

# The Effect of Network Topology on the Spread of Epidemics

A. Ganesh<sup>1</sup>, L. Massoulié<sup>1</sup>, D. Towsley<sup>2</sup>

<sup>1</sup>Microsoft Research  
7 J.J. Thomson Avenue  
CB3 0FB Cambridge,  
United Kingdom  
{ajg, lmassoul}@microsoft.com

<sup>2</sup>Dept. of Computer Science  
University of Massachusetts  
Amherst, MA 01003  
USA  
towsley@cs.umass.edu

**Abstract**—Many network phenomena are well modeled as spreads of epidemics through a network. Prominent examples include the spread of worms and email viruses, and, more generally, faults. Many types of information dissemination can also be modeled as spreads of epidemics. In this paper we address the question of what makes an epidemic either weak or potent. More precisely, we identify topological properties of the graph that determine the persistence of epidemics. In particular, we show that if the ratio of cure to infection rates is smaller than the spectral radius of the graph, then the mean epidemic lifetime is of order  $\log n$ , where  $n$  is the number of nodes. Conversely, if this ratio is bigger than a generalization of the isoperimetric constant of the graph, then the mean epidemic lifetime is of order  $e^{n^a}$ , for a positive constant  $a$ . We apply these results to several network topologies including the hypercube, which is a representative connectivity graph for a distributed hash table, the complete graph, which is an important connectivity graph for BGP, and the power law graph, of which the AS-level Internet graph is a prime example. We also study the star topology and the Erdős-Rényi graph as their epidemic spreading behaviors determine the spreading behavior of power law graphs.

**Index Terms**—Graph theory, Stochastic processes

## I. INTRODUCTION

Many network phenomena are well modeled as spreads of epidemics through a network. Prominent examples include the spread of worms and email viruses, and, more generally, faults. The spread of information can also often be modeled as the spread of an epidemic. An epidemic spreads along the underlying network topology from an initial set of *infected* nodes to *susceptible* nodes. In many cases infected nodes can be cured, reverting back to being susceptible and possibly

being reinfected. It is imperative, then to understand how the topology affects the spread of an epidemic, in other words how the topology either impedes or facilitates its spread and maintenance. Such an understanding can lead to better techniques for preventing and fighting worms and viruses, avoiding cascading failures, and for improving the dissemination of information.

In this paper we take a step towards developing an understanding of how the topology affects epidemic spread. We model the spread of an epidemic as a contact process over a finite undirected graph (with  $n$  nodes) where a node can be infected by its infected neighbors at a rate that is proportional to their number, and a node can be cured when it is infected. For such a system, given any set of initially infected nodes, the epidemic dies out in a finite amount of time. We develop general, topology dependent conditions under which the epidemic dies out either quickly or slowly. More precisely, if  $\tau$  denotes the length of time that the epidemic is active (i.e., at least one node is infected), then a sufficient condition for quick die out, defined as  $E[\tau] = O(\log n)$ , is that the ratio of infection rate to cure rate not exceed the spectral radius of the adjacency matrix of the underlying topology graph. A sufficient condition for slow die out, defined as  $E[\tau] = \Omega(e^{n^\alpha})$  for some  $\alpha > 0$ , is that the ratio of infection rate to cure rate be larger than the isoperimetric constant associated with the graph. These conditions are not necessary. For some topologies, such as a hypercube, clique, and Erdős-Rényi graph, the conditions are quite close. In others, such as a star and a power law graph, the conditions are farther apart. We narrow the gap through supplementary analysis for these latter cases. We focus on real world topologies, namely the hypercube, clique and power law graph. We also extensively study the star topology and the Erdős-Rényi graph since their properties play prominent roles in the analysis of the power law graph. In all of these

D. Towsley was a visitor at the Cambridge Microsoft Research Lab at the time this work was completed.

applications, the condition for fast die out based on the spectral radius appears to be tight.

Several papers have touched on different aspects of the problem studied here. Much of the work has focused on infinite scale-free graphs, establishing conditions under which epidemics either die out or sustain themselves forever; see, e.g., [15]. Some work has been done on finite graphs, using primarily heuristic arguments to obtain conditions under which epidemics spread quickly or not. For example, [16] uses a mean field approximation for establishing conditions for quick and slow die out in scale-free graphs and [12] provides an approximate analysis for the case of an Erdős-Rényi graph. We rigorously establish similar conditions in Section V. An exception is the work of Durrett and Liu [8], which presents a rigorous analysis leading to conditions for fast and slow die out on a finite one dimensional linear network. Also relevant is the work of Coffman et al., [7], which models cascading BGP failures on a fully connected topology. They identify regimes where BGP routers recover quickly (corresponding to fast die out) and others where they go through long periods of repeated failures (corresponding to slow die out). We treat this in a rigorous manner in Section V. Last, Wang et al. [18] show, through simulation and approximate analysis, that the condition for fast die out relates to the spectral radius of the adjacency matrix of the underlying graph.

Our model is an example of what is termed an SIS (Susceptible-Infective-Susceptible) model in the epidemic literature. Related work has been concerned with the so-called SIR (Susceptible-Infective-Removed) model; see Ball et al. [2] for recent references as well as a study of vaccination strategies in that context.

The paper is structured as follows. We introduce the epidemic spreading model in Section II. Sufficient conditions for the fast die out and slow die out of an epidemic are derived in Sections III and IV respectively. Applications of these results to the star, hypercube, clique, Erdős-Rényi graph, and power law graph are found in Section V. Section VI summarizes the paper and describes further directions to pursue.

## II. MODEL

Consider the following continuous time version of the epidemic spread model in Wang et al. [18]. We represent the system by a connected graph  $G = (V, E)$ . Here  $V$  is a set of  $n$  sites,  $E \subseteq V^2$  and  $(i, j) \in E$  represents a link between the pair of sites  $i, j \in V$ . The state at time  $t$  is represented by a vector  $X(t)$ ; site  $i$  is infected (respectively, healthy) at time  $t$  iff  $X_i(t) = 1$  (respectively,  $X_i(t) = 0$ ). Assume that infected nodes

contaminate neighbours at rate  $\beta$  and recover at rate  $\delta$ . This defines a Markov process with transition rates:

$$\begin{aligned} X_i : 0 \rightarrow 1 & \quad \text{at rate } \beta \sum_{(j,i) \in E} X_j, \\ X_i : 1 \rightarrow 0 & \quad \text{at rate } \delta. \end{aligned}$$

Henceforth, without loss of generality, we set  $\delta = 1$ . Denote by  $A$  the adjacency matrix of the graph structure on the set of sites. Denote by  $\rho(A)$  the spectral radius of  $A$ , namely, its largest eigenvalue. Observe that the Markov process  $X$  is such that one can reach the absorbing state 0 starting from any state  $x$ . Thus, it is the case that epidemics always die out. Even more is true: the probability that they have not died out by time  $t$  will decay exponentially with  $t$ . This fact follows from standard theory of Markov processes with absorbing states, reviewed for instance in [4]. The question of interest is then: how quickly do the epidemics die out, or how quickly does the system recover from the epidemic? More precisely, define  $\tau$  to be the time until the epidemic dies out provided there is at least one infected node initially. We are interested in determining the behavior of  $E[\tau]$  as a function of the system size  $n$ , whether it dies out quickly, i.e.,  $E[\tau] = O(\log n)$ , or slowly, i.e.,  $\log E[\tau] = \Omega(n^\alpha)$  for some  $\alpha > 0$ . Although parameters of interest such as  $\beta$  may depend on the system size  $n$ , we do not make this dependency explicit in our notation.

## III. A SUFFICIENT CONDITION FOR FAST RECOVERY

We show that the following condition,

$$\rho(A) < \frac{1}{\beta}, \quad (1)$$

implies that the epidemic dies out fast, and the characteristic time before extinction is related to the difference between the two terms in (1). Note that the threshold condition (1) is the one proposed in [18].

More precisely, we have the following result.

*Theorem 3.1:* Suppose condition 1 holds. Then, the probability that the epidemic has not died out by time  $t$ , given the initial condition  $X(0) \in \{0, 1\}^V$ , admits the following upper bound:

$$\mathbf{P}(X(t) \neq 0) \leq \sqrt{n \|X(0)\|_1} e^{(\beta\rho(A)-1)t}, \quad (2)$$

where  $\|X(0)\|_1 = \sum_{i=1}^n X_i(0)$ . In addition, under the condition (1), the time to extinction  $\tau$  verifies

$$\mathbf{E}(\tau) \leq \frac{\log(n) + 1}{1 - \beta\rho(A)} \quad (3)$$

for any initial condition,  $X(0)$ .

*Proof:* Consider the continuous time Markov process  $Y = \{Y_i\}_{i \in V}$ , with values in  $\mathbb{N}^V$ , and transition rates

$$\begin{aligned} Y_i : k &\rightarrow k + 1 && \text{at rate } \beta \sum_{(i,j) \in E} Y_j, \\ Y_i : k &\rightarrow k - 1 && \text{at rate } Y_i. \end{aligned}$$

Standard coupling arguments yield  $X(t) \leq_{st} Y(t)$  for all  $t \geq 0$  when starting from the same initial conditions; here,  $X \leq_{st} Y$  denotes that  $Y$  stochastically dominates  $X$ . This implies that

$$\mathbf{P}\left(\sum_i X_i(t) = 0\right) \geq \mathbf{P}\left(\sum_i Y_i(t) = 0\right).$$

Moreover, it holds that:

$$\mathbf{P}\left(\sum_i Y_i(t) > 0\right) \leq \sum_i \mathbf{E}(Y_i(t)).$$

However, the transition rates for process  $Y$  are such that

$$\frac{d}{dt} \mathbf{E}(Y(t)) = (\beta A - I) \mathbf{E}(Y(t)),$$

where  $I$  denotes the identity matrix. Hence,

$$\mathbf{E}(Y(t)) = \exp(tM) Y(0),$$

where  $M = \beta A - I$ . Note that  $\exp(tM)$  is a symmetric matrix (since  $A$  is) with spectral radius  $e^{(\beta\rho(A)-1)t}$ . Hence,

$$\|\mathbf{E}(Y(t))\|_2 \leq e^{(\beta\rho(A)-1)t} \|\mathbf{E}(Y(0))\|_2,$$

where  $\|\mathbf{E}(Y(t))\|_2 = \sqrt{\sum_{i=1}^n \mathbf{E}(Y_i(t))^2}$ . But, by the Cauchy-Schwarz inequality,

$$\sum_{i \in V} \mathbf{E}(Y_i(t)) \leq \|\mathbf{E}(Y(t))\|_2 \|\mathbf{1}\|_2,$$

where  $\mathbf{1}$  denotes the vector of ones, so  $\|\mathbf{1}\|_2 = \sqrt{n}$ . Thus, we obtain:

$$\begin{aligned} \mathbf{P}(X(t) \neq 0) &\leq \sum_i \mathbf{E}(Y_i(t)) \\ &\leq \sqrt{n} e^{(\beta\rho(A)-1)t} \sqrt{\sum_{i=1}^n Y_i(0)^2}. \end{aligned}$$

Equation (2) follows, upon replacing  $Y_i(0)$  by  $X_i(0)$ ; note that  $X_i(0)^2 = X_i(0)$  since  $X_i(0)$  takes values in  $\{0, 1\}$ .

Now, write

$$\mathbf{E}(\tau) = \int_0^\infty \mathbf{P}(\tau > t) dt = \int_0^\infty \mathbf{P}(X(t) \neq 0) dt.$$

Replacing  $\mathbf{P}(X(t) \neq 0)$  by the trivial upper bound 1 on  $[0, (\log(n))/(1 - \beta\rho(A))]$ , and by the upper bound  $n \exp((\beta\rho(A) - 1)t)$  on the interval  $[(\log(n))/(1 - \beta\rho(A)), \infty)$ , in this expression yields the claimed result (3).  $\blacksquare$

In other words, under condition (1), the probability that the epidemic hasn't died out by time  $\frac{\log(n)}{1 - \beta\rho(A)} + t$  decays exponentially in  $t$ .

However, Condition (1) may not be necessary for ensuring exponentially fast death of epidemics, with small characteristic time before extinction. We will observe this in the case of a star-shaped network in Section V-A.

#### IV. SUFFICIENT CONDITION FOR LASTING INFECTION

We now provide a condition under which an epidemic will survive for a long time. To this end, we introduce the so-called *generalized isoperimetric constant* of the graph  $G$ , defined as:

$$\eta(G, m) = \inf_{S \subset \{1, \dots, n\}, |S| \leq m} \frac{E(S, \bar{S})}{|S|}, \quad 0 < m \leq \lfloor n/2 \rfloor. \quad (4)$$

In the above,  $E(S, \bar{S})$  denotes the number of edges connecting the set of vertices  $S$  to the complementary set,  $\bar{S}$ . When  $m = \lfloor n/2 \rfloor$ , this corresponds to the standard isoperimetric constant  $\eta(G)$ . Henceforth we omit the reference to  $G$ . Define now a Markov process  $Z$  on  $\{0, \dots, m\}$ , with transition rate:

$$\begin{aligned} z &\rightarrow z + 1 : && \text{at rate } \eta(m)\beta z, \\ z &\rightarrow z - 1 : && \text{at rate } z. \end{aligned}$$

Now, standard coupling arguments yield  $\{Z(t)\} \leq_{st} \{\sum_{i \in V} X_i(t)\}$  due to the fact that (i)  $Z(t)$  can never exceed  $m$ , (ii) their downward transition rates are identical, and (iii) the upwards transition rate for  $\{Z(t)\}$  is  $\eta(m)\beta z$ , that for  $\{\sum_{i \in V} X_i(t)\}$  is  $E(S, \bar{S})\beta$  where  $S = \{i \in V : X_i(t) = 1\}$ , and we have  $E(S, \bar{S}) \geq \eta(m)z$  whenever  $|S| = z \leq m$ .

This is useful for establishing the following result.

*Theorem 4.1:* Assume that the following inequality holds:

$$r := \frac{1}{\beta\eta(m)} < 1. \quad (5)$$

Then for any initial condition  $X(0)$  with  $\sum_{i=1}^n X_i(0) > 0$ , it holds that,

$$\mathbf{P}\left(\tau > \frac{\lfloor r^{-m+1} \rfloor}{2m}\right) \geq \frac{1-r}{e} (1 + O(r^m)). \quad (6)$$

*Proof:* First consider the embedded discrete time Markov chain associated with process  $Z$ , which tracks the successive states visited by  $Z$ . We denote it by  $Z_t$ , where now  $t \in \mathbb{N}$ . Its transition matrix has as non-negative terms:

$$\begin{aligned} p(i, i+1) &= \frac{\beta\eta(m)}{\beta\eta(m)+1}, && i = 1, \dots, m-1, \\ p(i, i-1) &= \frac{1}{\beta\eta(m)+1}, && i = 1, \dots, m, \\ p(0, 0) &= 1, \\ p(m, m-1) &= 1. \end{aligned}$$

Let  $q(k)$  denote the probability that, starting from state  $k$ , the Markov chain  $Z_n$  will enter state  $m$  before entering state 0. The evaluation of  $q(k)$  is a standard problem, known as the gambler's ruin problem. It has the following solution:

$$q(k) = \frac{1 - r^k}{1 - r^m}, \quad k = 1, \dots, m - 1,$$

where  $r = 1/(\beta\eta(m))$ . Let  $T$  denote the number of steps before  $X_t$  is absorbed at 0. The solution of the gambler's ruin problem implies that:

$$\mathbf{P}(T > t) \geq \frac{1 - r}{1 - r^m} \left( \frac{1 - r^{m-1}}{1 - r^m} \right)^t. \quad (7)$$

Indeed,  $T$  is going to be larger than  $t$  if  $Z$  pays at least  $t$  distinct visits to  $m$  before being absorbed. The probability of the first visit taking place before absorption is the first term in the above expression; after a visit to  $m$ , the probability of having a subsequent visit prior to absorption is given by  $(1 - r^{m-1})/(1 - r^m)$  as can be seen by considering what happens in the step just after the visit to  $m$ , where the process is necessarily in state  $m - 1$ . The formula above follows. Taking  $t = \lfloor r^{-m+1} \rfloor$  in expression (7), the right-hand side is then at least  $(1 - r)/e (1 + O(r^m))$ . Conditioned on the event that there are indeed  $t$  visits to  $m$  before absorption at zero, the time before absorption is then larger than the sum of  $t$  exponentially distributed random variables with parameter  $m$ , describing the sojourn times in state  $m$ . When the number of such visits is large, the total sojourn time is then at least  $t/(2m)$ , with a probability  $1 - \epsilon$  where the error  $\epsilon$  is exponentially small in  $t$ , and the result (6) follows. ■

We have the following corollary.

*Corollary 4.1:* Consider a sequence of graphs indexed by  $n$ . Suppose there is an  $a > 0$  and a sequence  $m = \Theta(n^a)$  such that  $r < 1$  uniformly in  $n$ , for  $r$  defined as in (5). Then  $\log(\mathbf{E}[\tau]) = \Omega(n^a)$ .

*Proof:* We have

$$\begin{aligned} \mathbf{E}[\tau] &= \int_0^\infty \mathbf{P}(\tau > x) dx, \\ &\geq \int_0^{\lfloor r^{-m+1} \rfloor / (2m)} \mathbf{P}(\tau > x) dx, \\ &\geq \frac{\lfloor r^{-m+1} \rfloor}{2m} \frac{1 - r}{e} (1 + O(r^m)). \end{aligned}$$

Upon taking logarithms, this yields

$$\log(\mathbf{E}[\tau]) \geq \Theta(n^a) - O(\log(n)),$$

which establishes the corollary. ■

We now use Theorem 4.1 to provide sufficient conditions for exponentially long survival of the epidemics

stated in terms of the spectral structure of the underlying graph  $G$ . This will allow comparison with the former sufficient condition for fast extinction, (1).

The Laplacian matrix  $L$  of the graph  $G$  is by definition  $D - A$ , where  $D$  is the diagonal matrix with as entries the degrees of the vertices in  $G$ , and  $A$  is the adjacency matrix of  $G$ . We shall denote the eigenvalues of  $L$  by  $\lambda_1(L) \leq \lambda_2(L) \leq \dots \leq \lambda_n(L)$ . We now have:

*Corollary 4.2:* Let

$$r' := \frac{2}{\beta\lambda_2(L)} < 1. \quad (8)$$

Then (6) is still valid, with  $r$  replaced by  $r'$  and  $m = \lfloor n/2 \rfloor$ .

The proof follows from Corollary 3.8 in [14], which states that for any graph  $G$ , the following inequality holds:

$$\eta(G) \geq \frac{\lambda_2(L)}{2}.$$

*Corollary 4.3:* Assume that the graph  $G$  is regular, i.e. all its vertices have the same degree, say  $d$ . Denote by  $\lambda_1(A) \leq \dots \leq \lambda_n(A)$  the eigenvalues of its adjacency matrix  $A$ . Then a sufficient condition for exponentially long survival of the epidemics is that:

$$2 < \beta(d - \lambda_{n-1}) = \beta(\rho(A) - \lambda_{n-1}). \quad (9)$$

Indeed, Condition (9) coincides with (8) in the case of a  $d$ -regular graph. Ignoring the factor 2 in (9), this condition amounts to comparing the ratio  $1/\beta$  not to  $\rho(A)$ , as is done in Condition (1), but rather to the gap between the two largest eigenvalues of  $A$ , namely  $\rho(A) - \lambda_{n-1}$ .

## V. SOME EXAMPLES

In this section we apply the previous results to several important classes of topologies: star-shaped networks, hypercubes, cliques, Erdős-Rényi random graphs, and random power law graphs.

### A. Star-shaped networks

The star-shaped network is of interest for several reasons. First, it illustrates that neither of the preceding conditions are sufficient. We are able to get much tighter conditions because the star is simple enough to study exhaustively. Second, the spreading behavior of a large class of power law graphs is determined by the spreading behavior of stars embedded within them, as we will observe in Section V-E.

Consider a star-shaped network, with  $n + 1$  nodes, where the only edges are  $(0, i)$ ,  $i = 1, \dots, n$ . Let us first

identify the spectral radius of the corresponding matrix  $A$ . An eigenvector  $x$  associated with eigenvalue  $\lambda$  must satisfy

$$\lambda x_0 = \sum_{i=1}^n x_i, \quad \lambda x_i = x_0.$$

The only solutions for  $\lambda$  are 0 and  $\pm\sqrt{n}$ , and thus  $\rho(A) = \sqrt{n}$ . The results from the previous two sections tell us that

- 1) epidemics die out quickly provided that  $\beta < 1/\sqrt{n}$ , and
- 2) epidemics die out slowly provided that  $\beta > 1$ .

These results are clearly quite loose. In this section we derive tighter thresholds for fast and slow die out.

We describe the state of the epidemic on a star-shaped network by a vector  $(X(t), N(t))$  of length two. Here  $X(t)$  is the indicator that the hub is infected, and  $N(t)$  is the number of infected leaves, at time  $t$ . The pairwise infection rate is  $\beta$ , while each infected node recovers at rate 1. Thus, the process  $(X(t), N(t))$  is Markovian, with transition rates

$$\begin{aligned} (0, i) &\rightarrow (1, i), \text{ rate } \beta i, & (1, i) &\rightarrow (0, i), \text{ rate } 1, \\ (0, i) &\rightarrow (0, i-1), \text{ rate } i & (1, i) &\rightarrow (1, i-1), \text{ rate } i, \\ (1, i) &\rightarrow (1, i+1), \text{ rate } \beta(n-i); \end{aligned}$$

all other transition rates equal zero. Standard coupling arguments establish  $(X(t), N(t))$  to be stochastically nondecreasing in  $(X(0), N(0))$ . During each interval over which the state of the hub is fixed at 1, the leaves evolve as independent Markov chains on  $\{0, 1\}$  with transition rate  $\beta$  from 0 to 1, and 1 from 1 to 0. Denote the transition rate matrix by  $Q^1$ . A straightforward calculation yields that the transition probability matrix is

$$\begin{aligned} P^1(t) &= e^{Q^1 t} \\ &= \frac{1}{1+\beta} \begin{pmatrix} 1 + \beta e^{-(1+\beta)t} & \beta(1 - e^{-(1+\beta)t}) \\ 1 - e^{-(1+\beta)t} & \beta + e^{-(1+\beta)t} \end{pmatrix} \end{aligned} \quad (10)$$

Define the process  $\tilde{N}(t)$  to be the second component of the Markov process  $(X(t), N(t))$ , conditioned on  $X(s) = 1$  for all  $s \in (0, \infty)$ . It is easy to see by a coupling argument that, for any  $t > 0$ ,  $N(t) \leq_s \tilde{N}(t)$  provided  $N(0) \leq_s \tilde{N}(0)$ ; for random variables  $X, Y$ , we write  $X \leq_s Y$  to mean that  $X$  is stochastically dominated by  $Y$ . But conditioning on  $\tilde{X}(s) \equiv 1$  ensures that the transition probabilities for each leaf are given by (10). Hence, if we take  $\tilde{N}(0) = n$ , then  $\tilde{N}(t)$  is binomial with parameters  $n$  and  $(\beta + e^{-(1+\beta)t})/(1+\beta)$ . Since the latter is a decreasing function of  $t$ , we also note that  $\tilde{N}(s)$  stochastically dominates  $\tilde{N}(t)$  for any  $s \leq t$ . In fact, if  $T \geq s$  is a random time which is independent of the  $N(\cdot)$  process, then  $N(s)$  stochastically dominates  $N(T)$ . The

following theorem shows that (1) is not sufficient for fast die out.

*Theorem 5.1:* Suppose  $\beta = C/\sqrt{n}$  for a fixed  $C > 0$ . Then,  $\mathbf{E}[\tau] = O(\log n)$  when either  $X(0) = 1$  and/or  $N(0) \geq 1$ .

*Proof:* As  $(X(t), N(t))$  is nondecreasing in the size of the initial population of infectives, it suffices to consider the initial state  $(1, n)$ . Fix  $\epsilon > 0$ , and take  $t = -\frac{\log(\beta\epsilon)}{1+\beta}$ , so that  $\frac{\beta + e^{-(1+\beta)t}}{1+\beta} = \frac{(1+\epsilon)\beta}{1+\beta}$ . Define

$$T_1 = \inf\left\{s \geq -\frac{\log(\beta\epsilon)}{1+\beta} : X(s) = 0\right\} \quad (11)$$

to be the first time after  $t$  that the central hub becomes uninfected in the original Markov process  $(X(t), N(t))$ . By the arguments above,  $N(T_1)$  is stochastically dominated by  $\tilde{N}(t)$ . Note that  $T_1 - t$  is stochastically dominated by the time for the hub to become uninfected (assuming that it is infected at time  $t$ ), which is an exponentially distributed random variable with mean 1, denoted  $U_1$ . Since  $\beta = C/\sqrt{n}$ , we have

$$T_1 \leq \frac{\log n}{2} - \log(\epsilon C) + U_1, \quad \mathbf{E}[T_1] \leq \frac{\log n}{2} - \log(\epsilon C) + 1. \quad (12)$$

Next, observe that, conditioned on  $N(T_1) = k$ , the probability that the infection dies out before the hub becomes infected again is  $1/(1+\beta)^k$ . This is because each leaf becomes uninfected before infecting the hub with probability  $1/(1+\beta)$ , and the infection attempts of different leaves are independent. Let  $S_1 = \inf\{t > T_1 : X(t) = 1\}$ , with the convention that  $S_1 = \infty$  if the infection dies out before the centre is re-infected. Then,

$$\mathbf{P}(S_1 = \infty) = \mathbf{E}\left[\left(\frac{1}{1+\beta}\right)^{N(T_1)}\right] \geq \mathbf{E}\left[\left(\frac{1}{1+\beta}\right)^{\tilde{N}(t)}\right],$$

Recall that  $t = -\frac{\log(\beta\epsilon)}{1+\beta}$  and that  $\tilde{N}(t)$  is binomial with parameters  $n$  and  $\frac{(1+\epsilon)\beta}{1+\beta}$ . Hence, using the generating function for the binomial, we have

$$\begin{aligned} \mathbf{P}(S_1 = \infty) &\geq \left(1 - \frac{(1+\epsilon)\beta}{1+\beta} + \frac{(1+\epsilon)\beta}{1+\beta} \frac{1}{1+\beta}\right)^n \\ &= \left(1 - \frac{(1+\epsilon)\beta^2}{(1+\beta)^2}\right)^n. \end{aligned}$$

Substituting  $\beta = C/\sqrt{n}$ , we have for all  $n$  sufficiently large, that

$$q := \mathbf{P}(S_1 = \infty) \geq \frac{1}{2} \exp(-(1+\epsilon)C^2). \quad (13)$$

Next, we bound the time by which either the hub becomes re-infected or the infection dies out. Associate with each leaf  $i$  an exponential random variable of unit mean,  $Z_i$ , denoting the time at which this node would become uninfected if it was initially infected. Define

$\tilde{S}_1 = T_1 + \max_{i=1}^n Z_i$ . Then, either the infection has died out before  $\tilde{S}_1$ , which happens with probability at least  $q$  (as given by (13)), or the hub was re-infected at some time in  $(T_1, \tilde{S}_1)$ . On the latter event, the evolution of the process after  $\tilde{S}_1$  can be stochastically dominated by the evolution which starts at  $\tilde{S}_1$  in the state  $(1, n)$ . But this is the state we considered at time 0, so  $\tilde{S}_1$  is a regeneration time for the dominating process. We have thus shown that the time to extinction is stochastically dominated by the sum of  $1 + V$  independent copies of  $\tilde{S}_1$ , where  $V$  is geometric with mean  $(1 - q)/q$ .

Now consider  $\tilde{S}_1 - T_1 = \max_{i=1}^n Z_i$ . We have

$$\mathbf{P}\left(\max_{i=1}^n Z_i \geq \log n + u\right) = 1 - \left(1 - \frac{e^{-u}}{n}\right)^n \leq e^{-u}.$$

In other words,  $\tilde{S}_1 - T_1 - \log n$  is stochastically dominated by an exponential random variable with mean 1, denoted  $U_2$ . We now have from (12) that

$$\begin{aligned} \tilde{S}_1 &\leq \frac{3}{2} \log n - \log(\epsilon C) + U_1 + U_2, \\ \mathbf{E}[\tilde{S}_1] &\leq \frac{3}{2} \log n - \log(\epsilon C) + 2, \end{aligned} \quad (14)$$

where  $U_1$  and  $U_2$  are independent  $Exp(1)$  random variables defined on the same probability space as  $\tilde{S}_1$ . Finally, the time to extinction,  $\tau$ , satisfies

$$\tau \leq \sum_{j=1}^{1+V} \tilde{S}_j, \quad \mathbf{P}(V = j) = q(1 - q)^j, \quad j = 0, 1, 2, \dots \quad (15)$$

where  $\tilde{S}_j$  are iid and independent of  $V$ . Here  $q$  is given by (13), and is bounded below by a constant that doesn't depend on  $n$ . It is now immediate from (14) and (15) that  $\mathbf{E}[\tau] = O(\log n)$ . This completes the proof of the theorem.  $\blacksquare$

Next, suppose  $\beta = n^{\alpha - \frac{1}{2}}$ , for some  $\alpha \in (0, \frac{1}{2})$ . Define  $K = \frac{\beta n}{4(1+\beta)}$ , and let

$$\mathcal{C} = \bigcup_{k=K}^n \{(0, k) \cup (1, k)\},$$

be the set of states in which  $K$  or more leaves are infected. We shall show that, starting from any state in  $\mathcal{C}$ , the Markov chain  $(X(t), N(t))$  returns to the set  $\mathcal{C}$  before hitting  $(0, 0)$  with high probability.

**Theorem 5.2:** Suppose  $\beta = n^{\alpha - \frac{1}{2}}$ , for some  $\alpha \in (0, \frac{1}{2})$  and  $\mathcal{C}$  is defined as above. Then:

1) There is a constant  $\kappa > 0$  such that, for all  $n$  sufficiently large and all  $x \in \mathcal{C}$ ,

$$\begin{aligned} \mathbf{P}_x((X(t), N(t)) \text{ hits } (0, 0) \text{ before returning to } \mathcal{C}) \\ \leq \exp(-\kappa n^\alpha). \end{aligned}$$

Here,  $\mathbf{P}_x$  denotes probabilities for the Markov chain  $(X(t), N(t))$ , conditioned on  $(X(0), N(0)) = x$ .

2) For all  $x \neq (0, 0)$ ,  $\log \mathbf{E}_x[\tau] = \Omega(n^\alpha)$ .

The proofs of 1) and 2) rely on the following three lemmas, whose proofs can be found in the Appendix.

The first lemma states that, for the Markov chain started in  $(0, k)$ , with high probability the hub becomes infected before the number of infected leaves decreases by  $\sqrt{n}$  or more.

**Lemma 5.1:** Suppose  $(X(0), N(0)) = (0, k)$ , with  $k > \sqrt{n}$ . Let  $\sigma = \inf\{t > 0 : X(t) = 1\}$ . Then,

$$\mathbf{P}(N(\sigma) \leq k - \sqrt{n}) \leq \exp\left(-\frac{n^\alpha}{1 + \beta}\right).$$

Recall that  $K = \frac{\beta n}{4(1+\beta)}$ . Fix  $k < K$  and consider the Markov chain  $(X(t), N(t))$  started in the state  $(1, k)$ . The following lemma states that, with high probability, either  $X(t) = 0$  or  $N(t) = K$  before  $N(t) = k - \sqrt{n}$ . Note that, if  $N(t) = K$ , then  $(X(t), N(t)) \in \mathcal{C}$ .

**Lemma 5.2:** Suppose  $(X(0), N(0)) = (1, k)$ , with  $k \leq K$ . Let  $\sigma = \inf\{t > 0 : (X(t), N(t)) \in \mathcal{C} \text{ or } X(t) = 0\}$ . Then,

$$\mathbf{P}(N(t) \leq k - \sqrt{n} \text{ for some } t \in [0, \sigma]) \leq 4^{-\sqrt{n}}.$$

The last lemma states that the Markov chain  $(X(t), N(t))$ , starting in state  $(1, k)$  for any  $k$ , has probability at least one third of hitting the set  $\mathcal{C}$  before the hub becomes uninfected.

**Lemma 5.3:** Suppose  $(X(0), N(0)) = (1, k)$ , with  $k < K$ . Let  $\sigma = \inf\{t > 0 : X(t) = 0\}$ . Then,

$$\mathbf{P}((X(\sigma), N(\sigma)) \in \mathcal{C}) \geq \frac{1}{3},$$

for all  $n$  sufficiently large.

*Proof of Theorem 5.2:* Let  $x$  be an arbitrary state in the set  $\mathcal{C}$ . Define  $\tau_{\mathcal{C}}$  to be the return time to the set  $\mathcal{C}$ ;  $\tau_{\mathcal{C}} = 0$  if the first transition out of state  $x$  takes the chain to another state in  $\mathcal{C}$ . We want to show that  $\mathbf{P}(\tau \leq \tau_{\mathcal{C}}) \leq \exp(-\kappa n^\alpha)$ .

We recursively define the times  $T_i$  and  $S_i$  as follows:  $T_0 = \inf\{t > 0 : X(t) = 0\}$ , which could possibly be zero; for  $i \geq 1$ ,  $S_i = \inf\{t > T_{i-1} : X(t) = 1\}$  and  $T_i = \inf\{t > S_i : X(t) = 0\}$ . Note that the state  $(0, 0)$  is absorbing. Hence, if  $N(T_i) = 0$  for some  $i$ , then we define  $S_j$  and  $T_j$  to be  $\infty$  for all  $j > i$ . Now, by Lemmas 5.1 and 5.2, we have

$$\mathbf{P}\left(N(T_0) \geq \frac{\beta n}{4(1+\beta)} - \sqrt{n}\right) \geq 1 - \exp(-\sqrt{n} \log 4),$$

and, for all  $k \geq 1$ ,

$$\begin{aligned} P\left(N(S_k) \geq \frac{\beta n}{4(1+\beta)} - 2k\sqrt{n} \mid \right. \\ \left. N(T_{k-1}) \geq \frac{\beta n}{4(1+\beta)} - (2k-1)\sqrt{n}\right) \\ \geq 1 - \exp\left(-\frac{n^\alpha}{1+\beta}\right), \\ P\left(N(T_k) \geq \frac{\beta n}{4(1+\beta)} - (2k+1)\sqrt{n} \mid \right. \\ \left. N(S_k) \geq \frac{\beta n}{4(1+\beta)} - 2k\sqrt{n}\right) \\ \geq 1 - \exp(-\sqrt{n} \log 4). \end{aligned}$$

Now,  $\beta = n^{\alpha-\frac{1}{2}}$ , so  $\frac{\beta n}{4(1+\beta)} \frac{1}{2\sqrt{n}} = \frac{n^\alpha}{8(1+\beta)} \geq \frac{n^\alpha}{9}$  for all  $n$  sufficiently large. Thus, we obtain by the union bound that

$$P(T_{n^\alpha/9} < \infty) \geq 1 - \left(\frac{2n^\alpha}{9} + 1\right) \exp\left(-\frac{n^\alpha}{1+\beta}\right). \quad (16)$$

Note that, on this event,  $\tau > T_{n^\alpha/9}$ . Moreover, on each interval  $[S_k, T_k]$ ,  $k \leq n^\alpha/9$ , there is probability at least  $\frac{1}{3}$  of hitting the set  $\mathcal{C}$ , by Lemma 5.3. Hence,

$$\begin{aligned} P(\tau_{\mathcal{C}} \geq T_{n^\alpha/9} | T_{n^\alpha/9} < \infty) &\leq \left(\frac{2}{3}\right)^{n^\alpha/9} \\ &= \exp\left(-\frac{\log(3/2)}{9} n^\alpha\right). \end{aligned}$$

Combining this with (16), we get

$$\begin{aligned} P(\tau_{\mathcal{C}} < \tau) &\geq P(\tau_{\mathcal{C}} < T_{n^\alpha/9} \leq \tau) \\ &\geq 1 - P(\tau_{\mathcal{C}} \geq T_{n^\alpha/9}) - P(T_{n^\alpha/9} > \tau) \\ &= P(T_{n^\alpha/9} < \infty) - P(\tau_{\mathcal{C}} \geq T_{n^\alpha/9}) \\ &\geq 1 - \left(\frac{2n^\alpha}{9} + 1\right) \exp\left(-\frac{n^\alpha}{1+\beta}\right) \\ &\quad \times \exp\left(-\frac{\log(3/2)}{9} n^\alpha\right). \end{aligned}$$

Taking  $\kappa = \frac{1}{2} + \frac{\log(3/2)}{9}$  for instance, we see that the first claim of the theorem is satisfied.

To establish the second claim, observe that the Markov chain started in any state of the form  $(0, k)$  with  $k > 0$  dominates the chain started in the state  $(0, 1)$ . Likewise, the chain started in  $(1, k)$  dominates that started in  $(1, 0)$ . Thus, we need only consider the initial conditions  $(0, 1)$  and  $(1, 0)$ . Now, if the chain is started in  $(0, 1)$ , it hits  $(1, 1)$  before  $(0, 0)$  with probability  $\beta = n^{\alpha-\frac{1}{2}}$ , and subsequently dominates the chain started in  $(1, 0)$ . Moreover, by Lemma 5.3, the chain started in  $(1, 0)$  hits the set  $\mathcal{C}$  before hitting  $(0, 0)$ , with probability at least a third. Consequently,  $P_x(\tau_{\mathcal{C}} < \tau) \geq n^{\alpha-\frac{1}{2}}/3$  for all  $x \neq (0, 0)$ . Next, starting from any state  $y \in \mathcal{C}$ , we have shown that  $P_y(\tau < \tau_{\mathcal{C}}) \leq \exp(-\kappa n^\alpha)$ . Consequently,

the number of returns to  $\mathcal{C}$  before hitting  $(0, 0)$  stochastically dominates a geometric random variable with mean  $\exp(\kappa n^\alpha)$ . The time for each return to  $\mathcal{C}$  dominates the time for a single transition from state  $y$ ; for all  $y \in \mathcal{C}$ , this time dominates an exponential random variable with mean  $1/n$ . Consequently,

$$\begin{aligned} \min_{y \in \mathcal{C}} E_y[\tau] &\geq \frac{1}{n} e^{\kappa n^\alpha}, \\ \min_{x \neq (0,0)} E_x[\tau] &\geq \frac{1}{3} n^{\alpha-\frac{1}{2}} e^{\kappa n^\alpha} \geq e^{\kappa' n^\alpha}, \end{aligned}$$

for large enough  $n$  and suitably chosen  $\kappa' > 0$ ;  $\kappa' = \kappa/2$  will suffice. Thus  $\log \mathbf{E}_x[\tau] = \Omega(n^\alpha)$ . This completes the proof of the theorem.  $\blacksquare$

*Remark 5.1:* It is interesting to note that the system behavior is greatly affected by the initial set of infected hosts in the following way. If  $X(0) = 0$  and  $N(0) = o(n^{1/2-\alpha})$ , then the probability that the hub becomes infected goes to zero as  $n \rightarrow \infty$ . In particular, if  $T = \inf\{s : X(s) = 1\}$ , then  $\mathbf{P}(T < \infty) < A e^{-N(0)/n^{1/2-\alpha}}$  for  $A > 0$  and large  $n$ . Furthermore, these epidemics are of short duration,  $O(\log n)$ . However, even though nearly all epidemics die out (with high probability), the rare occasions that they do not yield epidemics of such long durations that the average epidemic length is exponential in  $n$ . On the other hand, if either  $X(0) = 1$  or  $N(0) = \Omega(n^{1/2-\alpha})$ , then a non-zero fraction of the epidemics do not die out quickly, even as  $n \rightarrow \infty$ .

## B. Hypercubes

The hypercube is of interest because of the widespread and growing interest in distributed hash tables and applications, such as file sharing [17], being built on top of them. Already worms and viruses have appeared in some applications, [11]. As many DHT structures are hypercubic in nature, it is important to understand the spreading behavior of such worms on a hypercube. Here we represent a hypercube as a graph  $G$  with vertex set  $\{0, 1\}^\ell$  for some  $\ell \in \mathbb{N}$ , and where the edge  $(v, w)$  is present if and only if the Hamming distance  $d_H(v, w)$  equals 1. As a hypercube is a regular graph, its spectral radius is  $\rho(G) = \log_2 n$ . Hence, from Section III we know that the epidemic dies out provided that  $\beta < 1/\log_2 n$ . We now determine conditions for the epidemic to die out slowly. In particular, we establish that the epidemic dies out slowly provided  $\beta > \frac{1}{(1-a)\log_2 n}$ , for any  $0 < a < 1$ . Thus, unlike the star topology, there is essentially no gap for the hypercube.

We first establish the following result.

*Theorem 5.3:* Let  $m = 2^k$ , for some  $k \in \{0, \dots, \ell - 1\}$ . Assume that the following inequality holds:

$$r := \frac{1}{\beta(\ell - k)} < 1. \quad (17)$$

Assume further that  $r^m$  goes to zero as  $n \rightarrow \infty$ . Then the time to extinction satisfies identity (6), with our current choices for  $r$  and  $m$ .

The result follows by a direct application of Theorem 4.1. In order to prove the theorem, we need the following result established by Harper [10] (see also [1] for background and recent extensions).

*Lemma 5.4:* (Harper [10]) Let  $S$  be a set of  $m$  vertices of the hypercube  $\{0, 1\}^\ell$ . Then the edge-boundary size  $E(S, \bar{S})$  is larger than or equal to  $E(S^*, \bar{S}^*)$ , where  $S^*$  denotes the set of the  $m$  smallest vertices according to lexicographic order.

*Proof:* (of Theorem 5.3) Let  $m = 2^k$ , for some  $k < \ell$ . Let us first establish the following inequality:

$$\eta(m) \geq \ell - k. \quad (18)$$

Indeed, in view of the lemma it is the case that the set  $S^*$  achieving the minimum isoperimetric ratio  $\eta(m)$  consists of all points  $0^{\ell-j+1}x_1^{j-1}$ , where  $x_1^{j-1}$  spans the whole set  $\{0, 1\}^{j-1}$ , plus some points of the type  $0^{\ell-j}1x_1^{j-1}$ , for some  $j \leq k$ . It turns out that any point  $y_1^{\ell-j}z_1^j$ , where  $\sum_{p=1}^{\ell-j} y_p = 1$  and  $0^{\ell-j}z_1^j$  belongs to  $S^*$ , contributes exactly one edge to  $E(S^*, \bar{S}^*)$ . Hence the lower bound

$$E(S^*, \bar{S}^*) \geq (\ell - j)|S^*|,$$

from which (18) directly follows.

The result follows by applying Theorem 4.1.  $\blacksquare$

Now reverting to the notation  $n$  and  $m$ , condition (17) can be expressed as

$$\beta > \frac{1}{\log_2 n - \log_2 m}.$$

Take  $m = n^a$  for some  $a$  such that  $0 < a < 1$ , we obtain

$$\beta > \frac{1}{(1 - a)\log_2 n}.$$

### C. Complete graph

The complete graph also plays a prominent role in networks. For example, the BGP routers belonging to the top level autonomous systems of the Internet form a completely connected component. In addition, large ISPs often organize their internal BGP (iBGP) routers into a set of *route reflectors* that are completely connected. Because of the importance of BGP to the Internet, it is important to understand what effect BGP router failures can have on each other. At a very high level, the behavior

of routers failing and coming up can be modeled as the spread of an epidemic within a complete graph, [7].

Consider a complete graph with  $n$  vertices, namely the graph where an edge is present between each pair of nodes. The adjacency matrix of this graph is  $A = \mathbf{1}\mathbf{1}^T - I$ , where  $\mathbf{1}$  denotes the column vector of ones, and  $\mathbf{1}^T$  is its transpose.  $A$  has the spectral radius  $\rho(A) = n - 1$ . The isoperimetric constant  $\eta(m)$  is easily shown to be  $n - m$ . Application of Theorem 3.1 and Corollary 4.1 tell us that epidemics die out quickly when  $\beta < 1/(n - 1)$  and slowly when  $\beta > 1/(n - m)$ , when  $m = n^a$  for some  $a > 0$ . Thus  $1/n$  is essentially the correct threshold and, as in the case of the hypercube, there is no gap.

### D. Erdős-Rényi random graphs

The Erdős-Rényi graph  $G(n, p)$  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. 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 [3]. Second, it is an important component of the class of power law random graphs that model the Internet AS graph. Thus if we are to understand the robustness of the Internet AS-level graph, we need to characterize the robustness of the Erdős-Rényi graph.

We shall consider a sequence of such graphs indexed by  $n$ . Denote by  $d$  the corresponding average degree, i.e.  $d = (n - 1)p$ . Note that  $p$  and  $d$  depend on  $n$ , but this is suppressed in the notation. We say that a property holds with high probability if its probability goes to 1 as  $n \rightarrow \infty$ .

We consider the regime  $\log(n) \ll d$ , i.e.,  $\log(n)/d \rightarrow 0$  as  $n \rightarrow \infty$ . In this case, it is known that the graphs are connected with high probability. Moreover, by the Perron-Frobenius theorem, the spectral radius  $\rho(A)$  lies between the smallest and largest node degree. Since the node degrees are binomial,  $B_{n-1, p}$ , it is easy to see using the Chernoff bound that  $\rho(A) = (1 + o(1))np = (1 + o(1))d$ , with high probability.

*Theorem 5.4:* Let  $m$  be such that  $m/n \rightarrow \alpha$  as  $n \rightarrow \infty$ , for a fixed  $\alpha \in (0, 1)$ . Assume further that  $\log(n) \ll d$ . Then it holds that, with high probability,

$$\eta(G, m) = (1 + o(1))(1 - \alpha)d. \quad (19)$$

*Proof:* Fix some  $k \in (0, 1 - \alpha)$ . By the union bound,

$$\mathbf{P}(\eta(G, m) < kd) \leq \sum_{i=1}^m \sum_{S: |S|=i} \mathbf{P}(E(S, \bar{S}) < kdi).$$



Note that  $E(S, \bar{S})$  has a binomial distribution with parameters  $i(n-i)$  and  $p$ . Denote by  $B_{i(n-i),p}$  a generic Binomial random variable with this distribution. We thus have

$$\mathbf{P}(\eta(G, m) < kd) \leq \sum_{i=1}^m \binom{n}{i} \mathbf{P}(B_{i(n-i),p} < kdi).$$

Since  $k < 1 - \alpha$  and  $m = (1 + o(1))\alpha n$ , we have for  $n$  large enough that  $(n-i)p > kd$  for all  $i \in \{1, \dots, m\}$  and that  $\frac{n-1}{n-i} \leq 1/(1-\alpha)$ . We now apply the following Chernoff bound,

$$P(B < (1-\delta)\mathbf{E}(B)) < e^{-\mathbf{E}(B)\delta^2/2}, \quad (20)$$

valid for any Binomial random variable  $B$ , [13, Theorem 4.2], to  $B_{i(n-i),p}$  in the above expression. Taking  $\delta = 1 - k\frac{n-1}{n-i} \geq 1 - \frac{k}{1-\alpha}$ , this yields

$$\mathbf{P}(\eta(G, m) < kd) \leq \sum_{i=1}^m \binom{n}{i} e^{-i(n-i)p\epsilon}, \quad (21)$$

where  $\epsilon = \frac{1}{2} \left(1 - \frac{k}{1-\alpha}\right)^2 > 0$ . We thus have the upper bound

$$\begin{aligned} \mathbf{P}(\eta(G, m) < kd) &\leq \sum_{i=1}^m \frac{n^i}{i!} e^{-i(n-i)p\epsilon}, \\ &= \sum_{i=1}^m \frac{1}{i!} e^{-i((n-i)p\epsilon - \log n)} \end{aligned}$$

where we have also used the upper bound  $\binom{n}{i} \leq n^i/i!$  to obtain the first inequality. By assumption,  $\log(n) \ll d = (n-1)p$ , and  $\epsilon$  is a positive constant that doesn't depend on  $n$ . Thus, for any constant  $K > 0$ , it holds that, for  $n$  large enough,

$$\mathbf{P}(\eta(G, m) < kd) \leq \sum_{i=1}^m \frac{1}{i!} n^{-Ki} \leq \exp(n^{-K}) - 1.$$

This suffices to conclude that for any  $k < 1 - \alpha$ , with high probability  $\eta(G, m) \geq kd$ .

In order to obtain an inequality in the opposite direction, choose any set  $S$  of cardinality  $m$ . Then one certainly has

$$\eta(G, m) \leq \frac{E(S, \bar{S})}{m}.$$

Let  $k > (1 - \alpha)$  be fixed. One then has that

$$\mathbf{P}(\eta(G, m) > kd) \leq \mathbf{P}(B_{m(n-m),p} > km(n-1)p).$$

It is readily seen using a Chernoff bound that the right-hand side goes to zero as  $n \rightarrow \infty$ . This concludes the proof of the theorem. ■

*Theorem 5.5:* Consider Erdős-Rényi random graphs  $G(n, p)$  with  $\log(n) \ll d = np$ . The following claims

hold with high probability: an epidemic on  $G(n, p)$  dies out quickly,  $E[\tau] = O(\log n)$ , provided  $\beta < (1 - u)/d$  for  $0 < u < 1$ . On the other hand, the epidemic dies out slowly,  $\log E[\tau] = \Omega(n)$  provided that  $\beta > (1 + v)/d$  for  $v > 0$ .

*Proof:* The theorem follows from Theorem 3.1 and Corollary 4.1, the expression  $\rho(A) = (1 + o(1))d$  for the spectral radius, and Theorem 5.4. ■

Notice that there is very little gap between the two thresholds as  $u$  and  $v$  can be chosen to be arbitrarily close to zero.

### E. Power law 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, [9]. 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. The Internet AS-level graph is characterized by  $\gamma \approx 2.1$

In this section we consider a class of random power law graphs first introduced in [6]. Let  $w_1, w_2, \dots, w_n$  denote the expected degrees of the nodes in the graph. An edge is assigned to a pair of vertices with probability  $w_i w_j / \sum_{k=1}^n w_k$ . Let  $d$  denote the average degree and  $m$  the maximum degree. If  $w_i = c(i_0 + i)^{-\frac{1}{\gamma-1}}$  ( $1 \leq i \leq n$ ) where  $c = \frac{\gamma-2}{\gamma-1} d n^{\frac{1}{\gamma-1}}$ ,  $i_0 = n \left(\frac{d(\gamma-2)}{m(\gamma-1)}\right)^{\gamma-1}$  then the number of nodes with degree  $k$  is proportional to  $k^{-\gamma}$ . Theorem 4 from [6] states under mild conditions that, with high probability, the spectral radius of the graph is

$$\rho(A) = \begin{cases} (1 + o(1))\sqrt{m}, & \gamma > 2.5, \\ (1 + o(1)) \frac{d(\gamma-2)^2}{(\gamma-1)(3-\gamma)} \left(\frac{(\gamma-1)m}{(\gamma-2)d}\right)^{3-\gamma} & 2 < \gamma < 2.5 \end{cases} \quad (22)$$

We have the following result on the fast/slow die out of epidemics on the above class of power law graphs. Here we will take  $m$  to be the following increasing function of  $n$ ,  $m = n^\lambda$  where  $0 < \lambda < \frac{1}{\gamma-1}$ .

*Theorem 5.6:*

- 1) For  $\gamma \geq 2.5$ ,
  - (a) the infection dies out quickly,  $\mathbf{E}[\tau] = O(\log n)$ , provided that  $\beta < (1 - u)/\sqrt{m}$  for some  $1 > u > 0$ ,
  - (b) the infection dies out slowly,  $\log \mathbf{E}[\tau] = \Omega(n^\lambda)$ , provided that  $\beta > m^{\alpha-1/2}$  for some  $\alpha \in (0, 1)$ .
- 2) For  $\gamma \in (2, 2.5)$ ,

- (a) the infection dies out quickly,  $\mathbf{E}[\tau] = O(\log n)$ , provided that  $\beta\rho < (1 - u)$  where

$$\rho = d \frac{(\gamma - 2)^2}{(\gamma - 1)(3 - \gamma)} \left( \frac{(\gamma - 1)m}{(\gamma - 2)d} \right)^{3-\gamma}$$

for some  $0 < u < 1$ ,

- (b) the infection dies out slowly,  $\log \mathbf{E}[\tau] = \Omega(n^{1-\lambda(\gamma-1)})$  provided that (i)  $\beta\eta > (1 + u)$  where

$$\eta = d \frac{(\gamma - 2)^2 (3 - \gamma)^{\frac{3-\gamma}{\gamma-1}}}{2^{2/(\gamma-1)} (\gamma - 1)} \left( \frac{(\gamma - 1)m}{(\gamma - 2)d} \right)^{3-\gamma}$$

for some  $u > 0$ , and (ii) the following condition holds:

$$\eta \gg \log(i_0). \quad (23)$$

*Proof:* The first part of each of the claims follows from Theorem 3.1 and (22). Consider claim 1.b. We can bound the time to die out of an epidemic that starts in either the maximum degree node or one of its neighbors by the time to die out on a star of size  $m + 1$ . Now consider the case that some other randomly chosen node is initially infected. Theorem 4 in [5] states that, with high probability, the diameter of the power law graph is  $\Theta(\log n)$  when  $\gamma > 2$ . Hence the probability that an epidemic starting from a randomly infected node spreads to either the maximum degree node or its immediate neighbors is  $\Theta(1/n^{a \log n})$  for some  $a > 0$  independent of  $n$ . This combined with arguments used to establish Theorem 5.2 establishes the claim.

We focus on claim 2.b. We shall now identify a subgraph of the original power-law random graph that is an Erdős-Rényi random graph. Consider the first  $N$  nodes, where  $N < n$  is to be chosen, with respective weights  $w_{i_0}, \dots, w_{i_0+N}$ . Then the probability of an edge being present between any two of these nodes is at least as large as

$$\begin{aligned} p_N &:= \frac{(w_{i_0+N})^2}{nd} \\ &= \frac{c^2 (N + i_0)^{-2/(\gamma-1)}}{nd} \\ &= d \left( \frac{\gamma - 2}{\gamma - 1} \right)^2 \frac{1}{n} \left( \frac{n}{N + i_0} \right)^{2/(\gamma-1)}. \end{aligned}$$

The original graph thus contains an Erdős-Rényi graph with parameters  $N$  and  $p_N$ . Survival of the epidemic on the original graph will be at least as long as on the Erdős-Rényi subgraph provided that at least one node within the Erdős-Rényi component is initially infected. The average degree of the subgraph is then given by

$$(N - 1)p_N \sim d \left( \frac{\gamma - 2}{\gamma - 1} \right)^2 \frac{N}{n} \left( \frac{n}{N + i_0} \right)^{2/(\gamma-1)}.$$

Let us now take  $N = \epsilon i_0$ , for some positive  $\epsilon$ . The average degree of the induced Erdős-Rényi graph is then equivalent to

$$\begin{aligned} &(N - 1)p_N \\ &\sim d \left( \frac{\gamma - 2}{\gamma - 1} \right)^2 \frac{\epsilon}{(1 + \epsilon)^{2/(\gamma-1)}} \left( \frac{n}{i_0} \right)^{(3-\gamma)/(\gamma-1)} \\ &\sim d \left( \frac{\gamma - 2}{\gamma - 1} \right)^2 \frac{\epsilon}{(1 + \epsilon)^{2/(\gamma-1)}} \left( \frac{m(\gamma - 1)}{d(\gamma - 2)} \right)^{3-\gamma}. \end{aligned}$$

Simple calculus shows that this is maximized by setting  $\epsilon = (\gamma - 1)/(3 - \gamma)$ , for which choice we arrive at

$$\begin{aligned} (N - 1)p_N &\sim \\ \eta &:= d \frac{(\gamma - 2)^2 (3 - \gamma)^{\frac{3-\gamma}{\gamma-1}}}{2^{2/(\gamma-1)} (\gamma - 1)} \left( \frac{(\gamma - 1)m}{(\gamma - 2)d} \right)^{3-\gamma}. \end{aligned}$$

The die out result is a direct application of Theorem 5.4 and Corollary 4.1, which is applicable under the condition (23).

If the initially infected nodes are not within the Erdős-Rényi subgraph, then because the diameter of the power law graph is  $\Theta(\log n)$  with high probability, the probability that the epidemic dies out before infecting a node within the Erdős-Rényi component is  $\Theta(1/n^{a \log n})$  for some  $a > 0$  independent of  $n$ . This combined with arguments used to establish Theorem 5.2 establishes the claim. ■

**Remark.** Note that in the case of  $\gamma \in (2, 2.5)$ , the thresholds for fast and slow recovery are almost the same, since  $\eta$  in the statement of Theorem 5.6 is smaller than  $\rho(A)$  by a factor of

$$\left( \frac{3 - \gamma}{2} \right)^{2/(\gamma-1)}.$$

Put together, our results determine the outcome of the epidemics for such graphs in the ranges  $\beta\rho(A) < 1$  (fast extinction) and  $\beta\rho(A) > [2/(3 - \gamma)]^{2/(\gamma-1)}$  (long survival). Our techniques do not cover the range in between at the present time.

## VI. SUMMARY

We have presented a preliminary investigation of how topology affects the spread of an epidemic, motivated by networking phenomena such as worms and viruses, cascading failures, and dissemination of information. We have developed sufficient conditions under which epidemics either die out quickly (logarithmically in the size of the network) or slowly (exponentially in the size of the network). These conditions are tight for several network topologies, such as hypercubes, complete

graphs, and Erdős-Rényi random graphs. They are not tight for topologies such as stars and power law graphs. We provided a supplementary analysis of the star that significantly tightened the condition for slow die out so that it is close to that for fast die out. In the case of a power law graph, we presented tight conditions by drawing on results for the star and Erdős-Rényi random graph. In all cases, the condition for fast die out appears to be tight.

There are several interesting directions to pursue this. As we remarked within the paper, the behavior of the epidemic in the case of the star and power law graph is sensitive to the initial set of infected nodes. It would be useful to understand this relationship better. In addition, as pointed out in [12], there is a notion of a *metastable* set of nodes that are infected given that one is infected in the regime of slow die out. They present an approximate analysis of the distribution of the number of nodes that belong to this set. It would be useful to pursue this in a more rigorous manner and for other classes of network topologies.

#### APPENDIX

##### PROOF OF LEMMAS 5.1, 5.2, 5.3

*Proof of Lemma 5.1:* When the Markov chain is in state  $(0, k)$ , the probability that it hits  $(0, k-1)$  before  $(1, k)$  is  $k/(k + \beta k) = 1/(1 + \beta)$ ; no other transitions are possible from  $(0, k)$ . Hence, by the Markov property,

$$\begin{aligned} P(N(\sigma) \leq k - \sqrt{n}) &= \prod_{j=k-\sqrt{n}+1}^k P_{(0,j)}(\text{hit } (0, j-1) \text{ before } (1, j)) \\ &= \left(\frac{1}{1+\beta}\right)^{\sqrt{n}}. \end{aligned}$$

Using the inequality  $\log x \leq x - 1$ , we obtain

$$P(N(\sigma) \leq k - \sqrt{n}) \leq \exp\left(-\frac{\sqrt{n}\beta}{1+\beta}\right) = \exp\left(-\frac{n^\alpha}{1+\beta}\right), \quad (24)$$

as claimed.  $\blacksquare$

*Proof of Lemma 5.2:* Observe that, while  $X(t) = 1$ ,  $N(t)$  evolves as a birth-death Markov chain with birth rate  $\beta(n - N(t))$  and death rate  $N(t)$ . Now, over the time interval  $[0, \sigma)$ ,  $X(t) = 1$  and  $N(t) \leq K$ . Hence, over  $[0, \sigma)$ ,  $N(t)$  stochastically dominates the Markov chain  $\tilde{N}(t)$  with birth rate  $\beta(n - K)$  and death rate  $K$ , having the same initial condition  $\tilde{N}(0) = k$ . Hence, it suffices to show that, with probability at least  $1 - 4^{-\sqrt{n}}$ ,  $\tilde{N}(t)$  hits  $K$  before it hits  $k - \sqrt{n}$ . Defining

$$\begin{aligned} M(t) &= \left(\frac{K}{\beta(n - K)}\right)^{\tilde{N}(t)}, \\ \tilde{\sigma} &= \inf\{t > 0 : \tilde{N}(t) = K \text{ or } \tilde{N}(0) - \sqrt{n}\}, \end{aligned}$$

we observe that  $M(t)$  is a martingale and that  $\tilde{\sigma}$  is a stopping time. Define  $\gamma = P(\tilde{N}(\tilde{\sigma}) = \tilde{N}(0) - \sqrt{n})$ . By the optional stopping theorem,  $EM(\tilde{\sigma}) = EM(0)$ , i.e.,

$$\begin{aligned} \gamma \left(\frac{K}{\beta(n - K)}\right)^{\tilde{N}(0) - \sqrt{n}} + (1 - \gamma) \left(\frac{K}{\beta(n - K)}\right)^K \\ = \left(\frac{K}{\beta(n - K)}\right)^{\tilde{N}(0)}. \end{aligned}$$

Now,  $K = \frac{\beta n}{4(1+\beta)}$ , so  $n - K = \frac{(4+3\beta)n}{4(1+\beta)} = \frac{(4+3\beta)K}{\beta} \geq \frac{4K}{\beta}$ . Hence, it is immediate from the above that

$$\gamma \leq \left(\frac{K}{\beta(n - K)}\right)^{\sqrt{n}} \leq 4^{-\sqrt{n}}. \quad (25)$$

This establishes the claim of the lemma.  $\blacksquare$

*Proof of Lemma 5.3:* Clearly, the Markov chain started in  $(1, k)$  stochastically dominates the chain started in  $(1, 0)$ , so it suffices to establish the claim of the lemma for  $k = 0$ . By conditioning on the time  $\sigma$  that it takes the hub to become uninfected, which is exponentially distributed with unit mean, we get

$$\begin{aligned} P(N(\sigma) \geq K) &= \int_0^\infty e^{-t} P(N(t) \geq K | N(0) = 0, \\ &\quad X(s) = 1 \forall s \in [0, t]) dt. \quad (26) \end{aligned}$$

Now, conditional on  $N(0) = 0$  and  $X(s) = 1$  on  $[0, t]$ ,  $N(t)$  is binomial with parameters  $n$  and  $\beta(1 - e^{-(1+\beta)t})/(1 + \beta)$ . Applying the Chernoff bound (20) with  $\delta = 1 - \frac{1}{4(1 - e^{-(1+\beta)t})}$  yields for  $t > \frac{\log 2}{1+\beta}$

$$\begin{aligned} P(N(t) < \frac{\beta n}{4(1+\beta)}) &< \exp\left(-n \frac{\beta(3 - 4e^{-(1+\beta)t})^2}{32(1+\beta)((1 - e^{-(1+\beta)t}))^2}\right), \\ &< \exp\left(-\frac{9n^{\alpha+1/2}}{16(1+\beta)}\right) \end{aligned}$$

Here  $t = \frac{\log 2}{1+\beta}$  minimizes the denominator and  $t = \infty$  maximizes the numerator of the exponent in the right hand expression of the first inequality thus yielding the second inequality.

$$\begin{aligned} P(N(\sigma) \geq K) &\geq \left(1 - \exp\left(-\frac{9n^{\alpha+1/2}}{16(1+\beta)}\right)\right) \int_{\frac{\log 2}{1+\beta}}^\infty e^{-t} dt \\ &\geq \frac{1}{3}, \quad (27) \end{aligned}$$

for all  $n$  sufficiently large. This establishes the claim of the lemma.  $\blacksquare$

## REFERENCES

- [1] M. C. Azizoglu and Ö. Egecioglu, "Extremal sets minimizing dimension-normalized boundary in Hamming graphs", *SIAM J. Discrete Math.*, **17** (2003), 219–236.
- [2] F. Ball, T. Britton and O. Lyne, "Stochastic multitype epidemics in a community of households: estimation and form of optimal vaccination schemes", *Mathematical Biosciences*, **191** (2004), 19–40.
- [3] B. Bollobás. *Random Graphs, 2nd Edition*. Cambridge University Press, 2001.
- [4] P. Brémaud, *Markov Chains, Gibbs fields, Monte Carlo simulation, and queues*, Springer, New York 1999.
- [5] F. Chung and L. Lu. "The average distances in random graphs with given expected degrees", *Internet Mathematics*, **1** (2003), 91–114.
- [6] F. Chung, L. Lu and V. Vu. "Eigenvalues of random power law graphs", *Annals of Combinatorics*, **7** (2003), 21–33.
- [7] E.G. Coffman Jr., Z. Ge, V. Misra and D. Towsley. "Network resilience: exploring cascading failures within BGP", *Proc. 40th Annual Allerton Conference on Communications, Computing and Control*, October 2002.
- [8] R. Durrett and X.-F. Liu. "The contact process on a finite set," *Ann. Prob.*, **16** (1988), 1158–1173.
- [9] M. Faloutsos, P. Faloutsos and C. Faloutsos. "On power-law relationships of the Internet topology", *Proc. ACM Sigcomm*, 1999.
- [10] L. Harper, "Optimal assignment of numbers to vertices", *J. Soc. Ind. Appl. Math.*, **12** (1964), 131–135.
- [11] [http://www.kazaa.com/us/help/known\\_virus.htm](http://www.kazaa.com/us/help/known_virus.htm)
- [12] J.O. Kephart and S.R. White. "Directed-graph epidemiological models of computer viruses," *Proc. 1991 IEEE Computer Society Symposium on Research in Security and Privacy* (1991), 343–359.
- [13] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
- [14] B. Mohar, "Some applications of Laplace eigenvalues of graphs", in *Graph Symmetry: Algebraic Methods and Applications*, Eds. G. Hahn and G. Sabidussi, NATO ASI Ser. C 497, Kluwer (1997), 225–275.
- [15] R. Pastor-Satorras and A. Vespignani. "Epidemic spreading in scale-free networks," *Physical Review Letters*, **86** (2001), 3200–3203.
- [16] R. Pastor-Satorras and A. Vespignani. "Epidemic dynamics in finite scale-free networks," *Physical Review E*, **65** (2002).
- [17] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems", *Proc. Middleware 2001*, Germany, November 2001.
- [18] Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint", *Proc. IEEE SRDS*, 2003.