Polynomial Growth in Branching Processes with Diverging Reproductive Number

Alexei Vazquez

Department of Physics and Center for Complex Network Research, University of Notre Dame, Notre Dame, Indiana 46556, USA

(Received 4 May 2005; published 27 January 2006)

We study the spreading dynamics on graphs with a power law degree distribution $p_k \sim k^{-\gamma}$, with $2 < \gamma < 3$, as an example of a branching process with a diverging reproductive number. We provide evidence that the divergence of the second moment of the degree distribution carries as a consequence a qualitative change in the growth pattern, deviating from the standard exponential growth. First, the population growth is extensive, meaning that the average number of vertices reached by the spreading process becomes of the order of the graph size in a time scale that vanishes in the large graph size limit. Second, the temporal evolution is governed by a polynomial growth, with a degree determined by the characteristic distance between vertices in the graph. These results open a path to further investigation on the dynamics on networks.

DOI: 10.1103/PhysRevLett.96.038702 PACS numbers: 89.75.Hc, 05.70.Ln, 87.19.Xx, 87.23.Ge

Branching processes model the evolution of populations whose elements reproduce generating new elements [1,2], such as a population of physical particles [3,4], cells [2], or infected individuals [5]. A key magnitude determining the dynamical evolution of the population size is the average reproductive number $\tilde{R}$, giving the number of secondary particles generated by a primary particle. When $\tilde{R} < 1$, the average number of new elements decreases exponentially, while it grows exponentially when $1 < \tilde{R} < \infty$ [1]. On the other hand, it has been found recently that $\tilde{R}$ may be unbounded for branching processes taking place on graphs with a power law degree distribution [5–9], where by unbounded we mean that $\tilde{R}$ diverges with increasing graph size. This observation is extremely important, since several graphs representing interactions among humans or computers are characterized by a power law degree distribution [10–14], requiring us to consider branching processes with an unbounded average reproductive number.

Barthélemy *et al.* [15] have recently studied the spreading dynamics of an infectious disease on a graph with a power law degree distribution. Using a mean-field approach, they obtained that the average number of infected vertices grows exponentially in time with a characteristic time $\tau \sim \langle k^2 \rangle^{-1}$, where $\langle k^2 \rangle$ is the second moment of the degree distribution p_k. For graphs where $p_k \sim k^{-\gamma}$, with $2 < \gamma < 3$, the second moment diverges and $\tau \to 0$ with increasing graph size, predicting that all vertices will be instantaneously infected [15]. A disease that spreads at a constant rate, however, cannot spread to all vertices in a time scale much smaller than the inverse of the spreading rate, indicating that the predicted exponential growth should not dominate the system's dynamics.

In this Letter, we study branching processes with an unbounded average reproductive number using a spreading process on a graph as a case study. When the degree distribution has the power law tail $p_k \sim k^{-\gamma}$, with $2 < \gamma < 3$, we obtain that the exponential regime is followed by a polynomial growth in time, a result that is completely unexpected based on previous mathematical studies. We also show that both the characteristic time separating the exponential and polynomial regimes and the polynomial degree depend on the characteristic distance between vertices. More important, in the limit of infinite graph sizes, the exponential regime is virtually absent, indicating that the polynomial regime is a novel and characteristic feature of the spreading dynamics on graphs with degree exponent $2 < \gamma < 3$ and, more generally, of branching processes with an unbounded average reproductive number.

Consider a spreading process on a graph with a treelike structure. At $t = 0$, a vertex selected at random is infected by a "virus," which can then propagate to other vertices through the graph edges. The causal tree representing the spreading process can be modeled as a branching process. Each vertex in the causal tree represents an infected vertex in the original graph, and each arc in the causal tree represents the generation of a secondary infected vertex from a primary infected vertex. The out-degree of a vertex in the causal tree gives the number of other vertices it infects, i.e., its reproductive number. In turn, the length of an arc $A \to B$ in the causal tree gives the generation time, the time elapsed from the infection of the primary case A to the infection of the secondary case B. Finally, the vertex generation coincides with the topological distance from the first infected vertex, the root, in the original graph.

We assume that the reproductive numbers are independent random variables with the probability distribution $q_k^{(d)}$ and average reproductive number $R^{(d)} = \sum_k q_k^{(d)} k$, parametrized by the generation d. The parametrization by d is introduced to take into account that the degree distribution may change significantly from generation to generation [7,16]. We also assume that the generation times are independent random variables with the distribution $G^{(d)}(\tau)$ and the probability density $g^{(d)}(\tau) = dG^{(d)}(\tau)/d\tau$. Let $P_N^{(d)}(t)$ be the probability distribution of the number of vertices N that are found at time t in a branch of the causal tree, given

that the branch is rooted at a vertex at generation d. Because of the tree structure, we can write the recursive relation

$$P_N^{(d)}(t) = \sum_{k=0}^{\infty} q_k^{(d)} \sum_{N_1=0}^{\infty} \cdots \sum_{N_k=0}^{\infty} \delta_{\sum_{i=1}^{k} N_i + 1, N} \times \prod_{i=1}^{k} \Big[\int_0^t dG^{(d)}(\tau) P_{N_i}^{(d+1)}(t-\tau) + \delta_{N_i,0}(1 - G^{(d)}(t)) \Big], \quad (1)$$

with the boundary condition $P_N^{(D)}(t) = \delta_{N,1}$, where D is the maximum distance between two vertices on the graph. The sum over k runs over the possible reproductive numbers of the reference vertex, while the sum over N_i, $i = 1, \ldots, k$, runs over the possible number of infected vertices in the branch rooted at the ith neighbor of the reference vertex. These sums are then restricted by the Kronecker delta to configurations satisfying $1 + \sum_{i=1}^{k} N_i = N$. Finally, within the $[\cdots]$ we have the probability that the branch rooted at the ith neighbor has N_i infected vertices at time $t - \tau$, averaged over the generation time distribution $G(\tau)$. The product structure in (1) suggests the use of the generating functions

$$H^{(d)}(x) = \sum_{k=0}^{\infty} q_k^{(d)} x^k, \quad (2)$$

$$F^{(d)}(x, t) = \sum_{N=0}^{\infty} P_N^{(d)}(t) x^N, \quad (3)$$

for the reproductive number and the number of infected vertices, respectively. From (1)–(3), we obtain

$$F^{(d)}(x, t) = x H^{(d)}\Big(\int_0^t dG^{(d)}(\tau) F^{(d+1)}(x, t-\tau) + 1 - G^{(d)}(t) \Big), \quad (4)$$

with the boundary condition $F^{(D)}(x) = x$. From this equation, we obtain the average number of infected vertices on the branch rooted at a vertex at generation d,

$$N^{(d)}(t) = \frac{\partial F^{(d)}(1, t)}{\partial x}, \quad (5)$$

with the boundary condition $N^{(D)}(t) = 1$. Iterating this equation from $d = D$ to $d = 0$, we obtain the average number of infected vertices at time t, $N^{(0)}(t)$, and the average number of new vertices infected between t and $t + dt$, $n(t) = dN^{(0)}(t)/dt$, resulting in

$$n(t) = \sum_{d=1}^{D} z_d (g^{(0)} * g^{(1)} * \cdots * g^{(d)}(t)), \quad (6)$$

where

$$z_d = \prod_{l=0}^{d} R_l \quad (7)$$

is the average number of vertices at generation d, and the second factor is the probability that the infection has reached a vertex at generation d, where $*$ denotes the convolution operation, for instance, $g^{(0)} * g^{(1)}(t) = \int_0^t d\tau g^{(0)}(\tau) g^{(1)}(t-\tau)$.

Next we consider the cases when (i) the reproductive number of vertices other than the root has the same statistical properties, i.e., $R^{(0)} = R$ and $R^{(d)} = \tilde{R}$ for $d \geq 1$, and (ii) the infection is transmitted from an infected vertex to a susceptible (not yet infected) vertex at constant rate $1/T_G$. This last assumption corresponds to an exponential distribution of generation times $G^{(d)}(\tau) = 1 - \exp(-t/T_G)$, with average generation time T_G. Under these approximations from (6) and (7), we obtain

$$n(t) = \frac{R}{T_G} \exp\Big(-\frac{t}{T_G}\Big) \sum_{d=1}^{D} \frac{1}{(d-1)!} \Big(\frac{\tilde{R} t}{T_G}\Big)^{d-1}. \quad (8)$$

The sum in (8) is the Taylor series expansion of $\exp(\tilde{R}t/T_G)$, up to the $D - 1$ order. It actually approximates an exponential function depending on the ratio of t/τ_0, where

$$\tau_0 = T_G \frac{D}{\tilde{R}}. \quad (9)$$

When $t \ll \tau_0$, we obtain

$$n(t) \approx \frac{R}{T_G} \exp\Big((\tilde{R} - 1)\frac{t}{T_G}\Big), \quad (10)$$

becoming an exponential growth for $\tilde{R} > 1$ [1,15]. In contrast, when $t \gg \tau_0$ we obtain a polynomial growth followed by an exponential decay:

$$n(t) \approx \frac{R\tilde{R}^{D-1}}{T_G(D-1)!} \Big(\frac{t}{T_G}\Big)^{D-1} \exp\Big(-\frac{t}{T_G}\Big). \quad (11)$$

In general, the time scale τ_0 depends on the graph size N_0. For random graphs with an arbitrary degree distribution, $q_k^{(0)} = p_k$ and $q_k^{(d)} = (k-1)p_{k-1}/\langle k \rangle$ for $d > 0$ [7], resulting in $R \sim \langle k \rangle$ and $\tilde{R} \sim \langle k^2 \rangle$, where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments of the degree distribution, respectively. In this case, we obtain the following scenarios.

(i) When the tail of the degree distribution decays faster than $p_k \sim k^{-3}$, the diameter scales as $D \sim \log N_0$ [7], while $\tilde{R}$ is constant or approaches a constant in the large graph size limit. Thus, from (9) it follows that

$$\tau_0 \sim \frac{T_G}{\tilde{R}} \log N_0. \quad (12)$$

In this case, the exponential growth lasts until $t \sim \tau_0$, where $\tau_0 \to \infty$ when $N_0 \to \infty$.

(ii) When the degree distribution has the power law tail $p_k \sim k^{-\gamma}$, with $2 < \gamma < 3$, the diameter D increases at most as $\log N_0$ [17–19], while $\tilde{R} \sim N_0^{(3-\gamma)/(\gamma-1)}$. Thus, from (9) it follows that

$$\tau_0 \sim T_G \frac{\log N_0}{N_0^{(3-\gamma)/(\gamma-1)}}. \quad (13)$$

The initial exponential growth is, thus, a finite size effect restricted to $t \ll \tau_0$, where $\tau_0 \to 0$ when $N_0 \to \infty$. Following this vanishing time window, the number of infected vertices is already of the order of the graph size N_0 ($R\tilde{R} \sim N_0$), and its temporal evolution is polynomial (11), with a degree determined by the characteristic distance between vertices in the underlying graph.

To check the validity of our calculations, we perform numerical simulations of the susceptible infected (SI) model on random graphs with a power law degree distribution $p_k = Ak^{-\gamma}$. Within this model, vertices can be in two states, susceptible or infected, and infected vertices transmit the infection to each of their neighbors at a constant rate $1/T_G$ [20]. We generate random graphs with a power law degree distribution using the algorithm proposed in Ref. [21]. Then we generated single outbreaks on these graphs, starting from one infected vertex. Finally, we take averages over 10 000 outbreaks, starting from randomly selected vertices, and over 100 graph realizations.

When $\gamma > 3$, the spreading dynamics is better described by an initial exponential growth [Fig. 1(a)], in agreement with (10) and previous mathematical approaches [15,20,22]. In contrast, when $2 < \gamma < 3$, the spreading dynamics is better described by (11) [Fig. 1(b)], and the exponent D resulting from the fit to the numerical data scales linearly with the average distance between nodes [see inset in Fig. 1(b)]. In a more realistic scenario, we use the SI model to simulate the spreading of a routing table error on the autonomous system (AS) network representation of the Internet [23]. This network is characterized by a power law degree distribution with $\gamma \approx 2.1$ [11], but it also exhibits degree-degree correlations [11] and a large degree dependent clustering coefficient [24]. Yet the average number of new infections is well fitted by (11), indicating that our predictions are also valid for graphs that are not random as well (see Fig. 2).

With relevance to the spreading of computer virus and worms among email users, there is empirical evidence indicating that email networks are characterized by a power law degree distribution with $2 < \gamma < 3$ [13,14]. The transmission rates of computer viruses are, however, of the order of their typical detection times, making difficult the empirical observation of the initial epidemic growth. With relevance to sexually transmitted diseases, there are several reports indicating that the network of sexual contacts is characterized by a power law degree distribution [12,25,26], with an exponent $\gamma > 3$ for some communities and $2 < \gamma < 3$ for others. This fact together with the results obtained in this work represent a possible explanation for the observation of both exponential and polynomial HIV epidemic growth in different populations [22,27,28]. The available data is, however, not sufficient to make a definitive conclusion.

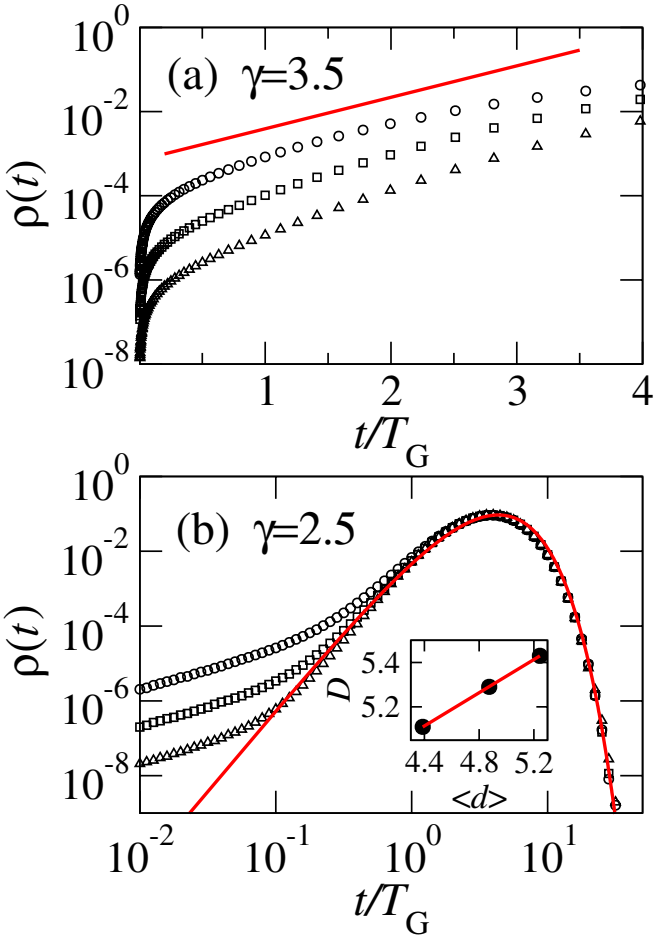

FIG. 1 (color online). Fraction of infected nodes $\rho(t) = n(t)/N_0$ as a function of time resulting from SI model simulations on random graphs with a power law degree distribution $p_k = Ak^{-\gamma}$, with (a) $\gamma = 3.5$ and (b) $\gamma = 2.5$. Different symbols correspond to different graph sizes: $N_0 = 1000$ (circles), 10 000 (squares), and 100 000 (triangles). (a) For $\gamma = 3.5$, the spreading dynamics is characterized by an exponential growth (line), as predicted by (10). (b) For $\gamma = 2.5$, the number of new infections is better described by (11) (line). There are some deviations at short times, but they get reduced with increasing the graph size. The inset shows the exponent D resulting from the fit of (11) as a function of the average distance $\langle d \rangle$ between two nodes in the graph. The increase in $\langle d \rangle$ is obtained by increasing the network size from $N_0 = 1000$ to 10 000 and 100 000. The line emphasizes the linear scaling between D and $\langle d \rangle$.

In a more general perspective, our results indicate that the degree statistics is not sufficient to characterize the spreading dynamics and probably other dynamical processes, taking place on graphs with a power law degree distribution with exponent $2 < \gamma < 3$. To determine the

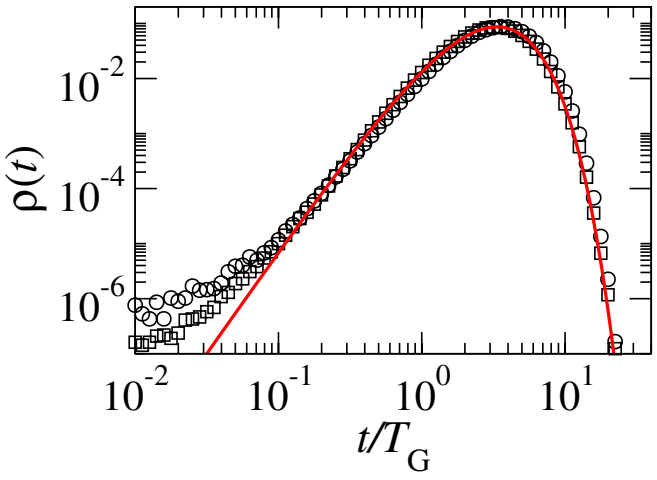

FIG. 2 (color online). Fraction of infected nodes $\rho(t) = n(t)/N_0$ as a function of time resulting from SI model simulations on AS networks, of September 1997 with $N_0 = 3015$ (circles) and of October 2001 with $N_0 = 10\,515$ (squares). The line is a fit to (11) resulting in $D = 4.7 \pm 0.1$.

characteristic time τ_0 and the polynomial degree, we need the characteristic distance between vertices in the graph as well. The amount of information needed to determine the distance between vertices is, however, more difficult to collect, in principle requiring the complete mapping of the graph. In this respect, the development of realistic graph models that can accurately represent the real graphs will be extremely valuable, allowing us to characterize the distance statistics from the degree statistics.

I thank A.-L. Barabási, A. Vespignani, and E. Almaas for helpful comments and suggestions. This work was supported by NSF Grants No. ITR 0426737 and No. ACT/SGER 0441089.

[1] T. E. Harris, *The Theory of Branching Processes* (Springer-Verlag, Berlin, 2002).
[2] M. Kimmel and D. E. Axelrod, *Branching Processes in Biology* (Springer-Verlag, New York, 2002).
[3] A. T. Bharucha-Reid, Phys. Rev. **96**, 751 (1954).
[4] A. Z. Mekjian, Phys. Rev. Lett. **86**, 220 (2001).
[5] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[6] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).
[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[8] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[9] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[10] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[11] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[12] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Berg, Nature (London) **411**, 907 (2001).
[13] J.-P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. U.S.A. **101**, 14 333 (2004).
[14] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103 (2002).
[15] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004).
[16] R. Cohen, D. Dolev, S. Havlin, T. Kalisky, O. Mokryn, and Y. Shavitt, cond-mat/0305582.
[17] F. Chung and L. Lu, Proc. Natl. Acad. Sci. U.S.A. **99**, 15 879 (2002).
[18] B. Bollobás and O. M. Riordan, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Weinheim, 2003), pp. 1–34.
[19] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
[20] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, New York, 1991).
[21] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[22] T. M. May and R. M. Anderson, Phil. Trans. R. Soc. B **321**, 565 (1988).
[23] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[24] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[25] J. H. Jones and M. S. Handcock, Nature (London) **423**, 605 (2003).
[26] A. Schneeberger *et al.*, Sex. Transm. Dis. **31**, 380 (2004).
[27] B. Szendriöi and G. Csányi, Proc. R. Soc. B **271**, S364 (2004).
[28] R. Brookmeyer and M. H. Gail, *AIDS Epidemiology: A Quantitative Approach* (Oxford University Press, New York, 1994).