

# Phase Transitions for Information Spreading in Random Clustered Networks with Arbitrary Contact Rates

Sungsu Lim  
KAIST  
ssungssu@kaist.ac.kr

Namju Kwak  
KAIST  
beatu@kaist.ac.kr

Kyomin Jung  
KAIST  
kyomin@kaist.edu

Information spreading plays an essential role in numerous human interactions, including the diffusion of innovations, rumors, fads, and epidemics. Phase transitions for information spreading have been observed in many complex systems, such as social, economic, and biological systems [3,8]. That is, there exists a *tipping point* at which information suddenly spreads rapidly and globally so that the final cascade size becomes a non-zero fraction of the total number of individuals. Identifying under what conditions such a large-scale spreading happens and analyzing spreading patterns are crucial problems because they are closely related to the eruption of an epidemic in epidemiology, the initiation of a trend in marketing, and so on. In this work, we study the effect of community structure on the tipping point, and interestingly we prove that the condition needed to induce a phase transition is strictly easier for networks with more clustered structures. More precisely, we prove that in the random clustered network with higher clustering coefficient (a measure of the degree to which the nodes in a network tend to cluster together) has strictly lower tipping point than random unclustered network with the same expected degree distribution.

The independent cascade model (essentially the stochastic SIR model), which is one of the most popular information diffusion models, has been widely used to explain phase transitions [12]. In this model, when node  $i$  becomes active, it has a single chance of activating each currently inactive neighbor  $j$  with probability  $f(i, j)$ , which is called the *contact rate* between information sender  $i$  and receiver  $j$ . We assume that the diffusion process begins with a randomly chosen set of initial adopters with a finite size. Many previous works on information spreading have focused on the case that a network has a locally tree-like structure, such as Erdős-Rényi random networks [5] and configuration models [4,9]. However, *triadic closure* (friends of friends are more likely to become friends) occurs in real-world networks [13]. That is, a community structure is found in many complex networks, and the locally tree-like assumption can be misleading. To demonstrate this, Newman [14] and Miller [10] simultaneously proposed a model for *random clustered networks* by considering the given degree distribution and the number of triangles each node participates in. This model displays the clustered structure of a network.

For the case in which  $f(i, j)$ 's are identical to some constant  $T$ , the spread of information over networks has been well-studied and this process can be mapped exactly onto a bond percolation procedure on the same network with the edge occupation probability  $T$  [7,10,12]. In [14] and [10], the authors analyzed bond percolation on random clustered networks, but only for the constant  $T$ . In addition, Prakash et al. [16] provided a formula for finding the tipping points for arbitrary virus propagation models on any networks, for the case in which the transmission probability is a constant value. However, the contact rates almost never be the constant on information spreading [1]. In this work, we analyze phase transitions for information spreading and obtain some conditions for tipping point occurrence in random clustered networks with *arbitrary* contact rates  $f(i, j)$ , and provide a formula to compute the probability and the size of the occurrence of the large-scale spreading.

In Online Social Networking (OSN) services, such as Twitter and Facebook, each user can only pay attention to a limited number of messages from user's followees (people who are subscribed to the user) [17]. Assume that each user retweets (repost another user's message)  $c$  messages in a specific time period from her followees equally likely. Then, the contact rate can be modeled according to  $f(i, j) = c/w_j$ , where  $w_j$  is the number of followees of user  $j$ . On the other hand, suppose that each informed individual selects  $c$  neighbors uniformly at random and transmits messages to them (i.e., an individual informed about a specific message has  $c$  tokens to transmit the message), the contact rate can be modeled by  $f(i, j) = c/w_i$ . Moreover, we consider the multiple contact scenario. If several types of diffusion mechanisms can co-exist in a network, then we union them to represent a contact rates  $f(i, j)$  for each pair of individuals  $i$  and  $j$ . When the above two contact rates are accumulated, we can approximate a contact rate according to  $c/w_i + c/w_j$ . We identify conditions for phase transition under random unclustered and clustered networks, when the contact rates are heterogeneous; we consider three different types of basic contact rates (i)  $c/w_i$ , (ii)  $c/w_j$ , and (iii)  $c/w_i + c/w_j$ . Interestingly, under these various contact rates, we prove that the condition to induce a phase transition is strictly easier for highly clustered networks. This means that clustered networks improve the appearance of the large-scale spreading. Our proof can be applied to other general contact rates as well.

**Theorem 1** *Let  $\mathcal{G}_1$  and  $\mathcal{G}_2$  be random clustered networks with the same expected degree distribution  $\mathbf{d} = (d_1, \dots, d_n)$ . If the clustering coefficient of  $\mathcal{G}_1$  is larger than that of  $\mathcal{G}_2$ , then the  $c$  value corresponding to the tipping point is lower for  $\mathcal{G}_1$ .*

We now describe the details of the random clustered network and provide a method for computing the probabilities and the sizes of the occurrence of the large-scale spreading on random clustered networks. Suppose that we have two sequences  $\mathbf{s} = (s_1, \dots, s_n)$  and  $\mathbf{t} = (t_1, \dots, t_n)$ , where  $s_i$  is the number of incident edges of a node  $i$  that are not included in any triangle and  $t_i$  is the number of triangles that contain a node  $i$ . The random clustered network  $G(\mathbf{s}, \mathbf{t})$ , with two given sequences  $\mathbf{s} = (s_1, \dots, s_n)$  and  $\mathbf{t} = (t_1, \dots, t_n)$ , is a probability space over the set of networks on the node set  $V = \{1, \dots, n\}$ . The procedure for generating a random clustered network is as follows: Consider a random unclustered network  $\mathcal{G}_s$  with a given expected degree sequence  $\mathbf{s}$ , and a random network  $\mathcal{G}_{(i,j,k)}$  for each  $1 \leq i < j < k \leq n$ , so that it has only three edges  $\{i, j\}$ ,  $\{i, k\}$ , and  $\{j, k\}$  that form a triangle with probability  $\frac{t_i t_j t_k}{\sum_{x < y} t_x t_y}$ , and that has no edges otherwise. We get a random clustered network by taking the union of a network generated by  $\mathcal{G}_s$  and a network generated by  $\mathcal{G}_{(i,j,k)}$  for all  $1 \leq i < j < k \leq n$ . If we take  $\mathbf{s} = \mathbf{d}$  and  $\mathbf{t} = \mathbf{0}$ , then  $G(\mathbf{s}, \mathbf{t})$  is equivalent to the configuration model with a given expected degree sequence  $\mathbf{d}$  [15].

To prove Theorem 1, we first consider the following. For any  $i, j \in V$ , let  $P_i$  be the probability that an initial infectious node  $i$  induces a large-scale spreading, and let  $S_j$  be the probability that a node  $j$  is contained in the large-scale spreading. Then, we obtain the following formula using the branching process method at the macroscopic level of the structure of the network,  $1 - P_i = \prod_{j \neq i} (1 - k(i, j)P_j)$ , where  $k(i, j) = \frac{s_i s_j}{\sum_x s_x} f(i, j) + \frac{t_i t_j}{\sum_{x < y} t_x t_y} (f(i, j) \sum_k t_k - \frac{1}{2} f(i, j) \sum_k t_k f(i, k) P_k + (1 - f(i, j)) \sum_k t_k f(i, k) (1 - P_k) f(k, j))$  is the probability that an infectious node  $i$  infects a susceptible node  $j$  in a random unclustered network. Similarly, we obtain that  $1 - S_j = \prod_{i \neq j} (1 - k(i, j)S_i)$  using the backward branching process method. Then, we compute the probability and the size of the occurrence of the large-scale spreading  $P = \frac{1}{n} \sum_{i=1}^n P_i$  and  $S = \frac{1}{n} \sum_{j=1}^n S_j$ .

Our argument can be applied to any network with specific substructure of finite size, including cliques of size larger than 3, cycles, and motifs, rather than triangles in the above. By a suitable mapping from a clustered structure to a locally tree-like structure, we can apply a branching process

method at the macroscopic level of the structure of a network. More specifically, when there is a modular structure of a network that is locally tree-like, we can lead to the values of  $k(i, j)$ 's and analyze the occurrence of the large-scale spreading using our formula.

We conducted Monte Carlo experiments of information spreading on real-world social network topologies such as Facebook (63K nodes) and MySpace (100K nodes) to compare the computed values of the expected sizes and the probabilities of the occurrence of the large-scale spreading based on our proposed method with the Monte Carlo estimates. We performed simulations for various contact rates  $f(i, j) = c/w_i$ ,  $c/w_j$ , and  $c/w_i + c/w_j$ , where  $c$  is a certain constant varying from 0 to 3. The simulations show that the values obtained using our formula with random clustered networks are more accurate than those obtained with configuration models.

## References

- [1] A.-L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207–211, 2005.
- [2] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Trans. on Information and System Security*, 10(4), 2008.
- [3] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Epidemic spreading in scale-free networks. *Rev. Mod. Phys.*, 80:1275–1335, 2008.
- [4] R. M. D’Souza. Complex networks: Structure comes to random graphs. *Nature Phys.*, 5:627–628, 2009.
- [5] P. Erdős and A. Rényi. On random networks. *Publicationes Mathematicae*, 6:290–297, 1959.
- [6] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *IEEE INFOCOM*, pages 1455–1466, 2005.
- [7] E. Kenah and J. M. Robins. Second look at the spread of epidemics on networks. *Phys. Rev. E*, 76:036113, 2007.
- [8] J. Kleinberg. The convergence of social and technological networks. *Commun. ACM*, 51:66–72, 2008.
- [9] L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham. Network theory and sars: predicting outbreak diversity. *Journal of Theoretical Biology*, 232(1):71–81, 2005.
- [10] J. C. Miller. Percolation and epidemics in random clustered networks. *Phys. Rev. E*, 80:020901(R), 2009.
- [11] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–180, 1995.
- [12] M. E. J. Newman. The spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.
- [13] M. E. J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [14] M. E. J. Newman. Random graphs with clustering. *Phys. Rev. Lett.*, 103:058701, 2009.
- [15] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- [16] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. In *IEEE ICDM*, pages 537–546, 2011.
- [17] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2:335, 2012.