Propagation and immunization of infection on general networks with both homogeneous and heterogeneous components

Zonghua Liu,[1] Ying-Cheng Lai,[1,2] and Nong Ye[3]
[1]*Department of Mathematics, Arizona State University, Tempe, Arizona 85287*
[2]*Department of Electrical Engineering and Physics, Arizona State University, Tempe, Arizona 85287*
[3]*Department of Industrial Engineering, Arizona State University, Tempe, Arizona 85287*
(Received 11 October 2002; published 19 March 2003)

We consider the entire spectrum of architectures of general networks, ranging from being heterogeneous (scale-free) to homogeneous (random), and investigate the infection dynamics by using a three-state epidemiological model that does not involve the mechanism of self-recovery. This model is relevant to realistic situations such as the propagation of a flu virus or information over a social network. Our heuristic analysis and computations indicate that (1) regardless of the network architecture, there exists a substantial fraction of nodes that can never be infected and (2) heterogeneous networks are relatively more robust against spreads of infection as compared with homogeneous networks. We have also considered the problem of immunization for preventing wide spread of infection, with the result that targeted immunization is effective for heterogeneous networks.

DOI: 10.1103/PhysRevE.67.031911 PACS number(s): 87.23.Ge, 89.75.Hc, 05.10.−a

Complex networks have become one of the most rapidly developing areas in statistical and nonlinear physics [1] since the seminal work on scale-free networks by Barabási and Albert [2]. Most large networks in nature are sparse, that is, the average number of links per node is much smaller than the total number of nodes in the network. In a scale-free network, the number of links of various nodes follows a power-law (or an algebraic) probability distribution, indicating that nodes in the network are organized into a hierarchy of connected clusters. Scale-free networks are thus *heterogeneous*. Random networks, on the other hand, are *homogeneous* [3].

An issue of broad interest concerns how an infection or virus propagates on growing, complex networks and whether there exist effective control strategies to prevent or to suppress the spread of the infection. This problem has caught attention only recently [4–9]. A two-state (SI) model is first considered [4], where nodes in the networks can be either *susceptible* (S) or *infected* (I). A susceptible node can become infected and an infected node can recover and return to the susceptible state. The key issue of interest is whether an initially localized infection can spread to the entire network. The work in Ref. [4] on strictly heterogeneous (scale-free) networks reveals that, within the SI framework, infection can spread and persist to the entire network even when the probability of transmission is infinitesimally small. This result, however, is in sharp constrast to the well-known threshold phenomenon in epidemiology [10]: the spread of infection or virus requires that the transmission probability exceed a threshold. A more general, three-state (SIR) infection model incorporating a recovery stage is presented in Ref. [7] with the finding that, for strictly scale-free networks of finite size, the threshold phenomenon does exist in the sense that infections or viruses cannot spread for arbitrarily low transmission probabilities [7]. We mention that for completely random networks, there is an early rigorous work [11] indicating the existence of a fraction of nodes which can never be infected. The fraction approaches asymptotically the value of about 0.2 as $N\to\infty$, where N is the total number of nodes. More recent work [9] suggests that similar results hold for regular and small-world [12] networks but the percentage of the nodes that are never infected is larger than that for the random network. In fact, the infection tends to be localized for regular networks [9].

The focus of this paper is on the SIR dynamics on *general* networks that contain both heterogeneous and homogeneous components. We utilize a recent model for such general networks, in which the relative weights of heterogeneity versus homogeneity can be adjusted by a physically meaningful parameter. We then construct a simplified SIR model for general networks based on a similar model for homogeneous networks [9]. To account for the heterogeneity of the network, we introduce a structural factor in the population model. The model is simplified in that it does not include the mechanism of self-recovery. The contributions of this paper are two.

(1) We give theoretical analysis and numerical support which indicate that, in general, the fundamental SIR-dynamics stipulates the existence of a fraction of nodes that can never be infected, regardless of the relative weights of heterogeneity and homogeneity in the network connectivity. The fraction for homogeneous networks tends to be slightly smaller than that for heterogeneous networks.

(2) We investigate the role of immunization. The idea is that a node can be immunized for any particular type of virus so that it will not be infected. In situations where only a small number of nodes can be immunized, it is important to assess whether immunization should be conducted for randomly selected nodes or targeted toward certain key nodes in the network.

Our computations indicate that random immunization appears to have similar effect for general networks with different architectures, but targeted immunization is much more effective for heterogeneous networks. We believe these results are relevant to practically important problems such as

the control of spread of virus or the prevention of disease propagation on epidemic networks. In particular, our result on the effect of immunization suggests that the network architecture should be a key factor in making policies for delivering immunization.

The mechanism of self-recovery, which is neglected in our model, may in fact be important for some realistic networks [7]. For instance, a computer, after being infected by a virus, can usually recover and become susceptible to the same virus. Our model is not applicable to this type of networks. However, there are many networks in reality for which self-recovery is not important or even is irrelevant. For example, an individual can be infected by a flu virus, but the same virus can never infect the individual again. This usually means that the state of this ''node'' can go from being infected to being refractory, but it will not be susceptible to the same virus. Another example is information propagation on a social network. After hearing a particular piece of information, a person will likely lose interest in it because it is not new anymore. Thus, while this person can be ''infected'' by this piece of information, he or she will usually be ''refractory'' (R) to it after the infection. Our model applies to these networks where recovery to a susceptible state is relatively not important. As we will show, there are universal features governing the SIR dynamics in the absence of self-recovery, regardless of the architecture of the underlying network on which the infection propagates.

Dynamically, to have a scale-free network, growth and preferential attachment are required [2,13]. An earlier model [2] assumes that at a given time the probability Π_i for a node in the network with k links to acquire a new link is $\Pi_i \sim k$. This form of preferential-attachment rule yields the universal scaling exponent $\gamma = 3$ [2]. A completely random network, on the other hand, can be generated by requiring that the attachment probability be a constant independent of k. Many realistic networks are scale-free or random only to certain extent. For these, the connectivity distribution is neither algebraic nor exponential, but a mixture of the two, examples of which include the scientific collaboration network [14]. This means that the attachment probability Π_i should contain both a preferential and a constant component. A natural hypothesis is then [15] $\Pi_i \sim (1-p)k_i + p$, where $0 \leqslant p \leqslant 1$ is the parameter characterizing the relative weights between the preferential and random contributions to Π_i. That is, p is the probability that a new node is *randomly* connected to the existing node i and $(1-p)$ is the probability that the new node is *preferentially* attached to node i. For $p=0$, the model generates strictly scale-free networks, while for $p=1$, it generates completely random networks. For $0<p<1$, if a new node with m links are added to the network at each time step, the resulting connectivity distribution is shown to be [15] $P(k) \sim [k + p/(1-p)]^{-\gamma}$, where the scaling exponent γ is $\gamma = 3 + p/[m(1-p)]$. We see that the power-law scaling for scale-free networks is recovered for $p=0$ and the distribution becomes exponential $P(k) \sim e^{-k/m}$ for $p \to 1$.

The SIR model for infection propagation is well known in mathematical epidemiology [16,10]. Consider a general network containing N nodes. Each node in the network can be in one of the three states (S, I, or R). Infection is spread along the links among the nodes. A typical infection dynamics starts by having a small subset of infected nodes, acting as ''seed,'' and the remaining nodes are susceptible. As time goes, the seed can affect the susceptible nodes that are connected to it and make them infected, and some infected nodes can become refractory. An infected node will have no effect on connected nodes which are in the refractory state. For convenience in analysis, we assume the absence of self-recovery. That is, an infected node cannot recover to a state that is susceptible to the *same* virus. As we have discussed, our model still applies to a variety of realistic networks.

Let $n_I(t)$, $n_S(t)$, and $n_R(t)$ be the numbers of infected, susceptible, and refractory nodes at time t, respectively. For instance, for the SIR model [16] concerning information propagation on a social network, n_S is the number of people who have not received the information, n_I the number of individuals who have received it and wish to spread it, and n_R the number of people who have no interest to spread the information. The dynamical rules are as follows. At each time step, every infected node contacts one of its neighbors. If the node contacted is in the susceptible state, it will become infected; otherwise, the original infected node, which is transmitting the information, will lose interest in the information and become refractory to it. Initially, n_I tends to increase, so does n_R. The manner by which $n_I(t)$ changes is determined by n_I, n_S, and n_R: the larger n_I and n_S, the faster n_I increases, but the larger n_R, the more slowly n_I increases. At sufficiently large times when $n_S(t) < n_I(t) + n_R(t)$, n_I will decrease and eventually becomes zero. A typical infection dynamics thus lasts for only a finite amount of time: the ending time T is defined to be the time when there is no longer any infected nodes in the network, i.e., $n_I(T) = 0$. At time t during the evolution, a susceptible node can become infected and an infected node can become refractory, so the number of susceptible nodes keeps decreasing. For a homogeneous network, $f_I(t) = n_S(t)/N$ is the probability for a susceptible node to become infected at time $t+1$. Since only infected nodes can affect susceptible nodes, the reduced number of susceptible nodes at $t+1$ is $n_I(t)[n_S(t)/N]$. We thus have the following simple relation between $n_S(t)$ and $n_S(t+1)$:

$$\begin{aligned} n_S(t+1) &= n_S(t) - n_I(t) f_I(t) \\ &= n_S(t) - n_I(t)\{1 - [n_R(t) + n_I(t)]/N\}, \end{aligned}$$

where the conservation law

$$n_S(t) + n_I(t) + n_R(t) = N$$

is used. The probability for a node to be refractory at time $t+1$ is $f_R(t) = 1 - n_S(t)/N = [n_R(t) + n_I(t)]/N$. Since only infected nodes can become refractory, the increment in the number of refractory nodes at $t+1$ is $n_I(t) f_R(t)$. We thus have

$$\begin{aligned} n_R(t+1) &= n_R(t) + n_I(t) f_R(t) \\ &= n_R(t) + n_I(t)[n_R(t) + n_I(t)]/N. \end{aligned}$$

The conservation law gives

$$n_I(t+1)=n_I(t)+n_I(t)n_S(t)/N-n_I(t)[n_R(t)+n_I(t)]/N.$$

For a heterogeneous network, nodes are not on equal footing because of the hierarchy of structures characterized by the algebraic connectivity distribution. In particular, there can be a small subset of nodes with relatively large numbers of links. It is more likely for such a heavily linked node to be infected and then to infect the nodes that are connected to it. As a result, more nodes can become refractory through these heavily linked nodes. A heuristic way to take into account this is to assign a weighted function to $n_R(t)$ when considering the probability for a node to be refractory. Let $\epsilon(t) \geqslant 1$ be the weighting function associated with $n_R(t)$. We assume

$$f_I(t)=1-[\epsilon(t)n_R(t)+n_I(t)]/N,$$

$$f_R(t)=[\epsilon(t)n_R(t)+n_I(t)]/N, \tag{1}$$

and obtain the following SIR model governing the infection dynamics on a general network with both a heterogeneous and a homogeneous components:

$$n_S(t+1)=n_S(t)-n_I(t)\{1-[\epsilon(t)n_R(t)+n_I(t)]/N\},$$

$$n_I(t+1)=n_I(t)+n_I(t)\{1-2[\epsilon(t)n_R(t)+n_I(t)]/N\},$$

$$n_R(t+1)=n_R(t)+n_I(t)[\epsilon(t)n_R(t)+n_I(t)]/N. \tag{2}$$

Some properties of the weighting function are the following. For homogeneous networks, a natural assumption is that the function is unity: $\epsilon(t)=1$, because nodes are considered equivalent although there are statistical fluctuations in the number of links. For heterogeneous networks, there typically exists a subset of nodes with relatively large number of links. To account for the influence on infection propagation of those heavily linked nodes, a simple but reasonable way is to assume that the weighting function $\epsilon(t)$ can be greater than unity during evolution. In particular, initially we have $\epsilon(0)=1$, but as time goes it increases, decreases, and eventually returns to unity at time T when the cycle of the infection dynamics is completed. Roughly, the rates of the increase and decrease of $\epsilon(t)$ are determined by the weight of the heterogeneous component relative to that of the homogeneous one. The form of the weighting function depends on the structural detail of the network and cannot be written done explicitly.

To solve for our SIR model, we utilize the continuous-time approximation and write

$$\dot{n}_S=-n_I(t)\left[1-\frac{\epsilon(t)n_R(t)+n_I(t)}{N}\right],$$

$$\dot{n}_I=n_I(t)\left[1-2\frac{\epsilon(t)n_R(t)+n_I(t)}{N}\right],$$

$$\dot{n}_R=n_I(t)\frac{\epsilon(t)n_R(t)+n_I(t)}{N}. \tag{3}$$

This equation governs the population evolution of the whole network but not the behavior of individual nodes. It is a mean-field description. Let $N_R \equiv n_R(T)$ and $N_S \equiv n_S(T)$ be the numbers of refractory and susceptible nodes at the end of the infection cycle $[N_I \equiv n_I(T)=0]$. The quantity N_R is thus the number of nodes that are infected in the cycle. The goals of our analysis are (1) to show $N_R<N$ (so the infection cannot be spread to the whole network) and (2) to understand how N_R is affected by the network architecture. To proceed, we follow the approach in Ref. [9] by introducing the auxiliary variable $s \equiv \int_0^t n_I(t')dt'$. Assuming that the seed of infection at $t=0$ contains only one node, we have the initial conditions: $n_I(0)=1$, $n_S(0)=N-1$, and $n_R(0)=0$. Under these, we obtain the following solution to Eq. (3):

$$n_S(s)=a(s)(N-1)e^{-s/N},$$

$$n_I(s)=b(s)[1-s+2(N-1)(1-e^{-s/N})],$$

$$n_R(s)=c(s)[s-(N-1)(1-e^{-s/N})], \tag{4}$$

where the functions $a(s)$, $b(s)$, and $c(s)$ depend on the weighting function $\epsilon(t)$. In fact, an explicit expression for $a(s)$ can be written down, as follows:

$$a(s)=1+\int_0^t dt[\{\epsilon(t)-1\}n_I(t)n_R(t)e^{s/N}]/[N(N-1)]>1.$$

Setting $n_I(S)=0$, we obtain the following relation giving implicitly the completion time T of the infection cycle:

$$S-1=2(N-1)(1-e^{-S/N}),$$

where $S \equiv \int_0^T n_I(t)dt$. For large N, we have $S/(2N) \equiv l \approx 1-e^{-2l}$, which gives $l \approx 0.8$ and $S \approx 1.6N$. The quantity of interest N_R is given by

$$N_R=c(S)[S-(N-1)(1-e^{-S/N})] \approx c(S)(0.8N).$$

For homogeneous networks $\epsilon=1$, $c(S)=1$, hence $N_R^{homogeneous} \approx 0.8N<N$ (in the large-N limit). For heterogeneous networks for which $\epsilon(t)>1$, note from Eq. (3) that the product $\epsilon(t)n_R(t)$ appears as a single entity. An immediate result is that $N_R^{general}<N_R^{homogeneous}$. That is, in a general network with a heterogeneous component, fewer nodes will be infected as compared with a homogeneous network. As we will show, numerical computations give strong credance to the validity of this result. We thus conclude that

$$N_R^{general}<N_R^{homogeneous} \approx 0.8N<N, \tag{5}$$

indicating that the spread of infection in a general network cannot be global because there exists a fraction of nodes that can never be infected. This is a universal property of the SIR dynamics without recovery because this feature is independent of the details of the network.

We now provide numerical support for Eq. (5). Note that, because of the random components involved in the construction of the network and because the theoretical prediction (5) is valid only in the large N and continuous-time limits, the

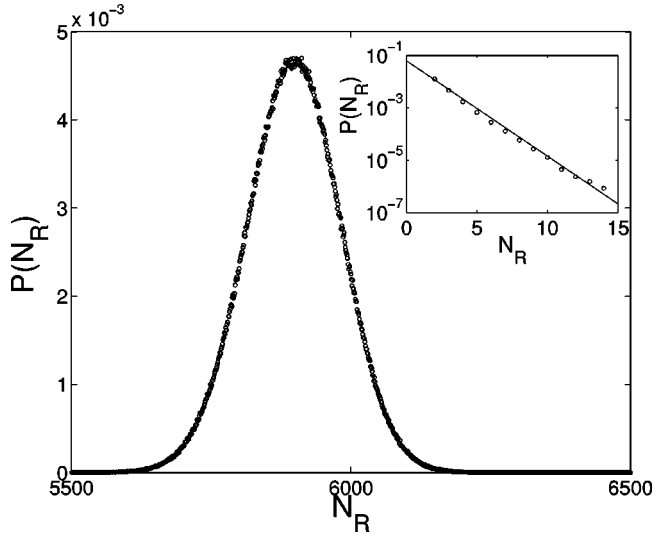

FIG. 1. Histogram of N_R for $p=0.5$ and $m=5$ obtained using 10^6 realizations. The inset shows the exponential distribution of N_R for $N_R \approx 0$.

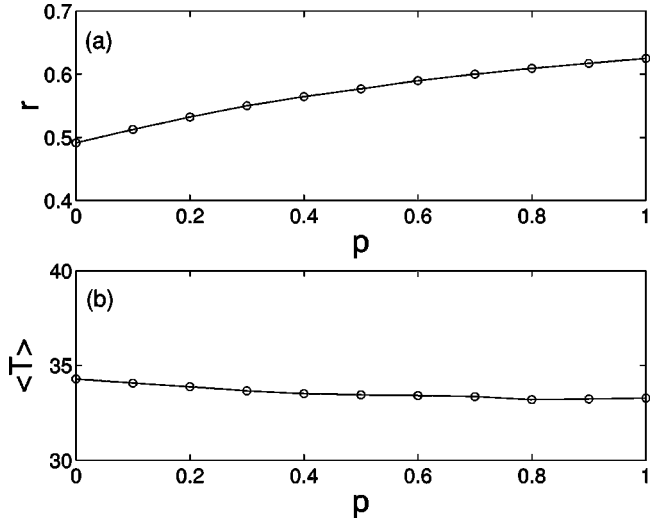

FIG. 2. For $m=5$, $N=10^4$, (a) r versus p, which indicates that the number of infected nodes for a network containing a heterogeneous (or scale-free) component is smaller than that for a random network; and (b) the average time $\langle T \rangle$ required to complete an infection cycle versus p. The dependence of $\langle T \rangle$ on p is weak.

key quantity N_R should be interpreted in a statistical sense. To compute the probability distribution of N_R for a given network architecture, we construct a network with $N=10^4$ nodes and randomly choose a seeding node at $t=0$ from which the infection starts. At each time step, every infected node infects one of the nodes that are connected to it. If this node is susceptible, it will be infected; otherwise, the original infected node itself will become refractory. The whole process continues until time T at which there is no longer any infected node $[n_I(T)=0]$. The number N_R of refractory nodes is then recorded. The process is repeated for 10^6 times, each time starting with a *new* network configuration and a *new* seeding node, and a histogram of N_R is constructed. One example is shown in Fig. 1 for $p=0.5$ and $m=5$, where we observe an approximately Gaussian distribution with peak around $N_R \approx 5900$. The probability $P(N_R > 6500)$ is zero, indicating that about one third of the nodes in the network are not be infected, which is consistent with the theoretical prediction (5). We also find an exponential peak for $N_R < 10$, as shown in the inset. This is caused by the placement of the seed of infection at nodes which locally are connected to other nodes in a regular fashion. Infection starting from these nodes are most likely to remain localized [9]. Regarding the evolutions of n_S and n_I, our analysis indicates that n_S decreases exponentially with s and $n_I(t)$ increases initially, reaches a maximum, and then decreases. These behaviors have been verified numerically.

To compare the infection dynamics on general networks with different architectures, a convenient quantity is the average fraction of infected nodes, or the *order parameter* [9]

$$r = \langle N_R/N \rangle = \int N_R P(N_R) dN_R, \tag{6}$$

where $P(N_R)$ is the normalized probability satisfying $\int P(N_R) dN_R = 1$. Our analysis predicts that N_R grows linearly with N and, hence, the ratio remains constant as N is increased, which is indeed observed numerically. Figure 2(a) shows, for $m=5$ and $N=10^4$, the ratio r versus the key parameter p of the network. Recall that p determines the relative weights between the heterogeneous and homogeneous components of the network, where $p=0$ corresponds to a scale-free network, and $p=1$ to a completely random network. We see that the number of infected nodes for a network containing a heterogeneous (or scale-free) component is indeed smaller than that for a random network, as predicted by Eq. (5). For $p=1$ (random network), the percentage of the number of infected nodes is about 0.62. We find that this percentage approachs the theoretical value of 0.8 as m is increased. Figure 2(b) shows the time T required to complete an infection cycle versus p, which exhibits little variation. This is somewhat expected from the evolution equations in Eq. (3), where the structural factor $\epsilon(t)$ of the network is assumed to be associated with the variable $n_R(t)$. Thus, although the final value of N_R depends on the architecture of the network, there is little dependence of the time it takes to reach the final state on the network structure.

We next consider two types of external immunization within the framework of our SIR model: random versus intentional, where the former means applying immunization to a randomly selected subset of nodes and the latter means making refractory a particularly chosen set of nodes, such as heavily connected ones. In the case of intentional immunization, we can assume that the probability of a node being immunized is proportional to the number of links that it carries. The immunized nodes are chosen according to the probabilities. As a result, when resources are limited, only nodes with relatively large numbers of links are immunized. Figure 3(a) shows, for three different network architectures [$p=0$ (scale-free, open circles), $p=0.5$ (general, stars), and $p=1$ (random, open squares)], the order parameter r versus the fraction f of the randomly immunized nodes, where other parameters are $m=5$ and $N=10^4$. Apparently, the value of r

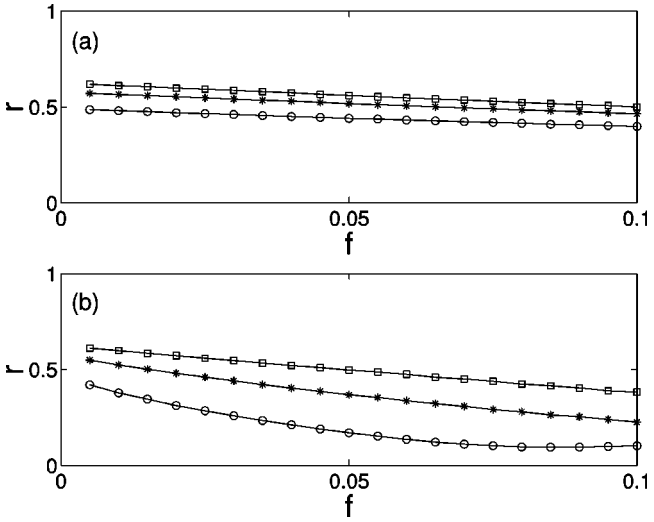

FIG. 3. For three different network architectures [$p=0$ (scale-free, open circles), $p=0.5$ (general, stars), and $p=1$ (random, open squares)], the order parameter r versus the fraction f of the immunized nodes for (a) random and (b) intentional immunizations. Other network parameters are $m=5$ and $N=10^4$.

decreases slowly as f is increased, indicating that random immunization has little effect on reducing the spread of the infection, regardless of the network architecture. The situation is somewhat different for intentional immunization targeted at nodes with relatively large number of links, as shown in Fig. 3(b), where the legends are the same as for Fig. 3(a). We see that for scale-free networks (the lower trace), the value of the order parameter r decreases more rapidly as f is increased, demonstrating that intentional immunization can be relatively more effective to prevent the spread of infection for scale-free networks. The intuitive explanation is the existence of a small subset of heavily connected nodes. They act as local ''centers'' for transferring the infection. Immunizing those nodes can block the passage of the infection relatively more effectively.

In summary, we have investigated the SIR dynamics on general growing networks and the role of immunization, which are important issues for realistic networks. Our findings are (1) there exists a substantial fraction of nodes that can never be infected, regardless of the network architecture, (2) heterogeneous networks are more robust against spread of infection, comparing with homogeneous networks, and (3) targeted immunization can be quite effective for heterogeneous networks with the implication that, by immunizing a small subset of nodes with many connections, infection spread can be greatly suppressed. These results are relevant to a host of problems in many areas of natural science, engineering, and social science. For instance, blockage of information spread in a social network may be achieved by making a few heavily connected individuals indifferent to the information. Or, in a sexual partnership network which is typically highly heterogeneous, targeted immunization on a group of highly active individuals (such as prostitutes) may effectively prevent the wide spread of sexually transmitted diseases [7].

This work was supported by AFOSR under Grant No. F49620-01-1-0317.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999); **281**, 69 (2000).
[3] P. Erdös and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960); B. Bollobaás, *Random Graphs* (Academic, London, 1985).
[4] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001); **65**, 035108 (2002).
[5] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).
[6] Z. Dezsö and A.-L. Barabási, Phys. Rev. E **65**, 055103(R) (2002).
[7] R.M. May and A.L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[8] M. Kuperman and G. Abramson, Phys. Rev. Lett. **86**, 2909 (2001).
[9] D.H. Zanette, Phys. Rev. E **64**, 050901(R) (2001); **65**, 041908 (2002).
[10] R.M. Anderson and R.M. May, *Infectious Diseases in Humans* (Oxford University Press, Oxford, 1992).
[11] A. Sudbury, J. Appl. Probab. **22**, 443 (1985).
[12] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[13] Z. Liu, Y.-C. Lai, and N. Ye, Phys. Rev. E **66**, 036112 (2002).
[14] M.E.J. Newman, Phys. Rev. E **64**, 016131 (2001).
[15] Z. Liu, Y.-C. Lai, N. Ye, and P. Dasgupta, Phys. Lett. A **303**, 337 (2002).
[16] J.C. Frauenthal, *Mathematical Modelling in Epidemiology* (Springer-Verlag, Berlin, 1980).