULIÉ, L. and TOWLEY, D. (2005). The effect of network topology on the spread of epidemics. In *Proceedings IEEE Infocom*.

[17] JANSON, S., ŁUCZAK, T. and RUCIŃSKI, A. (2000). *Random Graphs*. Wiley, New York. MR1782847

[18] KERMACK, W. O. and MCKENDRICK, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond. A* **115** 700–721.

[19] LEFEVRE, C. and UTEV, S. (1995). Poisson approximation for the final state of a generalized epidemic process. *Ann. Probab.* **23** 1139–1162. MR1349165

[20] LOVÁSZ, L. and SZEGEDY, B. (2004). Limits of dense graph sequences. Technical Report TR-2004-79, Microsoft Research.

[21] MOLLOY, M. and REED, B. A. (1995). A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6** 161–180. MR1370952

[22] MOLLOY, M. and REED, B. A. (1998). The size of the largest component of a random graph on a fixed degree sequence. *Combin. Probab. Comput.* **7** 295–306. MR1664335

[23] PASTOR-SATORRAS, R. and VESPIGNANI, A. (2001). Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86** 3200–3203.

[24] PASTOR-SATORRAS, R. and VESPIGNANI, A. (2002). Epidemic dynamics in finite scale-free networks. *Phys. Rev. E* **65**.

[25] WEAVER, N., PAXSON, V., STANIFORD, S. and CUNNINGHAM, R. (2003). A taxonomy of computer worms. In *ACM Workshop on Rapid Malcode (WORM)* 11–18. ACM Press, New York.

[26] WILLIAMSON, M. M. (2002). Throttling viruses: Restricting propagation to defeat malicious mobile code. In *Proc. 18th Annual Computer Security Applications Conference* 1–61. IEEE Comput. Soc., Washington.

M. DRAIEF
INTELLIGENT SYSTEMS AND NETWORKS
IMPERIAL COLLEGE LONDON
LONDON SW7 2AZ
UNITED KINGDOM
E-MAIL: m.draief@imperial.ac.uk

A. GANESH
MICROSOFT RESEARCH
7 J. J. THOMSON AVENUE
CAMBRIDGE CB3 0FB
UNITED KINGDOM
E-MAIL: ajg@microsoft.com

L. MASSOULIÉ
THOMSON RESEARCH LAB
92648 BOULOGNE CEDEX
PARIS
FRANCE
E-MAIL: laurent.massoulie@thomson.net