Information Cascades on Arbitrary Networks

Hyuna Kim, Seulki Lee and Kyomin Jung

Among various fields of complex networks, information diffusion has recently attracted huge attention due to the growing popularity social network services and increasing accessibility of their data. Indeed, analysis of information diffusion patterns not only can be widely applicable from epidemiology to viral marketing, but also can greatly benefit to society by offering predictions of information flows.

To shed light on how new ideas, technologies, and epidemics diffuse through networks, numerous models have been established [5, 9, 4]. Kempe, Kleinberg, and Tardos (2003) integrated those previous models to propose two fundamental models, (the linear threshold model and the independent cascade model) and their generalizations (the general threshold model and the general cascade model) [6]. In particular, the threshold model assumes that an individual needs to have enough proportion (or threshold) of his neighbors who have previously influenced by the same information. This mechanism is based on utility maximization of individuals in game theoretic consideration. Despite its significance, because of its difficulties, analysis of the threshold models have entirely focused on limited conditions such as the submodular influence (by Mossel-Roch (STOC 07)), homogeneous thresholds (by Whitney(Phys. Rev. E. 10)), and locally tree-like networks (by Watts(PNAS 02)) [7, 11, 10].

Under the general threshold model and arbitrary networks, we proved that only if all nodes are of degrees $\omega(\log n)$, the final cascade size is highly concentrated around its mean with high probability. In other words, for any $\varepsilon > 0$ and $\delta > 0$,

$$Pr(|\mathscr{X} - \overline{\sigma}| > n\varepsilon) = o(n^{-\delta})$$

Hyuna Kim
KAIST e-mail: hyunak@kaist.ac.kr

Seulki Lee
KAIST e-mail: sklee19@kaist.ac.kr

Kyomin Jung
KAIST e-mail: kyomin@kaist.edu

where $\mathscr{X}$ is the final cascade size, $\overline{\sigma} = E[\mathscr{X}]$ and n is the number of individuals constituting a network. our result is also valid for the independent cascade models, the linear threshold models and Katz-Shapiro models – a special case of the linear threshold model. Our proof implies that the influence, or the expected cascade size, is asymptotically invariant to the network structures. Along with this proof, we provide a formula under the linear threshold model and an efficient algorithm under the general threshold model that estimate the average cascade size and the probability of being influenced for each individual.

Being comparable with one trial of Monte-Carlo experiments, our algorithm can be used for a tipping point prediction problem [3] and an influence maximization problem [2, 8]. The former problem refers to predicting a phase transition in the final cascade size. Because a special case of our diffusion process can be regarded as diffusion of innovation initiated by public marketing, our formula can provide the minimum intensity of public marketing to trigger a large cascade on a network so that the marketing succeeds. The latter problem indicates identifying a most influential set of initial adopters of a specific size which maximizes the final cascade size. Closely related to the targeted marketing with word-of-mouth effects, this problem is known as NP-hard under the linear threshold model and the independent cascade model [6]. Our algorithm can be employed as a subroutine of many approximation algorithms such as [1].

On top of this, we confirmed by performing extensive experiments that the final cascade size is actually concentrated around its mean and a tipping point appears at the predicted point.

References

1. Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. pages 1029–1038, 2010.
2. Pedro Domingos and Matthew Richardson. Mining the network value of customer. *Seventh International Conference on Knowledge Discovery and Data Mining*, 2001.
3. Malcom Gladwell. The tipping point: How little things can make a big difference. *Back Bay Books*, 2000.
4. J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 2001.
5. Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
6. David Kempe, Jon Kleinberg, and Eva Tardos. Maximizing the spread of influence through a social network. *KDD*, 2003.
7. Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proc. 39th ACM Symposium on Theory of Computing*, 2007.
8. Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. *Eighth International Conference on Knowledge Discovery and Data Mining*, 2002.
9. Thomas C. Schelling. Micromotives and macrobehavior. *Norton*, 1978.
10. Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99:5766–5771, 2002.
11. Daniel E. Whitney. Dynamic theory of cascades in finite clustered random networks with a threshold rule. *Physical Review E*, 82:066110–1–10, 2010.