

Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint

Yang Wang, Deepayan Chakrabarti, Chenxi Wang, Christos Faloutsos
yangwang, chenxi, deepayan, christos@andrew.cmu.edu
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA, 15232
Corresponding Author: Chenxi Wang

Abstract

How will a virus propagate in a real network? Does an epidemic threshold exist for a finite power-law graph, or any finite graph? How long does it take to disinfect a network given particular values of infection rate and virus death rate?

We answer the first question by providing equations that accurately model the virus propagation in *any* network including real and synthesized network graphs. We propose a general epidemic threshold condition that applies to *arbitrary* graphs: we prove that, under reasonable approximations, the epidemic threshold for a network is closely related to the largest eigenvalue of its adjacency matrix. Finally, for the last question, we show that infections tend to zero exponentially below the epidemic threshold.

We show that our epidemic threshold includes many known thresholds for special-case graphs (e.g., Erdős-Rényi, BA power-law, homogeneous); we show that the threshold tends to zero for infinite power-law graphs. Finally, we illustrate the predictive power of our model with extensive experiments on real and synthesized graphs. We show that our threshold condition holds for arbitrary graphs.

Keywords: Computer Virus, Security, Network Modeling, Epidemiological Models

1 Introduction

Computer viruses remain a significant threat to today’s networks and systems. Existing defense mechanisms typically focus on local scanning of virus signatures. While these mechanisms can detect and prevent the spreading of known viruses, they do little for globally optimal defenses. The recent proliferation of malicious code that spreads with virus code exacerbates the problem [10, 24, 25]. From a network dependability standpoint, the propagation of malicious code represents a particular form of fault propagation, which may lead to the ultimate demise of the network (consider distributed denial-of-service attacks). With the exception of a few specialized modeling studies [7, 8, 16, 19, 26], much still remains unknown about the propagation characteristics of computer viruses and the factors that influence them.

In this paper, we investigate epidemiological modeling techniques to reason about computer viral propagation. Epidemiological models have been used in several virus and worm studies. Kephart and White [7, 8] are among the first to propose epidemiology-based analytic models. Their studies, however, are based on topologies that do not represent modern networks. Staniford et al. [23] reported a study of the code red worm propagation, but did not attempt to create an analytic model. The more recent studies by Pastor-Satorras et al. [16, 17, 18, 19, 20] and Barabási et al. [2, 4] focused on mathematical models for power-law networks.

This work aims to develop a general analytic model of virus propagation. Specifically, we are interested in models that can capture the impact of the underlying topology but are not limited by it. We found that, contrary to prior beliefs, viral propagation is largely determined by intrinsic characteristics of the network. Our model holds for arbitrary graphs and renders surprisingly simple but accurate predictions.

The layout of this paper is as follows: section 2 gives a background review of previous models. In section 3, we describe our proposed model for virus propagation. We show that our model conforms better to simulation results than previous models over real-world networks. In section 4, we revisit the issue of epidemic threshold and present a new theory for arbitrary graphs—the epidemic threshold of a given network is related intrinsically to the first eigenvalue of its adjacency matrix. We summarize in section 6.

2 Earlier models and their limitations

The class of epidemiological models that is most widely used is the so-called *homogeneous models* [1, 11]. A homogeneous model assumes that every individual has equal contact to every one else in the population, and the rate of infection is largely determined by the density of the infected population. Kephart and White adopted a modified homogeneous model, which models the communication pattern among individuals as a directed graph [7]. Nodes in the graph represent the individuals in the population, and a directed edge from node i to node j denotes that i can directly infect j . A rate of infection, called the birth rate, β , is associated with each edge. A virus death rate, δ , is associated with each infected node.

If we denote the infection population at time t as η_t , a deterministic time evolution of η_t in the Kephart-White model (hereafter referred to as the KW model) can be represented as

$$\frac{d\eta_t}{dt} = \beta \langle k \rangle \eta_t (1 - \eta_t) - \delta \eta_t \quad (1)$$

where $\langle k \rangle$ is the average connectivity. The steady state solution is $\eta = 1 - \delta / (\beta \langle k \rangle) * N$, where N is the number of nodes.

An important prediction of Equation 1 is the epidemic threshold. An epidemic threshold, τ , is the critical $\frac{\beta}{\delta}$ ratio beyond which epidemics ensue. In a homogeneous network, the epidemic threshold is,

$$\tau_{hom} = \frac{1}{\langle k \rangle} \quad (2)$$

where $\langle k \rangle$ is the average connectivity.

These earlier models provide a good approximation of virus prevalence in networks where the contact among individuals is sufficiently homogeneous. However, there is overwhelming evidence that real networks (including social networks [21], router and AS networks [6], and Gnutella overlay graphs [22]) deviate from homogeneity, following power law structure instead. Computer viruses, therefore, are likely to propagate among nodes with a high variance in connectivity.

Pastor-Satorras and Vespignani studied epidemic spread for power-law networks where the connectivity distribution is characterized as $P(k) = k^{-\gamma}$ ($P(k)$ is the probability that a node has k

links) [14, 16, 18, 19]. Power-law networks have a highly skewed connectivity distribution and for certain values of γ resemble the Internet topology [6]. Pastor-Satorras et al. developed an analytic model (we refer to their model as the SV model) for the Barabási-Albert (BA) power-law topology ($\gamma = 3$). Their steady state prediction is,

$$\eta = 2e^{-\delta/m\beta} \quad (3)$$

where m is the minimum connection in the network. The SV model, however, critically depends on the assumption $\gamma = 3$, which is not true for real networks [9, 6]. Their model yields less than accurate predictions for networks that deviate from the BA topology, as we show later in the paper. Pastor-Satorras et al. [18] also proposed an epidemic threshold condition

$$\tau_{SV} = \frac{\langle k \rangle}{\langle k^2 \rangle} \quad (4)$$

where $\langle k \rangle$ is the expected connectivity and $\langle k^2 \rangle$ signals the connectivity divergence.

Following [19], Boguñá and Satorras studied epidemic spreading in networks where the connectivity of a node is related to the connectivity of its neighbors [3]. These *correlated networks* include Markovian networks where, in addition to $P(k)$, a function $P(k|k')$ determines the probability that a node of degree k is connected to a node of degree k' .

While some degree of correlations may exist in real networks, it is often difficult to characterize connectivity relationships with a simple $P(k|k')$ function. Indeed, prior studies on real networks [6, 15] have not found any conclusive evidence to support the type of correlation as defined in [3]. Hence, we will not discuss models for correlated networks further in this paper.

We present a new analytic model that does not assume any particular propagation topology. We will show later that our model subsumes previous models that are tailored to fit special case graphs (homogeneous, BA power-law, etc.).

3 The Proposed Model

In this section, we describe a model that does not assume homogeneous connectivity or any particular topology. We assume a connected network $G = (N, E)$, where N is the number of nodes in the network and E is the set of edges. Like the traditional models, we assume a universal infection rate β for each edge connected to an infected node, and a virus death rate δ for each infected node. Table 1 lists the symbols used.

β	Virus birth rate on a link from an infected neighbor
δ	Virus death rate on an infected node
t	Timestamp
$p_{i,t}$	Probability that node i is infected at time t
$\zeta_{k,t}$	Probability that a k -linked node will not receive infections from its neighbors at time t
η_t	infection population at time t
$\langle k \rangle$	Average degree of nodes in a network
$\langle k^2 \rangle$	Connectivity divergence

Table 1: Table of Symbols

3.1 Model

Our model assumes discrete time. During each time interval, an infected node i will try to infect its neighbors with probability β . At the same time, i may be cured with probability δ . We denote the probability that a node i is infected at time t as $p_{i,t}$. We define $\zeta_{k,t}$, the probability that a k -linked node will not receive infections from its neighbors at time t as,

$$\begin{aligned}
 \zeta_{k,t} &= \prod_{j:\text{neighbor of } i} (p_{j,t-1}(1 - \beta) + (1 - p_{j,t-1})) \\
 &= \prod_{j:\text{neighbor of } i} (1 - \beta * p_{j,t-1})
 \end{aligned} \tag{5}$$

In this model, a node i is healthy at time t if

- i was healthy before t and not infected at t OR
- i was at $t - 1$, cured at t and not infected at t OR

- i was at $t - 1$, infected, but cured immediately afterwards

The healthy probability of a k -linked node i at time t , $1 - p_{i,t}$, can be defined as,

$$1 - p_{i,t} = (1 - p_{i,t-1})\zeta_{k,t} + \delta p_{i,t-1}\zeta_{k,t} + \frac{1}{2}\delta p_{i,t-1}(1 - \zeta_{k,t}) \quad i = 1 \dots N \quad (6)$$

Note that for the last term on the right hand side of this equation, we assume that the probability that a curing event at node i takes place before infection from neighbors is roughly 50%.

Given a particular network topology and values of β and δ , we can solve Equation 6 numerically and obtain the time evolution of infected population, η_t , where $\eta_t = \sum_{i=1}^N p_{i,t}$

In this paper, we conduct simulation experiments on a variety of real and synthesized graphs. We use a real network graph collected at the Oregon router views¹. This dataset contains 31180 links among 10900 AS peers. All synthesized power-law graphs used in this study are generated using BRITE [12]. Unless otherwise specified, each simulation plot is averaged over 15 individual runs.

3.2 Experiments

In this section, we present a set of simulation results. The simulations are conducted to answer the question—how does our model perform in *real*, *BA power law*, and *homogeneous* graphs?

Figure 1 shows the time evolution of η as predicted by our model (see Equation 6) on the 10900-node Oregon AS graph, plotted against simulations and the steady state prediction of the SV model in Equation 3 (Since the SV model does not estimate the transients, we plot the steady state only.) As shown, our model yields a strictly more precise result than the SV model.

Figure 2 depicts the predictions of our model against the SV model for Barabási-Albert networks (see Equation 3). The topology used in Figure 2 is a synthesized 1000-node BA network. Since the SV model in Equation 3 is specifically tailored for BA networks, we expect the comparison to be very much a sanity check. As shown, both models conform nicely to the simulation results, though our model appears to be slightly more precise.

Figure 3 shows simulation results of epidemic spreading on an 1000-node random network,

¹<http://topology.eecs.umich.edu/data.html>

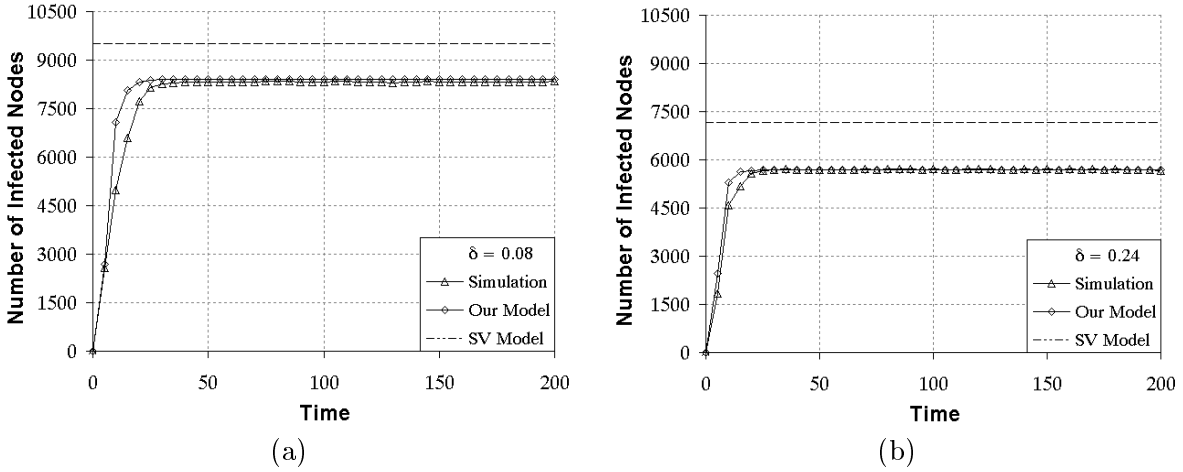


Figure 1: Experiments show the time evolution of infection in an 10900-node power-law network. Both simulations were performed on an Oregon network graph, with $\langle k \rangle = 5.72$ and $\beta = 0.14$. In both cases, our model conforms much closer to the simulation results than the SV model.

plotted against the KW model [7] and our model. The random network is constructed according to the Erdős-Rényi model [5]. Since an Erdős-Rényi network is sufficiently close to being homogeneous as far as epidemiological models are concerned, the results in Figure 3 suggest that our model is as precise as a model designed specifically for homogeneous networks. In one case where β is 0.2 and δ is 0.72, simulations appear to follow our prediction more closely than that of the KW model.

The experiments we show here, conducted on a real network, a synthesized BA power-law network, and an Erdős-Rényi network, illustrate the predictive power of our model—as a general model, it subsumes prior models and produces predictions that equal or outperform predictions that are designed for specific topologies.

4 Epidemic Threshold and Eigenvalues

As described previously, an epidemic threshold is a critical state beyond which infections become endemic. Predicting the epidemic threshold is an important part of an epidemiological model. The epidemic threshold of a graph depends fundamentally on the graph itself. The challenge therefore is to capture the essence of the graph in as few parameters as possible. We present one such model here that predicts the epidemic threshold with a single parameter—the largest eigenvalue of the

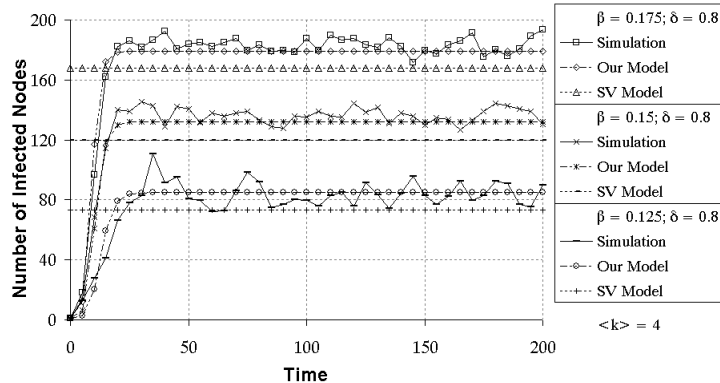


Figure 2: Experiments on BA topology: shows time evolution of infected population in a 1000-node power-law network. Our model outperforms the SV model in its steady state prediction.

adjacency matrix of the graph—for arbitrary graphs.

We note that it is possible to develop threshold conditions for special case but unrealistic graphs. For instance, the epidemic threshold for a homogeneous network is the inverse of the average connectivity, $\langle k \rangle$. Similarly, threshold for infinite power-law networks is zero. But a unifying model for *arbitrary, real* graphs has not appeared in the literature yet. The closest model thus far is the one put forth by Pastor-Satorras et al. (see Equation 4. But we show later that their model is not accurate for arbitrary graphs.

In this section, we describe a general theory for epidemic threshold that holds for arbitrary graphs. We observe that the epidemic threshold is the condition linking the birth and death rates to the adjacency matrix of the graph, such that an infection becomes an epidemic if the condition holds, and dies out if it does not. Our theory is surprisingly simple yet accurate at the same time. We show later in this section that this new threshold condition subsumes prior models for special case graphs. Table 2 lists the symbols used in this section.

Next, we will show that our estimate for the epidemic threshold τ is

$$\tau = \frac{1}{\lambda_{1,A}} \quad (7)$$

where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix \mathbf{A} .

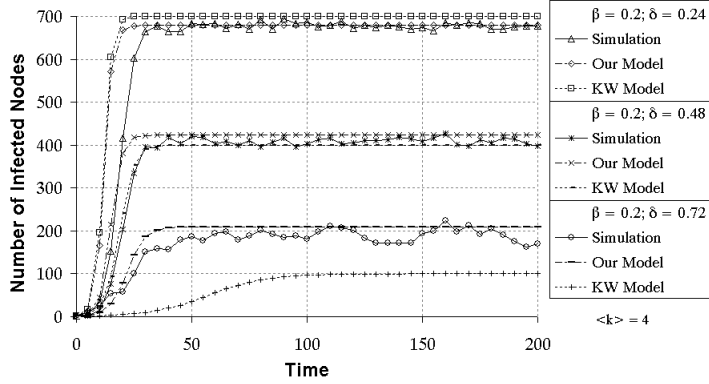


Figure 3: Experiments on ER topology: shows time evolution of infected population in a 1000-node random Erdős network. Our model generally yields similar predictions to the KW model, but outperforms it when δ is high.

\mathbf{A}	Adjacency matrix of the network
$\text{tr}\mathbf{A}$	The transpose of matrix \mathbf{A}
$\lambda_{i,A}$	The i -th largest eigenvalue of \mathbf{A}
$u_{i,A}$	The eigenvector of \mathbf{A} corresponding to $\lambda_{i,A}$
\mathbf{S}	The ‘system’ matrix describing the equations of infection
$\lambda_{i,S}$	The i -th largest eigenvalue of \mathbf{S}

Table 2: Symbols used in the eigenvalue analysis

Theorem 1 (Epidemic Threshold) *When an epidemic dies out, we should have $\frac{\beta}{\delta} < \tau = \frac{1}{\lambda_{1,A}}$, where β is the birth rate, δ is the death rate and $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix \mathbf{A} .*

Proof: Restating Equation 6,

$$1 - p_{i,t} = (1 - p_{i,t-1})\zeta_{k,t} + \delta p_{i,t-1}\zeta_{k,t} + \frac{1}{2}\delta p_{i,t-1}(1 - \zeta_{k,t}) \quad i = 1 \dots N \quad (8)$$

Rearranging the terms,

$$\begin{aligned} 1 - p_{i,t} &= \frac{1}{2}\delta p_{i,t-1} + \left(1 + \left(\frac{1}{2}\delta - 1\right) p_{i,t-1}\right) \zeta_{k,t} \\ &= \frac{1}{2}\delta p_{i,t-1} + 1 + \left(\frac{1}{2}\delta - 1\right) p_{i,t-1} - \beta \sum_j p_{j,t-1} \end{aligned} \quad (9)$$

$$= 1 + \delta p_{i,t-1} - p_{i,t-1} - \beta \sum_j p_{j,t-1} \quad (10)$$

This uses the approximation

$$(1 - a)(1 - b) \approx 1 - a - b \quad (11)$$

when $a \ll 1$, $b \ll 1$.

Thus, we have

$$\begin{aligned} -p_{i,t} &= \delta p_{i,t-1} - p_{i,t-1} - \beta \sum_j p_{j,t-1} \\ \text{so, } p_{i,t} &= (1 - \delta)p_{i,t-1} + \beta \sum_j p_{j,t-1} \end{aligned} \quad (12)$$

Converting to matrix notation (\mathbf{P}_t is the column vector $(p_{1,t}, p_{2,t}, \dots, p_{N,t})$),

$$\mathbf{P}_t = ((1 - \delta)\mathbf{I} + \beta\mathbf{A})\mathbf{P}_{t-1} \quad (13)$$

Thus, \mathbf{P}_t is of the form

$$\mathbf{P}_t = \mathbf{S}\mathbf{P}_{t-1} \quad (14)$$

$$= \mathbf{S}^t \mathbf{P}_0 \quad (15)$$

where $\mathbf{S} = (1 - \delta)\mathbf{I} + \beta\mathbf{A}$ is the ‘system’ matrix.

As we show in Lemma 1 in the Appendix, the matrices \mathbf{A} and \mathbf{S} have the same eigenvectors $\mathbf{u}_{i,S}$, and their eigenvalues, $\lambda_{i,A}$ and $\lambda_{i,S}$, are closely related:

$$\lambda_{i,S} = 1 - \delta + \beta\lambda_{i,A} \quad \forall i \quad (16)$$

Using the spectral decomposition, we can say

$$\begin{aligned} \mathbf{S} &= \sum_i \lambda_{i,S} \mathbf{u}_{i,S} \mathbf{tr}(\mathbf{u}_{i,S}) \\ \text{and, } \mathbf{S}^t &= \sum_i \lambda_{i,S}^t \mathbf{u}_{i,S} \mathbf{tr}(\mathbf{u}_{i,S}) \end{aligned} \quad (17)$$

Using this in Equation 15,

$$\mathbf{P}_t = \sum_i \lambda_{i,S}^t \mathbf{u}_{i,S} \text{tr}(\mathbf{u}_{i,S}) \mathbf{P}_{t-1} \quad (18)$$

Without loss of generality, order the eigenvalues such that $\lambda_{1,A} \geq \lambda_{2,A} \dots$. For an infection to die off and not become an epidemic, the vector \mathbf{P}_t should go to zero for large t , which happens when $\forall i, \lambda_{i,S}^t$ tends to 0. That implies $\lambda_{1,S} < 1$. So,

$$1 - \delta + \beta \lambda_{1,A} < 1 \quad (19)$$

which means that, $\boxed{\tau = \frac{1}{\lambda_{1,A}}}$ □

Theorem 2 (Exponential Decay) *When the condition above is met, the probability of infection decays exponentially over time.*

Proof: We have:

$$\begin{aligned} \mathbf{P}_t &= \mathbf{S}^t \mathbf{P}_0 \text{ (from Equation 15)} \\ &\approx \sum_i \lambda_{i,S}^t \mathbf{u}_{i,S} \text{tr}(\mathbf{u}_{i,S}) \text{ (from Equation 17)} \\ &\approx \lambda_{1,S}^t * \mathbf{C} \end{aligned} \quad (20)$$

where \mathbf{C} is some constant vector depending on the initial conditions. The value of $\lambda_{1,S}$ is less than 1 (because of the no-epidemic condition). Thus, the values of $p_{i,t}$ are exponentially decreasing over time. □

Corollary 1 *When the network is below the epidemic threshold, the number of infected nodes decays exponentially over time.*

Proof: Let the number of infected nodes at time t be denoted by n_t .

$$\begin{aligned} n_t &= \sum_{i=1}^N p_{i,t} \\ &= \sum_i \lambda_{1,S}^t * C_i \end{aligned}$$

$$= \lambda_{1,S}^t * \sum_i C_i \tag{21}$$

where C_i are the individual elements of the matrix \mathbf{C} in Equation 20 above. Now, $\sum_i C_i$ is a constant and $\lambda_{1,S} < 1$ (from Theorem 1). Thus, we see that n_t decays exponentially with time. \square .

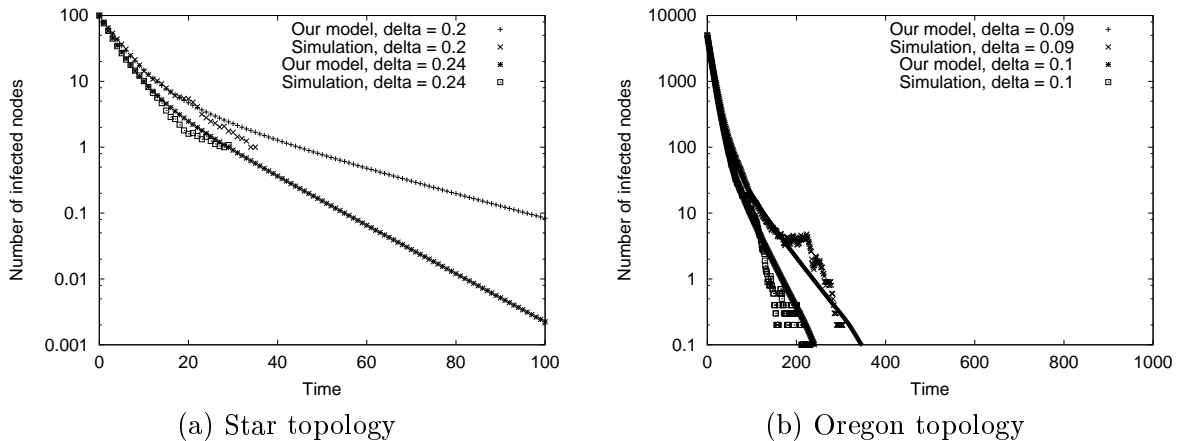


Figure 4: This figure shows the exponential decay in the number of infected nodes over time, when we are under the epidemic threshold. Plot (a) compares the logarithm of the number of infected nodes over time for a 100-node star topology; plot (b) shows the same for the Oregon topology. In both cases, the plot becomes linear for large t , meaning that the decay is exponential.

The exponential decay in the number of infected nodes can be seen clearly in Figure 4, where we plot the logarithm of the number of infected nodes, η_t , versus t . Two plots are shown: one for the star topology, and one for the Oregon dataset. In both cases, we observe that for large values of time t , the plots become linear, implying that the number of nodes decays exponentially.

5 Discussion—generality of our threshold condition

We now turn to show that our threshold condition is general and holds for other graphs. In particular, we show that the threshold condition holds for a) homogeneous, b) star, c) infinite power-law, and d) finite power-law graphs. We do that with the following corollaries.

Corollary 2 *The new threshold model holds for homogeneous or random Erdős-Rényi graphs.*

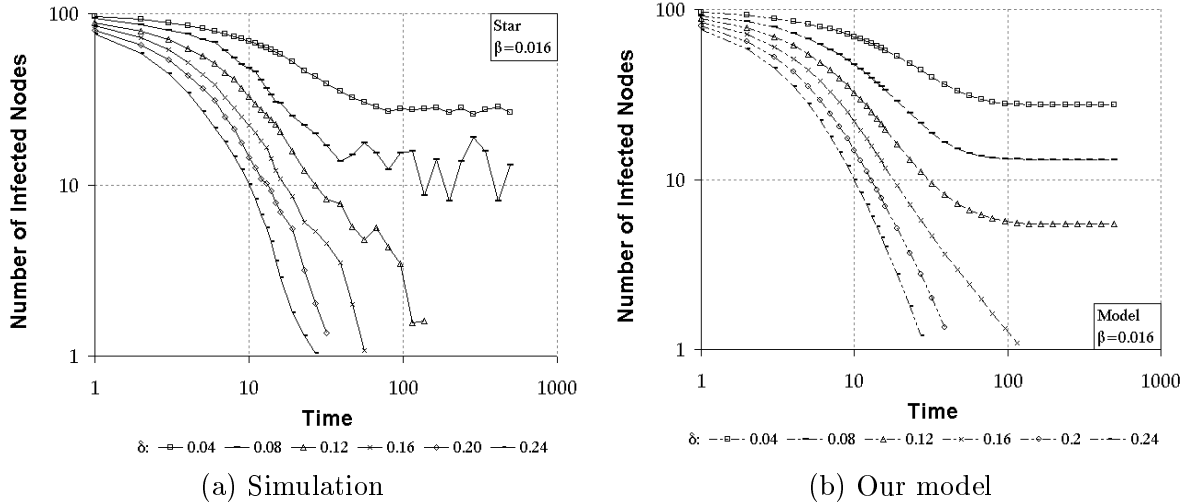


Figure 5: Critical δ for an 100-node star topology: number of infected nodes versus time in log-log scales, given $\beta = 0.016$. Our model predicts critical δ at 0.16. (Triangles at left and crosses at right plot)

Proof: As reported previously, the epidemic threshold in a homogeneous network or a random Erdős-Rényi graph is $\tau_{hom} = 1/\langle k \rangle$ where $\langle k \rangle$ is the average connectivity. It is easily shown that, in a homogeneous or random network, the largest eigenvalue of the adjacency matrix is $\langle k \rangle$. Therefore, our model of epidemic threshold yields the same threshold condition as the homogeneous models [11]. \square

Corollary 3 *The epidemic threshold, τ , for a star topology is exactly $\frac{1}{\sqrt{d}}$, where \sqrt{d} is the square root of the degree of the central node.*

Proof: In a star topology, we have two types of nodes, the center node and the satellite nodes. Suppose that we have d satellites, the first eigenvalue of the adjacency matrix, λ_1 , is \sqrt{d} . The stability condition then becomes

$$\lambda_1 = 1 - \delta + \beta * \sqrt{d} = 1 \quad (22)$$

which means that $\delta = \beta * \sqrt{d}$ to achieve stability, thus rendering $\tau = \frac{1}{\sqrt{d}}$. \square

Figure 5 shows the infection spread over time in a 100-node star graph with $\beta = 0.016$. Given $\tau = \frac{1}{\sqrt{99}}$, the critical δ on the threshold is approximately 0.16. This is confirmed by our propagation model as shown in Figure 5(b). Notice that our prediction for critical δ holds. More specifically, our model closely reflects the simulation, for $\delta > 0.16$. For $\delta < 0.16$, there is no epidemic. For $\delta = 0.16$, a very interesting setting appears.

For the case of $\delta = 0.16$, our model seems to show that the expected number of infected nodes η_t drops approximately at the rate of t^{-1} , which is qualitatively different from the other two cases: for $\delta > 0.16$, $\eta_t \approx \lambda_1^t$; for $\delta < 0.16$, η_t stabilizes. This suggests a phase transition phenomenon, which indicates a very interesting future research question.

Figure 6(c) and (d) depict a further example for the star topology, plotting the number of infected nodes η_{200} at time $t=200$, for several values of the β/δ ratio. We plot both the theoretical, as well as simulation results. We also show the two epidemic thresholds with vertical lines: our threshold with “crosses”, at $\beta/\delta = 1/\lambda_1, A = 0.1$ and the SV threshold with “squares”, at $\beta/\delta = 0.02$. The simulation results indicate that our threshold is clearly in the correct region, while the SV threshold prediction is not accurate.

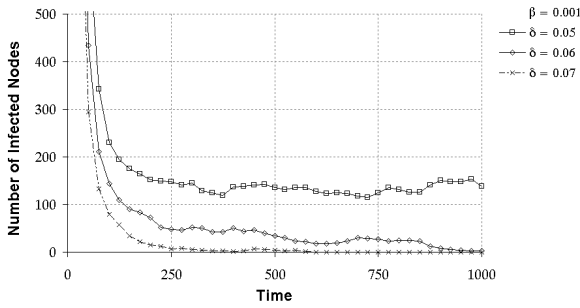
Corollary 4 *The epidemic threshold for an infinite power-law network is zero.*

Proof: In a power-law network, the first eigenvalue of the adjacency matrix, λ_1 , is $\sqrt{d_{max}}$ (according to [13]). Since $d_{max} \propto \ln(N)$ and N is infinite, λ_1 is infinite. Our epidemic threshold condition states that δ must be greater than $\beta * \lambda_1$ in order for there not be any epidemic. Therefore, the epidemic threshold is effectively zero for infinite power-law networks. This result concurs with previous work, which finds that infinite power-law networks lack epidemic thresholds. \square

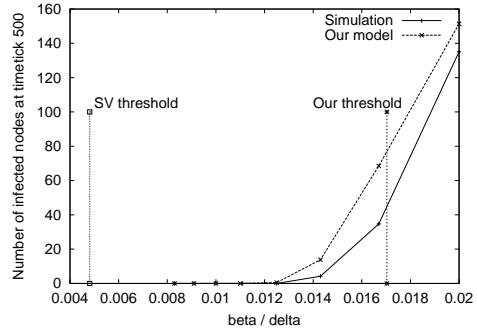
Corollary 5 *The epidemic threshold, τ , for finite power-law networks is more precisely indicated by $\frac{1}{\lambda_1}$, where λ_1 is the first eigenvalue of the adjacency matrix.*

Proof: This follows directly from the Theorem 1 shown above. \square

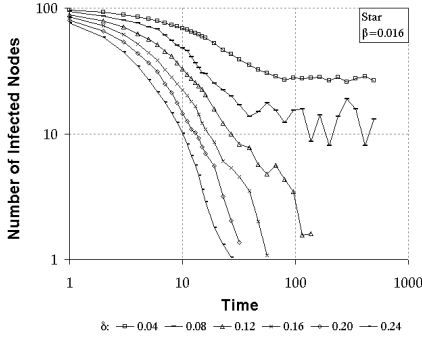
We compare our threshold prediction with the threshold model by Pastor-Satorras et al. in Equation 4. Their model, $\tau_{SV} = \langle k \rangle / \langle k^2 \rangle$, where k is the average connectivity, is put forth as a general model for arbitrary graphs. Figure 6(a) and (b) show simulated epidemic spreading on the



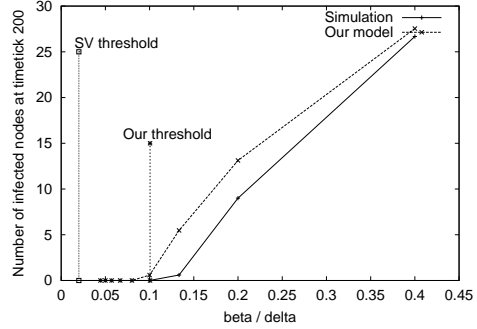
(a)infected population vs. time for Oregon



(b)infection at timetick 500 vs. β/δ for Oregon



(c)infected population vs. time for Star



(d)infection at timetick 500 vs. β/δ for Star

Figure 6: Epidemic threshold on the Oregon and Star topology. Plot (a) shows that the critical δ at 0.06 is very close to our predicted epidemic threshold critical $\delta \approx 0.0587211$. The SV model predicts critical $\delta \approx 0.207796$. Plot (b) shows that our predicted τ at 0.0167 approximates the behavior of the infection at timetick 500 where the system state has stabilized. As shown, the threshold predicted by the SV model does not accurately reflect reality. Plots (c) and (d) show the same information for the Star topology. Again, our estimate of the threshold is better than that of the SV model.

Oregon network. The largest eigenvalue λ_1 of the adjacency matrix for this network is approximately 58.7211.

We structured the experiment such that 5000 nodes are infected initially. Simulations proceed with $\beta = 0.001$ and δ ranging from 0.05 to 0.14. For the particular values of β and λ_1 , our epidemic threshold model predicts a critical δ at 0.0587211, while the SV threshold prediction puts the critical δ at 0.2078. As shown in Figure 6(a), the simulation with $\delta = 0.05$ reaches equilibrium while the run with $\delta = 0.07$ approaches zero at approximately timetick 600. The run with $\delta = 0.06$ steadily approaches zero, but has yet to reach it at timetick 1000. These results closely mirror our threshold prediction, which shows a critical δ at approximately 0.06.

Figure 6(b) shows an alternate view of the experiment result, plotting the number of infected nodes η_{500} at time $t=500$, for several values of the β/δ ratio. We plot both the theoretical, as well as simulation results. We also show the two epidemic thresholds with vertical lines: our threshold with “crosses”, at $\beta/\delta = 1/\lambda_1, A = 0.0167$ and the SV threshold with “squares”, at $\beta/\delta = 0.0048$. Notice that our threshold is clearly in the correct region, while the SV threshold is way off.

6 Conclusions - Contributions

How will a virus propagate in a real computer network? What is the epidemic threshold for a finite graph, if any? How long does it take for a viral outbreak to reach steady state? These questions have for decades intrigued researchers. In this paper we attempt to answer these questions by providing a new analytic model that accurately models the propagation of viruses on *arbitrary graphs*. The primary contributions of this paper are:

- We propose a new model for virus propagation in networks (Equation 6), and show that our model is more precise and general than those used before. We demonstrate the accuracy of our model in both real and synthetic networks.
- We are the first to show that we can capture the virus-propagation properties of a graph in a single parameter, namely the eigenvalue $\lambda_{1,A}$. We propose a precise epidemic threshold, $\tau = 1/\lambda_{1,A}$, which holds irrespective of the network topology; an epidemic is prevented when

$\delta > \delta_c = \beta * \lambda_{1,A}$. We show that our epidemic threshold is more general, and more precise than previous models for special-case graphs (e.g., Erdős-Rényi, homogeneous, BA power-law); we show that it tends to zero for infinite power-law graphs.

- We show that, below the epidemic threshold, the number of infected nodes in the network decays exponentially.

Future research directions abound, both for theoretical as well as experimental work. One could examine phase transition phenomena, when we are exactly on the epidemic threshold. Another promising direction is to enhance the model with a “vigilance” parameter to model environmental factors that affect viral propagations.

7 Appendix

Lemma 1 (Eigenvalues of the system matrix) *The i -th eigenvalue of \mathbf{S} is of the form $\lambda_{i,S} = 1 - \delta + \beta\lambda_{i,A}$, and the eigenvectors of \mathbf{S} are the same as those of \mathbf{A} .*

Proof: Let $\mathbf{u}_{i,A}$ be the eigenvector of \mathbf{A} corresponding to eigenvalue $\lambda_{i,A}$. Then, by definition, $\mathbf{A}\mathbf{u}_{i,A} = \lambda_{i,A}\mathbf{u}_{i,A}$ (because \mathbf{A} is symmetric in our case). Now,

$$\begin{aligned} \mathbf{S}\mathbf{u}_{i,A} &= (1 - \delta)\mathbf{u}_{i,A} + \beta\mathbf{A}\mathbf{u}_{i,A} \\ &= (1 - \delta)\mathbf{u}_{i,A} + \beta\lambda_{i,A}\mathbf{u}_{i,A} \\ &= (1 - \delta + \beta\lambda_{i,A})\mathbf{u}_{i,A} \end{aligned} \tag{23}$$

Thus, $\mathbf{u}_{i,A}$ is also an eigenvector of \mathbf{S} , and the corresponding eigenvalue is $(1 - \delta + \beta\lambda_{i,A})$. \square

References

- [1] Norman Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, London, 1975.

- [2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 15 October 1999.
- [3] Marián Boguñá and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66:047104, 2002.
- [4] Zoltán Dezső and Albert-László Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65:055103(R), 21 May 2002.
- [5] Paul Erdős and Alfred Rényi. On the evolution of random graphs. In *Publication 5*, pages 17–61. Institute of Mathematics, Hungarian Academy of Sciences, Hungary, 1960.
- [6] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationship of the internet topology. In *Proceedings of ACM Sigcomm 1999*, September 1999.
- [7] Jeffrey O Kephart and Steve R White. Directed-graph epidemiological models of computer viruses. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, May 1991.
- [8] Jeffrey O Kephart and Steve R White. Measuring and modeling computer virus prevalence. In *Proceedings of the 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 2–15, May 1993.
- [9] S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Trawling the web for emerging cyber-communities. *Computer Networks*, 31(11-16):1481–1493, 1999.
- [10] Helen Martin, editor. *The Virus Bulletin: Independent Anti-Virus Advice*. World Wide Web, <http://www.virusbtn.com>, 2002. Ongoing.
- [11] A G McKendrick. Applications of mathematics to medical problems. In *Proceedings of Edin. Math. Society*, volume 14, pages 98–130, 1926.
- [12] Alberto Medina, Anukool Lakhina, Ibrahim Matta, and John Byers. Brite: Universal topology generation from a user’s perspective. Technical Report BUCS-TR2001-003, Boston University, 2001. World Wide Web, <http://www.cs.bu.edu/brite/publications/>.

- [13] Milena Mihail and Christos H Papadimitriou. On the eigenvalue power law. In *RANDOM 2002*, Harvard University, Cambridge, MA, 15 September 2002.
- [14] Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B*, 26:521–529, 4 February 2002.
- [15] Mark E J Newman, Stephanie Forrest, and Justin Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66:035101(R), 10 September 2002.
- [16] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.
- [17] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2 April 2001.
- [18] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65:035108, 2002.
- [19] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemics and immunization in scale-free networks. In Stefan Bornholdt and Heinz Georg Schuster, editors, *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, Berlin, May 2002.
- [20] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Physical Review E*, 65:036104, 2002.
- [21] M. Richardson and P. Domingos. Mining the network value of customers. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, pages 57–66, San Francisco, CA, 2001.
- [22] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal*, 6(1), 2002.
- [23] Stuart Staniford, Vern Paxson, and Nicholas Weaver. How to own the internet in your spare time. In *Proceedings of the 11th USENIX Security Symposium*, August 2002.

- [24] CERT Advisory CA-1999-04.
Melissa macro virus. World Wide Web, <http://www.cert.org/advisories/CA-1999-04.html>, 1999.
- [25] CERT Advisory CA-2001-23. Continued threat of the "code red" worm. World Wide Web, <http://www.cert.org/advisories/CA-2001-23.html>, 2001.
- [26] Chenxi Wang, John C Knight, and Matthew C Elder. On computer viral infection and the effect of immunization. In *Proceedings of the 16th ACM Annual Computer Security Applications Conference*, December 2000.