Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks

Marc Barthélemy,[1] Alain Barrat,[2] Romualdo Pastor-Satorras,[3] and Alessandro Vespignani[2]

[1]*CEA-Centre d'Etudes de Bruyères-le-Châtel, Département de Physique Théorique et Appliquée, BP12, 91680 Bruyères-Le-Châtel, France*

[2]*Laboratoire de Physique Théorique (UMR du CNRS 8627), Batiment 210, Université de Paris-Sud, 91405 Orsay, France*

[3]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord 08034 Barcelona, Spain*

(Received 17 November 2003; published 27 April 2004)

We study the effect of the connectivity pattern of complex networks on the propagation dynamics of epidemics. The growth time scale of outbreaks is inversely proportional to the network degree fluctuations, signaling that epidemics spread almost instantaneously in networks with scale-free degree distributions. This feature is associated with an epidemic propagation that follows a precise hierarchical dynamics. Once the highly connected hubs are reached, the infection pervades the network in a progressive cascade across smaller degree classes. The present results are relevant for the development of adaptive containment strategies.

DOI: 10.1103/PhysRevLett.92.178701 PACS numbers: 89.75.Hc, 05.70.Ln, 87.19.Xx, 87.23.Ge

The connectivity pattern of the network of individual contacts has long been acknowledged as a relevant factor in determining the properties of epidemic spreading phenomena [1,2]. This issue assumes a particular relevance in the case of networks characterized by complex topologies and very heterogeneous structures [3–7] that in many cases present us with new epidemic propagation scenarios [8–13]. A striking example of this situation is provided by scale-free networks characterized by large fluctuations in the number of connections (degree) k of each vertex. This feature usually finds its signature in a heavy-tailed degree distribution with power-law behavior of the form $P(k) \sim k^{-\gamma}$, with $2 \leq \gamma \leq 3$, that implies a nonvanishing probability of finding vertices with very large degrees [3–5,14]. The latter are the ''hubs'' or ''superspreaders'' [1] responsible for the proliferation of infected individuals, whatever the rate of infection characterizing the epidemic, eventually leading to the absence of any epidemic threshold below which the infection cannot initiate a major outbreak [10]. This new scenario is of practical interest in computer virus diffusion and the spreading of diseases in heterogeneous populations [4–7]. It also raises new questions on how to protect the network and find optimal strategies for the deployment of immunization resources [15–17]. Thus far, however, studies of epidemic models in complex networks have been focused on the stationary properties of endemic states or the final prevalence (number of infected individuals) of epidemics. The dynamical evolution of the outbreaks has been instead far less investigated and a detailed inspection of the temporal behavior in complex topologies is still missing.

In this Letter, we intend to fill this gap by providing a first analysis of the time evolution of epidemic outbreaks in complex networks with highly heterogeneous connectivity patterns. We consider the time behavior of epidemic outbreaks and find that the growth of infected individuals is governed by a time scale τ proportional to the ratio between the first and second moment of the network's degree distribution, $\tau \sim \langle k \rangle / \langle k^2 \rangle$. This implies an instantaneous rise of the prevalence in very heterogeneous networks where $\langle k^2 \rangle \to \infty$ in the infinite size limit. In particular, this result shows that scale-free networks with $2 \leq \gamma \leq 3$ exhibit, along with the lack of an intrinsic epidemic threshold, a virtually infinite propagation velocity of the infection. Furthermore, we study the detailed propagation in time of the infection through the different degree classes in the population. We find a striking hierarchical dynamics in which the infection propagates via a cascade that progresses from higher to lower degree classes. This infection hierarchy might be used to develop dynamical *ad hoc* strategies for network protection.

In order to study the dynamical evolution of epidemic outbreaks, we shall focus on the susceptible-infected (SI) model in which individuals can be in two discrete states, either susceptible or infected [2,18]. Each individual is represented by a vertex of the network and the edges are the connections between individuals along which the infection may spread. The total population (the size of the network) N is assumed to be constant and, if $S(t)$ and $I(t)$ are the number of susceptible and infected individuals at time t, respectively, then $N = S(t) + I(t)$. In the SI model, the infection transmission is defined by the spreading rate λ at which susceptible individuals acquire the infection from an infected neighbor. In this model, infected individuals remain always infective, an approximation that is useful to describe early epidemic stages in which no control measures are deployed. We shall see in the following that the results obtained for the SI model can be readily generalized to more complex schemes.

A first analytical description of the SI model can be undertaken within the homogeneous mixing hypothesis [2,18], consisting of a mean-field description of the system in which all vertices are considered as being

equivalent. In this case, the reaction rate equation for the average density of infected individuals $i(t) = I(t)/N$ (the prevalence) reads as

$$\frac{di(t)}{dt} = \lambda\langle k\rangle i(t)[1 - i(t)]. \tag{1}$$

The above equation states that the growth rate of infected individuals is proportional to the spreading rate, λ, the density of susceptible vertices that may become infected, $s(t) = 1 - i(t)$, and the number of infected individuals in contact with any susceptible vertex. The homogeneous mixing hypothesis considers that this last term is simply the product of the number of neighbors $\langle k\rangle$ and the average density $i(t)$. Obviously, this approximation neglects correlations among individuals and considers that all vertices have the same number of neighbors $\langle k\rangle$; i.e., it assumes a perfectly homogeneous network. At small times, when the density of infected vertices is very small, we can neglect terms of order $\mathcal{O}(i^2)$ and obtain the leading behavior $i(t) \simeq i_0 e^{t/\tau_H}$, where i_0 is the initial density of infected individuals and $\tau_H = (\lambda\langle k\rangle)^{-1}$ is the time scale governing the growth of the infection in a homogeneous network.

The above calculations recover that the outbreak's time scale τ_H is inversely proportional to the epidemic reproduction rate, i.e., the mean number of infections generated by each infected individual. However, in populations with a heterogeneous connectivity pattern, it is known that the reproduction rate depends upon the contacts' fluctuations [2], and it is natural to expect that also the outbreaks' time scale is analogously affected. Indeed, in heterogeneous networks the degree k of vertices is highly fluctuating and the average degree is not anymore a meaningful characterization of the network properties. In order to take fully into account the degree heterogeneity in the dynamical evolution, it is possible to write down the whole set of reaction rate equations for the average densities of infected vertices of degree k, $i_k(t) = I_k(t)/N_k$, where N_k and $I_k(t)$ are the number of vertices and infected vertices within each degree class k, respectively [10]. For the SI model, the evolution equations read as

$$\frac{di_k(t)}{dt} = \lambda k[1 - i_k(t)]\Theta_k(t), \tag{2}$$

where the creation term is proportional to the spreading rate λ, the degree k, the probability $1 - i_k$ that a vertex with degree k is not infected, and the density Θ_k of infected neighbors of a vertex of degree k. The latter term is thus the average probability that any given neighbor of a vertex of degree k is infected. In the case of uncorrelated networks [19], $\Theta_k \equiv \Theta$ is independent of the degree of the emanating edge. The probability that each edge of a susceptible is pointing to an infected vertex of degree k' is proportional to the fraction of edges emanated from these vertices. By considering that at least one of the edges of each infected vertex is pointing to another infected vertex, from which the infection has been transmitted, one obtains

$$\Theta(t) = \frac{\sum_{k'}(k' - 1)P(k')i_{k'}(t)}{\langle k\rangle}, \tag{3}$$

where $\langle k\rangle = \sum_{k'} k'P(k')$ is the proper normalization factor dictated by the total number of edges. A reaction rate equation for $\Theta(t)$ can be obtained from Eqs. (2) and (3). In the initial epidemic stages, we neglect terms of order $\mathcal{O}(i^2)$, obtaining the following set of equations:

$$\frac{di_k(t)}{dt} = \lambda k\Theta(t), \tag{4}$$

$$\frac{d\Theta(t)}{dt} = \lambda\left(\frac{\langle k^2\rangle}{\langle k\rangle} - 1\right)\Theta(t). \tag{5}$$

These equations can be solved with the uniform initial condition $i_k(t = 0) = i_0$ yielding for the total average prevalence $i(t) = \sum_k P(k)i_k(t)$:

$$i(t) = i_0\left[1 + \frac{\langle k\rangle^2 - \langle k\rangle}{\langle k^2\rangle - \langle k\rangle}(e^{t/\tau} - 1)\right], \tag{6}$$

where

$$\tau = \frac{\langle k\rangle}{\lambda(\langle k^2\rangle - \langle k\rangle)}. \tag{7}$$

This readily implies that the growth time scale of an epidemic outbreak is related to the graph heterogeneity. Indeed, the ratio $\kappa = \langle k^2\rangle/\langle k\rangle$ is the parameter defining the level of heterogeneity of the network, since the normalized degree variance can be expressed as $\kappa/\langle k\rangle - 1$ and therefore high levels of fluctuations correspond to $\kappa \gg \langle k\rangle$. In networks with a Poisson degree distribution in which $\kappa = \langle k\rangle + 1$, we recover the result $\tau \simeq (\lambda\langle k\rangle)^{-1}$. Instead, in networks with very heterogeneous connectivity patterns, κ is very large and the outbreak time scale τ is very small, signaling a very fast diffusion of the infection. In particular, in scale-free networks characterized by a degree exponent $2 \leq \gamma \leq 3$, we have $\kappa \sim \langle k^2\rangle \to \infty$ with the network size $N \to \infty$. Therefore, while τ is a function of the finite size N, for large scale-free networks we face a virtually instantaneous rise of the epidemic incidence.

It is worth stressing that these results can be easily extended to the susceptible-infected-susceptible and the susceptible-infected-removed models (SIR) [2], in which Eq. (2) contains an extra term $-\mu i_k(t)$ defining the rate at which infected individuals of degree k recover and become again susceptible or removed from the population, respectively. In addition, in the SIR model, the normalization imposes that $s_k(t) = 1 - i_k(t) - r_k(t)$, where $r_k(t)$ is the density of removed individuals of degree k. The inclusion of the decaying term $-\mu i_k$ does not change the picture obtained in the SI model. By using the same approximations, the time scale is found to behave as

$\tau \sim \langle k\rangle/[\lambda\langle k^2\rangle - (\mu+1)\langle k\rangle]$. In the case of diverging fluctuations, the time-scale behavior is therefore still dominated by $\langle k^2\rangle$, and τ is always positive whatever the spreading rate λ. This allows one to recover the absence of an epidemic threshold, i.e., the lack of a decreasing prevalence region in the parameter space.

In order to check these analytical results, we have performed numerical simulations of the SI model on networks generated with the Barabási-Albert (BA) algorithm [20]. These networks have a scale-free degree distribution $P(k) \sim k^{-3}$, and we use different network sizes N and minimum degree values m in order to change the level of heterogeneity, that in this case is given by $\kappa \sim m \ln N$. Simulations use an agent-based modeling strategy in which at each time step the SI dynamics is applied to each vertex by considering the actual state of the vertex and its neighbors. It is then possible to measure the evolution of the average number of infected individuals and other quantities, by averaging over 10^3 realizations of the dynamics. In addition, given the stochastic nature of the model, different initial conditions and network realizations can be used to obtain averaged quantities. In Fig. 1, we report the early time behavior of outbreaks in networks with different heterogeneity levels and the behavior of the measured τ with respect to $\langle k\rangle/\lambda(\langle k^2\rangle - \langle k\rangle)$. The numerical results recover the analytical prediction with great accuracy. Indeed, the BA network is a good example of a heterogeneous network in which the approximations used in the calculations are satisfied [19]. In networks with correlations, we expect to find different quantitative results but a qualitatively similar framework as happens in the case of the epidemic threshold evaluation [19].

The previous results show that the heterogeneity of scale-free connectivity patterns favors epidemic spreading not only by suppressing the epidemic threshold, but also by accelerating the virus propagation in the population. The velocity of the spreading leaves us with very short response times in the deployment of control measures, and a detailed knowledge of the way epidemics propagate through the network could be very valuable in the definition of adaptive strategies. Indeed, the epidemic diffusion is far from homogeneous. The simple formal integration of Eq. (2) written for $s_k = 1 - i_k$ yields $s_k(t) = s_k^0 e^{-\lambda k \Phi(t)}$, where $\Phi(t) = \int_0^t dt' \, \Theta(t')$. This result is valid for any value of the degree k and the function $\Phi(t)$ is positive and monotonically increasing. This last fact implies that s_k is decreasing monotonically towards zero when time grows. For any two values $k > k'$, and whatever the initial conditions s_k^0 and $s_{k'}^0$ are, there exists a time t^* after which $s_k(t) < s_{k'}(t)$. This means that vertices belonging to higher degree classes are generally infected more quickly. A more precise characterization of the epidemic diffusion through the network can be achieved by studying some convenient quantities in numerical spreading experiments in BA networks. First, we measure the average degree of the newly infected nodes at time t, defined as

$$\overline{k}_{\rm inf}(t) = \frac{\sum_k k[I_k(t) - I_k(t-1)]}{I(t) - I(t-1)}. \tag{8}$$

In Fig. 2(a), we plot this quantity for BA networks as a function of the rescaled time t/τ. The curves show an initial plateau that can be easily understood by considering that, at very low density of infected individuals i, each vertex will infect a fraction of its neighbors without correlations with the spreading from other vertices. In this case, each edge points to a vertex with degree k with probability $kP(k)/\langle k\rangle$ and the average degree of newly infected vertices is given by $\overline{k}_{\rm inf}(t) = \langle k^2\rangle/\langle k\rangle$. After this initial regime, $\overline{k}_{\rm inf}(t)$ decreases smoothly when time increases. The dynamical spreading process is therefore clear; after the hubs are very quickly infected, the spread is going always towards smaller values of k. This is confirmed by the large time regime that settles in a plateau $\overline{k}_{\rm inf}(t) = m$; i.e., the vertices with the lowest degree are the last to be infected.

Further information on the infection propagation is provided by the inverse participation ratio $Y_2(t)$ [21]. We first define the weight of infected individuals in each class degree k by $w_k(t) = I_k(t)/I(t)$. The quantity Y_2 is then defined as

$$Y_2(t) = \sum_k w_k^2(t). \tag{9}$$

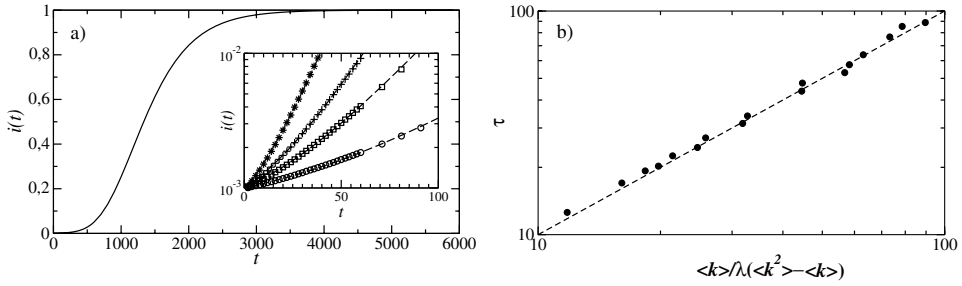

FIG. 1. (a) Average density of infected individuals versus time in a BA network of $N = 10^4$ with $m = 2$. The inset shows the exponential fit obtained in the early times (lines) and the numerical curves $i(t)$ for networks with $m = 4$, 8, 12, and 20 (from bottom to top). (b) Measured time scale τ in BA networks as obtained from exponential fitting versus the theoretical prediction for different values of m and N corresponding to different levels of heterogeneity.

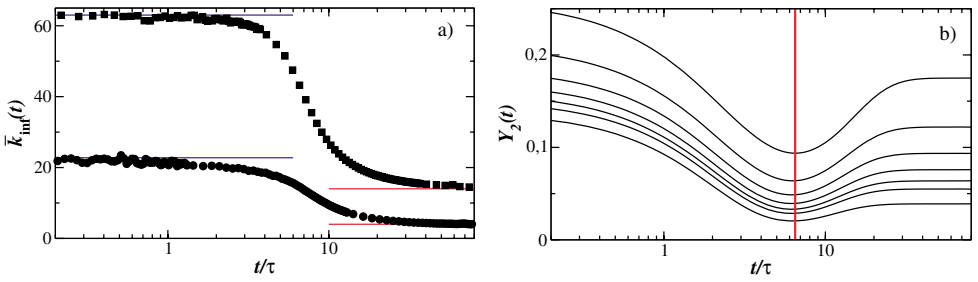

FIG. 2 (color online). (a) Time behavior of the average degree of the newly infected nodes for SI outbreaks in BA networks of size $N = 10^4$. Time is rescaled by τ. Reference lines are drawn at the asymptotic values $\langle k^2\rangle/\langle k\rangle$ for $t \ll \tau$ and m for $t \gg \tau$. The two curves are for $m = 4$ (bottom) and $m = 14$ (top). (b) Inverse participation ratio Y_2 versus time for BA network of size $N = 10^4$ with minimum degree $m = 4$, 6, 8, 10, 12, 14, and 20, from top to bottom. Time is rescaled with τ. The reference line indicates the minimum of Y_2 around $t/\tau \simeq 6.5$.

If $Y_2 \sim 1/k_{\max}$, infected vertices are homogeneously distributed among all degree classes. In contrast, if Y_2 is not small (of order unity) then the infection is localized on some specific degree classes that dominate the sum of Eq. (9). In Fig. 2(b), we report the behavior of Y_2 versus time for BA networks with different minimum degree. The function Y_2 has a maximum at the early time stage, indicating that the infection is localized on the large k classes, as we infer from the plot of $\langle k^{\inf}(t)\rangle$ [see Fig. 2(a)]. Afterwards Y_2 decreases, with the infection progressively invading lower degree classes, and providing a more homogeneous diffusion of infected vertices in the various k classes. Finally, the last stage of the process corresponds to the capillary invasion of the lowest degree classes which have a larger number of vertices and thus provide a larger weight. In the very large time limit, when the whole network is infected, $Y_2(t = \infty) = \sum_k P(k)^2$. Noticeably, curves for different levels of heterogeneity have the same time profile in the rescaled variable t/τ. This implies that, despite the various approximations used in the calculations, the whole spreading process is dominated by the time scale defined in the early exponential regime of the outbreak.

The presented results provide a clear picture of the infection propagation in heterogeneous networks. First, the infection takes control of the large degree vertices in the network. Then it rapidly invades the network via a cascade through progressively smaller degree classes. The dynamical structure of the spreading is therefore characterized by a hierarchical cascade from hubs to intermediate k, and finally to small k classes. This result, along with the very fast growth rate of epidemic outbreaks, could be of practical importance in the setup of dynamic control strategies in populations with heterogeneous connectivity patterns. In particular, targeted immunization strategies that evolve with time might be particularly effective in epidemics control.

A. B., A. V., and R. P.-S. are partially funded by the EC-Fet Open project COSIN IST-2001-33555. R. P.-S. acknowledges support from the MCyT (Spain), and from the DURSI, Generalitat de Catalunya (Spain).

[1] H. W. Hethcote and J. A. Yorke, Lect. Notes Biomath. **56**, 1 (1984).
[2] R. M. Anderson and R. M. May, *Infectious Diseases in Humans* (Oxford University Press, Oxford, 1992).
[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2000).
[4] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[5] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2003).
[6] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[7] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001).
[8] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[9] G. Abramson and M. Kuperman, Phys. Rev. Lett. **86**, 2909 (2001).
[10] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001).
[11] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[12] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **26**, 521 (2002).
[13] M. E. J. Newman, Phys. Rev. E **64**, 016128 (2002).
[14] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[15] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 036104 (2002).
[16] Z. Dezso and A.-L. Barabási, Phys. Rev. E **65**, 055103 (2002).
[17] R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
[18] J. D. Murray, *Mathematical Biology* (Springer, New York, 1993), 2nd ed.
[19] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Lect. Notes Phys. **625**, 127 (2003).
[20] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[21] B. Derrida and H. Flyvbjerg, J. Phys. A **20**, 5273 (1987).