Selectivity-based spreading dynamics on complex networks

Rui Yang,[1] Liang Huang,[1] and Ying-Cheng Lai[1,2]
[1]*Department of Electrical Engineering, Arizona State University, Tempe, Arizona 85287, USA*
[2]*Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287, USA*
(Received 17 December 2007; revised manuscript received 9 July 2008; published 19 August 2008)

Most previous studies on spreading dynamics on complex networks are based on the assumption that a node can transmit infection to any of its neighbors with equal probability. In realistic situations, an infected node can preferentially select a targeted node and vice versa. We develop a first-order correction to the standard mean-field theory to address this type of more realistic spreading dynamics on complex networks. Our analysis reveals that, when small-degree nodes are selected more frequently as targets, infection can spread to a larger part of the network. However, when a small set of hub nodes dominates the dynamics, spreading can be severely suppressed. Our analysis yields more accurate predictions for the spreading dynamics than those from the standard mean-field approach.

DOI: 10.1103/PhysRevE.78.026111 PACS number(s): 89.75.Hc, 05.70.Ln

I. INTRODUCTION

Spreading dynamics on complex networks are fundamental to many branches of science and engineering. In computer science, the propagation and spreading of a virus over the internet have always been of great concern. In biomedical science and engineering, the transmission of electrical signals over a neuronal network is critical to its function. In epidemiology, to understand the spreading dynamics of infections on networks is a basic task. Propagation of information over a friendship network is even relevant to political science. Because of the importance of spreading dynamics, it has been under extensive investigation since the beginning of modern network science as marked by the discoveries of the small-world [1] and the scale-free [2] topologies.

In the literature, special attention has been paid to scale-free networks [3–9] whose degree distributions follow a power law: $P(k) \sim k^{-\gamma}$, where γ is the degree exponent. A basic issue concerns the existence of a critical threshold in the transmission probability, above which spreading can occur on a global scale in the sense that a significant portion of the network can be infected. For a typical scale-free network with degree exponent between 2 and 3 [10], due to the divergence of the second moment of the degree distribution, the threshold transmission probability tends to zero for a standard two-state epidemic model [3], indicating that any virus can spread over the entire network, regardless of its infectiousness. Subsequent work reveals that certain clustering structure embedded in the network can lead to a finite threshold [4]. In addition to the threshold issue, detailed dynamics of virus or information spreading have also been investigated [11–13].

In most existing studies of spreading dynamics on complex networks, the underlying contact process is assumed to be completely random, or nonselective. That is, when a node becomes infected, it selects randomly one of its neighbors and infects it with certain probability. (Here, all nodes in the network that are connected to the original node are referred to as neighboring nodes.) There are, however, many realistic situations where the selection of a target node by an already infected node from its neighbors is not completely random but highly preferential. For instance, in a communication network with a hierarchical structure, once a node in a certain level acquires a piece of information, it is more likely for the information to be sent to some nodes in a higher level. This preferential selection of target node is only one aspect of spreading dynamics. Another equally important ingredient is the selection of an infected node by a susceptible node that has yet to be infected. For example, in a scientific citation network, the better known a paper, the higher the probability that this work will be cited. This is basically the preferential recognition (spreading) mechanism to be treated in this paper. Another example occurs in friendship networks where an individual is more likely to seek and accept his or her best friends' opinions. Such a phenomenon indeed appears to be common in human-relation networks [14]. To better understand spreading dynamics in real-world networks, the preferential selections of a target node by an infected node to pass on the infection and of an infected node by a susceptible node to receive infection must be taken into account.

In this paper, we investigate spreading dynamics with preferential selections. We first introduce suitable parameters to characterize the probabilities of the selections. We then consider a generic contact-process (CP) model and obtain analytic results for a fundamental quantity in any spreading dynamics: the fraction of nodes in the entire network that can be infected. Our theory predicts a surprising phenomenon: preferential selections in fact tend to hinder effective spreading. That is, in order to achieve efficient spreading so as to make the fraction of infected nodes as large as possible, both processes of selection should be made as uniform as possible, regardless of the degrees of the nodes. This is somewhat counterintuitive as many previous works have emphasized the role of hub nodes, nodes of unusually high degrees, in shaping the spreading dynamics [3–9,11–13]. Heuristically, this contradiction can be understood as follows. When the overall fraction of infected nodes is small, preference to select hub nodes helps the infection to survive and to spread by maintaining the hub nodes in a state with a high infection probability, leading to an increased density of infected nodes. However, when the generating rate is high and the fraction of infected nodes is large, the hub nodes are almost always

infected. Thus new attempts for the infection to be sent to the hubs are in fact wasted. In this case, infecting small-degree nodes that have lower probabilities of being infected can be more effective for increasing the overall fraction of infected nodes. From a different perspective, preferential selections tend to suppress the spreading dynamics if the infection or the virus is undesirable, and our theory provides specific scenarios for how selections should be done to achieve this goal. Our theoretical predictions are verified by extensive numerical simulations on scale-free networks.

A technical contribution of our work is the development of a set of new rate equations for spreading dynamics on networks of arbitrary topology. We shall present evidence that our approach yields results that agree with those from numerical experiments more accurately than the predictions from the standard mean-field approach. Our approach is thus appealing to the study of network spreading dynamics, considering that the applicability of the mean-field theory to highly heterogeneous networks has been an issue of recent debate [15,16].

In Sec. II, we introduce our spreading model based on preferential selections. In Sec. III, we present a detailed analysis leading to a set of rate equations from which the steady-state value of the fraction of infected nodes can be calculated. In Sec. IV, we provide numerical verifications. Concluding remarks are presented in Sec. V.

II. MODEL

We consider the general CP model, first proposed by Harris [17] and recently adapted to complex networks by Castellano and Pastor-Satorras [15] in the rate-equation framework. In the model, a CP process starts from a fraction ρ_0 of initially infected nodes. For convenience, we say that, when a node is infected, it carries a "particle" that can survive for a finite amount of time. During its lifetime, a particle can generate an "offspring" that can leave the "parent" node to infect one of its neighbors. At each time step t, each particle dies off with probability p, but with probability $1-p$, it generates an offspring. Assume that at time $t-1$ node i carries a particle. Let V_i be the set of neighbors of i. Then, at time t, one of the nodes in V_i, say node j, receives an offspring of the particle at node i and becomes infected. To incorporate selectivity in the CP model, we assume that the probability that node j is selected as the "target" node depends on its degree, as follows:

$$\Phi_{\text{out}} = k_j^{\beta} \Big/ \sum_{l \in V_i} k_l^{\beta}, \tag{1}$$

where β is an adjustable parameter. If $\beta > 0$ ($\beta < 0$), a node with large (small) degree is more likely to be selected as the target for the particle at an infected node to pass on its offspring to. Say node j is selected. If it is already occupied by a particle, the new particle from node i dies. If node j is not occupied, it accepts the new particle with the following receiving probability:

$$\Phi_{\text{in}} = k_i^{\alpha} / \max(k_m^{\alpha} | m \in V_j), \tag{2}$$

where V_j denotes the set of neighboring nodes of node j. For $\alpha > 0$, node j is more likely to accept a particle from a node with larger degree, i.e., a better-connected node (e.g., a better-known paper in a scientific citation network) has a higher probability to infect other nodes. The process repeats insofar as there are still particles in the network. The spreading dynamics terminate when there is no particle left.

III. THEORETICAL ANALYSIS

Our goal is to derive a set of rate equations governing the spreading dynamics on a complex network, which can be regarded as a first-order correction to the standard mean-field approach. We consider the quantity $\rho_k(t)$, the density of the infected nodes (nodes that host particles) of degree k at time t. The average density at time t is $\rho(t) = \Sigma_k \rho_k(t) P(k)$, where $P(k)$ is the degree distribution of the network. The steady-state solution ρ thus quantifies the spreading dynamics, and its value can be used to compare the extent of the dynamics under different parameters. In general, we find that the value of ρ can be significantly greater than zero, indicating that, at each time step, a susceptible node can be selected as a target by many infected nodes. This feature is usually unaccounted for in the standard mean-field treatment, where a susceptible node is assumed to be selected only once at each time step.

Our rate equation can be written as

$$\partial_t \rho_k(t) = -p\rho_k(t) + [1 - \rho_k(t)]\Theta_k, \tag{3}$$

where Θ_k is the probability that a node of degree k is selected and accepts a new particle. Note that the first term is the fraction of infected nodes that become susceptible at time t, and the second term is the fraction of susceptible nodes that become infected at time t. The steady-state solution is

$$\rho_k = \Theta_k / (\Theta_k + p). \tag{4}$$

The standard mean-field treatment neglects the density correlation among different nodes [3]. Generally, a susceptible node capable of accepting a new particle can be selected many times by infected nodes that are connected to it, but only one selection counts. Thus, to take into account the density correlation, we identify the set of nodes with degree k and consider the number of different nodes that have been selected, or equivalently, consider the probability that a node of degree k is selected at least once, which is Θ_k. The number of nodes of degree k is $NP(k)$. For each event of selection, the probability that one such node is not selected is $1 - [NP(k)]^{-1}$. Let N_k be the total number of the selection events. The probability that a node of degree k is not selected for all the N_k events is $\{1 - [NP(k)]^{-1}\}^{N_k}$. We thus have

$$\Theta_k = 1 - \{1 - [NP(k)]^{-1}\}^{N_k}. \tag{5}$$

To determine N_k, we note that Φ_{out} is the probability that, associated with each link, a node is selected as a target node. The probability that a node of degree k is selected as a target is thus proportional to $kk^{\beta} = k^{\beta+1}$. The total number of infected nodes of degree k' that can pass the infection onto other nodes is

$$NP(k')\rho_{k'}(t)(1-p).$$

Let $g(k,k')$ be the probability that a node of degree k accepts a particle from a node of degree k'. Taking into account these three factors and summing over k', we have

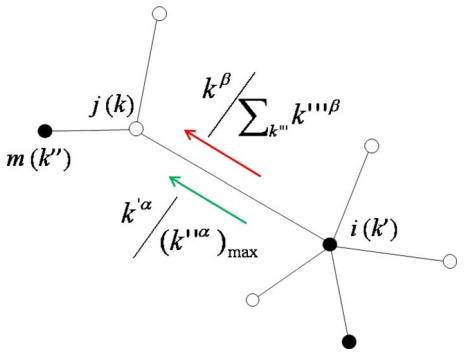

FIG. 1. (Color online) Schematic illustration of preferential selections of a target susceptible node (open circle) by an infected node (filled circle) and of an infected node by a node susceptible to receiving the infection.

$$N_k = \frac{P(k)k^{\beta+1}}{\langle k^{\beta+1}\rangle} N(1-p) \sum_{k'} [P(k')\rho_{k'}(t)] g(k,k'). \quad (6)$$

The effect of the network topological structure on spreading dynamics has been manifested in Eq. (6), i.e., how nodes of different degrees interact with each other in the process. In order to determine $g(k,k')$, we note that, for a node of degree k, the probability that it can be reached by following a random link is $Q(k)=P(k)k/\langle k\rangle$. Here we assume that the network has a constant correlation profile so that $Q(k)$ can be written in the above way. If the correlation profile is not constant, $Q(k)$ should be written explicitly in the form of $Q(k,k')=P(k|k')$ and the corresponding terms in Eqs. (7) and (8) should be changed accordingly to $Q(\cdot,k')$. Thus our results are expected to be valid for general correlation profiles.

First consider the case of $\alpha>0$. As shown in Fig. 1, for a susceptible node, say node j, we have $\Phi_{\text{in}}=(k'/k''_{\text{max}})^{\alpha}$, where k''_{max} denotes the maximum degree of all neighboring nodes of node j. For a given pair of nodes, one of degree k and another of degree k', the probability that the degree $k''_{\text{max}} \geqslant k'$ is the maximum degree among all neighboring nodes of node j is

$$C_g(k,k')Q(k'' \leqslant k''_{\text{max}})^{k-2} Q(k'' = k''_{\text{max}}),$$

where $C_g(k,k')$ is a normalization constant. Summing over all possible k''_{max} yields

$$g(k,k') = C_g(k,k') \sum_{k''_{\text{max}} \geqslant k'} \left[\left(\frac{k'}{k''_{\text{max}}} \right)^{\alpha} \times Q(k'' \leqslant k''_{\text{max}})^{k-2} Q(k'' = k''_{\text{max}}) \right]. \quad (7)$$

For $\alpha<0$, we need to consider the node of minimum degree among all neighboring nodes of a node of degree k'. Similar steps of reasoning give

$$g(k,k') = C_g(k,k') \sum_{k''_{\text{min}} \leqslant k'} \left[\left(\frac{k'}{k''_{\text{min}}} \right)^{\alpha} \times Q(k'' \geqslant k''_{\text{min}})^{k-2} Q(k'' = k''_{\text{min}}) \right]. \quad (8)$$

Equations (4)–(8) are the set of equations that can be used to calculate the steady-state solution ρ_k for any given set of parameters $[p,P(k),\beta,\alpha]$, from which $\rho=\Sigma_k \rho_k P(k)$ can be determined.

Our analysis assumes a network of finite size N. However, the final result ρ should be N independent if it is large enough. This is so because for large N we have

$$\Theta_k = 1 - \{1 - 1/[NP(k)]\}^{N_k} \approx 1 - \exp\{-N_k/[NP(k)]\},$$

where the ratio N_k/N is independent of N [Eq. (6)]. An interesting observation is that, if $N_k/[NP(k)] \ll 1$, we have $\Theta_k \approx N_k/[NP(k)]$, which reduces to the typical situation treatable by the standard mean-field theory [15,16].

IV. NUMERICAL VERIFICATION

A. Case I: $\alpha=0$

To test the applicability of Eqs. (4)–(8), we now consider two exemplary cases. For each case, we shall calculate ρ from Eqs. (4)–(8) as a function of some control parameter and compare the values of ρ with those from direct numerical simulations. The first case is $\alpha=0$ so that $g(k,k')=1$ and the summation in Eq. (6) is simply ρ. We obtain

$$N_k = \rho N(1-p)P(k)k^{\beta+1}/\langle k^{\beta+1}\rangle.$$

Substituting this relation in Eqs. (4) and (5) yields a nonlinear self-consistent equation in ρ, which can then be solved numerically, say, by an iterative procedure. To be specific, we use a scale-free network generated by the preferential attachment rule [2], which has an algebraic degree distribution $P(k)\sim k^{-3}$. The size of the network is $N=10^4$ and average degree $\langle k\rangle=6$. A typical example of the evolution of $\rho(t)$ and its approach to a steady-state value is shown in Fig. 2. Figure 3 shows, for $p=0.2$, values of ρ versus the parameter β from the theoretical self-consistent equation (solid curve) and from direct numerical simulations (open squares), which agree with each other very well. We observe that, for $\beta<0$, ρ has a weak dependence on β in the sense that, as β is decreased from zero, ρ varies only slightly. However, ρ decreases rapidly as β is increased from zero. Physically, this means that, for $\beta>0$ so that nodes with relatively large degrees (e.g., hubs) are more likely to receive new particles, the spreading dynamics is in fact suppressed, which is somewhat counterintuitive as the hub nodes might be regarded as highly effective for spreading the infection. This behavior can nonetheless be understood by noting that there is typically a high probability for a hub node to be infected. New particles, when being sent to a hub node, will then have a high probability to die off, effectively reducing the probability of infection generation. Nodes with smaller degrees are generally less likely to be infected. Thus, new particles sent to these nodes will be more likely to be accepted. In Fig. 3,

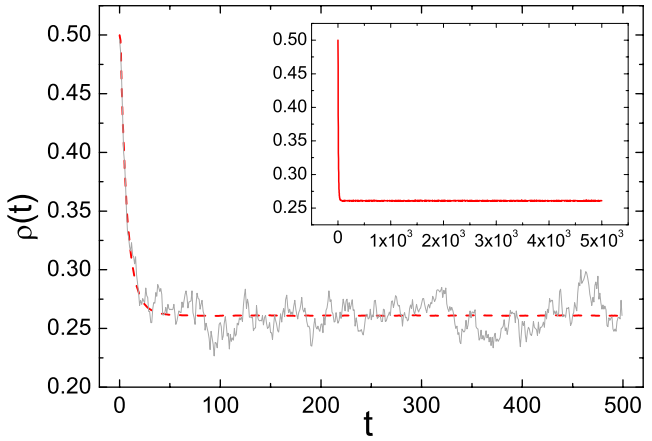

FIG. 2. (Color online) A typical example of the evolution of $\rho(t)$ toward steady state with $\alpha=0$ and $\beta=1$. The solid line indicates one network realization and the dashed one indicates the average $\langle\rho(t)\rangle$ over 100 network realizations. The inset shows the long-time behavior of averaged density. In the following simulation, for each case, the spreading process is evolved for sufficiently long time where $\rho(t)$ becomes nearly a constant (with small and time-independent fluctuations), indicating a steady state. Then ρ is taken as the average of $\rho(t)$ from the last 200 time steps. An additional average over 100 network realizations is taken in the calculation of ρ.

we have also shown the predicted values of ρ from the standard mean-field theory (dashed line). We observe that, although the standard mean-field result appears to be reasonable as compared with numerical results, our approach yields results that better agree with the numerics.

B. Case II: $\beta=0$

For $\beta=0$, the quantity $g(k,k')$, as given by Eqs. (7) and (8), depends on the degree distribution only. An example of the theoretically predicted relation between ρ and the control parameter α is shown in Fig. 4 (solid curve), together with results from direct numerical simulations. Also shown in Fig.

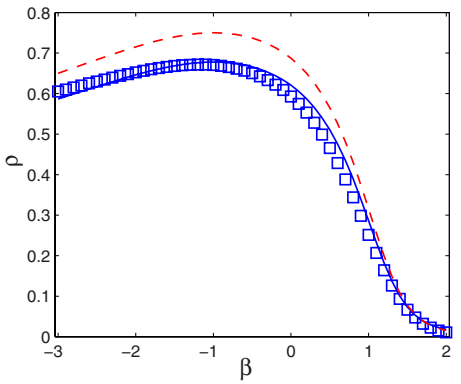

FIG. 3. (Color online) For a representative scale-free network, for $\alpha=0$, steady-state value ρ of the fraction of infected nodes versus parameter β (for $p=0.2$). The solid curve is from our theory. Dashed curve represents the standard mean-field prediction. Open squares are data points from direct numerical simulations.

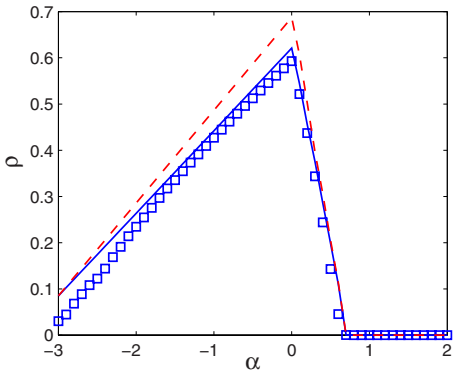

FIG. 4. (Color online) For the same scale-free network in Fig. 3 and for $\beta=0$, ρ versus α. The key is the same as in Fig. 3.

4 are predictions from the standard mean-field approach. Again we observe a better agreement between our theory and numerics. The nonsmooth behavior between ρ and α at $\alpha=0$ can be understood, as follows. For $\alpha>0$, a node is more likely to receive and accept offspring particles from its neighbors with relatively large degrees. The opposite occurs for $\alpha<0$. There is thus a change of the set of neighboring nodes from which a node can receive a new particle. We also observe that, for α not too close to zero, the probability $g(k,k')$ is generally much smaller than unity, resulting in effectively a smaller probability of generating offsprings and, consequently, a rapid decrease of ρ from its value for $\alpha=0$ as α either is decreased or is increased from $\alpha=0$.

C. General case: $\alpha\neq 0$ and $\beta\neq 0$

What happens when both α and β are not zero? Figure 5 shows, in the (α,β) plane, various contours of ρ. When both α and β are negative so that nodes with small degrees are preferably chosen as targets and senders, the value of ρ is high and it changes little. This is due to the particular topology of scale-free networks: there are substantially more nodes with small degrees in the network. For $\alpha<0$ but $\beta>0$, new particles are preferentially sent to hubs that represent, however, a small set of nodes in the network. For larger positive values of β, new particles are sent to nodes of larger degrees, the number of which is, however, even smaller. In

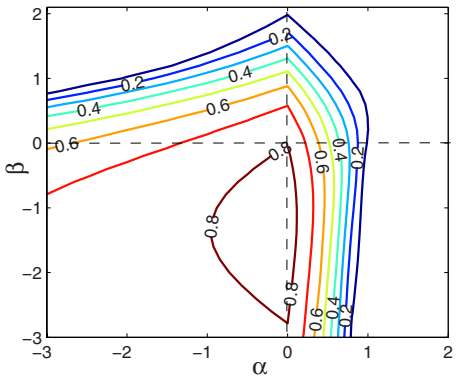

FIG. 5. (Color online) Numerically obtained contour plots of ρ in the (α,β) parameter plane for $p=0.1$.

this case, to maintain the same level of infection, α needs to be larger so as to increase the probability to receive new particles from nodes of larger degrees. A similar behavior occurs for the $\alpha>0$ and $\beta<0$ region. For $\alpha>0$ and $\beta>0$, new particles are more likely to be both sent to and received from hub nodes, effectively restricting the dominant spreading dynamics to a small subset of nodes in the network. In this case, the value of ρ is much smaller as compared with its values in the regions where both α and β are negative. Say there is an increase in β in the $\alpha>0$ and $\beta>0$ region. In order to achieve the same value of ρ, α needs to be reduced so that more nodes of smaller degrees can contribute to the spreading process. The general observation is that, to maximize the efficiency of spreading, the network should be designed so as to allow for more small-degree nodes to play an active role in the dynamical process.

V. CONCLUDING REMARKS

In summary, we have investigated spreading dynamics with preferential selection on complex networks with heterogeneous degree distributions. Our theoretical approach can be considered as a first-order correction of the standard mean-field theory in the field of spreading dynamics, i.e., we take into account the density correlations among different nodes. We have presented evidence that our theory yields more accurate predictions than the standard one. The key finding is that spreading can be severely suppressed when a small set of hub nodes dominates the dynamics. For efficient spreading, more nodes with relatively small degrees should be active in serving as "exchanging stations" for transferring the infection.

Spreading dynamics is fundamental to various networks arising in natural and social sciences. Efficient spreading, depending on applications, may be either desirable or undesirable. Our work yields insights into the dynamics of spreading that can be useful for designing networks and protocols to achieve either goal.

ACKNOWLEDGMENTS

This work was supported by an ASU-UA Collaborative Program on Biomedical Research, by AFOSR under Grant No. FA9550-07-1-0045, and by ONR through WVHTC (West Virginia High Technology Consortium Foundation).

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[4] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).
[5] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).
[6] Y. Moreno, M. Nekovee, and A. F. Pacheco, Phys. Rev. E **69**, 066130 (2004).
[7] V. Colizza and A. Vespignani, Phys. Rev. Lett. **99**, 148701 (2007).
[8] M. Small, D. M. Walker, and C. K. Tse, Phys. Rev. Lett. **99**, 188702 (2007).
[9] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási, Phys. Rev. Lett. **98**, 158702 (2007).
[10] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[11] Y.-C. Lai, Z. Liu, and N. Ye, Int. J. Mod. Phys. B **17**, 4045 (2003).
[12] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004).
[13] L. Huang, K. Park, and Y.-C. Lai, Phys. Rev. E **73**, 035103(R) (2006).
[14] G. Palla, A.-L. Barabási, and T. Vicsek, Nature (London) **446**, 664 (2007).
[15] C. Castellano and R. Pastor-Satorras, Phys. Rev. Lett. **96**, 038701 (2006); **98**, 029802 (2007); **100**, 148701 (2008).
[16] M. Ha, H. Hong, and H. Park, Phys. Rev. Lett. **98**, 029801 (2007).
[17] T. E. Harris, Ann. Probab. **2**, 969 (1974).