Influence Propagation in Modular Networks

Aram Galstyan and **Paul R. Cohen**
USC Information Sciences Institute
Center for Research on Unexpected Events (CRUE)
4676 Admiralty Way, Marina del Rey, CA 90292
{galstyan,cohen}@isi.edu.

Abstract

The objective of viral marketing is to utilize existing social interactions between customers for world–of–mouth advertising of products. In order to design effective marketing strategies, one needs to understand how influence is propagated across such social networks. Here we study a simple influence propagation process in a network composed of two loosely coupled communities. We find that for a certain range of network parameters, the dynamics of the influence propagation is characterized by a doubly–critical behavior. Our results also suggest that the presence of the community structure, or network modularity, might have important implications for choosing appropriate marketing strategies.

The idea behind viral marketing is to use existing social structures for world–of–mouth advertising of products or services (Domingos & Richardson 2001; Leskovec, Adamic, & Huberman 2006). Instead of targeting customers indiscriminately, efficient marketing strategies aim at targeting certain customers that will propagate the influence among many others. An important problem is then to decide what nodes to target so that the propagation of the influence will be maximized (Kempe, Kleinberg, & Éva Tardos 2003). Thus, to understand implications of specific targeting strategies, it is imperative to understand how the influence propagates through a social network. In recent years there has been an extensive amount of work on studying various dynamical processes on complex networks. Most of the studies have focused on the effect of the scale–free degree distribution on dynamical processes. In this paper, we focus on networks that have a modular structure, i.e., they are composed of clusters, or communities, that are loosely coupled with each other.

Here we focus on a network composed of only two communities. Specifically, 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 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.

Each node is in one of two states: passive and active. Initially, all but a small fraction of *seed* nodes are passive. During the activation process, a passive node will be activated with probability that depends on the state of its neighborhs. In Watt's original model (Watts 2004) 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 use 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. We will assume that all nodes have the same activation threshold, $H_i = H$ for all i.

Below we examine a simplified scenario. Let us assume that seed nodes are chosen among a–nodes only, and let ρ_a^0 be the fraction of those seed nodes. 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. Then the fraction of active a nodes evolves according to the following equation (Galstyan & Cohen 2007):

$$\tau \frac{d\rho_a}{dt} = -\rho_a + g_a(z_{aa}\rho_a) \quad (1)$$

where we have defined

$$g_a(x) = 1 - (1 - \rho^0_{a,b})Q(H, x) \quad (2)$$

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

Thus, the fraction of the population that will be activated at the end of the cascading process is determined from the following steady state equation:

$$\rho_a^s = g_a(z_{aa}\rho_a^s) \quad (3)$$

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 1, where we plot both sides of Equation 3 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 $\rho \approx 1$, aside from exponentially small correction of order $\sim z^{H-1}e^{-z_{aa}}$, indicating that the activation will

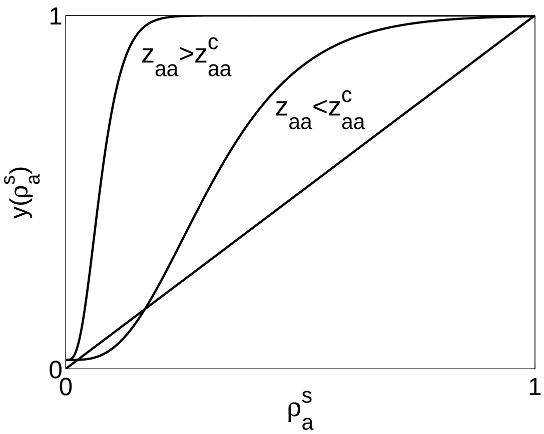

Figure 1: Graphical representation of the Equation 3. Plotted are the straight line $y = \rho^s_a$, and the function $y = g_a(z_{aa}\rho^s_a)$ for two different values of z_{aa}

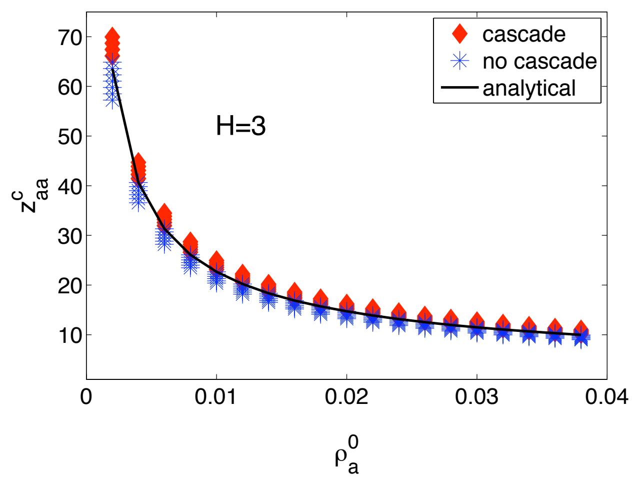

Figure 2: (Color online) The critical connectivity plotted against the fraction of seed nodes for the threshold parameter $H = 3$. The solid line shows the phase boundary obtained analytically.

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 1. Specifically, there is a critical value z^c_{aa} so that for $z_{aa} < z^c_{aa}$ the cascading dynamics dies out, while for $z_{aa} > z^c_{aa}$ it spreads throughout the system. Let us define $x = z_{aa}\rho^s_a$, and rewrite Equation 3 as $z^{-1}_{aa}x = g(x)$. At the critical point, the line $z^{-1}_{aa}x$ must be tangential to $g(x)$. It is then straightforward to demonstrate that the critical connectivity is given by

$$z^c_{aa} = [g'_a(x_0)]^{-1} \equiv \left[(1-\rho^0_a)e^{-x_0}\frac{x_0^{H-1}}{(H-1)!}\right]^{-1} \quad (4)$$

where x_0 satisfies the following equation:

$$x_0 g'_a(x_0) = g_a(x_0) \quad (5)$$

In the limit of small ρ^0_a one obtains the following scaling behavior:

$$z^c_{aa} \propto (\rho^0_a)^{-\frac{H-1}{H}}. \quad (6)$$

We also note that at the critical point the convergence time diverges as $T_{conv} \propto (z - z^c_{aa})^{-1/2}$.

In Figure 2 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^0_a, 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.

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. 1. In particular, the steady state fraction of active b nodes satisfies the following equation:

$$\rho^0_b = 1 - Q(H, z_{bb}\rho^0_b + \lambda) \equiv g_b(z_{bb}\rho^0_b + \lambda). \quad (7)$$

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

$$\frac{x-\lambda}{z_{bb}} = g_b(x) \quad (8)$$

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

$$\lambda^c = x_0 - z_{bb}g_b(x_0) \quad (9)$$

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

$$g'_b(x_0) = \frac{1}{z_{bb}} \quad (10)$$

Note that for $\rho^0_a = 1$ λ_c is simply the critical across–group connectivity $z^c_{ab}(z_{bb})$ for which the cascade will spread to b nodes, assuming that all a nodes have already been activated. Hence, equations 9 and 10 implicitly define a critical line $z^c_{bb}(z_{ba})$ 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^{max}_a be the fraction of active a nodes that corresponds to the maximum activation rate among a nodes. This can be found from Equation 1 by differentiating the right hand side with respect to ρ_a and setting it to zero, which yields $z_{aa}g'_a(z_{aa}\rho^{max}_a) = 1$. If the across—group connectivity is smaller than λ^c/ρ^{max}_a, 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 $\lambda^c < z_{ba} < \lambda^c/\rho_a^{max}$.

Thus, we have demonstrated that activity spreading in a network composed of two loosely coupled Erdosh–Renyi graphs is characterized by a doubly–critical behavior. We believe that this phenomenon has potential implications for viral marketing strategies. Indeed, 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. To show this, consider a scenario where the size of the group A is much smaller than group, but with higher link density among the group members. In real settings, group A might represent a small but devoted group of a product followers (e.g., Mac users), while group B represents the rest of the consumer market. Then, according to our results, it might be more optimal to target the members of group A. This is because it is easier to cause a global cascade in group A, that will later "spill" into the second, larger group B. We intend to examine this issue more thoroughly in out future work.

References

Domingos, P., and Richardson, M. 2001. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 57–66. New York, NY, USA: ACM Press.

Galstyan, A., and Cohen, P. 2007. Cascading dynamics in modular networks. *Phys. Rev. E* 75:036109.

Kempe, D.; Kleinberg, J.; and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146. New York, NY, USA: ACM Press.

Leskovec, J.; Adamic, L. A.; and Huberman, B. A. 2006. The dynamics of viral marketing. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, 228–237. New York, NY, USA: ACM Press.

Watts, D. 2004. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. U.S.A.* 99:5766.