Tipping Point of Information Spreading in Random Clustered Networks with Heterogeneous Contact Rates

Tipping point phenomena for information spreading in complex networks have been studied for decades in various disciplines. A tipping point is a moment at which information suddenly spreads rapidly and globally. Understanding how tipping points occur in networks is an important problem, which is closely related to the eruption of an epidemic in epidemiology, the initiation of a trend in marketing, and so on. In this work, we identify condition for tipping points to occur and analyze spreading behaviors in random clustered networks with heterogeneous contact rates.

The Susceptible-Infected-Recovered (SIR) model and its variants have been widely used to explain the tipping point phenomena [3]. Much of the previous research on the SIR model have concentrated on locally tree-like networks such as the configuration model. However, triadic closure (friends of friends are more likely to become friends) occurs in social networks. To demonstrate this, Newman and Miller proposed simultaneously a model for random clustered networks by considering the degree distribution and the number of triangles each node participates [4, 2].

Miller introduced a method to calculate the probabilities and the sizes of the large-scale spreading in random clustered networks for constant contact rates [2]. However, contact rates almost never be the constant in social networks. Let $f(i,j)$ be the contact rate between an infectious node i and its susceptible neighbor j. For example, if we assume that each user can read c messages per a day from his or her friends equally likely, then $f(i,j) = c/d_j$ where d_j is the degree of j. Likewise, if we assume that each user can send c messages to his or her friends equally likely, then $f(i,j) = c/d_i$. In this work, we consider arbitrary contact rate $f(i,j)$, and obtain a necessary and sufficient condition for the tipping point occurrence under the random clustered network. We also obtain formulas to compute the probability and the size of the large-scale spreading.

Suppose that we have two sequences $\mathbf{s} = (s_1,\ldots,s_n)$ and $\mathbf{t} = (t_1,\ldots,t_n)$, 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 general random clustered network $G(\mathbf{s},\mathbf{t})$ is a probability space over the set of networks on the node set $\{1,\ldots,n\}$, so that it is determined by adding an edge with probability $s_i s_j / \sum_x s_x$

for each pair of nodes i and j, and also joining three nodes to form a triangle with probability $t_i t_j t_k / \sum_{x<y} t_x t_y$ for each triple of nodes i, j and k. If we take $\mathbf{s} = \mathbf{d}$ and $\mathbf{t} = \mathbf{0}$, it is equivalent to the configuration model with a given degree sequence $\mathbf{d}$ [5].

Let S be the probability that a randomly chosen node is contained in a large-scale spreading and S_j be the probability that a node j is contained in a large-scale spreading. Then, we obtain the following formula for each S_j,

$$1 - S_j = \prod_{i \neq j} (1 - k(j,i) f(j,i) S_i), \tag{1}$$

where $k(j,i) = s_j s_i / \sum_x s_x + t_j t_i \sum_k t_k / \sum_{x<y} t_x t_y$ is the probability that there is an edge from i to j in a random clustered network. We solve (1) by taking the logarithm of both sides and applying the first-order Taylor series approximation, and show that it computes asymptotically correct values of S_j for random clustered networks with any power-low degree distribution. Then, we compute the size of the large-scale spreading $S = \frac{1}{n} \sum_{j=1}^{n} S_j$. Note that a large-scale spreading can occur if and only if $S = \Omega(1)$. The above formula is a very general framework. For example, when $f(i,j) = c/(s_i + 2t_i)$ or $c/(s_j + 2t_j)$, we prove that for any $(\mathbf{s}, \mathbf{t})$ with power-law degree distribution the large-scale spreading occurs if $c > 1$, and it does not occur if $c < 1$. If $s_i = np$ and $t_i = 0$ for all i, our result is identical to the well-known result for the Erdős-Rényi random graph $G(n,p)$ [1].

Using a similar argument, we also compute the probability that a large-scale spreading occurs when the spreading is initiated by a single node. Our argument can be applied to any specific substructure of the network of finite size including cliques, motifs, and chains rather than triangles. Furthermore, we provide a formula to estimate the probability and the size of the large-scale spreading when the network topology is given and not random.

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 sizes and the probabilities of the large-scale spreading based on our proposed method with the Monte Carlo estimates. We performed simulations of three different contact rates $f(i,j) = c/d_i$, c/d_j, and $c/d_i + c/d_j$ where c is some constant varying from 0 to 3. The simulations show that the values obtained by our formula with random clustered networks ($0 \sim 5\%$ error) are more accurate than with configuration models ($0 \sim 7\%$ error). Moreover, the accuracy of predictions using our formula for the actual network topology ($0 \sim 1\%$ error) is better than the accuracy of predictions using random clustered networks.

References

1. P. Erdős and A. Rényi, Publicationes Mathematicae **6**, 290-297 (1959).
2. J. C. Miller, Phys. Rev. E **80**, 020901(R) (2009)
3. M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002)
4. M. E. J. Newman, Phys. Rev. Lett. **103**, 058701 (2009)
5. M. E. J. Newman, Netwoks: An Introduction (Oxford University Press, 2010).