

Cascading dynamics in modular networks

Aram Galstyan and Paul Cohen

Information Sciences Institute, University of Southern California, 4676 Admiralty Way, Marina del Rey, California 90292-6695, USA

(Received 5 August 2006; published 23 March 2007)

In this paper we study a simple cascading process in a structured heterogeneous population, namely, a network composed of two loosely coupled communities. We demonstrate that under certain conditions the cascading dynamics in such a network has a two-tiered structure that characterizes activity spreading at different rates in the communities. We study the dynamics of the model using both simulations and an analytical approach based on annealed approximation and obtain good agreement between the two. Our results suggest that network modularity might have implications in various applications, such as epidemiology and viral marketing.

DOI: [10.1103/PhysRevE.75.036109](https://doi.org/10.1103/PhysRevE.75.036109)

PACS number(s): 89.75.Hc

I. INTRODUCTION

Networks are a useful paradigm for studying complex systems composed of large numbers of interconnected components [1–3]. There has been a growing interest in applying network analysis to examine various biological [4,5], ecological [6–8], technological [9–11], and even political [12] systems. Research on various statistical properties of such networks has revealed many interesting phenomena. For instance, the scale-free degree distributions observed in many real-world networks have significant implications for various dynamical processes on such networks. In particular, the dynamics of SIR (susceptible-infected-removed) epidemic processes in certain scale-free networks are characterized by a vanishing threshold for epidemics [13], in sharp contrast with results for the random Erdos-Renyi networks.

Another interesting property of networks is modularity, the tendency of nodes to partition themselves into *communities* [14,15]. Loosely speaking, a community is a group of nodes for which the density of links within a group is higher than across the groups. Much recent research has focused on methods for detecting and analyzing community structure in networks (for a recent review of existing approaches see [16] and references therein). However, the dynamical properties of modular networks have received relatively little attention despite the potential importance of the subject to problems such as epidemiology, viral marketing, and so on. Consider, for instance, word-of-mouth (or viral) marketing of a new product. If different consumer groups have different rating criteria for the product, or different reaction to marketing strategies, then one needs to model how influence propagates within and across communities to predict whether the product will be a hit, or confined to a small subset of consumers. A similar argument holds for epidemic models where nodes are heterogeneous with respect to their susceptibilities and/or interactions patterns [17–19]. For instance, Gupta *et al.* showed that the transient dynamics of sexually transmitted infection can be very different, depending on whether the network of sexual contacts is assortative or disassortative [17]. More recent work has addressed the role of the modularity on dynamics of cascading failures in scale-free networks [20], and synchronization patterns of networked oscillators [21]. In particular, Ref. [21] demonstrated that the

modularity of networks is strongly reflected in their synchronization dynamics, so that communities emerge as connected groups of synchronized oscillators.

The goal of this paper is to further improve our understanding of the connection between network modularity and its dynamics. Specifically, we examine the effects of modularity on a simple, threshold-based activation process on networks. Starting with a modified version of Watts' cascading model [22], we study its dynamical properties for networks composed of two loosely coupled communities. Our main observation is that if the initially active nodes (*seeds*) are contained in one of the communities, then under certain conditions the cascading process has a two-tiered structure, that is, the peaks of the activation dynamics in each community are well-separated in time. We present results of simulations as well as analytical results based on annealed approximation, and observe a good agreement between the two.

II. MODEL

Let us consider a network where each node is in one of two states: passive and active. Initially, all but a small fraction of *seed* nodes are passive. During the cascade process, a passive node will be activated with probability that depends on the state of its neighbors. In Watt's original model [22] this probability is $p = \Theta(h_i/k_i - \phi)$, where Θ is the step function, h_i and k_i are the number of active neighbors and the total number of the neighboring nodes, respectively, and ϕ_i is the activation threshold for the i th node. Here we consider a slight modification of the original model by using a threshold condition on the *number* of active neighbors rather than their fraction: $p = \tau^{-1} \Theta(h_i - H_i)$, where τ determines the time-scale of the activation process. For the sake of simplicity, we assume that all nodes have the same activation threshold, $H_i = H$ for all i .

Clearly, the dynamics of the cascade process will depend on both network structure and the threshold parameter H . Here we are interested in the case when the network is composed of two loosely coupled communities. Namely, we consider a random graph consisting of $N = N_a + N_b$ nodes of two different types, a and b . The probabilities of edges between nodes of different types are γ_{aa} , γ_{bb} , and $\gamma_{ab} = \gamma_{ba}$, and the average connectivity between nodes of the respective types

are then $z_{aa} = \gamma_{aa} N_a$, $z_{bb} = \gamma_{bb} N_b$, $z_{ab} = \gamma_{ab} N_b$, and $z_{ba} = \gamma_{ab} N_a$. We want to find out how the modularity of the network, as described by the coupling between the groups, affects the cascading process.

Let ρ_a^0 and ρ_b^0 be the fraction of seed nodes in each population. Further, let $P_a(h;t)$ and $P_b(h;t)$ be the probability distribution that a randomly chosen node of corresponding type is connected with exactly h active nodes at time t . It is easy to see that $P_a(h;t=0)$ and $P_b(h;t=0)$ are Poisson distributions with means $z_{aa}\rho_a^0 + z_{ab}\rho_b^0$ and $z_{bb}\rho_b^0 + z_{ba}\rho_a^0$, respectively. To study the dynamics of the process, we need to estimate these distributions for later times. This is particularly straightforward to do within the *annealed approximation*, e.g., by “rewiring” the network after each iteration [23]. Indeed, since all edges of corresponding type are equally likely, it is easy to see that $P_a(h;t)$ and $P_b(h;t)$ are still given by Poisson distribution, with the means that now depend on the fraction of active nodes $\rho_a(t)$ and $\rho_b(t)$: $P_{a,b}(h;t) = \mathcal{P}(\lambda_{a,b}(t))$, where $\lambda_a = z_{aa}\rho_a(t) + z_{ab}\rho_b(t)$ and $\lambda_b = z_{bb}\rho_b(t) + z_{ba}\rho_a(t)$.

On the first step of the cascading process, the fraction of activated nodes of each type is given by $\tau^{-1} \sum_{h \geq H} P_{a,b}(h;t=0)$. In later iterations, we can calculate the fraction of active nodes as follows. Let us consider, for instance, a nodes. There are $N_a[1 - \rho_a(t)]$ passive nodes at time t , and each one of these nodes will be activated with the rate $\tau^{-1} \sum_{h \geq H} P_a(h;t)$. Also, due to the rewiring, some of the $N_a[\rho_a(t) - \rho_a^0]$ active nodes will switch to a passive state with the rate $\tau^{-1} \sum_{h < H} P_a(h;t)$. We note that *the seed nodes never deactivate*. Combining these together, and using the normalization condition $\sum_{h=0}^{\infty} P_{a,b}(h;t) = 1$, we obtain in the continuous time limit

$$\tau \frac{d\rho_{a,b}}{dt} = 1 - \rho_{a,b} - (1 - \rho_{a,b}^0) Q(H; \lambda_{a,b}), \quad (1)$$

where $Q(n, x) = \sum_{k < n} e^{-x} x^k / k!$ is the regularized gamma function.

Equation (1) determines the time evolution of the cascading process in each group. Let $\rho(t) = \alpha \rho_a(t) + (1 - \alpha) \rho_b(t)$, $\alpha = N_a / (N_a + N_b)$ be the fraction of active nodes in the whole network. In Fig. 1 we compare the solutions obtained from Eq. (1) with the results of simulations on randomly generated graphs for the same network parameters but two different values of the threshold parameter. The parameters of the network are $N_a = 5000$, $N_b = 15\,000$, $z_{aa} = z_{bb} = 15$, $z_{ab} = 4$. The fraction of seed nodes is $\rho_a^0 = 0.1$ and $\tau^{-1} = 0.1$. The simulations are averaged over 100 random realizations.

The agreement between the analytical prediction and results of the simulations is quite good. The network settles to the same steady state for both values of the threshold parameter H : that is, all of the nodes are activated at the end of the cascading process. However, the transient dynamics depend on the threshold parameter H . For $H=2$, activation spreads very quickly through both communities and after a short interval all of the nodes are activate. For $H=4$, on the other hand, the fraction of active nodes seems to saturate, then, in later iterations, $\rho(t)$ increases rapidly and eventually all the nodes become active. In Fig. 1(b) we plot the rate of activa-

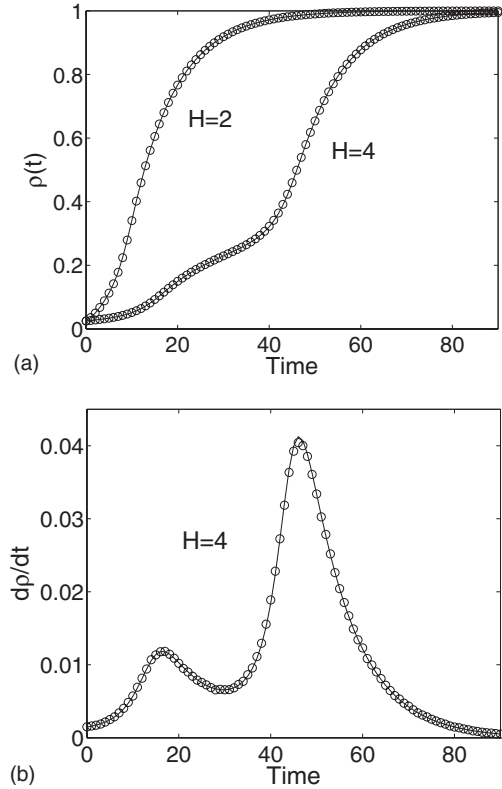


FIG. 1. Analytical (solid lines) and simulation (circles) results for the activation dynamics. The upper panel shows the fraction of active nodes vs time for threshold parameters $H=2$ and 4. The lower panel shows the activation rate dp/dt vs time for $H=4$.

tion process dp/dt vs time for $H=4$. Apparently, the peak rates of activation in the two communities are separated in time. We call this phenomenon *two-tiered dynamics*. We would like to note that previously such a multipeak structure has been observed in Ref. [17], where the authors studied the impact of different mixing patterns on the spread of sexually transmitted infection.

To better understand how two-tiered dynamics arises, we will examine a simplified scenario. Let us assume that seed nodes are chosen among a nodes only, so that $\rho_b^0 = 0$. Further, let us assume that the coupling between two populations is not very strong, so that the cascading process among a nodes is not affected by cross-group links. Hence the fraction of active a nodes evolves according to the following equation:

$$\tau \frac{d\rho_a}{dt} = -\rho_a + g_a(z_{aa}\rho_a), \quad (2)$$

where we have defined

$$g_a(x) = 1 - (1 - \rho_{a,b}^0) Q(H, x). \quad (3)$$

The fraction of the population that will be activated at the end of the cascading process is determined from the following equation:

$$\rho_a^s = g_a(z_{aa}\rho_a^s). \quad (4)$$

Note that for sufficiently dense networks (i.e., the connectivity of all nodes is greater than the threshold H) $\rho_a^s = 1$ is

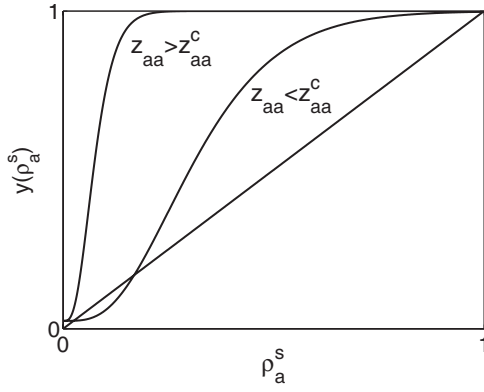


FIG. 2. Graphical representation of Eq. (4). Plotted are the straight line $y = \rho_a^s$ and the function $y = g_a(z_{aa}\rho_a^s)$ for two different values of z_{aa} .

always a solution. However, it is not always the *only* solution. This is shown graphically in Fig. 2, where we plot both sides of Eq. (4) as a function of ρ_a^s for two different connectivities. For a given fraction of seed nodes the connectivity z_{aa} . In particular, for sufficiently large values of z_{aa} , the only intersection of the curve with the line happens at $\rho \approx 1$, aside from exponentially small correction of order $\sim z^{H-1}e^{-z_{aa}}$, indicating that the activation will spread globally. If one decreases z_{aa} , however, other solutions appear as shown by the two intersections of ρ_a and $g_a(z_{aa}\rho_a)$ in Fig. 2. Specifically, there is a critical value z_{aa}^c so that for $z_{aa} < z_{aa}^c$ the cascading dynamics dies out, while for $z_{aa} > z_{aa}^c$ it spreads throughout the system. Let us define $x = z_{aa}\rho_a^s$, and rewrite Eq. (4) as $z_{aa}^{-1}x = g(x)$. At the critical point, the line $z_{aa}^{-1}x$ must be tangential to $g(x)$. It is then straightforward to demonstrate that the critical connectivity is given by

$$z_{aa}^c = [g'_a(x_0)]^{-1} \equiv \left[(1 - \rho_a^0) e^{-x_0} \frac{x_0^{H-1}}{(H-1)!} \right]^{-1}, \quad (5)$$

where x_0 satisfies the following equation:

$$x_0 g'_a(x_0) = g_a(x_0). \quad (6)$$

Equations (5) and (6) determine that critical connectivity needed to cause a global cascade among a nodes for a given fraction of seed nodes and the threshold parameter H . In Fig. 3 we compare the analytical prediction with simulation results for $H=2$. The simulations were done for a graph with 5×10^4 nodes, and for 100 random trials. Each parameter pair (ρ_a^0, z_{aa}) was considered to be above the critical line if a global cascade was observed in the majority of trials for that parameter. Again, the agreement of analytical prediction and the simulation results are excellent.

Let us examine the behavior of the critical connectivity in the limit of small ρ_a^0 . Equation (6) can be rewritten as

$$e^{-x_0} \left(\frac{x_0^H}{(H-1)!} + \sum_{k=0}^{H-1} \frac{x_0^k}{k!} \right) = \frac{\rho_a^0}{1 - \rho_a^0}. \quad (7)$$

Assuming $\rho_a^0, x_0 \ll 1$ we obtain in the leading order

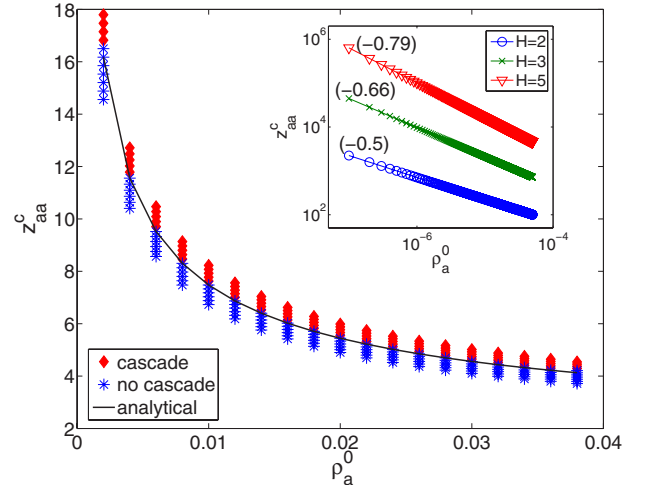


FIG. 3. (Color online) The critical connectivity vs fraction of seed nodes for threshold parameter $H=2$. The inset shows the scaling behavior of z_{aa}^c for different H (the numbers in parenthesis are the slopes of corresponding lines).

$$x_0 \approx \left[\frac{H!}{H-1} \rho_a^0 \right]^{1/H}. \quad (8)$$

Finally, using Eq. (5) we obtain the following scaling behavior:

$$z_{aa}^c \propto (\rho_a^0)^{-(H-1)/H}, \quad (9)$$

which is demonstrated in the inset of Fig. 3. We also note that at the critical point the convergence time diverges as $T_{conv} \propto (z - z_{aa}^c)^{-1/2}$.

Now consider the cascading dynamics in the second group. Initially, there are no active nodes in this group. As more and more a nodes are activated, the activation will spread to the b nodes for sufficiently large across-group connectivity z_{ba} . The activation dynamics is again governed by an equation similar to Eq. (2). In particular, the steady state fraction of active b nodes satisfies the following equation:

$$\rho_b^0 = 1 - Q(H, z_{bb}\rho_b^0 + \lambda) \equiv g_b(z_{bb}\rho_b^0 + \lambda). \quad (10)$$

where $\lambda = z_{ba}\rho_a^0$. Clearly, if λ is sufficiently large, then the cascade will propagate among b nodes independent of the within-group connectivity z_{bb} ; and vice versa, however large the connectivity z_{bb} , there is a critical value of λ_a^c so that for $\lambda < \lambda^c$ there will be no cascade among the b nodes. Let us define $x = z_{bb}\rho_b^0 + \lambda$ and rewrite the steady state equation as follows

$$\frac{x - \lambda}{z_{bb}} = g_b(x). \quad (11)$$

Using the same reasoning as for the a nodes, it is easy to show that the critical point is given by

$$\lambda^c = x_0 - z_{bb}g_b(x_0), \quad (12)$$

where x_0 is the smaller of the roots of the following equation:

$$g'_b(x_0) = \frac{1}{z_{bb}}. \quad (13)$$

Note that for $\rho_a^0=1$ λ_c is simply the critical across-group connectivity $z_{ab}^c(z_{bb})$ for which the cascade will spread to b nodes, assuming that all a nodes have already been activated. Hence Eqs. (12) and (13) implicitly define a critical line $z_{bb}^c(z_{ba})$ on the $z_{bb}-z_{ba}$ plane. Note that on this critical line the convergence time of the cascading process among the b -nodes, and consequently the separation of two activity peaks, is infinite. For a fixed within-group connectivity z_{bb} the two-tiered structure will be present provided that z_{ba} is only slightly above the critical line. To be more precise, let ρ_a^{max} be the fraction of active a nodes that corresponds to the maximum activation rate among a nodes. This can be found from Eq. (2) by differentiating the right-hand side with respect to ρ_a and setting it to zero, which yields $z_{aa}g'_a(z_{aa}\rho_a^{max})=1$. If the across-group connectivity is smaller than λ^c/ρ_a^{max} , then the cascade will not spread to b nodes until the rate of activation spreading among a nodes starts to decline from its peak. Consequently, the two-tiered pattern will be present for the range $\lambda^c < z_{ba} < \lambda^c/\rho_a^{max}$.

III. SUMMARY

To summarize, we have considered a simple cascading model on a random network consisting of two loosely

coupled communities. For a sufficiently weak coupling between two communities the dynamics of the activity spreading demonstrates a two-tiered structure, that is, the peak rates of the cascading processes in two communities are separated in time. This pattern is reminiscent of multipeak structure of sexually transmitted infection dynamics, previously reported in Ref. [17]. We studied this phenomenon both experimentally and theoretically using annealed approximation and obtained a good agreement between analytical results and simulations. Although our model is for undirected binary graphs, generalizations to directed and/or weighted graphs is straightforward. Directed models can be relevant if the interactions between two nodes are not symmetric.

The results presented here might have implications in problems such as epidemiology, viral marketing, and so on. Consider, for example, the problem of minimizing the number of seed nodes that will cause a global cascade in a given network, or more generally, the problem of maximizing a certain utility function $f(N_0, N_s)$, where N_0 is the number of seed nodes and N_s is the expected size of the cascade. Our results suggest that simple strategies that are suitable for homogenous networks (e.g., choosing nodes with high connectivity, or at random), might lead to a suboptimal solution for networks with a strongly modular structure. We note, however, that in order to assess the implication of our findings in real world problems, one needs to generalize the approach developed here to more complex networks.

[1] R. Albert and A.-L. Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).
 [2] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
 [3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006).
 [4] S. A. Kauffman, *The Origins of Order* (Oxford University Press, New York, 1993).
 [5] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, *Nature (London)* **407**, 651 (2000).
 [6] J. M. Montoya and R. V. Sole, *J. Theor. Biol.* **214**, 405 (2002).
 [7] J. Camacho, R. Guimera, and L. A. Nunes Amaral, *Phys. Rev. Lett.* **88**, 228102 (2002).
 [8] P. Jordano, J. Bascompte, and J. M. Olesen, *Ecol. Lett.* **6**, 6981 (2003).
 [9] R. Albert, H. Jeong, and A.-L. Barabasi, *Nature (London)* **401**, 130131 (1999).
 [10] A.-L. Barabasi, R. Albert, and H. Jeong, *Physica A* **281**, 6977 (2000).
 [11] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, in *Proceedings of the International Conference on Combinatorics and Computing*, Lecture Notes in Comput. Sci. Vol. 1627 (Springer-Verlag, Berlin, 1999), p. 118.
 [12] M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7057 (2005).
 [13] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001); *Phys. Rev. E* **63**, 066117 (2001).
 [14] M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
 [15] M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577 (2006).
 [16] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, *J. Stat. Mech.: Theory Exp.* (2005) P09008.
 [17] S. Gupta, R. M. Anderson, and R. M. May, *AIDS* **3**, 807 (1989).
 [18] S. O. Aral *et al.*, *Am. J. Public Health* **89**, 825 (1999).
 [19] F. Liljeros, C. R. Edling, and L. A. N. Amaral, *Microbes Infect.* **5**, 189 (2003).
 [20] Jian-jun Wu, Zi-you Gao, and Hui-jun Sun, *Phys. Rev. E* **74**, 066111 (2006).
 [21] A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente, *Phys. Rev. Lett.* **96**, 114102 (2006).
 [22] D. J. Watts, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5766 (2004).
 [23] B. Derrida and Y. Pomeau, *Europhys. Lett.* **1**, 45 (1986); B. Derrida and D. Stauffer, *ibid.* **2**, 739 (1986).