

Information Dissemination Models in Generalized Social Networks

Srinidhi Kulkarni V
4thSem M.Tech,
APSCE, Bangalore

Mr. Somashekar T
Senior Lecturer,
APSCE, Bangalore

Abstract—Information dissemination dynamics is an important performance index for characterizing the speed and scope of information dissemination in a network. Several analytical models have been proposed to estimate the information dissemination dynamics for social networks adopting a susceptible-infected information propagation model and having a constant informed rate. Here we extend existing model to analyze the social networks adopting either a susceptible-infected or an independent-cascade information propagation model and having a time-varying informed rate. Validated by simulations, our model demonstrates its flexibility and applicability to approximate the complicated propagation behavior of advertisement-like or malware-like information.

Index Terms—Information dissemination dynamics, social networks, susceptible-infected model, independent-cascade model.

I. INTRODUCTION

A generalized social network is a network that describes the interdependency of individuals from the perspective of both *social network* and *communication network* [1,2]. The social network characterizes the interdependency of individuals in terms of the social relationships such as friendship, kinship, financial exchange, or relationships of beliefs, knowledge or prestige. The communication network describes interdependency of individuals in terms of the connections offered by the communication technologies (e.g., Bluetooth, WiFi, GSM, etc.), social websites (e.g., Facebook, YouTube, Twitter, etc.) or instant messaging tools (e.g., Skype, MSN, WhatsApp, etc.). The generalized social network plays an important role in spreading information. An individual equipped with a communication device such as a smart phone or notebook can easily exchange information with his/her friends in a voluntary or involuntary manner.

The *information dissemination dynamics* characterizes the fraction of active subpopulations in a network as a function of time [3,4]. It is a widely used performance index for information dissemination and is used in the rest of this paper. The characteristic of information dissemination and the topologies of the networks are two main factors that affect the information dissemination dynamics.

The characteristic of information dissemination depends on the behavior of individuals.

The topologies of the networks depend on the quantity and quality of the social or communication connections among individuals. Several information propagation models have been developed to mimic the behaviors of individuals in dealing with the information based on sociology, physics, or epidemiology. An

independent-cascade (IC) model [5], a linear-threshold (LT) model [5], and a heat-diffusion (HD) model were proposed to characterize the dissemination of product, idea, or innovation [6]. A susceptible-infected (SI) model was proposed to model the infection of malicious rumors or computer viruses [7]. In IC model, each active individual has a single chance to activate each of his inactive friends and the attempt succeeds with a user-specific probability. In LT model, each active individual has a weight to influence his inactive friends. An individual is activated when the total weights of his active friends exceed a user-specific threshold. In HD model, the influence of each active individual is modeled as a heat source. An individual is activated when the total amount of heat received from his friends exceeds a user-specific threshold. In SI model, each active individual has multiple chances to activate the same inactive individual and the attempt succeeds with a user-specific probability. Note that the influence of active individual (i.e., weight in LT model and heat in HD model) is a constant in LT model but is dynamically changed in HD model during the period of information dissemination.

The topology of a social network is characterized by its degree distribution and connection attribute. The degree distribution is the probability distribution of the number of friends for each individual. The connection attribute is the strength of the link between two individuals. Three main types of degree distribution are widely used in the literature: random networks, small-world networks, and scale-free networks [8]. In random networks, the connection between two individuals is generated at random. Therefore, any two individuals in the network have

the same probability to be friends. In small-world network, the distance between two randomly chosen individuals grows proportionally to the logarithm of the number of individuals in the network. In this network, most individuals are not friends of one another, but most individuals can be reached from

every other by a small number of hops or steps. In scale-free network, the degree distribution follows a power law. Differential-equation-based model has been proposed to analyze the spreading of email worm in computer networks [8]. The authors in [3] adopted the same concept to model the information epidemics in complex networks with opportunistic links and dynamic topology. The authors in [4] further extend the concept to model the malware propagation in generalized social networks. Existing models for information dissemination were developed based on the assumptions that the information propagation in the networks follows the SI model. Moreover, a constant informed rate [3] was considered. In the letter, we extend the analytical models to generalized networks

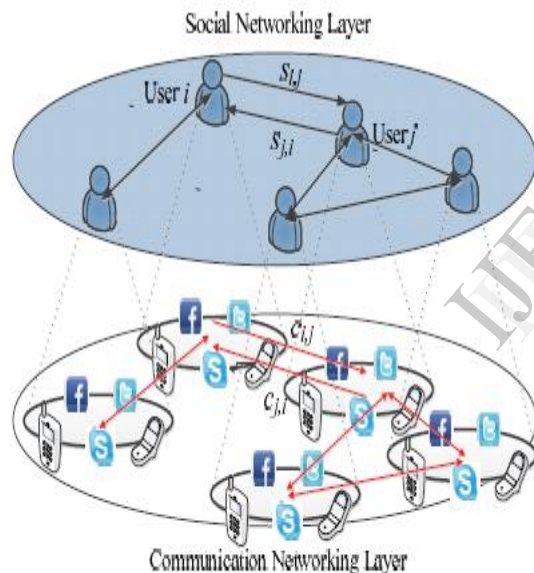


Fig. 1. A generalized social network model.

adopting both SI and IC information propagation models and a time-varying informed rate. The rest of this letter is organized as follows. Section II presents the system model considered in this letter. An analytical model is proposed to estimate the information dissemination dynamics in generalized social networks. Simulation results are shown in Section III. Finally, conclusions are summarized in Sec. IV.

II. SYSTEM MODEL

Figure 1 shows a generalized social network model comprising a social networking layer and a communication networking layer [1,2]. The two layers are used to describe the social relationship and the communication connectivity among

individuals. The social networking layer is formulated as a graph $G = (N, NE)$, in which N and NE denote the set of individuals and the set of connections among individuals, respectively.

Let $|N|$ and $|NE|$ denote the number of individuals and the number of connections among individuals, respectively. Let $s_{i,j}$ be the normalized social strength from individuals i to j in the social networking layer ($0 \leq s_{i,j} \leq 1$, $s_{i,j} = 0$). $s_{i,j}$ represents the probability that i successfully propagates information to j . In general, $s_{i,j} = s_{j,i}$; and individuals i and j are friends if $s_{i,j} > 0$.

In the communication networking layer, two friends can communicate with each other by using social websites or instant messaging tools carried by different communication technologies. Let $c_{i,j}$ be the contact rate [2] of individuals i and j ($0 \leq c_{i,j} \leq 1$, $c_{i,i} = 0$, $c_{i,j} = c_{j,i}$). $c_{i,j}$ represents the probability that an individual i communicates to his friend j using his communication devices and media in a unit slot time. In this paper, we assume that $s_{i,j}$ and $c_{i,j}$ are independently retrieved from the statistics of the social interactions and communication interactions between individuals i and j , respectively. In the generalized social network, each individual is in

either *active* or *inactive* state. An individual in active state implies that he/she has adopted the information. In the following, we present a discrete-time analytical model to analyze the information dissemination dynamics in a generalized social network. That is, we divided the time into fixed-length time slots, in which the slot length denotes the average duration that an individual spends to interact with his friend through the communication network.

Suppose in a general social network, M is the maximum degree of the network and $p(k)$ is the probability that a

randomly and uniformly selected user has degree k [1] ($1 \leq k \leq M$). The average degree of the network is

$$\bar{k} = \sum_{k=1}^M kp(k).$$

Let $ik(t)$ be the fraction of active individuals in the k -degree individual set at time t ($ik(t) = 0$ for $t < 0$). The information dissemination dynamics at time t , $I(t)$, is given by [3,4]

$$I(t) = \sum_{k=1}^M ik(t)p(k). \quad (1)$$

The differential equation for individuals with degree k is given by

$$d_{ik}(t) dt = k(1 - ik(t))\theta(t), \quad (2)$$

where $\theta(t)$ is the probability that an active individual activates an inactive friend at time t . Note that $\theta(t)$ in Eq. (2) is equal to $\lambda\Theta(t)$ in [Eq. (4), 9] since a time-varying informed rate

is considered in this paper. Using the discrete-time method suggested in [Eq. (6), 9], Eq. (2) can be rewritten as

$$i_k(t) = i_k(t-1) + k\theta(t-1)(1 - i_k(t-1)).$$

The probability that an inactive individual is activated at time

t , $\theta(t)$, is given by [Eq. (4), 9]

$$\theta(t) = \frac{\sum_{k=1}^M (k-1)p(k) \sum_{\tau=0}^t \lambda(\tau, t)(i_k(\tau) - i_k(\tau-1))}{\sum_{k=1}^M kp(k)} \quad (4)$$

In Eq. (4), $i_k(\tau) - i_k(\tau-1)$ represents the portion of k -degree inactive individuals which are activated at time τ , $\lambda(\tau, t)$ is the probability that an individual is activated at time τ and then further activates an inactive individual at time t ($\tau \leq t$) and $(k-1)p(k)$ is the probability that each individual contact with a k -degree individual [9]. Note that $\lambda(\tau, t)$ is known as informed rate [4] or infection rate [9]. $\lambda(\tau, t)$ depends on the social strength and the contact rate among individuals as well as the information propagation model.

Let v and u be the average values of the contact rate and the social strength, respectively. v and u can be obtained as

$$v = \frac{1}{|NE|} \sum_{i=1}^N \sum_{j=1}^N c_{i,j}$$

and

$$u = \frac{1}{|NE|} \sum_{i=1}^N \sum_{j=1}^N s_{i,j}$$

Note that v is the average probability that a user contact with a friend; and u is the average probability that a user propagates information to his friend.

In SI model, each active individual has multiple chances to activate the same friend. 'An individual i which is activated at time τ will activate a friend j at time t ' represents that

' i successfully activates j at time t but fails to do so during $(t - \tau - 1)$.

$$\lambda(\tau, t) = \begin{cases} (1 - vu)^{t-\tau-1}vu, & \text{if } t - \tau \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the product of u and v in Eq. (7) means that a user is informed if both its social neighbor and physical connection are being activated at the same time for successful information propagation.

In IC model, each active individual has a single chance to activate the same friend. 'An individual i which is activated at time τ will activate a friend j at time t ' represents that ' i contacts and activates j

at time t and does not contact with j during $(t - \tau - 1)$. (3)

$$\lambda(\tau, t) = \begin{cases} (1 - v)^{t-\tau-1}vu, & \text{if } t - \tau \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

III. SIMULATION RESULT

Simulation results are conducted on top of a C++ simulation platform to investigate $I(t)$. In the following, symbols and lines are used to represent the simulation and analytical results, respectively. In the simulation, $c_{i,j}$ and $s_{i,j}$ are independently generated. The same parameters setting used in [3] are used

herein for comparison. In the simulation, $|N| = 2500$ and $|NE| = 12500$; $c_{i,j} = 0.1$ for all i, j ; and $s_{i,j} = 0.1$

for all i, j , are used in both IC and SI models. The degree distribution of the network topology, $p(k)$, are generated by using Barabasi and Albert (BA) scale-free model, Erdos and Renyi (ER) random graph model, and Watts and Strogatz (WS) small-world model, respectively. We follow the procedures described in [8] to generate the degree distribution of the models. As in [3], we use $m = 5$ in BA model,

$PER = 2|NE|/(|N| \times (|N| - 1))$ in ER model, and $P_{ws} = 0.2$ in WS model, respectively.

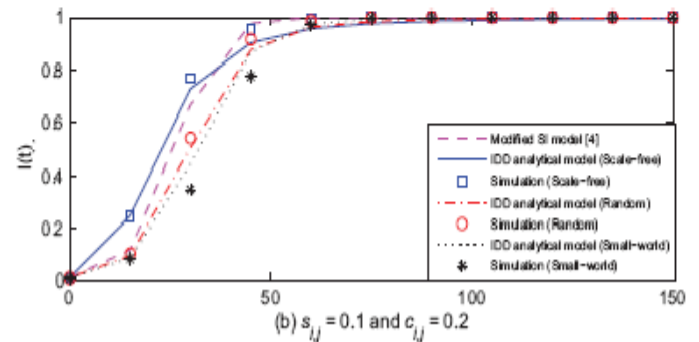
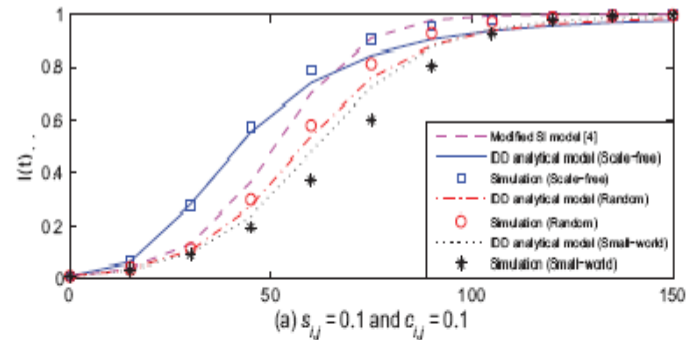
The complexity of the proposed analytical model and the simulation program are equal to $O(MT^2)$ and $O(R|N|2T)$, respectively, where M is the maximum degree value in the network, R is the number of repeated simulations, and T is the time constraint. In the simulation, the average of M was 190, 16, and 20, in BA, WS, and ER networks, respectively. $R = 50$, and $T = 150$. The time used by the simulation and the analysis is 27478 to 128828 ms and 17 to 217 ms, respectively. From the simulations, it is found that the curves of information dynamics closely match our analytical model. Limited discrepancy exists between the simulation and the analysis because the information may propagate to individuals who have already been activated and uncertain boundary conditions could not be considered in the analysis. Similar results can be

found in [9]. Figures 2 (a), (b), show the results of $I(t)$ based on SI model for $s_{i,j} = 0.1, c_{i,j} = 0.1$; $s_{i,j} = 0.1, c_{i,j} = 0.2$; and $s_{i,j} = 0.2, c_{i,j} = 0.1$, respectively. At $t = 0$, 1% of total populations is randomly selected as active individuals and thus, $I(0) = 0.01$. The modified SI model proposed in [4] is chosen as a benchmark for illustrating the effectiveness of the proposed model. It is found in Fig. 2 that $I(t)$ of the modified SI model proposed in [4] is irrelevant to the network topology since it is developed based on the assumption that all individuals in the network have identical degree.

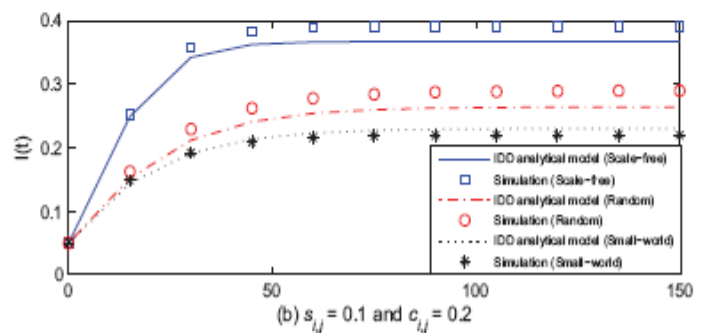
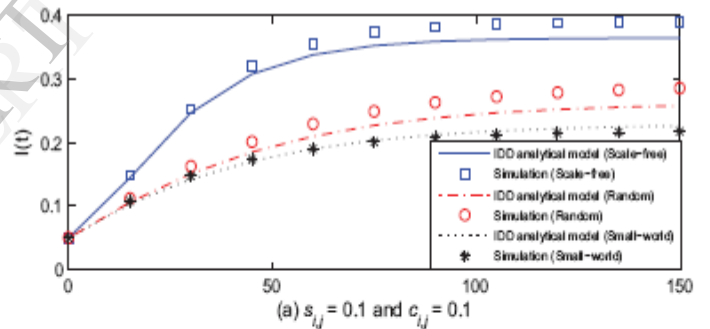
The effect of the contact rate ν on $I(t)$ of SI propagation model can be found in Figs. 2(a) and 2(b). Increasing $c_{i,j}$ results a higher contact rate and it reduces the time required to reach the final value of $I(t)$. Compare Figs. 2(b), it is found that increasing $c_{i,j}$ for $s_{i,j}$ by the same amount has identical effect on $I(t)$ in SI model. It is because that $\lambda(\tau, t)$ in Eq. (7) depends on the product of ν and u . However, the physical meanings of Figs. 2(b) is different. In Fig. 2(b), $I(t)$ increases because individuals have a higher probability to interact with the others. However, $I(t)$ increases in Fig. 2(c) since each individual has a higher probability to activate his friend.

Figures 3 (a), (b), and (c) show the results of $I(t)$ based on IC model for $s_{i,j}= 0.1, c_{i,j}= 0.1$; $s_{i,j}= 0.1, c_{i,j}= 0.2$; and $s_{i,j}= 0.2, c_{i,j}= 0.1$, respectively. At $t = 0, 5\%$ of total populations is randomly selected as active individuals and thus, $I(0) = 0.05$. No benchmark is chosen here because our model is the first approach developed for IC model. It can be found from Figs. 3(a) and 3(b) that increasing $c_{i,j}$ results in quickly increased $I(t)$ but the final value of $I(t)$ is not changed. From Figs. 3(a), it is found that increasing $s_{i,j}$ results in quickly increased $I(t)$; however, the final value of $I(t)$ is increased. It is because that, in IC model, an active individual i has a single chance to activate an inactive friend j and attempt succeeds with probability $s_{i,j}$. Increasing $c_{i,j}$ means that the individual i requires shorter time to contact with the inactive individual j . However, the final value of $I(t)$ depends only on the result of their first contact. From Figs. 2 and 3, it can be found that the information

Therefore, the final value of $I(t)$ is not increased if $c_{i,j}$ increases. However, a higher $s_{i,j}$ implies the individual i has a higher probability to activate the individual j at the first contact. Hence, the final value of $I(t)$ increases as $s_{i,j}$ increases. dissemination dynamics in generalized social networks with the SI model spreads faster than that with the IC model. It takes a shorter time for $I(t)$ to reach its final value in SI model than that in IC model. The final value of $I(t)$ in SI model is higher than that in IC model. It is because each user has multiple chances to activate the same friend in SI model.



Simulation result of $I(t)$ for SI model fig 2 (a) and 2(b)



Simulation result of $I(t)$ for IC model fig 3 (a) and 3(b)

REFERENCES

IV. CONCLUSION

This letter presents a general analytical model to capture the information dissemination dynamics of generalized social networks. To the best of our knowledge, this is the first analytical model that can be used for the networks adopting SI or IC information propagation models. Simulation results showed the accuracy of the proposed analytical model. The proposed analytical model can be used as a quick reference to gather approximate knowledge of propagation speed of information dynamics with various setting of social strengths, contact rates, and average degrees of the topology in generalized social networks. The advertiser could adopt such results to develop proper strategies and processes so as to adjust the speed of information propagation.

- [1] K. C. Chen, M. Chiang, H. V. Poor, "From technological networks to social networks," accepted and to be published in *IEEE J. Sel. Areas Commun.*
- [2] Y. F. Chou, R. G. Cheng, and H. H. Huang, "Performance evaluation of referral selection algorithms in human centric communication networks," in *Proc. 2012 International Symposium on Wireless Personal Multimedia Communications*, pp. 638–642.
- [3] P. Y. Chen and K. C. Chen, "Information epidemics in complex networks with opportunistic links and dynamic topology," in *Proc. 2010 IEEE Global Telecommunications Conference*, pp. 1–6.
- [4] S. M. Cheng, W. C. Ao, P. Y. Chen, and K. C. Chen, "On modelling malware propagation in generalized social networks," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 25–27, Jan. 2011.
- [5] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 2003 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146.
- [6] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," in *Proc. 2008 ACM Conference on Information and Knowledge Management*, pp. 233–242.
- [7] W. Kermack and A. McKendrick, "A contribution to the mathematical theory of epidemics," in *Proc. Roy. Soc.*, vol. A, no. 115, pp. 700–721, 1927.
- [8] X. F. Wang and G. Chen, "Complex networks: small-world, scale-free and beyond," *IEEE Circuits and Systems Mag.*, vol. 3, no. 1, pp. 6–20, 2003.