Cascade dynamics on clustered network

Y Ikeda\(^1\), T Hasegawa\(^2\), and K Nemoto\(^1\)

\(^1\) Department of Quantum and Condensed Matter Physics, Graduate School of Science, Hokkaido University, Kita 10-jo Nisi 8-tyome, Sapporo, JAPAN
\(^2\) Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, JAPAN

E-mail: hasegawa@stat.t.u-tokyo.ac.jp

Abstract. We investigate information cascades on a clustered network model generated from a projection of bipartite graphs by numerical simulations and compare it with that on a non-clustered network having the same degree distribution. Our result indicates that global cascades occur more easily on clustered networks than on non-clustered ones. Furthermore, we extend a recursive method by Gleeson and Cahalane to estimate the order parameter analytically and find that the result gives an excellent approximation to the observed transition point.

1. Introduction

In many complex networked systems [1–3], small initialized shocks are often propagated through a whole system. For example, the spread of rumors and the diffusion of norms and innovations are observed in social systems. These phenomena are called information cascades, during which individuals in a population exhibit collective behavior [4–17]. To explain the mechanism causing information cascades, Watts [4] proposed a network model (the Watts model) in which decisions of interacting agents are determined by the actions of their neighbors according to a simple threshold rule:

1. Consider a network consisting of nodes and edges, which correspond to individuals and relationships between them, respectively. The state of each node is either active or inactive.

2. As an initial state, we set a randomly-chosen node active. The node is called a seed.

3. At each time step, each inactive node updates its state active if at least a fraction \(\phi\) of its neighbors are active, otherwise it remains inactive. Once being active, a node cannot turn back to the inactive state (the permanently active property).

4. Step 3 is iterated until no node changes its state. Then cascade size \(\rho\) is evaluated as the final fraction of active nodes.

The dynamics of cascades depends on the structure of underlying network [4–8]. For random graphs it is well known that very large cascades, so called global cascades having non-vanishing \(\rho\), are triggered by small initialized shocks for small \(\phi\) and disappear as a discontinuous phase transition in \(\rho\) with increasing \(\phi\) [4, 5].

The dynamics of cascades has been well analyzed on locally tree-like networks, whereas many social networks have high clustering coefficient. Among such high-clustering networks, there are many real social networks represented as bipartite graphs, e.g., networks of CEOs, boards of directors, and collaborations of scientists [2]. A bipartite graph is defined as a network in which
nodes are divided into two sets in the way that no edge connects two nodes in the same set. In this paper, we analyze the Watts model on a clustered network generated from a bipartite graph to find the effect of clustering on cascade dynamics. Our numerical and analytical results demonstrate that global cascades on a clustered network occur in wider range of $\phi$ compared with those on a non-clustered one. Moreover our result shows that global cascades which are triggered by the effect of clustering structure are rare events, indicating that information cascades in our society are often hard to anticipate.

2. One-mode projected network

In this section, we introduce projected network and reconstructed network, as a clustered network and non-clustered one, respectively. Our clustered network model is generated by a projection of a bipartite graph [18]. Firstly, we consider a bipartite graph which consists of Person- and Group-nodes (Fig.1-(a)). Person-nodes follow a Poisson distribution $P_{\text{Person}}(k) \propto (z/2)^k e^{-z/2}/k!$ with mean $z/2$, and all Group-nodes have the same three degree $P_{\text{Group}}(k) = \delta_{k,3}$. Now the bipartite graph can be projected onto Person-nodes only (Fig.1-(b)). This resulting graph has a Poisson degree distribution $P(k) \propto z^k e^{-z}/k!$ ($k = 2, 4, 6, \cdots$), and simultaneously has a high density of triangles, the clustering coefficient $C = 1/(z+1)$ [18]. We call this graph the projected network.

Next, keeping the degree of each node, we reconnect edges at random. This operation produces a non-clustered network with the same degree distributions as the original one, which we call the reconstructed network.

We trigger information cascades on these two networks, and compare the results to investigate the effects of clustering structure on information cascades. In the next section, we propose an approximation to analyze the information dynamics on the projected network.

3. Random cactus approximation

Gleeson and Cahalane [5] proposed a general analytical method by using recursion relations to analyze the cascade dynamics for locally tree-like networks with arbitrary degree distributions. We extend their method to estimate the order parameter $\rho$ on the projected network analytically.

Cascade size $\rho$ is equal to the probability that a randomly-chosen node is active in a final state. To estimate this probability, we choose an arbitrary node as the root, and approximate the projected network by a random cactus tree with $P(k)$, starting from the root (see Fig.2). We
Figure 2. Random cactus approximation.

label the levels \( n \) of the nodes from the bottom of the tree \( (n = 0) \), with the root at an infinitely high level \( (n = \infty) \). We estimate the propagation of a cascade starting from a single seed \( (\rho_0 = 1/N) \) at the bottom by the following update equations. Firstly, we define the probabilities \( q_n^{(i)} \) \((i = 1, 2, 3, 4)\) that a node pair at level \( n \) is in the state \( i \), given that the parent node at level \( n + 1 \) is inactive. Here the pair-states \( i = 1, 2, 3, 4 \) correspond to inactive-inactive, active-inactive, inactive-active, and active-active node-pairs, respectively. The probabilities of the pair-states at level \( n \) are given by the probabilities \( \tilde{q}_n \) and \( \tilde{q}'_n \), where \( \tilde{q}_n \) \((\tilde{q}'_n) \) is the probability that an inactive node at the \( n \)-th level changes its state on the condition that the other of the pair being inactive (active). The update equations for \( \tilde{q}_n \) and \( \tilde{q}'_n \) are given as

\[
\tilde{q}_{n+1} = \rho_0 + (1 - \rho_0) \sum_{k=0}^{\infty} \frac{(k + 1)p_{k+1}}{z} \times \sum_{k_1,k_2,k_3,k_4} \frac{k!}{k_1!k_2!k_3!k_4!} q_n^{(1)k_1} q_n^{(2)k_2} q_n^{(3)k_3} q_n^{(4)k_4} F(k_2 + k_3 + 2k_4, 2k + 2),
\]

\[
\tilde{q}'_{n+1} = \rho_0 + (1 - \rho_0) \sum_{k=0}^{\infty} \frac{(k + 1)p_{k+1}}{z} \times \sum_{k_1,k_2,k_3,k_4} \frac{k!}{k_1!k_2!k_3!k_4!} q_n^{(1)k_1} q_n^{(2)k_2} q_n^{(3)k_3} q_n^{(4)k_4} F(k_2 + k_3 + 2k_4 + 1, 2k + 2).
\]

Here the response function \( F(m, k) \) is

\[
F(m, k) = \begin{cases} 
0 & \text{if } m < \phi k \\
1 & \text{if } m \geq \phi k,
\end{cases}
\]

where \( m \) is the number of active nodes in \( k \) nearest neighbors. Then, we obtain probabilities \( q_{n+1}^{(1)} \) at the \((n + 1)\)-th level as

\[
q_{n+1}^{(1)} = (1 - \tilde{q}_{n+1})^2.
\]
Figure 3. Phase diagram of $\phi - z$ plane. The symbols outline the region in which global cascades occur in simulation. The dashed line represents boundaries of cascade phase. The random graph shown here has a Poisson degree distribution $P(k) \propto 6^k e^{-6}/k!$ ($k = 1, 2, 3, \ldots$).

$$q^{(2)}_{n+1} = q^{(3)}_{n+1} = \tilde{q}_{n+1}(1 - \tilde{q}_{n+1}),$$
$$q^{(4)}_{n+1} = \tilde{q}_{n+1}^2 + 2\tilde{q}_{n+1}(\tilde{q}_{n+1} - \tilde{q}_{n+1}).$$

We iterate this level-by-level update to obtain $q^{(i)}_\infty$ which are the fixed points of the above recursion relations. Then, the probability of the root node being active, i.e., the cascade size $\rho$, is given as

$$\rho = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} p_k \times$$
$$\sum_{k_1, k_2, k_3, k_4} \delta_{k-k_1-k_2-k_3-k_4} \frac{k!}{k_1!k_2!k_3!k_4!} q^{(1)}_{k_1} q^{(2)}_{k_2} q^{(3)}_{k_3} q^{(4)}_{k_4} F(k_2 + k_3 + 2k_4, 2k).$$

4. Numerical simulations

In this section, we investigate numerically the cascade dynamics on both the projected network and the reconstructed network by Monte-Carlo simulation. We pick up the largest cascade in $10^2$ cascade trials on each network sample with $10^6$ nodes, and average those sizes over $10^2$ network samples. The global cascade phase is given as the region in which the averaged cascade size is proportional to the network size.

Figure 3 expresses the cascade condition of the two network models graphically as a boundary in the phase diagram of threshold $\phi$ and average degree $z$. Clearly, the global cascade phase of the projected networks is wider than that of the reconstructed networks, indicating that cluster structure enhances information cascades.

Figure 4 shows our numerical and analytical results for active node fraction $\rho$ as a function $\phi$. One finds that the random cactus approximation gives an excellent approximation to the observed transition point, and thus supports our numerical results. Both results show that global cascades appear as a discontinuous phase transition in $\rho$ with decreasing $\phi$ on both networks.
Finally, we compare the distributions of cascade sizes on both networks (Fig. 5). From the distribution of cascade sizes, we find that the Watts model on the projected network takes the following three regimes:

- For $\phi > \phi_2$, where $\phi_2 = \phi_2(z)$ is the critical threshold of the projected networks, no global cascades occur on both the projected network and the reconstructed network.
- For $\phi_1 < \phi < \phi_2$, where $\phi_1 = \phi_1(z)$ is the critical threshold of the reconstructed networks, global cascades occur rarely (1 time per 10 trials) on the projected network while no global cascades occur on the reconstructed network. This means that global cascades in this extended regime are triggered by the effect of the cluster structure.
- For $\phi < \phi_1$, global cascades occur frequently (9 times per 10 trials) on both networks.

In [4], the cascade size distributions on random graphs show us that the information cascades occur rarely at the critical threshold, implying a more extreme kind of instability that is harder to anticipate. Interestingly, our numerical result reveals that the cluster structure induces a bimodal distribution of cascade sizes, i.e., rarely occurring global cascades, in a wide range of $\phi$. This result may indicate that global information propagations in our human society, where two of your friends are likely to be friends of each other, will be harder to predict.

5. Summary
We studied an information cascade model proposed by Watts [4] on a clustered network to find the effect of clustering on cascade dynamics. We extended a recursive method by Gleeson and Cahalane [5] to estimate the order parameter $\rho$ analytically. Both analytical and numerical results showed that global cascades occur more easily on a clustered network than on a non-clustered one. Moreover, the clustered network has a nontrivial bimodal distribution of cascade sizes which indicates global cascades occur rarely in a wide range of $\phi$.

As indicated in [6], recursive method for cascade propagations can be applied to other dynamics, e.g., $k$-core percolations, by modifying the response function $F(m,k)$. It is very interesting to extend our method in this paper for general propagating dynamics.

Acknowledgments
The present work is supported by 21st Century COE program Topological Science and Technology.
References

[1] Albert R and Barabási A -L 2002 Rev. Mod. Phys. 74 47.
[2] Newman M E J 2003 SIAM Review 45 167.
[3] Boccaletti S, Latora V, Moreno Y, Chavez M, and Hwang D -U 2006 Phys. Rep. 424, 175.
[4] Watts D J 2002 PNAS 99 5766.
[5] Gleeson J P and Cahalane D J 2007 Phys. Rev. E 75 056103.
[6] Gleeson J P 2008 Phys. Rev. E 77 046117.
[7] Dodds P S and Payne J L 2009 Phys. Rev. E 79 066115.
[8] Payne J L, Dodds P S, and Eppstein M J 2009 Phys. Rev. E 80 026125.
[9] Kempe D, Kleinberg J and Tardos E 2003 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining 137.
[10] Dodds P S and Watts D J 2004 Phys. Rev. Lett. 92 218701.
[11] Duan W Q, Chen Z, Liu Z, and Jin W 2005 Phys. Rev. E 72 026133.
[12] Watts D J and Dodds P S 2007 J. Consum. Res. 34 441.
[13] Centola D, Eguiluz V M, and Macy M W 2007 Physica A 374 449.
[14] Galstyan A and Cohen P 2007 Phys. Rev. E 75 036109.
[15] Klimek P, Lambiotte R, and Thurner S 2008 Europhys. Lett. 82 28008.
[16] de Kerchove C, Krings G, Lambiotte R, Van Dooren P, and Blondel V D 2009 Phys. Rev. E 79 016114.
[17] Galstyan A, Musoyan V, and Cohen P 2009 Phys. Rev. E 79 056102.
[18] Newman M E J 2003 Phys. Rev. E 68 026121.