PHYSICAL REVIEW E, VOLUME 65, 055103(R)

Halting viruses in scale-free networks

Zoltán Dezső and Albert-László Barabási
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Received 7 March 2002; published 21 May 2002)

The vanishing epidemic threshold for viruses spreading on scale-free networks indicate that traditional methods, aiming to decrease a virus' spreading rate cannot succeed in eradicating an epidemic. We demonstrate that policies that discriminate between the nodes, curing mostly the highly connected nodes, can restore a finite epidemic threshold and potentially eradicate a virus. We find that the more biased a policy is towards the hubs, the more chance it has to bring the epidemic threshold above the virus' spreading rate. Furthermore, such biased policies are more cost effective, requiring less cures to eradicate the virus.

DOI: 10.1103/PhysRevE.65.055103 PACS number(s): 89.75.Fb, 89.75.Hc, 89.20.Hh

While most diffusion processes of practical interest, ranging from the spread of computer viruses to the diffusion of sexually transmitted diseases, take place on complex networks, the bulk of diffusion studies have focused on model systems, such as regular lattices or random networks [1–3]. A series of recent results indicate, however, that real networks significantly deviate from the structure of these model systems [4]—deviations that have a strong impact on the diffusion dynamics as well. In particular, the networks responsible for the spread of computer viruses, such as the Internet [5] or the email network [6], have a scale-free topology [7], exhibiting a power-law degree distribution $P(k) \sim k^{-\gamma}$, where γ ranges between 2 and 3. Similarly, a recent study indicates that the social network responsible for the spread of sexually transmitted diseases, such as AIDS, also exhibits a scale-free structure [8]. The topology of scale-free networks fundamentally deviate from the topology of both regular lattices and random networks [9], differences that impact the network's robustness and attack tolerance [10] or the dynamics of synchronization [11]. It is not unexpected, therefore, that the broad degree distribution leads to unexpected diffusion properties as well [12].

A simple model often used to study the generic features of virus spreading is the susceptible-infected-susceptible (SIS) model. In this model an individual is represented by a node, which can be either ''healthy'' or ''infected.'' Connections between individuals along which the infection can spread are represented by links. In each time step a healthy node is infected with probability ν if it is connected to at least one infected node. At the same time an infected node is cured with probability δ, defining an effective spreading rate $\lambda \equiv \nu/\delta$ for the virus.

The behavior of the SIS model is well understood if the nodes are placed on a regular lattice or a random network [1]. Diffusion studies indicate that viruses whose spreading rate exceeds a critical threshold λ_c will persist, while those under the threshold will die out shortly. Recently, however, Pastor-Satorras and Vespignani have shown [12] that for scale-free networks with $\gamma \leqslant 3$ the epidemic threshold vanishes, i.e., $\lambda_c = 0$. This finding implies that on such networks even weakly infectious viruses can spread and prevail. This vanishing threshold is a consequence of the hubs—nodes with a large number of links encoded by the tail of power law $P(k)$. Indeed, the hubs are in contact with a large number of nodes, and are therefore easily infected. Once infected, they pass on the virus to a significant fraction of the nodes in the system.

The finding that the epidemic threshold vanishes in scale-free networks has a strong impact on our ability to control various virus outbreaks. Indeed, most methods designed to eradicate viruses—biological or computer based—aim at reducing the spreading rate of the virus, hoping that if λ falls under the critical threshold λ_c, the virus will die out naturally. With a zero threshold, while a reduced spreading rate will decrease the virus' prevalence, there is little guarantee that it will eradicate it. Therefore, from a theoretical perspective viruses spreading on a scale-free network appear unstoppable. The question is, can we take advantage of the increased knowledge accumulated in the past few years about network topology to understand the conditions in which one can successfully eradicate viruses?

Here we study the spreading of a virus to which there is a cure, eradicating the virus from the node to which it is applied to, but which does not offer a permanent protection against the virus. If such a cure is available to all nodes, treating simultaneously all infected nodes will inevitably wipe the virus out. However, due to economic or policy considerations the number of available cures is often limited. This applies to AIDS, for which relatively effective but prohibitively expensive cures are available, unable to reach the most affected segments of population due to economic considerations [13]. But it also applies to computer viruses, where only a small fraction of users commit the time and investment to update regularly their virus protection system. We show that distributing the cures randomly in a scale-free network is ineffective, being unable to alter the fundamental properties of the threshold-free diffusion process. However, even weakly biased strategies, that discriminating between the nodes, curing with a higher probability the hubs than the less connected nodes, can restore the epidemic threshold. We find that such hub-biased policies are more cost effective as well, requiring fewer cures than those distributing the cures indiscriminately.

Curing the hubs. The vanishing epidemic threshold of a virus spreading in a scale-free network is rooted in the infinite variance of the degree distribution [12]. Indeed, the threshold λ_c depends on the variance as

 ©2002 The American Physical Society

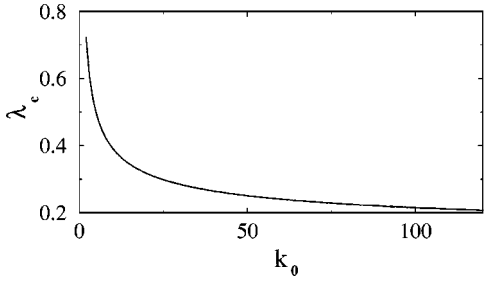

FIG. 1. The epidemic threshold as a function of k_0.

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{1}$$

On a regular lattice the degree distribution is a δ function, while on a random network it follows a Poisson distribution, in both cases resulting in a finite $\langle k^2 \rangle$, and therefore nonzero λ_c. In contrast, if the virus spreads on a scale-free network, for which $P(k)$ follows a power law with $\gamma \leqslant 3$, the variance is infinite and the epidemic threshold is $\lambda_c = 0$. Therefore, to restore a finite epidemic threshold, which would allow the infection to die out, one needs to induce a finite variance. As the origin of the infinite variance is in the tail of the degree distribution, dominated by the hubs, one expects that curing *all* hubs with degree larger than a given degree k_0 would restore a finite variance and therefore a nonzero epidemic threshold. Indeed, if on a scale-free network nodes with degree $k > k_0$ are always healthy, the epidemic threshold is finite and has the value [14]

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} = \frac{k_0 - m}{k_0 m} \left(\ln \frac{k_0}{m} \right)^{-1}. \tag{2}$$

This expression indicates that the more hubs we cure (i.e., the smaller k_0 is), the larger the value of the epidemic threshold (Fig. 1) [15]. Therefore, the most effective policy against an epidemic would be to cure as many hubs as economically viable. The problem is that in most systems of interest we do not have detailed network maps, thus we cannot effectively identify the hubs. Indeed, we do not know the number of sexual partners for each individual in the society, thus we cannot identify the social hubs that should be cured if infected. Similarly, on the email network we do not know which email accounts serve as hubs, as these are the ones that, for the benefit of all email users, should always carry the latest antivirus software.

Short of a detailed network map, no method aiming to identify and cure the hubs is expected to succeed at its goal of finding *all* hubs with degree larger than a given k_0. Yet, policies designed to eradicate viruses could attempt to identify and cure as many hubs as possible. Such biased policy will inevitably be inherently imperfect, as it might miss some hubs, and falsely identify some smaller nodes as hubs. The question is, however, would a policy biased towards curing the hubs, without a guarantee that it can identify all of them, succeed at restoring the epidemic threshold?

To investigate the effect of incomplete information about the hubs we assume that the likelihood of identifying and administering a cure to an infected node with k links in a given time frame depends on the node's degree as k^α, where α characterizes the policy's ability to identify hubs. In this framework $\alpha = 0$ corresponds to random cure distribution, which is expected to have zero epidemic threshold while $\alpha = \infty$ corresponds to an optimal policy that treats all hubs with degree larger than k_0. Within the framework of the SIS model we assume that each node is infected with probability ν, but each infected node is cured with probability $\delta = \delta_0 k^\alpha$, again becoming susceptible to the disease. We define the spreading rate as $\lambda = \nu / \delta_0$. As each healthy node is susceptible again to the disease, a node can get multiple cures during a simulation.

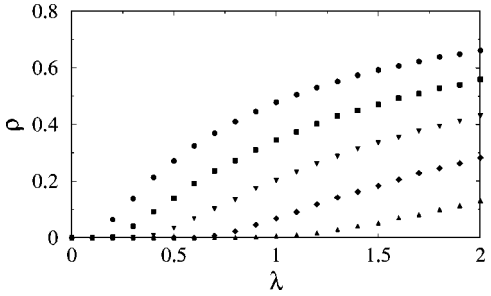

FIG. 2. Prevalence ρ measured as the fraction of infected nodes in function of the effective spreading rate λ for $\alpha = 0(\bigcirc)$, $0.25(\square)$, $0.50(\nabla)$, $0.75(\diamond)$, and $1(\triangle)$, as predicted by Monte Carlo simulations using the SIS model on a scale-free network with $N = 10\,000$ nodes.

We place the nodes on a scale-free network [16] and initially infect half of them. After a transient regime the system reaches a steady state, characterized by a constant average density of infected nodes ρ, which depends on both the spreading rate λ and α (Fig. 2). The $\alpha = 0$ limit corresponds to random immunization in which case the epidemic threshold is zero. As treating only the hubs will restore the nonzero epidemic threshold, for $\alpha = \infty$ we expect a nonzero λ_c. Yet, the numerical simulations indicate that we have a finite λ_c well before the $\alpha = \infty$ limit. Indeed, as Fig. 2 shows, λ_c is clearly finite for $\alpha = 1$ and so is for smaller value of α as well. The numerical simulations do not give an unambiguous answer to the crucial question: Is there a critical value of α at which a finite λ_c appears, or for any nonzero α we have a finite λ_c?

Mean-field theory. To interpret the results of the numerical simulations we studied the effect of a biased policy using the mean-field continuum approach [1,12]. Denoting by $\rho_k(t)$ the density of infected nodes with connectivity k, the time evolution of $\rho_k(t)$ can be written as [12]

$$\partial_t \rho_k(t) = -\delta_0 k^\alpha \rho_k(t) + \nu [1 - \rho_k(t)] k \theta(\lambda). \tag{3}$$

The first term in the right-hand side (rhs) describes the probability that an infected node is cured, and it is therefore proportional to the number of infected nodes $\rho_k(t)$ and the probability $\delta_0 k^\alpha$ that a node with k links will be selected for a cure. The second term is the probability that a healthy node with k links is infected, proportional to the infection rate (ν), the number of links (k), the number of healthy nodes with k

links $[1-\rho_k(t)]$, and the probability $\theta(\lambda)$ that a given link points to an infected node. The probability $\theta(\lambda)$ is proportional to $kP(k)$, therefore, it can be written as

$$\theta(\lambda)=\sum_k \frac{kP(k)}{\sum_s sP(s)}\rho_k. \tag{4}$$

Using $\lambda=\nu/\delta_0$ and imposing the $\partial_t\rho_k(t)=0$ stationary condition we find the stationary density as

$$\rho_k=\frac{\lambda\,\theta(\lambda)}{k^{\alpha-1}+\lambda\,\theta(\lambda)}. \tag{5}$$

Combining Eqs. (4) and (5) and using the fact that the connectivity distribution $P(k)=2m^2/k^3$ for the scale-free network [7], we obtain

$$m\lambda\int_m^\infty \frac{dk}{k^2[k^{\alpha-1}+\lambda\,\theta(\lambda)]}=1. \tag{6}$$

The average density of infected nodes is given by

$$\rho(\lambda)=\sum_k P(k)\rho(k)=2m^2\lambda\,\theta(\lambda)\int_m^\infty \frac{dk}{k^3[k^{\alpha-1}+\lambda\,\theta(\lambda)]}. \tag{7}$$

Equations (6) and (7) allow us to calculate the average density of infected nodes for any value of α. For $\alpha=0$ they reduce to the case studied in Ref. [12] giving $\lambda_c=0$. For $\alpha=1$ we can solve Eq. (6), and using Eq. (7) we obtain

$$\rho(\lambda)\big|_{\alpha=1}=\frac{\lambda-1}{\lambda}, \tag{8}$$

which indicates that for $\alpha=1$ the epidemic threshold is finite, having the value $\lambda_c(\alpha=1)=1$ [15].

To determine the epidemic threshold as a function of α we need to solve the $\rho(\lambda)=0$ equation. While we cannot get $\rho(\lambda)$ for arbitrary values of α, we can solve Eq. (6) in λ using that at the threshold $\lambda=\lambda_c$ we have $\theta(\lambda_c)=0$. In this case Eq. (6) predicts that the epidemic threshold depends on α as

$$\lambda_c=\alpha m^{\alpha-1}. \tag{9}$$

For $\alpha=0$ we recover $\lambda_c=0$, confirming that random immunization cannot eradicate an infectious disease. For $\alpha=1$ Eq. (9) predicts that the epidemic threshold is $\lambda_c=1$, in agreement with Eq. (8). Most important, however, Eq. (9) indicates that λ_c is nonzero for any positive α, i.e., any policy that is biased towards curing the hubs can restore a finite epidemic threshold. Furthermore, policies with larger α are expected to be more likely to lead to the eradication of the virus, as they result in larger λ_c values. Therefore, Eq. (9) indicates that a potential avenue to eradicating a virus is to increase the effectiveness of identifying and curing the hubs. Indeed, if the virus has a fixed spreading rate, increasing α could increase λ_c beyond λ, thus making it possible for the virus to die out naturally. To test the validity of prediction (9) we determined numerically the $\lambda(\alpha)$ curve from the simulations shown in Fig. 2. As Fig. 3 shows, we find excellent agreement between the simulations and the analytical prediction (9).

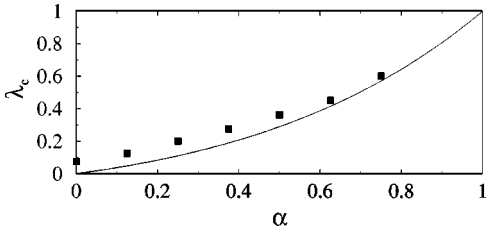

FIG. 3. The dependence of the epidemic threshold λ_c on α as predicted by our calculations (continuous line) based on the continuum approach, and by the numerical simulations based on the SIS model (boxes). The small deviation between the numerical results and the analytical prediction is due to the uncertainty in determining the precise value of the threshold in Monte Carlo simulations.

Cost effectiveness. A major criteria for any policy designed to combat an epidemic is its cost effectiveness. Supplying cures to all nodes infected by a virus is often prohibitively expensive. Therefore, policies that obtain the largest effect with the smallest number of administered cures are more desirable. To address the cost effectiveness of a policy targeting the hubs we calculated the number of cures administered in a time step per node for different values of α. Figure 4 indicates that increasing the policy's bias towards the hubs by allowing a higher value for α decreases rapidly the number of necessarily cures. Therefore, policies that distribute the cures mainly to the nodes with more links are more cost effective than those that spread the cures randomly, blind to the node's connectivity. We can understand the origin of the rapid decay in $c(\alpha)$ by noticing that the number of cures administered per unit time is proportional to the density of infected nodes. From Fig. 2 we see that for a given value of the spreading rate the prevalence is decreasing as α increases, decreasing the number of necessary cures as well.

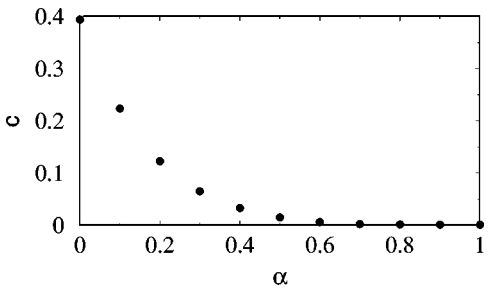

FIG. 4. The number of cures c administered in a unit time per node for different values of α. The rapidly decaying c indicates that the more successful a policy is in selecting and curing hubs (larger is α), the fewer the cures are required for a fixed spreading rate ($\lambda=0.75$). For $\alpha=0$ the number of cures is calculated by $c=\nu/(\nu+\delta)=\lambda/(1+\lambda)$ which gives $c=0.43$, which value is in good agreement with the numerical results.

In summary, our numerical and analytical results indicate that targeting the more connected infected nodes can restore the epidemic threshold, therefore making possible the eradication of a virus. Most important, however, is the finding that even moderately successful policies with small α can lead to a nonzero epidemic threshold. As the magnitude of λ_c rapidly decreases with α, the more effective a policy is at identifying and curing the hubs of a scale-free network, the higher are its chances of eradicating the virus. Finally, the simulations show that a biased treatment policy is not only more efficient but it is also less expensive than random immunization. These results, beyond improving our understanding of the basic mechanisms of virus spreading, could also offer important input into designing effective policies to eradicate computer or biological infections.

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991); J. D. Murray, *Mathematical Biology* (Springer-Verlag, Berlin, 1993); O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation* (Wiley, New York, 2000).

[2] D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, 2000).

[3] M. E. J. Newman, e-print cond-mat/0201433; C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000); M. Kuperman and G. Abramson, Phys. Rev. Lett. **86**, 2909 (2001); P. Grassberger, Math. Biosci. **63**, 157 (1983).

[4] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002); A. L. Barabasi, Phys. World **33** (July 2001).

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[6] H. Ebel, L. I. Mielsh, and S. Bornholdt, e-print cond-mat/0201476.

[7] A. L. Barabási and R. Albert, Science **286**, 509 (1999); A. L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).

[8] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).

[9] P. Erdös and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960); B. Bollobás, *Random Graphs* (Academic Press, London, 1985).

[10] R. Albert, H. Jeong, and A. L. Barabási, Nature (London) **406**, 378 (2000); D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000); R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *ibid.* **86**, 3682 (2001); **85**, 4626 (2000).

[11] X. F. Wang and G. R. Chen, IEEE Trans. Circuits Syst., I: Fundam. Theory Appl. **49**, 54 (2002); J. Jost and M. P. Joy, Phys. Rev. E **65**, 016201 (2002).

[12] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001); e-print cond-mat/0202298.

[13] P. Piot, M. Bartos, P. D. Ghys, N. Walker, and B. Schwartlander, Nature (London) **410**, 968 (2001).

[14] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001); R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).

[15] A result similar to immunizing all nodes with degree $k>k_0$ (i.e., $\alpha=\infty$) and the $\alpha=1$ case was studied independently by Pastor-Satorras and Vespignani [R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002)], who considered the permanent immunization of a $(1-1/\lambda k)$ fraction of nodes with k links, finding a finite epidemic threshold. Under permanent immunization nodes offered the cure will be immune to the virus, in contrast with the case studied here, in which cured nodes again become susceptible to the disease.

[16] To generate the scale-free network, we start from a small number of nodes (m_0) and for each time step we add a new node to the network, with m links that are connected to an old node i with k_i links according to the probability $k_i/\Sigma_j kj$. After iterating the system we obtain a network with connectivity distribution $P(k)\sim k^{-3}$ and average connectivity $\langle k\rangle=2m$ [7].