Modelling Information Dissemination under Privacy Concerns in Social Media

Hui Zhu[a], Cheng Huang[a], Rongxing Lu[b], and Hui Li[a]

[a]*State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an, China*
[b]*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.*

Abstract

Social media has recently become an important platform for users to share news, express views, and post messages. However, due to user privacy preservation in social media, many privacy setting tools are employed, which inevitably change the patterns and dynamics of information dissemination. In this study, a general stochastic model using dynamic evolution equations was introduced to illustrate how privacy concerns impact the process of information dissemination. Extensive simulations and analyzes involving the privacy settings of general users, privileged users, and pure observers were conducted on real-world networks, and the results demonstrated that user privacy settings affect information differently. Finally, we also studied the process of information diffusion analytically and numerically with different privacy settings using two classic networks.

Keywords: privacy concerns, information dissemination, social media, dynamic evolution equations

1. Introduction

With the development of mobile and web-based technologies, social media platforms, such as Facebook, Twitter, and LiveJournal, have become an important interactive platform allowing people to share information [1]. However, when increasing numbers of users publish news, ideas, and pictures/videos through social media, these platforms are capable of greatly influencing the real world (e.g., fake financial information may seriously lead

to turbulence of financial markets in nations worldwide [2]). Generally, social media users can be divided into two groups: *general users* and *special users*. *General users* transfer information through various social media tools, receive information from others, and have their friends share views and news in LiveJournal or other platforms. *Special users*, such as celebrities, use the same platforms to post information, news, and knowledge; however, their information may be more effectively and authoritatively transmitted due to their reputation and authority. For instance, general Twitter users may pay more attention to celebrities based on recognition, making their information and news more attractive and enabling it to spread faster relative to those of general users [3, 4]. Therefore, information diffusion in social media is a complex socio-psychological process, the analysis of which poses many challenges.

In order to accurately model information diffusion in social media, we need to know both a correct description of information dissemination and a quantitative formulation of various factors that motivate/restrict users to participate in social media. Based on the epidemic-spreading model [5], many information-dissemination models in social media have been proposed, including the susceptible-infected-susceptible (SIS) model and the susceptible-infected-removed (SIR) model, with several corresponding theoretical approaches for SIS and SIR having been presented, such as an individual-based mean-field approach [6, 7], a degree-based mean-field approach [8, 9, 10], and a generating-function approach [11, 12]. Pastor-Satorras et al. [13] and Vespignani [14] used the degree-based mean-field approach to describe information spreading using the SIS model, where the general methodology could be easily extended to almost all dynamic processes in social networks. Li et al. [15] and Ferreira et al. [16] also developed an SIS model using a degree-based mean-field approach for proposal of their models. Zhao et al. [17] and Wang et al. [18] improved the SIR model to a susceptible-infected-hibernator-removed model and a spreader-ignorant-stifler1-stifler2 model, respectively, using the individual-based mean-field approach to analyze rumour propagation in social networks. Additionally, many models are based on matrix calculations [19] that are used to describe information dissemination in social media. In order to analyze the dynamics of information dissemination thoroughly, the conditions of a given social network are usually changed in order to research different cases of information dissemination [20, 21, 22].

Most of the existing models discuss and analyze the process of information dissemination from various perspectives, such as different social media

tools, different sources of information, and different user reputations [23]. In this study, we focused on a new perspective, i.e., user privacy concerns, and attempted to reveal their effect on information dissemination in social media. Specifically, we considered whether information dissemination would be influenced by user privacy settings after the information had been produced, posted, and transferred. Users generally classify their information into three types: public information, half-privacy information, and privacy information, which describe whether the information can be viewed by anyone, friends, or no one, respectively. In order to protect their privacy, social media users may define privacy settings for different information using various social media tools; however, the effects of user privacy settings may lead to different dynamics of information spreading. Therefore, in this study and based on user privacy concerns, we proposed a general stochastic model to simulate the process of information diffusion under the constraints of user privacy concerns in social media. There exist some typical approaches [24, 25, 26] toward helping characterize user behaviour in social media; however, most of these are based on analysing detailed data obtained from crawling through user data from real-world sessions, based on their access to popular social networks. This data cannot be used to analyze the general process associated with information diffusion in social media. Therefore, considering our current model, the effects of individual behaviour on information diffusion are not included in our work and will be discussed in future work.

Here, we make several contributions to the study of information dynamics in social media. First, we provide specific definitions to the state of nodes and transition rules for nodes in social media. As a basis for analysing information diffusion with privacy concerns, we introduce a model of information spreading. Compared with previous models, we provide a simplified, but realistic description of this process. A LiveJournal dataset was used for simulation and analysis with different privacy concerns for general users, privileged users, and pure observers, collectively representing most users in social media. The simulation results showed that the different privacy settings of general users impacted the process of information dissemination, with privileged users having more rigorous privacy settings exerting greater effects on slowing information diffusion relative to general users, and pure observer privacy concerns having almost no influence on information dissemination, even under rigorous privacy settings. Finally, we numerically investigated two classic dynamics graphs, the directed erdos-renyi random (DER) graph and the directed scale-free (DSF) graph, with privacy concerns. With the DER

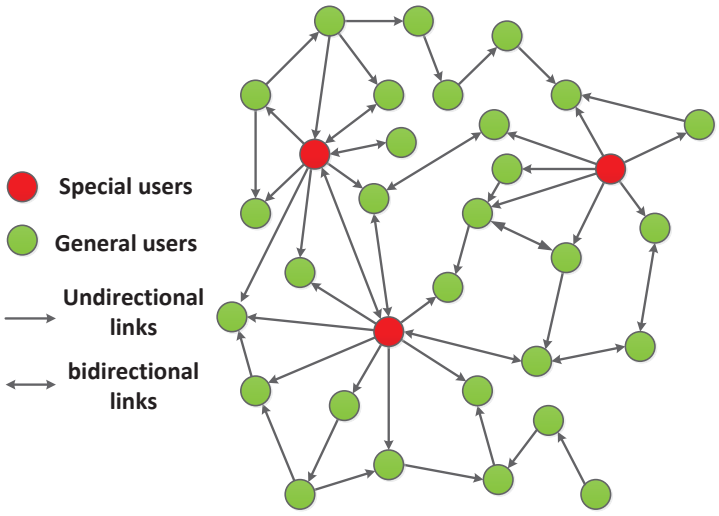


Figure 1: The network structure of social media.

graph, we found that privacy concerns had little influence on information diffusion, and with the DSF graph, the simulation results showed that privacy concerns exerted great influence on information dynamics, either promoting or blocking information diffusion.

This paper is organized as follows: In Section 2, we define the states and rules of information diffusion in social media according to user privacy concerns. Then, our proposed general stochastic model is presented to describe information dissemination with dynamic evolution equations in Section 3, followed by extensive simulations and discussions regarding various user privacy concerns in Section 4. In Section 5, we investigate two classic graphs with different privacy concerns. Finally, we present our conclusions in Section 6.

2. Definitions of and Privacy Concerns in Social Media

A network formed by social media can be viewed as a graph, $G = (V, E)$, where the nodes (vertices), V, are users and the links (edges), E, between nodes can be described as the interactions or interconnections between users [9]. Specifically, under different user privacy concerns, information may be diffused only via direct links from a user, and the diffused information will only influence some of the neighbouring nodes for this user. In order to accurately describe the relationships between users in social media, we defined the *out-degree* and *in-degree* of a node in G, where the *out-degree* captures

how information is transferred to neighbouring nodes via out-links, and the *in-degree* captures how information can be received from neighbouring nodes via in-links. Using *out-degree* and *in-degree*, we defined the network structure of a social media tool in Figure 1, where both general users and special users set their in-links and out-links based on their individual privacy concerns. In order to further distinguish individual privacy concerns, we classified them into two types: rigorous privacy settings and semi-rigorous privacy settings. The former indicates that users refuse to diffuse any information to their neighbours and the latter indicates that users only transfer information to some of their neighbours, e.g., their closer friends. In the following section, we consider the different privacy settings and evaluate their effects on information dissemination in social media.

- *Information dissemination with different general privacy settings.* In order to measure the effects of information dissemination with different general privacy settings, we randomly established the privacy settings of general users in social media (i.e. following the links, information will diffuse and influence other users having random restrictions according to different privacy settings in social networks).

- *Information dissemination with different special privacy settings.* In order to analyze the effects of special privacy concerns in social media, we chose the privileged users and pure observers as special users in order to evaluate their influence on the dynamics of information dissemination. The detailed definitions of privileged users and pure observers are illustrated below.

  - *Information dissemination with different privileged privacy settings.* In order to enhance activity, social media platforms invite and provide special privileges to some special users to encourage use of their platforms to post exclusive news and information. Therefore, there are some privileged users (celebrities) who are officially authenticated in social media. However, as privileged users, they would like increasing numbers of friends to receive their information in order to increase their reputation. Then, due to their rapid growth in reputation, they are able to attract additional users, enabling wider dissemination of their information (e.g., in Twitter, privileged users generally have many followers and always hope to increase this number). Therefore, we can

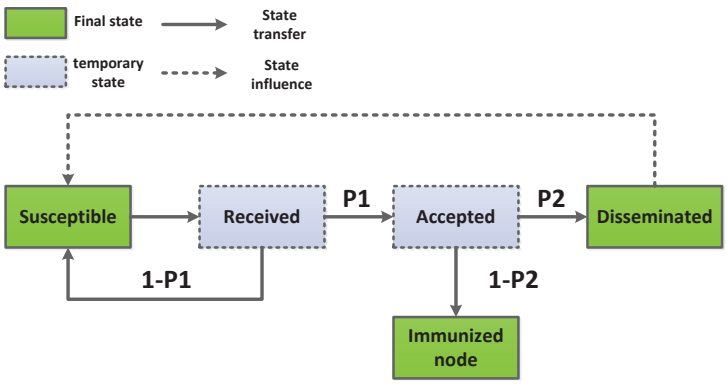


Figure 2: The rules of node state transition.

change privileged user privacy settings to analyze the effects of these special users on information diffusion.

– *Information dissemination with different pure observer privacy settings.* In social media, pure observers also inevitably exist. These users diffuse very little information, which defines their out-degrees as very low or even zero. Based on these characteristics, we call these users "divers". A wide array of divers may significantly change the dynamics of information dissemination. Therefore, we will also change the pure observer privacy settings to measure the effects of these special users on information diffusion.

In addition to the above considerations, in order to evaluate information diffusion with privacy concerns, different states of nodes and some transition rules should be defined (Fig. 2). According to the dynamics of infectious disease and traditional information-diffusion models, node states can be classified into two major categories: susceptible and accepted.

- Susceptible states (state S) represent nodes that are uninfected and can receive and accept information from their neighbours (this represents the initial state of all nodes except for the source node).

- Accepted states (state A) represent nodes that have received and accepted information from their neighbours (these nodes cannot be infected by neighbours if they have received information). The accepted state can also be divided into two categories: immunized and disseminated.

– Immunized states (state I) represent nodes that do not choose to spread information to their neighbours, even though they have accepted information (although they have accepted information, their neighbours cannot access it).

– Disseminated states (state D) represent nodes that choose to disseminate information to their neighbours, contrary to the immunized state (neighbours can obtain information from these nodes).

According to the state of nodes in social media, the transition state of each node will be impacted by its current state and the states of its neighbours. The transition rules are defined below.

- A node in state S may transfer to a node in state A with a rate of P_1 or remain constant with a rate of $1 - P_1$ when receiving information from neighbors.

- A node in state A may transfer to a node in state I with a rate of $1-P_2$ or a node in state D with a rate of P_2.

The above definitions of state and transition rules are simplified in social media. In the real world, there may exist more complicated situations that are not taken into consideration here. The aim of this study is to analyze the effects of user privacy concerns on information diffusion, and the above definitions are necessarily basic in order to allow analysis of information diffusion.

3. A General Model for Information Dissemination in Social Media

Table 1: Notations of different node number.

Notation	Description
$N(k,t)$	The number of k-degree nodes at time t
$S(k,t)$	$N(k,t)$ in state S
$I(k,t)$	$N(k,t)$ in state I
$D(k,t)$	$N(k,t)$ in state D
$A(k,t)$	$N(k,t)$ in state A

In order to evaluate the effects of user privacy concerns, we propose a general model based on interactive Markov chains [8]. As shown in Table 1, the notations of different node numbers are represented first. Then, we have

$$\begin{aligned} N(k,t) &= S(k,t) + A(k,t) \\ &= S(k,t) + I(k,t) + D(k,t). \end{aligned} \tag{1}$$

where N_x is a node in state S at time t, and the probability that N_x still stay in state S at time $[t, t+\Delta t]$ can be calculated as

$$P_{x(s\to s)} = (1 - P_1 \Delta t)^j, \tag{2}$$

where $j = j(t)$ denotes the number of disseminated neighbours of node N_x at time t. When N_x has k neighbours, $j = j(t)$ can be considered as a stochastic variable satisfying the following binomial distribution:

$$\Pr[j(t)] = \left(\begin{array}{c} k \\ j \end{array} \right) P(k,t)^j (1 - P(k,t))^{k-j}, \tag{3}$$

where $P(k,t)$ denotes the probability that a node lies in state D, and the node also connects another node in state S with k neighbours at time t. Therefore, $P(k,t)$ can be calculated as follows:

$$\begin{aligned} P(k,t) &= \sum_{k'} P(k'|k) P(d_{k'}|S_k) \\ &\approx \sum_{k'} P(k'|k) D_{dnew}(k',t). \end{aligned} \tag{4}$$

In Eq. (4), $P(k'|k)$ is the function that describes degree correlation, with $P(d_{k'}|S_k)$ representing the conditional probability that a node with degree k' is in state D ($d_{k'}$), given that its neighbour is a node in state S with degree k. In the above equation, the second equality should represent an approximation, which can be explained by the assumption that dynamic correlations can be ignored between the states of neighbouring nodes. $D_d(k',t)$ describes the density of nodes in state D with degree k' at time t. Therefore, according to the above assumption, $D_{dnew}(k',t)$ can be used to substitute $P(d_{k'}|S_k)$ as follows:

$$D_{dnew}(k',t) = D_d(k',t) - D_d(k',t-\Delta t). \tag{5}$$

Considering that a social media network is an uncorrelated inhomogeneous network, $P(k'|k)$ can be calculated with the following formula [27]:

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle}. \tag{6}$$

where $P(k')$ represents the probability that the node degree is k', and $\langle k \rangle$ represents the average degree of the network. Therefore, $P(k,t)$ can be described as follows:

$$\begin{aligned} P(k,t) &= \sum_{k'} \frac{k'P(k')}{\langle k \rangle} D_{dnew}(k',t) \\ &= P_2 \sum_{k'} \frac{k'P(k')}{\langle k \rangle} (D_a(k',t) - D_a(k',t-\Delta t)). \end{aligned} \tag{7}$$

We can now describe $D_{dnew}(k',t) = P_2(D_a(k',t) - D_a(k',t-\Delta t))$. By traversing all possible values of $j = j(t)$ in Eq. (3), the average probability $\overline{P}_{(s\to s)}(k,t)$ can be obtained as follows:

$$\begin{aligned} \overline{P}_{(s\to s)}(k,t) &= \sum_{j=0}^{k} \binom{k}{j} (1 - P_1\Delta t)^j P(k,t)^j \\ &\times \left(1 - P(k,t)\right)^{k-j} \\ &= \left(1 - P_1\Delta t P(k,t)\right)^k. \end{aligned} \tag{8}$$

Then, by substituting the value of $P(k,t)$ into the above equation, we can obtain

$$\begin{aligned} \overline{P}_{(s\to s)}(k,t) &= \left(1 - P_1P_2\Delta t \sum_{k'} \frac{k'P(k')}{\langle k \rangle} \right. \\ &\times \left. (D_a(k',t) - D_a(k',t-\Delta t)) \right)^k. \end{aligned} \tag{9}$$

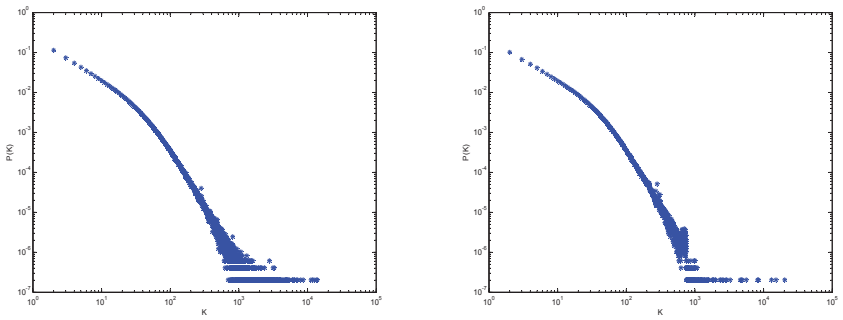
Therefore, based on our definitions of state transference, the average probability of a node transferring from state S to state A can be given by

$$\begin{aligned} \overline{P}_{(s\to a)}(k,t) &= 1 - \left(1 - P_1P_2\Delta t \sum_{k'} \frac{k'P(k')}{\langle k \rangle} \right. \\ &\times \left. (D_a(k',t) - D_a(k',t-\Delta t)) \right)^k. \end{aligned} \tag{10}$$

Using $\overline{P}_{(s\to a)}(k,t)$, the change of accepted node numbers with degree k during the time $[t, t+\Delta t]$ can be described as follows:

$$\begin{aligned} A(k,t+\Delta t) &= A(k,t) + S(k,t)\,\overline{P}_{(s\to a)}(k,t) \\ &= A(k,t) + [N(k,t) - A(k,t)] \\ &\times \left(1 - \left(1 - P_1 P_2 \Delta t \sum_{k'} \frac{k' P(k')}{\langle k \rangle}\right.\right. \\ &\times \left.\left.(D_a(k',t) - D_a(k',t-\Delta t))\right)^k\right). \end{aligned} \tag{11}$$

4. Simulation and Analysis



(a) (b)

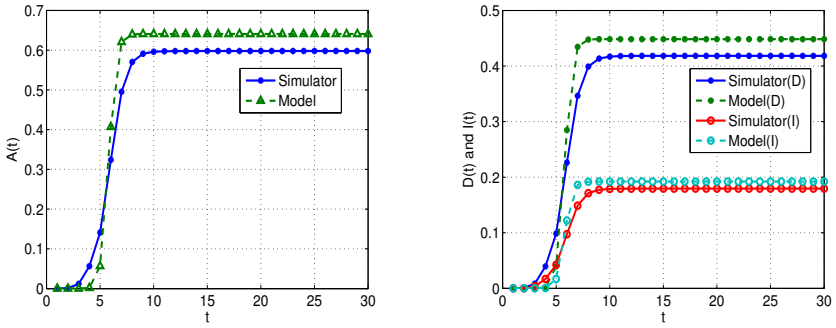
Figure 3: Degree distributions of a dataset. (A) In-degree distribution and (B) Out-degree distribution.

In this section, a dataset from LiveJournal [28] was selected for simulation and analysis. LiveJournal is a social media where users can share blogs, journals, or diaries through their website. The population may include a wide variety of political pundits, who also use the service for political commentary, enabling us to classify users in LiveJournal as special users and general users. Users in LiveJournal can be represented as nodes, where users can follow other users in order to obtain information. The following relationships can be represented as links between nodes. For example, if user A follows user B in LiveJournal, there exists a directed link from user B to user A. This link means that user A can obtain information from user B.

Statistically, the dataset consists of 4,847,571 nodes and 68,993,773 links, with the average degree being 14.23. Figure 3 shows the degree distributions of the dataset, including the in-degree and out-degree distributions. The privacy settings of general users, privileged users, and pure observers were adjusted by changing the corresponding node in-degree and out-degree to simulate information dissemination based on different user privacy concerns. Here, we considered the density of different nodes in state S, A, I, and D as $S(t)$, $D(t)$, $I(t)$, and $A(t)$, respectively, with time t.

According to Figure 3 and [3, 4], there are two ways to locate privileged users and pure observers. One is based on node degree and the other is based on centrality. The numbers of out-degree and in-degree nodes are used to locate the different kinds of users. The nodes with high out-degree can be seen as the privileged users and the nodes with no out-degrees and having high in-degrees can be seen as the pure observers.

4.1. Validation of the proposed model



(a) (b)

Figure 4: The density evolution of nodes in real-world and proposed models. (A) Nodes in state A; (B) nodes in state I and state D.

Before simulating user privacy concerns, our proposed model was validated by comparing the process of information dissemination in the real world with that of our model.

We used a simulator that we developed to simulate the process of information dissemination in the real-world network according to the LiveJournal dataset. Notably, there were no privacy concerns in our first comparison, and we used the same simulation conditions for both models for the comparison.

Detailed descriptions of the comparison, which included two processes, are below.

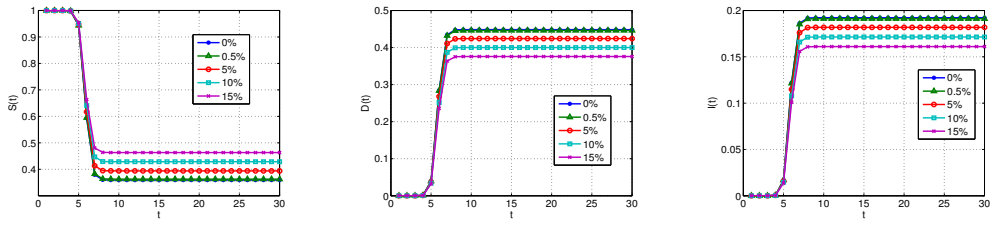
1. First, we classified all nodes in the dataset into state I or state D at random (70% and 30% in state I and state D, respectively). Then, we selected a node as the source of the information and recorded the density evolution of various nodes that accepted the message. This process was repeated 100 times in order to calculate the average result of the records. Finally, the process of information diffusion for this message using the LiveJournal dataset was described based on the average result. The simulation runs in time steps, with one time step meaning that some nodes in state D can accept information and in the next time step, these nodes begin to spread the information immediately and infect their neighbors.
2. Based on the density of various nodes in the simulation, we can calculate the corresponding parameters (e.g., $P_2 = 0.3$), then randomly select a node with the average degree as the source node of information and use our model to simulate the process of information diffusion to calculate the density evolution of nodes that accept the message. Finally, the corresponding process of information diffusion using our proposed model can be illustrated.

As shown in Figure 4, the density evolution, $A(t)$, $D(t)$, and $I(t)$, which describe the density of nodes in state A, I, and D, displayed similar results over time using our model and the real-world simulation. This means that the process of information dissemination using our model and the real-world simulation were similar. Therefore, our proposed model was accurate in describing the dynamics of information dissemination and can be used as a basis to evaluate the effects of user privacy concerns on information propagation in social media.

4.2. Information dissemination with general user privacy settings

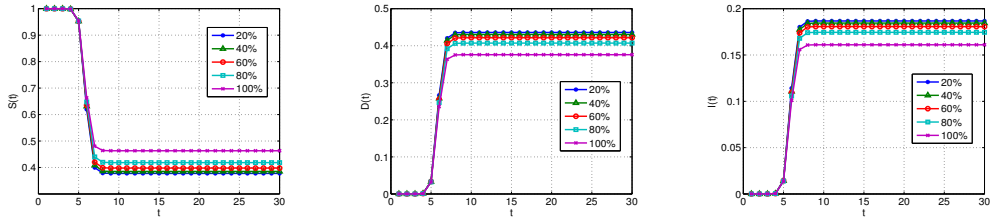
In order to simulate and analyze the influence of general user privacy concerns in social media, we changed the relationships between general users by adjusting the in-degree and out-degree of nodes.

When setting general user privacy settings to rigorous, all out-links are closed, i.e., the corresponding out-degree of nodes are set to 0. In our simulation, we randomly chose 0.5%, 5%, 10%, and 15% of general user nodes from the dataset, set their privacy settings to rigorous, and then set P_2 to



(a) (b) (c)

Figure 5: The density evolution of nodes with different general user privacy settings (rigorous privacy settings). (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I.



(a) (b) (c)

Figure 6: The density evolution of nodes with different general user privacy settings (semi-rigorous privacy settings). (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I.

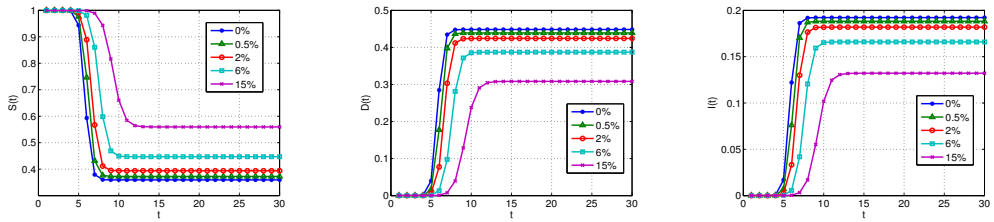
0.3. As shown in Figure 5, the privacy settings of general users had different effects on information dissemination. Specifically, the increase in nodes with rigorous privacy settings slowed information dissemination, i.e., fewer nodes received information and nodes were also required to wait longer to receive information. Furthermore, we choose 15% of general user nodes from the dataset and assigned them semi-rigorous privacy settings in order to investigate the effects of fewer privacy concerns on information diffusion. We also increased the out-links of the chosen nodes from 20% to 100%. As shown in Figure 6, semi-rigorous privacy settings for general users also affected the dynamics of information dissemination. The stricter the privacy settings for general users, the slower information spread.

Moreover, we found that the average clustering coefficient for the LiveJournal dataset was 0.274 [29]. Given that users in social media are closer to one another in LiveJournal, little variance was observed in information dissemination despite the relatively low number of users setting rigorous

or semi-rigorous privacy settings. Therefore, the privacy settings of a few users can impede information diffusion in social media. However, the final state of information diffusion, which describes how many users have received the information, cannot be affected. These results can be explained by the phenomenon of "social clustering", which indicates that when most of the neighbours of one node are real neighbours, eliminating a few links between nodes may have no effect. In other words, the information posted by a user with a semi-rigorous privacy setting will be received by all of their friends, unless the user has rigorous privacy settings. Nevertheless, if many users set rigorous privacy settings, information diffusion will be blocked.

4.3. Information dissemination with special user privacy settings

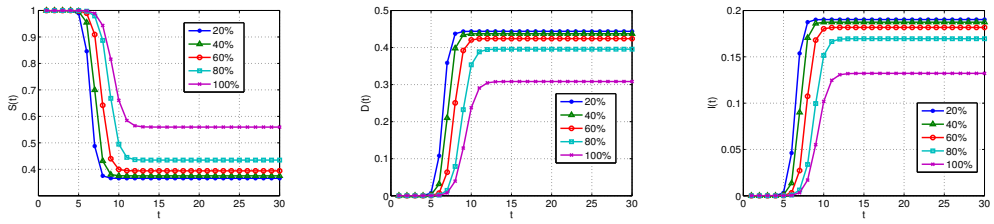
In order to simulate and analyze the influence of special user privacy concerns in social media, we also changed privileged user and pure observer privacy settings by adjusting the links of special nodes in the dataset.



(a) (b) (c)

Figure 7: The density evolution of nodes with different privileged user privacy settings (rigorous privacy settings). (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I.

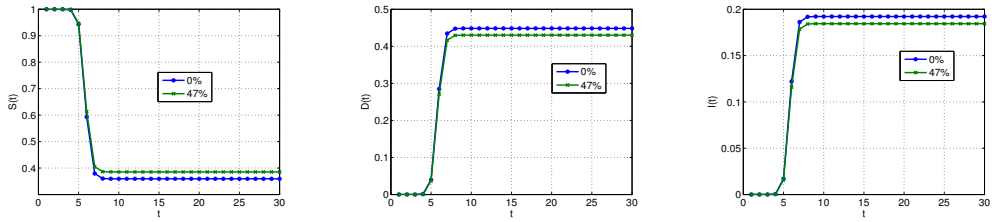
We first located the privileged nodes among the nodes in the LiveJournal dataset and chose 0%, 0.5%, 2%, 6%, and 15% of the nodes of all privileged nodes set their privacy setting to be rigorous and then set P_2 to 0.3. As shown in Figure 7, due to the large out-degree of the nodes, eliminating all of the out-links greatly affected information dissemination. The more privileged users under rigorous privacy settings resulted in slower information dissemination. Moreover, we chose 15% of all privileged nodes from the dataset and assigned them semi-rigorous privacy settings and then increased the out-links of these nodes from 20% to 100% in order to represent different privacy levels. As shown in Figure 8, the stricter privacy settings for privileged users resulted in slower information dissemination. It is clear that privileged users have more



(a) (b) (c)

Figure 8: The density evolution of nodes with different privileged user privacy settings (Semi-rigorous privacy setting). (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I.

power to change the process of information diffusion in social media, which is evident in their being regarded as core entities leading public opinion.



(a) (b) (c)

Figure 9: The density evolution of nodes with different pure observer privacy settings. (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I

Compared with privileged and general users, pure observers have fewer out-links, therefore, we only analyzed the effects of rigorous privacy settings on these users. First, we located pure observers among the nodes in the LiveJournal dataset, chose 0% and 47% of the nodes to be assigned rigorous privacy settings, and then set P_2 to 0.3. As shown in Figure 9, the increase in nodes with rigorous privacy settings did not change node density in state A, and the density evolutions were similar over time, despite the large number of pure observers under rigorous privacy settings. This finding indicates that the dynamics of information dissemination are almost entirely controlled by privileged users and general users in social media. Thus, pure observers with few relationships have almost no effect on the process of information diffusion.

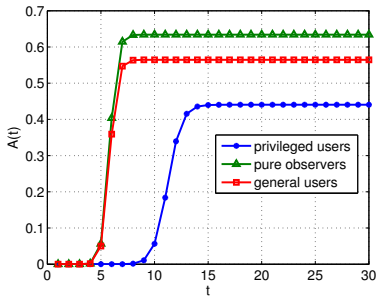


Figure 10: The density evolution of nodes in state A with the same privacy settings.

Finally, we chose 15% of nodes with rigorous privacy settings for all types of users and compared the density evolution of general users, privileged users, and pure observers in state A having the same privacy settings. As shown in Figure 10, privileged users had more profound effects on motivating or restricting information dissemination relative to general users and pure observers. Therefore, fewer privileged users were capable of exerting similar levels of influence, as compared with large numbers of general users. However, pure observers had almost no influence on changing the process of information dissemination in social media.

5. Numerical results of two classic networks

After analysing the effects of privacy concerns on information diffusion in LiveJournal, two classic network topologies (DER graphs [30] and the DSF graph [31]) have been proposed to help evaluate the effects of privacy concerns on information diffusion. This enables us to compare our simulation results involving the LiveJournal dataset. However, considering that the nodes in these generated networks do not involve privileged users and pure observers, we only simulated the effects of different general user rigorous privacy settings.

5.1. Directed ErdosRenyi Random Graph

We first considered a DER network, which is proposed as a classic model. The DER network can be represented as $G(n, p)$ and the graph is constructed by connecting nodes randomly. Each edge is included in the graph with probability, p, independent of every other edge. Equivalently, all graphs

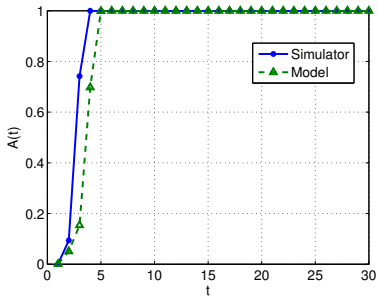


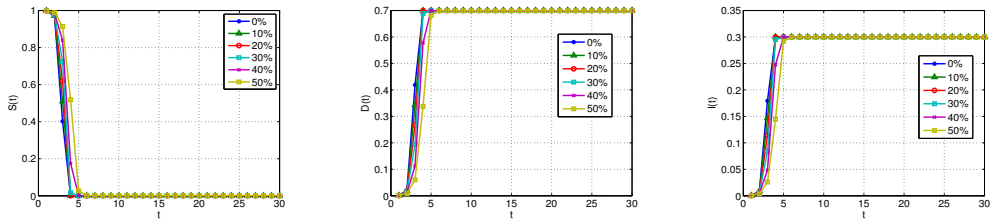
Figure 11: The density evolution of nodes in state A from the real-world simulation and our proposed model.

with n nodes and M edges have equal probability, as described by

$$p^M(1-p)^{\binom{n}{2}-M}. \tag{12}$$

According to the above definition, we can generate DER networks with different p. In our experiment, we used the networkx library [32] to generate 100 DER networks with parameter $p = 0.1$ and 10,000 nodes ($n = 10000$). After running 100 simulations, the average density evolution, $D(t)$, using our proposed model and the real-world simulation is shown in Figure 11. Based on the characteristics of DER networks, information was capable of spreading quickly because of the large average degree, and the results showed that our model accurately described information diffusion in this network. In order to evaluate the effects of different privacy settings on information diffusion in the DER network, we randomly eliminated different proportions of node out-links. To simulate information diffusion under different general user privacy settings, 0%, 10%, 20%, 30%, 40%, and 50% of nodes in the DER network were chosen to have rigorous privacy settings (these node out-links were eliminated). The simulation results are shown in Figure 12.

Our results indicated that different node privacy settings did not alter information diffusion. Furthermore, eliminating different node out-links and in-links did not affect information diffusion in the DER networks. Although p was set to 0.1, each node in the DER network had almost 1,000 neighbours. Therefore, even though 50% nodes did not forward information to their neighbours, the process of information diffusion was not affected.



(a) (b) (c)

Figure 12: The density evolution of nodes with different privacy settings. (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I.

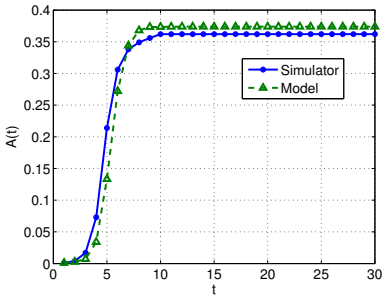


Figure 13: The density evolution of nodes in state A in a directed scale-free network and our proposed model.

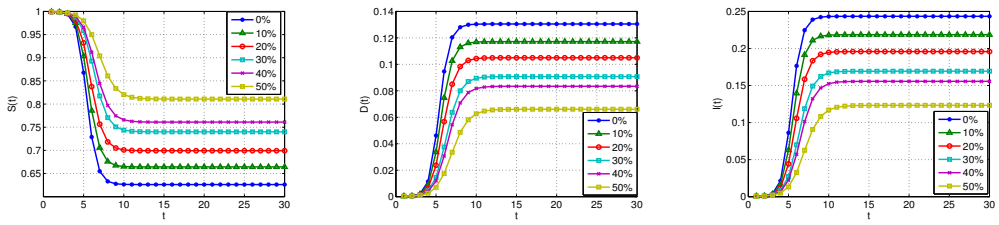
5.2. Directed Scale-free Graph

We then utilized a DSF network as a general classic barabasi-albert model [33], based on the addition of a single edge at discrete time steps in a directed graph. The parameters dictate that the random networks are α, β, γ, σ_{in}, and σ_{out}, with $\alpha + \beta + \gamma = 1$. The simplified rules are described below and detailed descriptions of DSF networks are described in [31].

- With probability α, add a new vertex, v, together with an edge from v to an existing vertex, w, where w is chosen according to $d_{in} + \sigma_{in}$, and where d_{in} is the in-degree of the vertex and σ_{in} is the bias value.
- With probability β, add a new vertex, w, and an edge from an existing vertex, v, to w, where v is chosen according to $d_{out} + \sigma_{out}$, and where d_{out} is the out-degree of the vertex and σ_{out} is the bias value.
- With probability γ, add an edge from an existing vertex, v, to an existing vertex, w, where v and w are chosen independently and determined

according to $d_{out} + \sigma_{out}$ and $d_{in} + \sigma_{in}$, respectively.

According to the above algorithm, we generated 100 DSF networks with parameters $\alpha = 0.3$, $\beta = 0.4$, and $\gamma = 0.3$ using the networkx library [32], with each network having 10,000 nodes. After running the simulation 100 times, the average density evolution, $D(t)$, for our proposed model and the real-world simulation are shown in Figure 13. The results showed that our proposed model accurately described information diffusion in this network. Compared with the validation of the LiveJournal dataset, we found that our model was more accurate in describing standard scale-free networks. Therefore, for non-standard scale-free networks, as represented by the LiveJournal dataset, there existed some errors in our proposed model. However, we believe these errors did not influence our analysis of the effects of privacy concerns.



(a) (b) (c)

Figure 14: The density evolution of nodes with different privacy settings. (A) Nodes in state S; (B) nodes in state D; (C) nodes in state I.

To evaluate the effects of different privacy settings on information diffusion in this network, we choose 0%, 10%, 20%, 30%, 40%, and 50% of general user nodes and assigned them rigorous privacy settings. The simulation results are shown in Figure 14. Compared with the effects of privacy settings on information diffusion using the LiveJournal dataset, the effects in this network were more obvious and significantly changed the process of information dissemination, as well as the final density of nodes in state A. For standard DSF networks, the privacy concerns had more profound effects on information dissemination; however, with limited computational capability, we were only able to generate sparse DSF graphs, which may have altered the evident influences of different privacy settings. In future work, we need to consider more complex situations.

6. Conclusion

In this study, we focused on the process of information dissemination under user privacy concerns in social media. Specifically, we first classified users in social media into general users, privileged users, and pure observers under various privacy concerns, and defined node states and the rules of node transition. Based on an epidemic-spread model, we then presented a general stochastic model with dynamic evolution equations, which accurately described the dynamics of information diffusion in a real-world network. Finally, we simulated and analyzed the effects of general user, privileged user, and pure observer privacy settings on information dissemination in social media. Statistical analysis of results showed that different privacy settings for general users impacted the process of information dissemination, while privileged users with more rigorous privacy settings had greater effects on slowing the speed of information diffusion relative to those general users. Additionally, pure observer privacy settings had almost no influence on information dissemination, even under the most rigorous privacy settings. Finally, we analyzed the effects of privacy settings using two classic network models. In future work, we will analyze the impact of privacy settings on information diffusion from the perspective of single individuals.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [No. 61303218, U1401251]; the Fundamental Research Foundations for the Central Universities of China [K5051301017] and the 111 Project [B08038].

7. Reference

[1] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, B. S. Silvestre, Social media? get serious! understanding the functional building blocks of social media, Business horizons 54 (3) (2011) 241–251.

[2] Social media:a security challenge and opportunity, `http://www.computerweekly.com/feature/Social-media-a-security-challenge-and-opportunity`, accessed: 2015-01-24.

[3] J. Borge-Holthoefer, A. Rivero, Y. Moreno, Locating privileged spreaders on an online social network, Physical Review E 85 (6) (2012) 066123.

[4] G. F. de Arruda, A. L. Barbieri, P. M. Rodriguez, F. A. Rodrigues, Y. Moreno, L. d. F. Costa, Role of centrality for the identification of influential spreaders in complex networks, Physical Review E 90 (17) (2014) 032812.

[5] S. Nizamani, N. Memon, S. Galam, From public outrage to the burst of public violence: An epidemic-like model, Physica A: Statistical Mechanics and its Applications 416 (0) (2014) 620–630.

[6] P. Van Mieghem, J. Omic, R. Kooij, Virus spread in networks, Networking, IEEE/ACM Transactions on 17 (1) (2009) 1–14.

[7] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, Y. Moreno, Discrete-time markov chain approach to contact-based disease spreading in complex networks, EPL (Europhysics Letters) 89 (3) (2010) 38009.

[8] J. Conlisk, Interactive markov chains, Journal of Mathematical Sociology 4 (2) (1976) 157–185.

[9] M. Nekovee, Y. Moreno, G. Bianconi, M. Marsili, Theory of rumour spreading in complex social networks, Physica A: Statistical Mechanics and its Applications 374 (1) (2007) 457–470.

[10] M. Boguñá, C. Castellano, R. Pastor-Satorras, Langevin approach for the dynamics of the contact process on annealed scale-free networks, Physical Review E 79 (3) (2009) 036110.

[11] H. S. Wilf, generatingfunctionology, Elsevier, 2013.

[12] R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Resilience of the internet to random breakdowns, Physical review letters 85 (21) (2000) 4626.

[13] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Physical Review Letters 86 (0) (2001) 3200–3203.

[14] A. Vespignani, Modelling dynamical processes in complex socio-technical systems, Nature Physics 8 (1) (2012) 32–39.

[15] C. Li, R. van de Bovenkamp, P. Van Mieghem, Susceptible-infected-susceptible model: A comparison of n-intertwined and heterogeneous mean-field approximations, Physical Review E 86 (2) (2012) 026116.

[16] S. C. Ferreira, C. Castellano, R. Pastor-Satorras, Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results, Physical Review E 86 (4) (2012) 041125.

[17] L. Zhao, J. Wang, Y. Chen, Q. Wang, J. Cheng, H. Cui, Sihr rumor spreading model in social networks, Physica A: Statistical Mechanics and its Applications 391 (7) (2012) 2444–2453.

[18] J. Wang, L. Zhao, R. Huang, Siraru rumor spreading model in complex networks, Physica A: Statistical Mechanics and its Applications 398 (2014) 43–55.

[19] J. Yang, J. Leskovec, Modeling information diffusion in implicit networks, in: Data Mining (ICDM), 2010 IEEE 10th International Conference on, IEEE, 2010, pp. 599–608.

[20] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A: Statistical Mechanics and its Applications 390 (6) (2011) 1150–1170.

[21] P.-Y. Chen, K.-C. Chen, Optimal control of epidemic information dissemination in mobile ad hoc networks, in: Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE, IEEE, 2011, pp. 1–5.

[22] I. Tunc, L. B. Shaw, Effects of community structure on epidemic spread in an adaptive network, Physical Review E 90 (8) (2014) 022801.

[23] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, no. 14 in IMC '07, ACM, 2007, pp. 29–42.

[24] F. Benevenuto, T. Rodrigues, V. A. F. Almeida, J. M. Almeida, M. A. Gonçalves, Detecting spammers and content promoters in online video social networks, in: Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009, 2009, pp. 620–627. doi:10.1145/1571941.1572047.
URL http://doi.acm.org/10.1145/1571941.1572047

[25] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media?, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 591–600. doi:10.1145/1772690.1772751.
URL http://doi.acm.org/10.1145/1772690.1772751

[26] L. Jin, Y. Chen, T. Wang, P. Hui, A. V. Vasilakos, Understanding user behavior in online social networks: a survey, IEEE Communications Magazine 51 (9). doi:10.1109/MCOM.2013.6588663.
URL http://dx.doi.org/10.1109/MCOM.2013.6588663

[27] A. Vazquez, M. Weigt, Computational complexity arising from degree correlations in networks, Physical Review E 67 (2) (2003) 027101.

[28] J. Leskovec, A. Krevl, SNAP Datasets: Stanford large network dataset collection, http://snap.stanford.edu/data (Jun. 2014).

[29] D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks, nature 393 (6684) (1998) 440–442.

[30] P. ERDdS, A. R&WI, On random graphs i, Publ. Math. Debrecen 6 (1959) 290–297.

[31] B. Bollobás, C. Borgs, J. Chayes, O. Riordan, Directed scale-free graphs, in: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2003, pp. 132–139.

[32] A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring network structure, dynamics, and function using NetworkX, in: Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA USA, 2008, pp. 11–15.

[33] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, science 286 (5439) (1999) 509–512.