# Analysis of Information Diffusion on Threshold Models in Arbitrary Social Networks


Kyomin Jung
KAIST
kyomin@kaist.edu

Seulki Lee
KAIST
sklee19@kaist.ac.kr

Sungsu Lim
KAIST
ssungssu@kaist.ac.kr


How information, new ideas, and rumors diffuse through a social network has been a fundamental question for many decades in various disciplines including computer science, social science, and economics [1, 5, 9, 14]. In recent decades, various models of information diffusion have been studied. In particular, Kempe et al. [10] proposed *the general threshold model*, a unified generalization of many diffusion models on networks, which is based on utility maximization of individuals in game theoretic consideration. In this model, each node $v$ has its own threshold function $f_v : \{0,1\}^{d_v} \to [0,1]$ that depends only on the decisions of its neighbors, where $d_v$ is the degree of $v$, and a threshold value $\theta_v \in [0,1]$, which is drawn independently and randomly from a certain probability distribution $\mu$. Despite the importance of such study, analysis on the threshold model has focused on special cases such as homogeneous thresholds [16], submodular influence [11], and locally tree-like networks [15].

In this work, we consider the general threshold model with arbitrary threshold distribution on an *arbitrary* network structure that may have many triangles and short cycles. First, we prove that, for any linear threshold model (i.e., $f_v$'s are linear functions) with arbitrary threshold distribution when the initial adopters are chosen uniformly at random, only if (essentially) all nodes are of degrees $\omega(\log n)$, then the final cascade size is sharply concentrated around its mean with high probability, where $n$ is the total number of nodes. Furthermore, somewhat surprisingly we prove that the expectation of the cascade size is asymptotically independent of the network structure under the same condition.

More precisely, we consider that the network structure is given and that each individual becomes an initial adopter independently with a given probability $x \in [0,1]$. This process can be regarded as diffusion process initiated by public marketing that affects each individual to be an initial adopter independently with probability $x$. Let $\mathcal{Z}(= \mathcal{Z}_x)$ be the final cascade size if an initial adopter set is given by the above process, and let $F$ be the cumulative density function of $\mu$. Let $\alpha_0 = x$, the fraction of initial adopters, and let $\alpha_t = F(\alpha_{t-1})$ for $t \in \mathbf{N}$. Let $\alpha^* = \lim_{t\to\infty} \alpha_t$. Then one can easily check that $\alpha^* = min\{\alpha | F(\alpha) = \alpha\} \leq 1$. We prove that the final cascade size is very close to $\alpha^* n$ with high probability and it is asymptotically independent of the network structure.

**Theorem 1** *Suppose that $\mu$ is continuous and $F$ is not tangential to $y = \alpha$ at $\alpha^*$. If $d_v = \omega(\log n)$ for all $v \in V$ and the weight of $u$'s influence on $v$ is $w_{uv} = O\left(\frac{1}{d_v}\right)$ for every $uv \in E$, then for any $\epsilon > 0$ and $\delta > 0$,*

$$Pr[|\mathcal{Z} - \alpha^* n| > n\epsilon] = o(n^{-\delta}).$$

The condition that $y = F(\alpha)$ is not tangential to the line of $y = \alpha$ at $\alpha^*$ is a very mild assumption and it only requires that there exists $\gamma > 0$ such that $F(\alpha^* + \Delta) < \alpha^* + \Delta$ for all

$\Delta \in (0, \gamma]$. Moreover, the condition that the edge weight of $uv \in E$ is of $O\left(\frac{1}{d_v}\right)$ simply means that there is no strong relationship by an edge $uv$ that is dominated the others which have the same end point $v$. Note that $\alpha^*$ can be computed in $O(1)$ and it is independent of the network structure, and only dependent on the threshold distribution $\mu$ and $x$. In general $\alpha^*$ is discontinuous at some values of $x$, and Theorem 1 enables us to compute those values of $x$ for which the cascade size $\alpha^*$ becomes suddenly large. Hence, we can compute such tipping points by Theorem 1. In the proof of Theorem 1, we prove that for any $t \in \mathbf{N}$, there exists $\epsilon_t$ (depending on $\epsilon$) such that any node $v$ with its threshold value $\theta_v \leq \alpha_t - \epsilon_t$ adopts information with high probability after $t$ iterations, while any node $v$ with $\theta_v > \alpha_t + \epsilon_t$ does not adopt information with high probability after $t$ iterations.

We then extend our results to arbitrary general threshold models with any initial adopter set and provide upper and lower bounds of the cascade size under certain mild conditions. We generalize so that each individual $v$ becomes an initial adopter independently with probability $x_v$. Furthermore, we provide an efficient algorithm under the general threshold model that estimates the final cascade on any network structure with any given set of initial adopters. Our algorithm can be employed as a subroutine for numerous algorithms for *the influence maximization problem* [4, 10]. The influence maximization problem is to find one of the most influential sets of initial adopters of a specific size that maximizes the final cascade size. Our work also can be applied to measuring *the resilience and the vulnerability of network structure* in the presence of cascading failures or attacks [2, 8]. This problem is closely related to financial crises, electrical blackouts of power stations, and the failure of Internet routers.

We conducted experiments with our algorithm for estimating the cascade size on real-world social networks including Epinions networks and political blogs, and synthetic networks including small-world networks and scale-free networks to validate our results on networks having nodes with small degrees. We verified at each iteration of the diffusion process that threshold values of individuals who are influenced at that time are actually concentrated around the mean, as suggested in Theorem 1. We also confirm by experiments that the final cascade size is actually concentrated around the value computed by our proposed algorithm, and that a tipping point appears at the values of $x$ computed by Thorem 1 even when the degrees of each individual are not so large (around 20 on average).

We now describe the diffusion process of the general threshold model in more details. In this model, each individual $v$ has its own threshold function $f_v : \{0, 1\}^{d_v} \to [0, 1]$ that depends only on the decisions of its neighbors, where $d_v$ is the number of neighbors of $v$, and on a threshold value $\theta_v \in [0, 1]$, which is drawn independently and randomly from a certain probability distribution $\mu$. Given a set of initial adopters $A$ at time $t = 0$, the diffusion process proceeds in discrete timesteps: in each iteration $t$, all nodes that were adopted in iteration $(t - 1)$ remain adopted, and each individual $v$ decides to adopt the information if the $f_v$ value becomes at least $\theta_v$. The process runs until no additional activation is possible. This model assumes that to adopt the information, an individual needs to have a large enough proportion of its neighbors who have adopted the same information. This consideration is rational when adoption requires some cost, or when the benefits of adoption increase as more people make the adoption.

One of the most widely studied threshold models is *the linear threshold model*, in which the threshold functions $f_v$ are linear [7, 12, 13]. For example, a local interaction game with selfish agents can be viewed as an example of the linear threshold model [3, 6]. In this game, each agent can choose between two choices, obtaining a payoff when her neighbors make the same decision. Another important class of diffusion model is *the independent cascade model* (essentially the SIR model). In this model, every time a node $u$ becomes infected, every uninfected neighbor $v$ of $u$ has a single chance of becoming infected, with probability $p_{uv}$. This can be regarded as a special case of the general threshold model that has the threshold functions $f_v(N_v(t)) = x_v + (1 - x_v)(1 - \prod_{u:uv \in E}(1 - p_{uv} Z_u(t)))$ where $x_v$ is the probability that $v$ becomes an initial adopter and $Z_u(t) = 1$ if $u$ has adopted the information at iteration $t$ and $Z_u(t) = 0$ otherwise.

The novelty of our analysis lies in its broad applicability to a general class of threshold models and network topologies. Our results are naturally applicable to broader settings, including non-submodular influence conditions for any network with node degrees that are (essentially) $\omega(\log n)$. As a future work, we plan to work on generalization of our results into more general network setups, for example when the states of nodes do not need to be monotone over time. It would be also our future direction to extend down our results to milder conditions than the current requirement of $\omega(\log n)$ degrees.

# References


[1] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *The Journal of Political Economy*, 100(5):992–1026, 1992.

[2] Larry Blume, David Easley, Jon Kleinberg, Robert Kleinberg, and Eva Tardos. Which networks are least susceptible to cascading failures? In *FOCS*, 2011.

[3] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[4] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. *ACM SIGKDD*, 2010.

[5] James Coleman, Elihu Katz, and Herbert Menzel. The Diffusion of an Innovation Among Physicians. *Sociometry*, 20(4):253–270, 1957.

[6] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.

[7] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[8] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65:056109, 2002.

[9] Joe A. Dodson Jr. and Eitan Muller. Models of new product diffusion through advertising and word-of-mouth. (182), 1976.

[10] David Kempe, Jon Kleinberg, and Eva Tardos. Maximizing the spread of influence through a social network. *ACM SIGKDD*, 2003.

[11] E. Mossel and S. Roch. On the submodularity of influence in social networks. In *STOC*, 2007.

[12] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. *ACM SIGKDD*, 2002.

[13] Thomas C. Schelling. Hotkey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. *The Journal of Conflict Resolution*, 17, 1973.

[14] Thomas W. Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18:69–89, 1996.

[15] Duncan J. Watts. A simple model of global cascades on random networks. *PNAS*, 99:5766–5771, 2002.

[16] Daniel E. Whitney. Dynamic theory of cascades in finite clustered random networks with a threshold rule. *Phys. Rev. E*, 82:066110–1–10, 2010.