

Cascading Dynamics in Modular Networks

Aram Galstyan and Paul Cohen
Information Sciences Institute
University of Southern California
4676 Admiralty Way, Marina del Rey, CA 90292-6695

(Dated: August 5, 2006)

In this letter 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. We also suggest a simple algorithm for detecting community boundaries that exploits two-tiered dynamics.

PACS numbers:

I. INTRODUCTION

Networks are useful paradigm for studying complex systems composed of large numbers of interconnected components [1]. There has been a growing interest in applying network analysis to examine various social, biological, technological, and even political 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 vanishing threshold for epidemics [6], 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* [4, 5]. 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 [2] and references therein). However, the dynamical properties of modular networks have received little attention, despite the importance of the subject to problems such as epidemiology, ecommerce, group-detection and so on. If individuals in different communities have different susceptibilities to a disease, then understanding the effect of the community structure on disease propagation can improve vaccination policies. Another example is a word-of-mouth (or viral) marketing of new products. If different consumer groups have different rating criteria for a 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.

This letter develops a model of the effects of modularity on a simple dynamical process on networks. Starting

with a modified version of Watts' cascading model [7], we study its dynamical properties for networks composed of two loosely coupled communities. Our main observation is that if the initial seed nodes 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 suggest that the observed two-tiered structure leads to algorithms for detecting community boundaries in networks.

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 [7] 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 type, 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 be-

tween 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 $P_a(k; t)$ and $P_b(k; t)$ be the probability distribution that a randomly chosen node of corresponding type is connected with exactly k active nodes at time t . It is easy to see that $P_a(k; t = 0)$ and $P_b(k; t = 0)$ 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. Indeed, since all edges of corresponding type are equally likely, it is easy to see that $P_a(k; t)$ and $P_b(k; t)$ are still given by Poisson distribution, with the means that now depend on the fraction of active nodes $\rho_a(t)$ and $\rho_b(t)$: $P_{a,b}(k; t) = \text{Poisson}(\lambda_{a,b}(t))$, 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 $\tau^{-1} \sum_{k \geq H} P_{a,b}(k; t = 0)$. 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 probability $\tau^{-1} \sum_{k \geq H} P_a(k; t)$. Also, due to the rewiring, some of the $N_a(\rho_a(t) - \rho_a^0)$ active nodes will switch to passive state with rate $\tau^{-1} \sum_{k < H} P_a(k; t)$. We note that the seed nodes never de-activate. Combining these together, we obtain the 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.

Equations 1 determine 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 Figure 1 we compare the solutions obtained from Equations 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 $N_a = 5000$, $N_b = 15000$, $z_{aa} = z_{bb} = 15$, $z_{ab} = 4$. 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

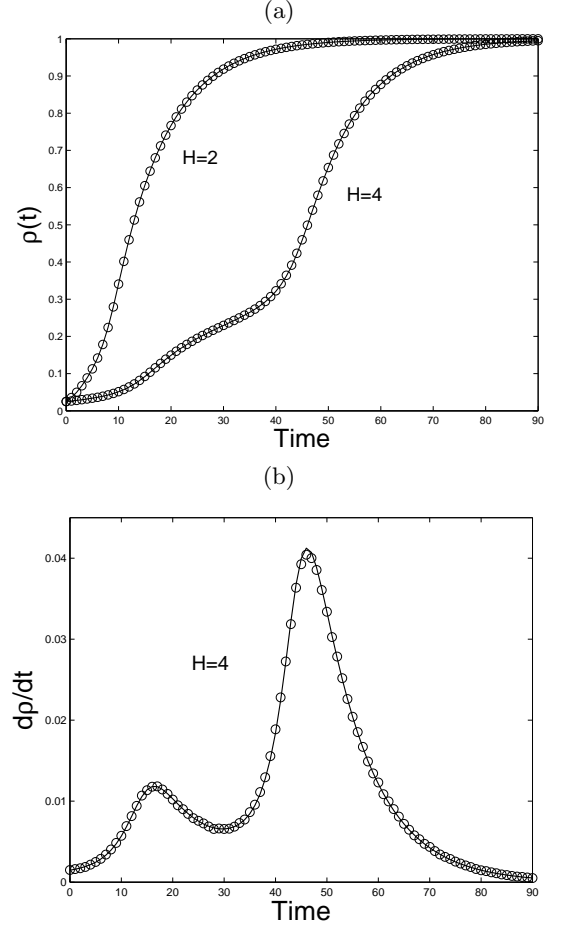


FIG. 1: Analytical (solid lines) and simulation (circles) results for the activation dynamics. The upper panel shows the fraction of active nodes vs time for threshold parameter $H = 2$ and $H = 4$. The lower panel shows the activation rate $d\rho/dt$ vs time for $H = 4$.

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 Figure 1(b) we plot the rate of activation process $d\rho/dt$ 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*.

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

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

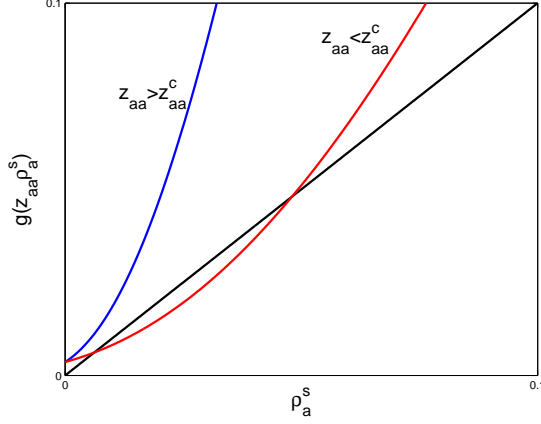


FIG. 2: Graphical representation of the Equation 4.

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 always a solution. However, it is not always the *only* solution. This is shown graphically in Figure 2, where we plot both sides of Equation 4 as a function of ρ_a^s for two different connectivities. For a given fraction of seed nodes the steady-state fraction of active nodes is determined by the connectivity z_{aa} . In particular, for sufficiently large values of z_{aa} , the only intersection of the curve with the line happens at $\rho \approx 1$ (not shown in the figure) 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 Figure 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. It is 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 Figure 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 parameters. 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 . The 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

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

Finally, using Equation 5 we obtain the following scaling behavior

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

which is demonstrated in the inset of Figure 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 the Equation. 2. In particular, the steady state fraction of active b nodes satisfies the following equation:

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

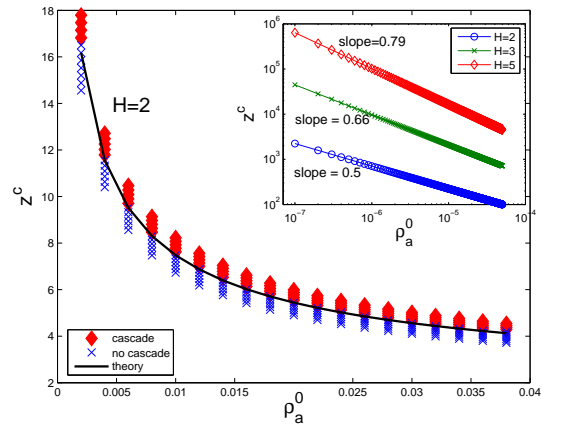


FIG. 3: The critical connectivity vs fraction of seed nodes for threshold parameter $H = 2$. The inset shows the scaling behavior of z_{aa}^c for different H .

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, equations 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 Equation 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 $\lambda^c < z_{ba} < \lambda^c/\rho_a^{max}$.

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 two-tiered structure, that is, the peak rates of the cascading processes in two communities are separated in time. 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 important implications in problems such as epidemiology, viral marketing, etc. Indeed, 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 sub-optimal solution for networks with strongly modular structure.

In conclusion, note that two-tiered dynamics can help detect the boundaries of a community given an initial set of known seed nodes from that community. The first step is to initiate a cascading process from the seed nodes. If the coupling between the communities is sufficiently weak, then the cascade will temporarily be largely confined within the community that contains the seed nodes, and only later will it “leak” into other nodes in the network. Thus, the boundary of a community can be detected by looking for two-tiered dynamics and identifying the onset of activation spreading in the second population [3].

-
- [1] R. Albert and A.-L. Barabasi. *Rev. Mod. Phys.*, 74:47, 2002.
 - [2] Leon Danon, Albert Díaz-Guilera, Jordi Duch, and Alex Arenas. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.
 - [3] A. Galstyan and P. R. Cohen. Inferring useful heuristics from the dynamics of iterative relational classifiers. In *Proceedings of IJCAI-05, 19th International Joint Conference on Artificial Intelligence*, 2005.
 - [4] M. Girvan and M. E. J. Newman. *PNAS*, 99:7821–7826, 2002.
 - [5] M. E. J. Newman. From the Cover: Modularity and community structure in networks. *PNAS*, 103(23):8577–8582, 2006.
 - [6] R. Pastor-Satorras and A. Vespignani. *Phys. Rev. Lett.*, 86:3200, 2001.
 - [7] D.J. Watts. *Proc. Natl. Acad. Sci. U.S.A.*, 99:5766, 2004.