Advance Access publication on 20 July 2014
doi:10.1093/comjnl/bxu062

Information Diffusion Model Based on Privacy Setting in Online Social Networking Services

HUI ZHU*, CHENG HUANG AND HUI LI

State Key laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an, China
**Corresponding author: xdzhuhui@gmail.com*

Social networking service (SNS) is one of the major technological applications based on Web 2.0, which can help users to express their views and share information with others. How to describe the information diffusion process in online SNS accurately has attracted considerable interest recently. However, almost all existing models focus on a single online SNS tool and face many challenges with multiple online SNS tools. In this paper, we turn to users' privacy setting policies and propose a general stochastic model with multiple diffusion mechanisms in online SNS, called DMPS. Specifically, we first define a privacy protection mechanism based on information sharing in online SNS and classify nodes according to different privacy setting policies; then, we define the states of the nodes and information dissemination rules with dynamics of infectious disease; finally, we describe the evolution process of different nodes by dynamic evolution equations. Detailed simulations and analysis show that the DMPS can precisely describe the diffusion process with multiple diffusion mechanisms and have the same characteristics as a diffusion process in real online SNS. As a result, DMPS can be used to identify the underlying diffusion mechanism of information and forecast its trend in online SNS.

Keywords: information diffusion; social networking service; privacy setting policies; dynamics of infectious disease; multiple diffusion mechanisms

Received 12 September 2013; revised 13 May 2014
Handling editor: Zhiyong Zhang

1. INTRODUCTION

Online social networking service (SNS) is one of the online platforms with hundreds of millions of registered users, which can be described as a graph representing the interactions or interconnections among individuals based on their common interests, activities and demographic identities [1]. With the pervasiveness of online SNS, such as blogs, Facebook and Twitter, users can express their views and share information with others without geographical restrictions [2], and the information diffusion process in online SNS has attracted considerable interest recently [3]. Tremendous efforts have been made in the areas of computer science and physics [4], and it is widely accepted that user-to-user exchanges can speed up the information spread process through the network [5]. One basic task in researching the information diffusion model is to identify the underlying mechanism of information diffusion and forecast its trend in online SNS [6]. However, most of existing information diffusion models focus on a signal online SNS tool or a single diffusion mechanism (i.e. different models are proposed to describe the information diffusion in different online SNS tools, respectively). Obviously, Internet users may use more than one online SNS tool, and each online SNS tool may have its own diffusion mechanism. For example, if someone posts information in WeChat, then only his/her friends in WeChat can read the information, but the information may be read by anyone if it is posted on Twitter. Therefore, almost all existing models cannot describe information diffusion in a real network accurately.

To describe the process of information diffusion with multiple online SNS tools, we take into account the users' privacy setting policies. In fact, a user usually disseminates with similar privacy setting polices using all of his/her online SNS tools for the same information (i.e. no matter what kind of online SNS tools are chosen to post the information, the

information diffusion mechanism is determined by the user's privacy setting polices for this information). Therefore, we propose an information diffusion model in online SNS based on users' privacy setting policies, called DMPS, which can describe the information diffusion process accurately in online SNS with multiple diffusion mechanisms. Specifically, the users are classified according to users' different privacy setting policies and a new architecture of online SNS is presented based on different features of information sharing in online SNS; then, the users' states and the information diffusion rules are defined according to dynamics of infectious disease; finally, the process of information diffusion is described by dynamic evolution equations. The main contributions of this paper include:

(i) *Representation of online SNS with users' privacy setting policies is constructed*. For any information in online SNS, various users may have different privacy setting policies, but each user usually has the similar privacy setting policies when using his/her multiple online SNS tools for the same information. Therefore, to describe individuals and information exchanges in online SNS with multiple diffusion mechanisms, we define different types of users and relationships among users based on users' privacy setting policies and features of information sharing in online SNS. The proposed classification and information dissemination framework can represent real online SNS faithfully.

(ii) *The process of information diffusion with multiple diffusion mechanisms is described*. Based on dynamics of infectious disease algorithm, we classify users' states into four final states and define the information dissemination rules. Then, the information diffusion model for multiple diffusion mechanisms is described by dynamic evolution equations.

(iii) *Information diffusion in DMPS is simulated and analyzed in detail with real data*. The results show that the process of information diffusion in DMPS and agree with the real data; the information can spread more easily without privacy setting of online SNS tools; the fewer privacy setting policies of the source, the faster information spreads; the processes of information diffusion have similar trends with different privacy setting policies. Besides, other factors which impact the information diffusion are also analyzed.

The rest of this paper is organized as follows. In Section 2, we introduce the requirements of information diffusion model in online SNS, and identify our design goal. Then, we formalize the online SNS with privacy setting polices in Section 3, and present our DMPS model in Section 4, followed by the simulation and analysis with real data in Section 5. We also discuss the related works in Section 6. Finally, we draw our conclusions in Section 7.

2. REQUIREMENTS AND DESIGN GOAL

In this section, we formalize requirements of the information diffusion model in online SNS, and identify our design goal as well.

2.1. Requirements

The information diffusion model is constructed to identify the underlying mechanism of information diffusion and forecast its trend. In real online SNS, more than one online SNS tool with various information diffusion mechanisms are applied by users. To diminish the impact of multiple diffusion mechanisms, we focus on users' privacy setting polices to analyze the information diffusion in online SNS. Therefore, in order to describe the process of information diffusion accurately and design the model as simple as possible, the following requirements should be satisfied:

(i) *The classification method for users should be simple and efficient*. Considering various online SNS tools, different classification methods of users are provided to manage the information sharing, which increase the complexity of online SNS representation. Therefore, a simple and efficient classification method is required.

(ii) *The representation of online SNS should be consistent with real online SNS environment*. The information diffusion model will not describe the process of information diffusion accurately, if the representation fails to reflect the features of multiple diffusion mechanisms.

(iii) *The impact factors of information diffusion should be considered comprehensively*. In online SNS, not all information has the same diffusion process, and the factors which influence information diffusion, such as users' habits and interest, should be considered comprehensively. For example, most of users just view the information pushed from online SNS tools, and others usually browse the information actively in public website.

2.2. Design goal

Under the aforementioned requirements, our design goal is to develop an information diffusion model with multiple diffusion mechanisms for online SNS. Specifically, the following three objectives should be achieved:

(i) *Reducing the complexity of representation of online SNS*. We will define the simple and efficient classification method for users and the relationships among users, and describe the representation of online SNS in a more compact form.

(ii) *Reflecting the process of information diffusion in online SNS accurately*. Even if users post information

using different online SNS tools, such as Twitter and Facebook, the underlying mechanism of this information diffusion should be described accurately.

(iii) *Analyzing the reasons of information diffusion mechanism*. Focusing on the key factors of the proposed model, such as users' privacy setting policies, habits and interest, we will simulate and analyze their effects on information diffusion in online SNS.

3. REPRESENTATION OF ONLINE SNS WITH PRIVACY SETTING

In this section, we propose an abstract representation of online SNS with users' privacy setting policies, which is described as a graph representing the interactions or interconnections.

For any users in online SNS, all of his/her information can be classified into three categories: *no privacy protection mechanism*, *half privacy protection mechanism* and *rigorous privacy protection mechanism*. The information with *no privacy protection mechanism* can be accessed by everyone, the information with *half privacy protection mechanism* can be accessed by his/her friends and the information with *rigorous privacy protection mechanism* can only be accessed by himself/herself. From the perspective of users, we can construct the graph of online SNS by the following nodes and relationships:

(i) N, the set of nodes, is represented as the collection of users' accounts. Each user in N has a unique account $N_i \in N$. Based on the above discussions, all nodes can be classified into three categories according to their privacy setting policies for the current information:

(a) N_{pub}, the public node, denotes that the user does not set any privacy-preserving policies for the information, and any users can access this information from his/her online SNS page. That is, this information in N_{pub} has *no privacy protection*.

(b) N_{pri}, the private node, denotes that the privacy-preserving policies are set by user for this information, and only the specified users can access this information. That is, this information in N_{pri} has *half privacy protection*.

(c) N_{r}, the exclusive node, denotes that the rigorous privacy-preserving policies are set by user, and others cannot access this information. That is, this information in N_{r} has *rigorous privacy protection*.

(ii) R, the set of relationships among users in online SNS, can also be classified into three categories according to the category of nodes for the current information:

(a) R_{nbr}, the neighbor relationship, represents that users can access each other bidirectionally for this information.

(b) $R_{\text{s-nbr}}$, the semi-neighbor relationship, represents that users can access this information directionally. For example, if user A has relationship $R_{\text{s-nbr}}$ with user B, then, user A can access this information in user B, but user B cannot access this information in user A.

(c) $R_{\text{r-nbr}}$, the non-neighbor relationship, represents that two users cannot access this information of each other, even if they are physically connected.

As shown in Fig. 1, each node in N_r has relationships $R_{\text{r-nbr}}$ or $R_{\text{s-nbr}}$ with other nodes; each node in N_{pub} only has relationship R_{nbr} with other nodes and each node in N_{pri} may have relationships R_{nbr}, $R_{\text{r-nbr}}$ or $R_{\text{s-nbr}}$ with other nodes. Therefore, we can get the following conclusions:

(i) Each node in N_{r} must have relationship $R_{\text{r-nbr}}$ with other nodes in N_{r} and may have relationship $R_{\text{s-nbr}}$ with nodes in N_{pub} and N_{pri}. Specifically, nodes in N_{r} can get information from other nodes, except nodes in N_{r}, but deny any requests of accessing their own information from other nodes.

(ii) Each node in N_{pub} has relationship R_{nbr} with other nodes in N_{pub} and may have relationship R_{nbr} with nodes in N_{pri}. Specifically, nodes in N_{pub} can get information from other nodes, except nodes in N_{r} and allow other nodes which have relationships R_{nbr} or $R_{\text{s-nbr}}$ to access their information.

(iii) Nodes in N_{pri} may have relationships R_{nbr} or $R_{\text{s-nbr}}$ with nodes in N_{pub} or N_{pri}, and can get information from these nodes. Besides, the nodes allow other nodes which have relationships R_{nbr} or $R_{\text{s-nbr}}$ to access their information.

DEFINITION 1. *For each node in social networks, it gets information from its neighbors and semi-neighbors, which can be defined as*

$$N_i = (i, S_{i\text{-}s\text{-}nbr}, S_{i\text{-}nbr}).$$

where i is the name of node N_i, $S_{i\text{-}s\text{-}nbr}$ is the set of all semi-neighbors of N_i and $S_{i\text{-}nbr}$ is the set of all neighbors of N_i.

Inference 1. Let D_{ni} be the degree of N_i, which means the total number of nodes in $S_{i\text{-nbr}}$ and $S_{i\text{-s-nbr}}$. That is, N_i has D_{ni} nodes as the potential source of information.

Inference 2. Any online SNS, $S = (N_1, N_2, \ldots, N_n)$, can be described as a directed graph in Fig. 1, whose nodes and edges are used to represent users in online SNS and relationships between users, respectively.

4. PROPOSED DMPS MODEL

In this section, we propose an information diffusion model with multiple diffusion mechanisms for online SNS, which mainly

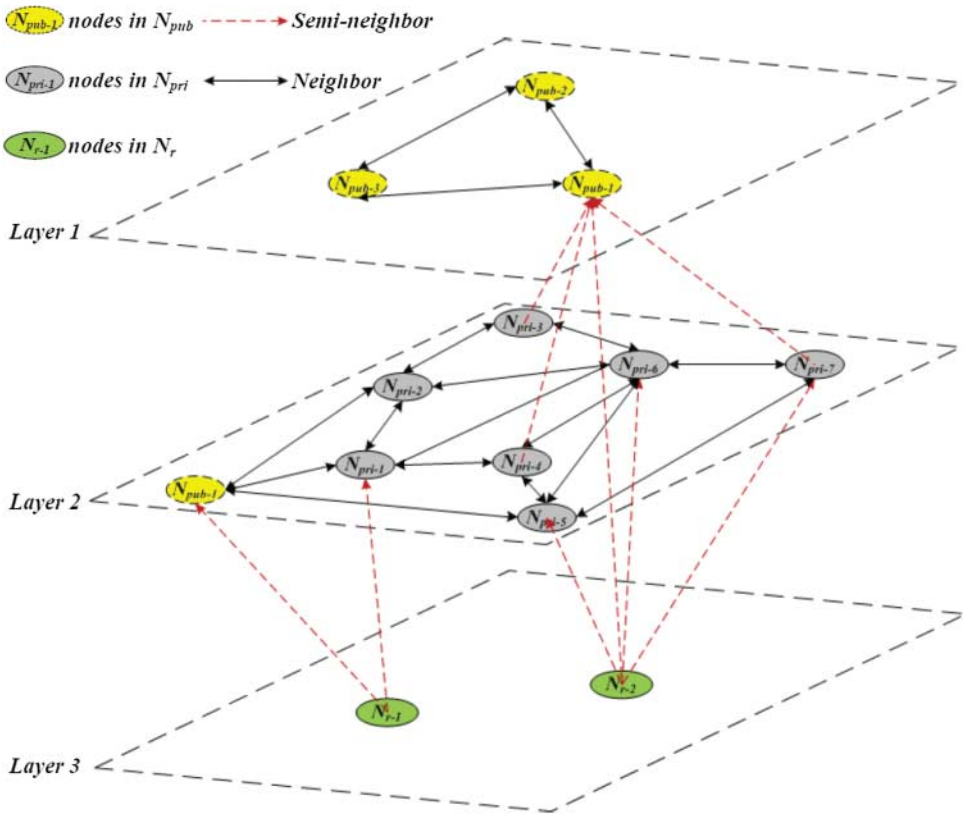

FIGURE 1. Representation of online SNS with privacy setting.

consists of the following three parts: states of nodes, rules of state transition and dissemination model.

4.1. States of nodes

According to the definitions of nodes in Section 3 and dynamics of infectious disease, the nodes' states in online SNS can be classified into two major categories: *susceptible state* and *accepted state*.

(i) *Susceptible state* means that the node can receive information from their neighbors, i.e. the node in *susceptible state* may be infected by the information. And all of nodes' states begin with *susceptible state* except the source node of information.

(ii) *Accepted state* means that the node has received and accepted the information from its neighbors or semi-neighbors. Accepted state can also be divided into two categories: *immunized state* and *disseminated state*.

(a) *Immunized state* represents that the node does not disseminate the information after it accepted information, i.e. other users cannot access the information from his/her online SNS page. Nodes with *immunized state* are corresponding to nodes in N_{r}.

(b) *Disseminated state* represents that the node will disseminate the information after accepting information, and disseminated state can also be divided into two sub-categories: *public disseminated state* and *privacy disseminated state*.

(1) *Public disseminated state* means that the node can disseminate the information to all nodes, i.e. all users can access the information from his/her online SNS page. Nodes with *public disseminated state* are corresponding to nodes in N_{pub}.

(2) *Privacy disseminated state* means that the node disseminates the information to special nodes, i.e. nodes which have the relationships R_{nbr} or $R_{\mathrm{s\text{-}nbr}}$ can access the information. Nodes with *privacy disseminated state* are corresponding to nodes in N_{pri}.

DEFINITION 2. *In the diffusion process of information, all nodes in online SNS have four final states* (*susceptible state*, *immunized state*, *public disseminated state and privacy disseminated state*) *and three temporary states* (*received state*, *accepted state and disseminated state*).

(i) *For any information, the node's state must be one of the final states at any time.*
(ii) *The temporary states are various collections of different final states (except susceptible state), which can help us to analyze the process of information diffusion.*

4.2. Rules of state transition

In online SNS, the state transition of each node will be impacted by its current state and the states of its neighbors and semi-neighbors. According to the dynamics of infectious disease, we define the state transition rules of nodes as follows:

(i) When a node with *disseminated state* (*public disseminated state* or *privacy disseminated state*) is accessed by a susceptible node, the susceptible node can receive the information and transfer to *received state*. Then, the node's state may transfer to *accepted state* by probability P_1. Obviously, P_1 is influenced by the features of the information, information credibility etc.
(ii) When a node with *susceptible state* accepts the information and transfers to *accepted state*, the node will disseminate the information and transfer to *disseminated state* by probability P_2.
(iii) A node with *disseminated state* will allow all nodes to access the information and transfers to *public disseminated state* by probability P_3.

Figure 2 shows the node's state transition process when information is disseminated. In the beginning, each node except the source node is in *susceptible state*. Then, we can calculate the probabilities of nodes' different final states when they received the information, which are shown in Table 1. Finally, as shown in Table 2, notations are defined to represent the numbers of nodes in different states, and the relationship of their numbers can be described as follows:

$$\begin{aligned} N(k,t) &= S(k,t) + A(k,t) \\ &= S(k,t) + D(k,t) + I(k,t) \\ &= S(k,t) + D_{\text{pri}}(k,t) + D_{\text{pub}}(k,t) + I(k,t). \end{aligned}$$

4.3. Dissemination model

In this section, we construct the dynamic evolution equations for information diffusion in online SNS.

Let N_x be a susceptible node in online SNS with n nodes at time t, and the probability that N_x is still in *susceptible state* at time $t + \Delta t$ can be calculated as follows:

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

TABLE 1. Probabilities of final states.

Final state	Probability
Susceptible state	$1 - P_1$
Immunized state	$P_1(1 - P_2)$
Public disseminated state	$P_1 P_2 P_3$
Privacy disseminated state	$P_1 P_2 (1 - P_3)$

TABLE 2. Notations of different nodes' number.

Notation	Description
$N(k,t)$	The number of k-degree nodes in time t
$S(k,t)$	The number of susceptible nodes in $N(k,t)$
$A(k,t)$	The number of accepted nodes in $N(k,t)$
$D(k,t)$	The number of disseminated nodes in $N(k,t)$
$D_{pri}(k,t)$	The number of privacy disseminated nodes in $N(k,t)$
$D_{pub}(k,t)$	The number of public disseminated nodes in $N(k,t)$
$I(k,t)$	The number of immunized nodes in $N(k,t)$

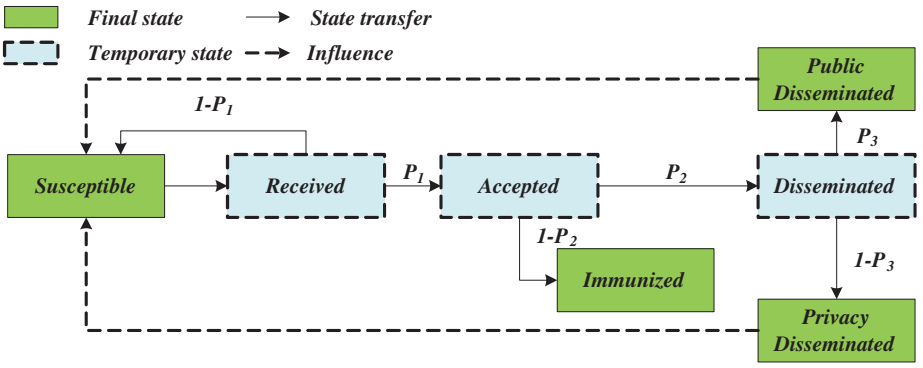

FIGURE 2. The procedure of nodes' state transition.

where $j = j(t)$ is the number of disseminated nodes which are N_x's neighbors or semi-neighbors at time t. Furthermore, if N_x has k neighbors or semi-neighbors, then j will satisfy Equation (1).

$$\prod(j,t) = \binom{k}{j}(P(k,t))^j(1-P(k,t))^{k-\mathrm{j}}. \tag{1}$$

Specifically, $P(k,t)$ is the probability that a disseminated node can connect a susceptible node with k neighbors or semi-neighbors. Let c be the average times that a user actively access the nodes which are not his/her neighbors or semi-neighbors, K' be the number of existing degrees in online SNS, and define notations in Table 3. Then, we can compute $P(k,t)$ as Equation (2).

$$\begin{aligned}P(k,t) &= \sum_{k'} P(k'|k)P(d_{k'}|s_k)\\ &= \sum_{k'}\left(P(k'|k)D_d(k',t) + \frac{cD_{pub}(k',t)P(k')}{K'}\right.\\ &\quad\left.\times\left(1-\frac{k+1}{n}\right)\right)\\ &= \sum_{k'}\left(P(k'|k)D_a(k',t)P_2 + \frac{c}{K'}D_a(k',t)P_2P_3P(k')\right.\\ &\quad\left.\times\left(1-\frac{k+1}{n}\right)\right)\\ &= P_2\sum_{k'}\left(P(k'|k) + \frac{c}{K'}P_3P(k')\left(1-\frac{k+1}{n}\right)\right)\\ &\quad\times D_a(k',t).\end{aligned} \tag{2}$$

Based on the conclusions in [7], we can obtain

$$P(k'|k) = \frac{k'P(k')}{\bar{k}},$$

TABLE 3. Notations of Equation (2).

Notation	Description
$P(k'\|k)$	The probability that a k-degree node is the neighbor of a k'-degree node
$P(k')$	The probability that the degree of node is k'
$P(d_{k'}\|s_k)$	The probability that a k'-degree node is in disseminated state when it connects a k-degree susceptible node
$D_d(k',t)$	The disseminated nodes' density in the k'-degree nodes at time t
$D_{pub}(k',t)$	The public disseminated nodes' density in the k'-degree nodes at time t
$D_a(k',t)$	The accepted nodes' density in the k'-degree nodes at time t

where $\bar{k}$ is the average degree of the network. Then, we can describe $P(k'|k)$ as follows:

$$\begin{aligned}P(k,t) &= P_2\sum_{k'}\left(\frac{k'P(k')}{\bar{k}} + \frac{c}{K'}P_3P(k')\left(1-\frac{k+1}{n}\right)\right)\\ &\quad\times D_a(k',t)\\ &= P_2\sum_{k'}\left(\frac{k'}{\bar{k}} + \frac{c}{K'}P_3\left(1-\frac{k+1}{n}\right)\right)P(k')D_a(k',t).\end{aligned} \tag{3}$$

The average probability $\bar{P}_{s\to s}(k,t)$ that the k-degree node is still in susceptible state at time $t+\Delta t$ can be calculated as follows:

$$\begin{aligned}\bar{P}_{s\to s}(k,t) &= \sum_{j=0}^{k}\binom{k}{j}(1-P_1\Delta t)^jP(k,t)^j(1-P(k,t))^{k-j}\\ &= (1-P_1\Delta tP(k,t))^k.\end{aligned} \tag{4}$$

Then, substituting the value of $P(k,t)$ into Equation (4), we can calculate

$$\begin{aligned}\bar{P}_{s\to s}(k,t) &= \left(1-P_1P_2\Delta t\sum_{k'}\left(\frac{k'}{\bar{k}} + \frac{c}{K'}P_3\left(1-\frac{k+1}{n}\right)\right)\right.\\ &\quad\left.\times P(k')D_a(k',t)\right)^k.\end{aligned} \tag{5}$$

The average probability $\bar{P}_{s\to a}(k,t)$ that the k-degree node is transferred from susceptible state to accepted state during the time $[t,t+\Delta t]$ can be obtained as follows:

$$\begin{aligned}\bar{P}_{s\to a}(k,t) &= 1-\left(1-P_1P_2\Delta t\sum_{k'}\left(\frac{k'}{\bar{k}} + \frac{c}{K'}P_3\right.\right.\\ &\quad\left.\left.\times\left(1-\frac{k+1}{n}\right)\right)P(k')D_a(k',t)\right)^k.\end{aligned} \tag{6}$$

Therefore, the change of k-degree accepted nodes' numbers during the time $[t,t+\Delta t]$ can be described as follows:

$$\begin{aligned}A(k,t+\Delta t) &= A(k,t) + S(k,t)\bar{P}_{s\to a}(k,t)\\ &= A(k,t) + (N(k,t) - A(k,t))\\ &\quad\times\left(1-\left(1-P_1P_2\Delta t\sum_{k'}\left(\frac{k'}{\bar{k}}\right.\right.\right.\\ &\quad\left.\left.\left.+\ \frac{c}{K'}P_3\left(1-\frac{k+1}{n}\right)\right)P(k')D_a(k',t)\right)^k\right),\end{aligned} \tag{7}$$

$$\begin{aligned}&\frac{A(k,t+\Delta t)-A(k,t)}{N(k,t)\Delta t}\\ &= \frac{N(k,t)-A(k,t)}{N(k,t)\Delta t}\left(1-\left(1-P_1P_2\Delta t\sum_{k'}\left(\frac{k'}{\bar{k}}\right.\right.\right.\\ &\quad\left.\left.\left.+\ \frac{c}{K'}P_3\left(1-\frac{k+1}{n}\right)\right)P(k')D_a(k',t)\right)^k\right).\end{aligned} \tag{8}$$

When $\Delta t \to 0$, we can obtain $\partial D_a(k,t)/\partial t$ as Equation (9) to describe the change of accepted nodes' density based on Taylor expansions formula.

$$\frac{\partial D_a(k,t)}{\partial t} = kP_1P_2(1 - D_a(k,t)) \times \sum_{k'} \left(\frac{k'}{\bar{k}} + \frac{c}{K'} P_3 \left(1 - \frac{k+1}{n} \right) \right) \times P(k') D_a(k',t). \quad (9)$$

In the same way, the formulas which describe the density's changes of susceptible nodes, disseminated nodes and immunized nodes with time t can be obtained as follows:

$$\frac{\partial D_{pub}(k,t)}{\partial t} = kP_1P_2P_2P_3(1 - D_a(k,t)) \times \sum_{k'} \left(\frac{k'}{\bar{k}} + \frac{c}{K'} P_3 \left(1 - \frac{k+1}{n} \right) \right) \times P(k') D_a(k',t). \quad (10)$$

$$\frac{\partial D_{pri}(k,t)}{\partial t} = kP_1P_2P_2(1 - P_3)(1 - D_a(k,t)) \times \sum_{k'} \left(\frac{k'}{\bar{k}} + \frac{c}{K'} P_3 \left(1 - \frac{k+1}{n} \right) \right) \times P(k') D_a(k',t). \quad (11)$$

$$\frac{\partial D_i(k,t)}{\partial t} = kP_1P_2(1 - P_2)(1 - D_a(k,t)) \times \sum_{k'} \left(\frac{k'}{\bar{k}} + \frac{c}{K'} P_3 \left(1 - \frac{k+1}{n} \right) \right) \times P(k') D_a(k',t). \quad (12)$$

5. SIMULATIONS AND ANALYSIS

In this section, we validate our proposed DMPS through experiments on real data set of online SNS, then, simulate and analyze the influences of the credibility of information, users' privacy setting polices and habits to describe the different situation of information diffusion in online SNS.

5.1. Effectiveness of DMPS

To verify the effectiveness of our proposed DMPS, we select a real Twitter data set [8], which has 81 306 nodes and 1 768 149 edges, its average degree is 29.77 and initial degree distribution is shown in Fig. 3. Then, we simulate the real online SNS environment to compare with DMPS model as follows:

(i) First, we classify all nodes of the data set into immunized nodes, privacy disseminated nodes and public disseminated nodes in random; secondly, we select a node as the source node of a message, and record the evolutions of various nodes' numbers which accepted the message, then this process is repeated 10 times to calculate the average result of the records; finally, the process of information diffusion for this message in the Twitter data with privacy setting policies can be described by the average result.

(ii) Based on the numbers of various nodes in the simulation above, we can calculate the corresponding parameters, such as P_1, P_2, P_3 and c; then, we randomly select a node with the average degree as the source node of information, and use DPMS to simulate the processes of information diffusion to calculate the densities of public disseminated nodes, privacy disseminated nodes and immunized nodes; finally, the corresponding process of information diffusion in our proposed DMPS can be illustrated.

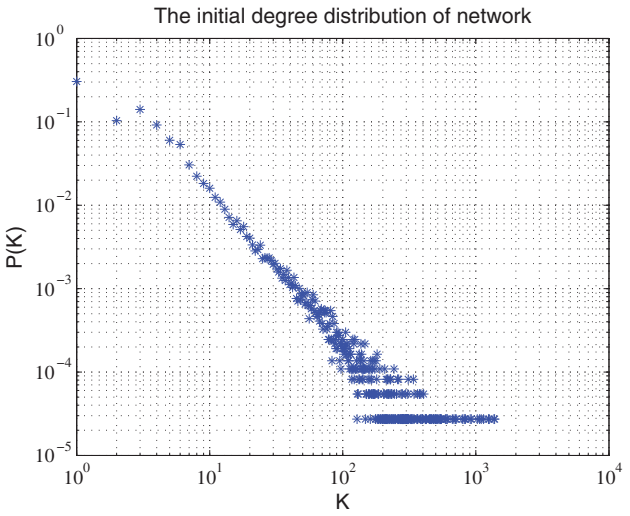

FIGURE 3. Initial degree distribution of data set.

Eventually, the density's evolutions of susceptible nodes, immunized nodes, privacy disseminated nodes and public disseminated nodes in Twitter with users' privacy setting policies and DMPS are illustrated in Fig. 4. Specifically, the numbers of immunized nodes, privacy disseminated nodes and public disseminated nodes of simulation are 65 044, 14 636 and 1626; the corresponding parameters in DMPS are $P_1 = 0.2$, $P_2 = 0.2$, $P_3 = 0.1$ and $c = 10$; in addition, x-axis represents the number of iterations t and y-axis represents the nodes' density D.

Obviously, the density's evolutions of various nodes are similar between the real Twitter data set and DMPS, and the density of various nodes will reach equilibrium with time t increasing. We can find the experiments' variance between real Twitter data set and DMPS is <5% according to Fig. 4. In specific, the density's evolution of Twitter is slightly faster than DPMS, since most disseminated nodes in Twitter data set have a large number of neighbors or semi-neighbors, while

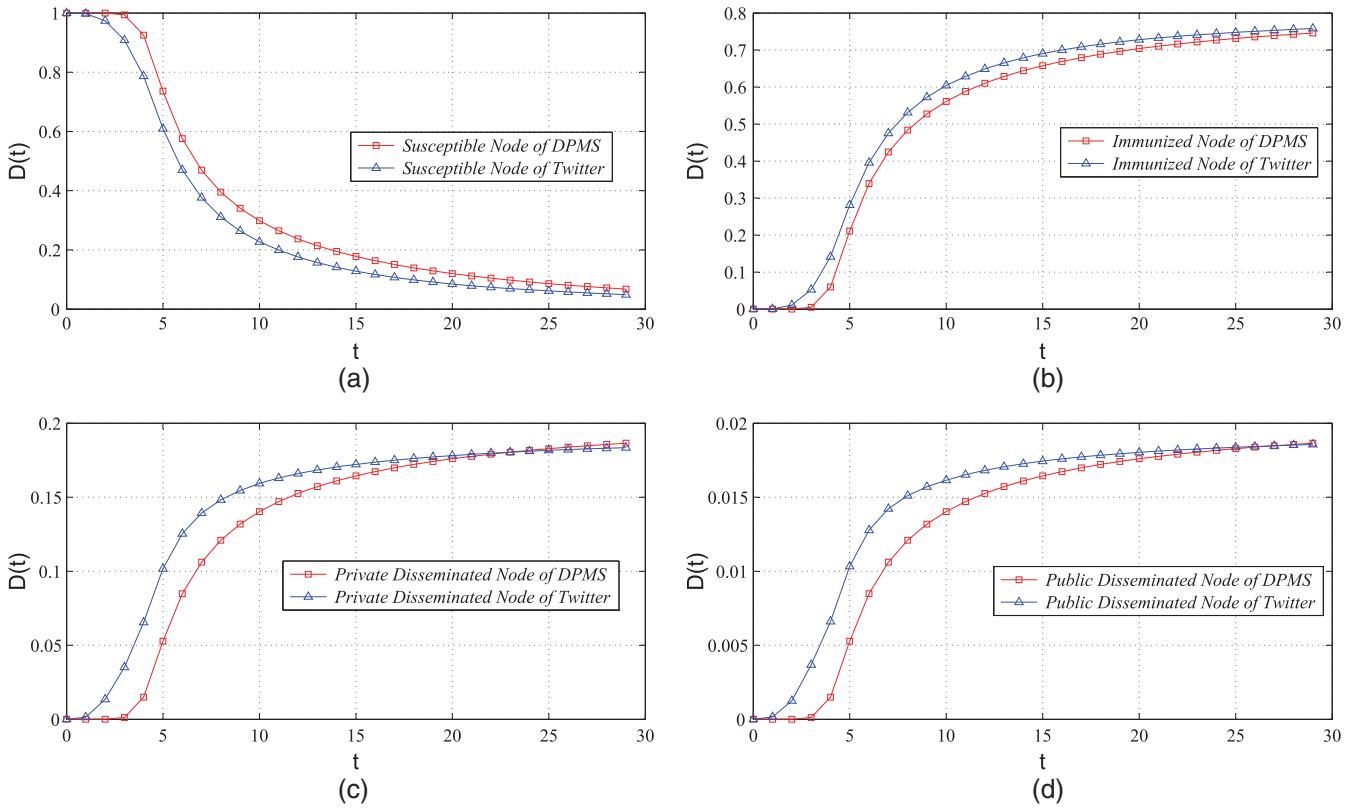

FIGURE 4. The density's evolution in Twitter and DMPS. (**a**) Susceptible nodes; (**b**) immunized nodes; (**c**) privacy disseminated nodes and (**d**) public disseminated nodes.

the disseminated nodes in DMPS have the average degree distribution in online SNS. Therefore, our proposed DMPS agrees with the characteristics of information dissemination in real online SNS [9].

5.2. Simulation and discussion

In this section, we concentrate on analyzing the impacts of various parameters' settings on the density's evolutions of various nodes in DMPS, such as privacy setting policies of nodes, the credibility of information and users' habits. We set the initial parameters of DMPS as follows:

- There is only one source node of average degree with information, and all other nodes are in susceptible state without information.
- The total number of iterations is 300.

5.2.1. Privacy setting policies of source node

According to the definition of k_0, different settings of k_0 can be used to evaluate the influence of the source node's privacy setting policies in DMPS. We set $P_1 = 0.2$, $P_2 = 0.2$, $P_3 = 0.1$, $c = 10$ and the total number of iteration $T = 300$, which is the round number of information dissemination, as initial parameters. Various values of k_0 (10, 100, 500, 1000 and 2000) are chosen to simulate the density's evolutions with t for susceptible nodes, immunized nodes, privacy disseminated nodes and public disseminated nodes. The density's evolutions of various nodes with different k_0 are shown in Fig. 5. We get the following results:

(i) The larger k_0 is, the faster the density of susceptible nodes decreases.
(ii) The densities of public disseminated nodes, privacy disseminated nodes and immunized nodes increase more quickly when k_0 increases from 10 to 2000.
(iii) The final densities of nodes turn to the same equilibrium in the end of simulation, even if k_0 is different.

The results illustrate that the larger source node's degree is, the faster information spreads in online SNS. However, when t tends to infinity, regardless of the source node's initial degree, the final densities of different settings of k_0 turn to the same equilibrium. This result is due to the high connectivity in online SNS, and the source node with high degree can obviously accelerate the speed of information diffusion in online SNS. At the same time, although the nodes with high degree can accelerate the dissemination of information, they can also block the dissemination of information effectively if they become

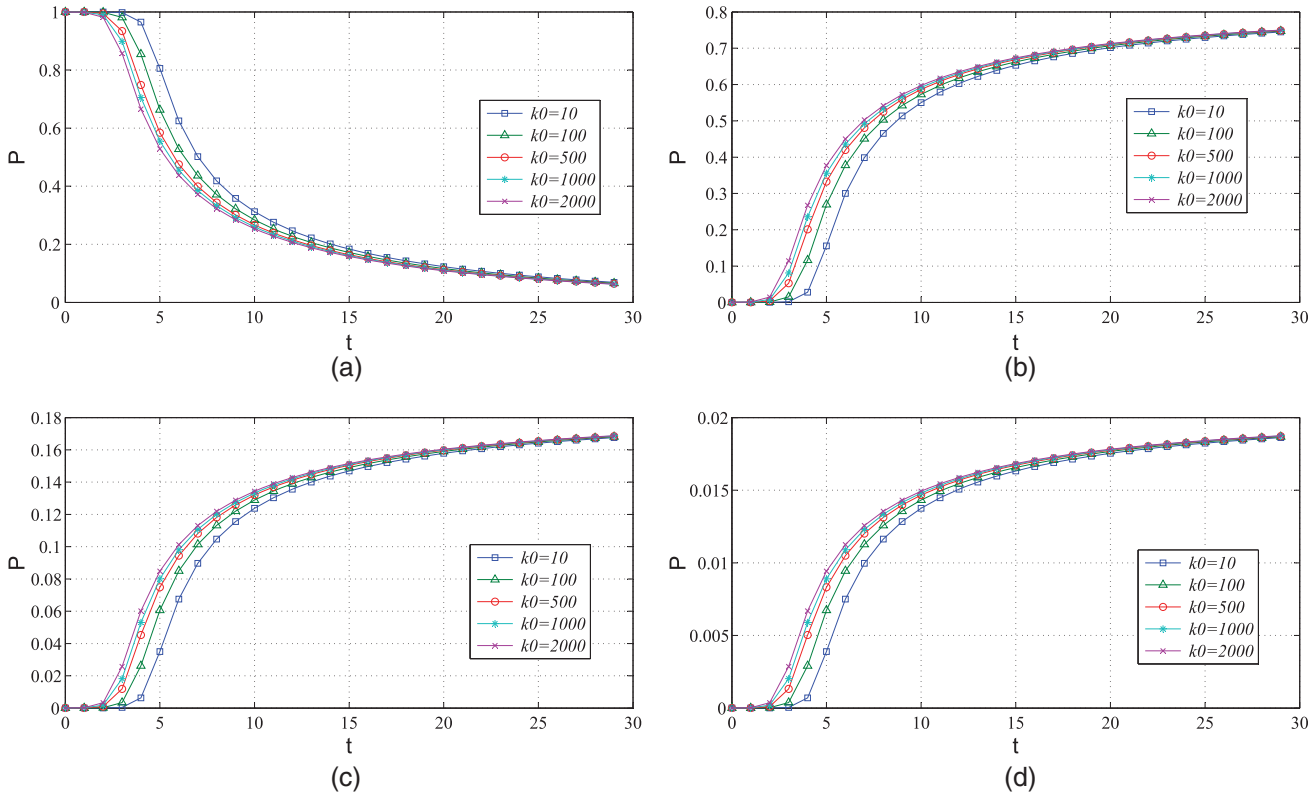

FIGURE 5. The density's evolution with t when k_0 choose 10, 100, 500, 1000 and 2000. (**a**) Susceptible nodes; (**b**) immunized nodes; (**c**) privacy disseminated nodes and (**d**) public disseminated nodes.

immunized nodes. Thus, the nodes with high degree have the large 'social influence'. And compared with real online SNS, the impacts of source node are similar.

5.2.2. *Privacy setting polices in dissemination*

In order to evaluate the influence of the privacy setting polices, and P_2 are defined as the probability of nodes transferring from accepted state to disseminated state, then various values of P_2 (0.1, 0.2, 0.5, 0.8 and 1) are chosen to simulate different users' privacy setting polices, and we set $k_0 = 29$, $P_1 = 0.2$, $P_3 = 0.1$, $c = 10$, the total number of iteration $T = 300$ as initial parameters. Simulation results about the density's evolutions of susceptible nodes, immunized nodes, privacy disseminated nodes and public disseminated nodes are shown in Fig. 6. We obtain the following results from the figure:

(i) The smaller P_2 is the densities of various nodes reach the stage of stabilization slower.
(ii) The larger P_2 is the density of disseminated nodes increases more quickly.

The results arrive at the conclusion that the more nodes with rigorous privacy setting policies exist in online SNS, the slower information spreads. Besides, the processes of information diffusion with nodes of different privacy setting polices have the similar trends, and the information can still be diffused in online SNS, even if P_2 is very small. Therefore, users' privacy setting polices really impact the process of information diffusion greatly in online SNS, which agree with the real online SNS.

5.2.3. *The credibility of information*

According to the definition of P_1 which represents the probability of nodes transferring from the susceptible state to the accepted state, P_1 can be used to evaluate the information's credibility. In real online SNS, the information is more credible if P_1 is larger, and the situation can lead to the result that more users accept the information.

The initial parameters are $k_0 = 29$, $P_2 = 0.2$, $P_3 = 0.1$, $c = 10$ and the total number of iteration $T = 300$. To simulate different information's credibility, various values of P_1 (0.1, 0.2, 0.5, 0.8 and 1) have been tested with the initial parameters. And the density's evolutions of various nodes with different P_1 are shown in Fig. 7. From this figure, we get the following results:

(i) The density of susceptible nodes decreases more quickly with the increase of P_1.
(ii) The densities of disseminated and immunized nodes increase more quickly with the increase of P_1.

The results show that the more credible information is, the faster information spreads in online SNS, and the simulation results accord with the characteristics of information's

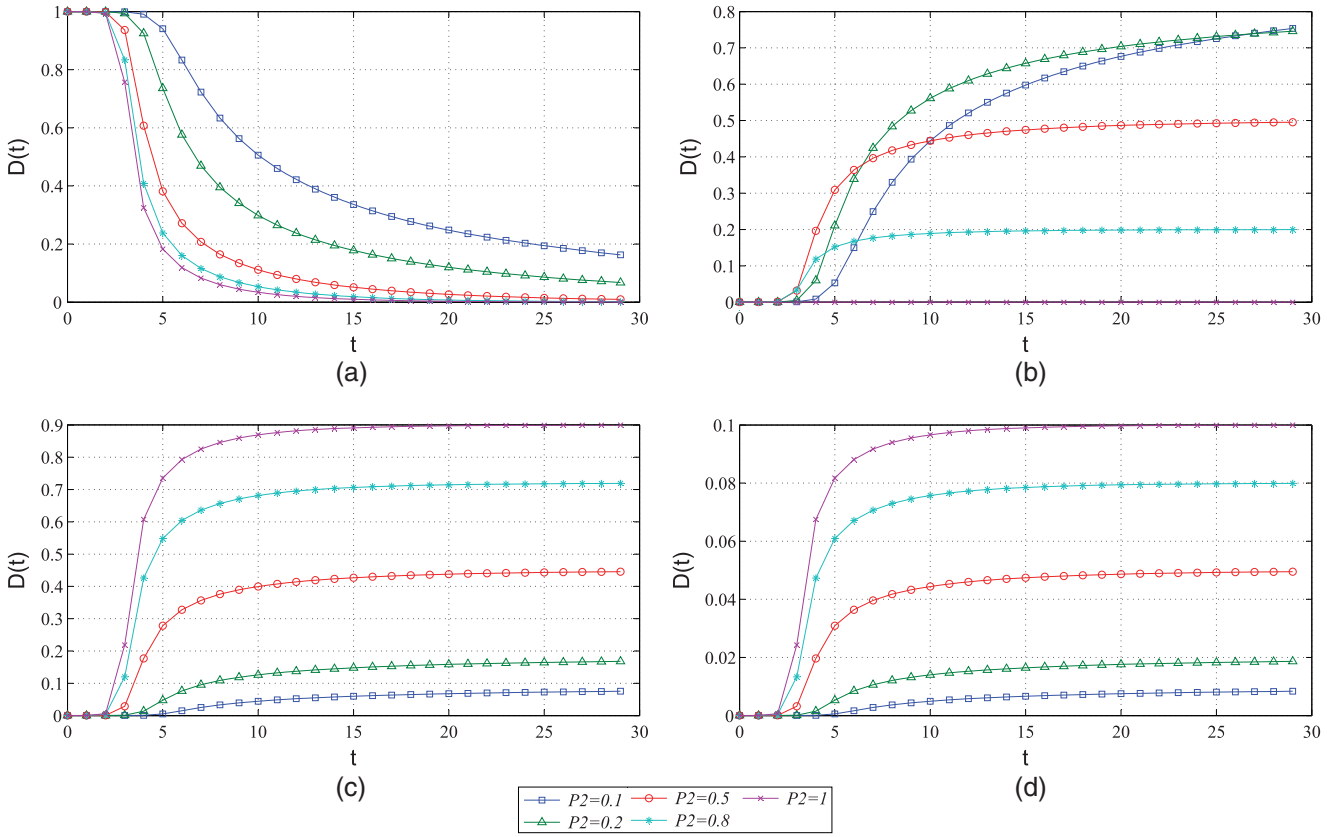

FIGURE 6. The density's evolution with t when P_2 choose 0.1, 0.2, 0.5, 0.8 and 1. (**a**) Susceptible nodes; (**b**) immunized nodes; (**c**) privacy disseminated nodes and (**d**) public disseminated nodes.

credibility in real online SNS. In addition, the results also illustrate that information with different credibility has the similar diffusion trends in online SNS, and this condition means the information can still be diffused in SNS although the information's credibility is low. Therefore, information diffusion in online SNS cannot be prevented easily, unless the information is untrusted to everyone.

5.2.4. *Users' habits*

Users' habits are defined as the habits of users' accessing the public information actively, so the habits can be evaluated by parameter P_3 and c of model. Therefore, we list four different situations for simulating different users' habits. $P_1 = 0.4$, $P_2 = 0.4, k_0 = 29$ and the total number of iteration $T = 300$ are the initial parameters of model; meanwhile, we change values of P_3 and c to simulate different users' habits in real online SNS.

The values are divided into two groups, we choose $P_3 = 0.2$ and 1, $c = 5$ and 100 to make a comparison of the influences of different users' habits. According to our definitions of P_3 and c, the users' habits may be impacted by both of the parameters of model. And the density's evolutions of various nodes with different users' habits are shown in Fig. 8. From this figure, we get the following results:

(i) The density's evolutions of various nodes increase very slowly with increase of P_3 and c in online SNS.
(ii) When $P_3 = 0.2$ or 1.0, their density's evolutions are almost the same with different settings of c.
(iii) When $c = 5$ or 100, their density's evolutions are different only in consideration of the density of disseminated nodes with different settings of P_3 .

The results illustrate that different users' habits can hardly change the process of information diffusion in SNS, and different P_3 can only change the evolutionary process of private and public disseminated nodes. According to Fig. 8, c has almost no effects in model. Compared with the real online SNS, the users' habits and public information indeed improve the information diffusion, but it's a tiny improvement.

Users normally do not spend too much time searching information in real online SNS and there are just a few public disseminated nodes. Nevertheless, in special situations, there are users searching information actively (the simulation that $c = 100$ and $P_3 = 1.0$ may illustrate the situation). The results also illustrate that the process of disseminated nodes' evolution may change, but the density's changes of disseminated nodes do not impact information diffusion in DMPS with the different P_3 and c. Therefore, we come to a conclusion that the defined

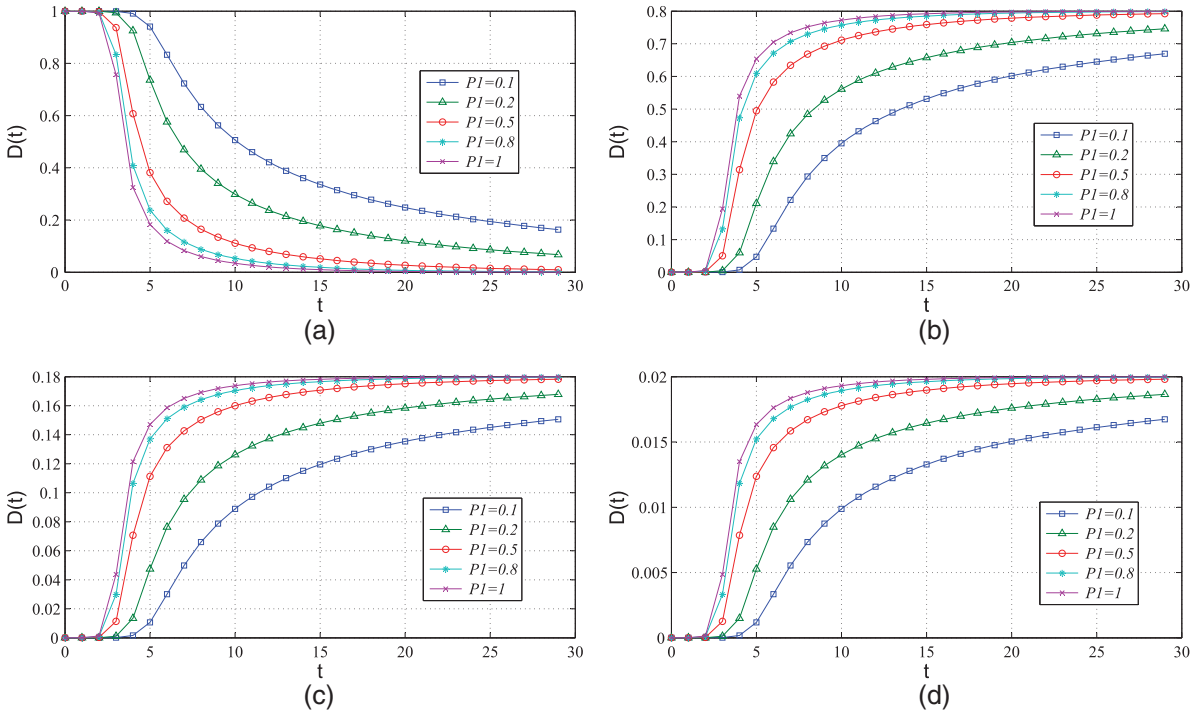

FIGURE 7. The density's evolution with t when P_1 choose 0.1, 0.2, 0.5, 0.8 and 1. (**a**) Susceptible nodes; (**b**) immunized nodes; (**c**) privacy disseminated nodes and (**d**) public disseminated nodes.

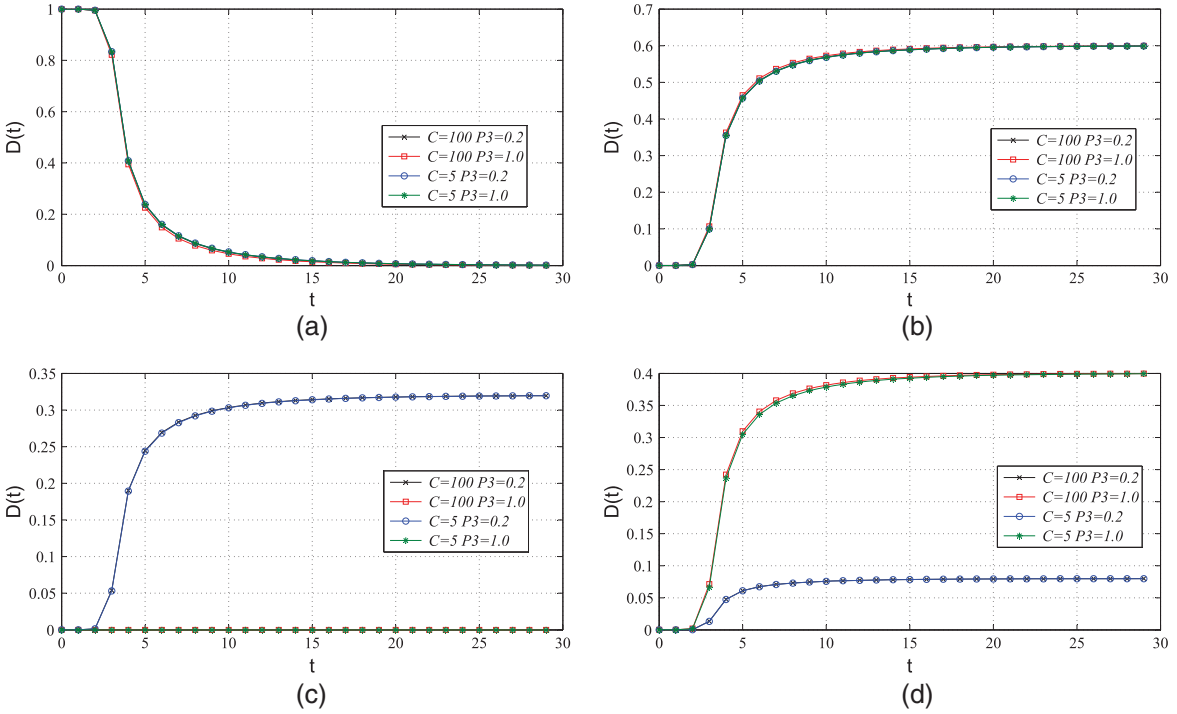

FIGURE 8. The density's evolution with t when P_3, c choose 0.2 and 100, 1.0 and 100, 0.2 and 5, 1.0 and 5. (**a**) Susceptible nodes; (**b**)immunized nodes; (**c**) privacy disseminated nodes and (**d**) public disseminated nodes.

users' habits just have slight impact on information diffusion in online SNS.

6. RELATED WORKS

With the pervasiveness of online SNS, the information diffusion has attracted considerable attentions and becomes an emerging research topic [10]. Recently, several models, such as independent cascade model [11] and linear threshold model, have been proposed to simulate and analyze the information diffusion in online SNS tools. For example, the pattern of cascades in large blog [12] and the dissemination of influence on Twitter data set [13] are discussed. Moreover, the different impacts of various topics [14], unknown situations [15] and competitive information [16] on information diffusion in online SNS are simulated and analyzed. At the same time, many analysis methods, such as mean-field analysis [17], structural trend analysis [18] and topology adaptation analysis [19], have been used to model and speculate the diffusion dynamics of online SNS. Besides, a lot of research works focused on various factors which influence information diffusion in online SNS. For instance, user's emotion and social media [20], individual behavior and social influence [21], trust metric [22] and locating privileged spreaders [23].

However, most of existing information diffusion models focus on the single diffusion mechanism, and all of them are not appropriate to discover the process of information diffusion in online SNS with multiple diffusion mechanisms. In specific, the individuals in most of existing information diffusion models are just classified into two categories: active or inactive, which cannot describe the features of factors which influence information diffusion in online SNS. Therefore, we focus on the users' privacy setting polices, and classify the users' states into three temporary states and four final states; then, we define the state transition rules based on the dynamics of infectious disease; finally, we use dynamic evolution equations to describe the process of information dissemination with multiple diffusion mechanisms in online SNS.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a general stochastic model with multiple diffusion mechanisms, called DMPS, to describe the information diffusion in online SNS. In specific, to represent online SNS, we first classify users into various types based on their privacy setting policies, and define the relationships among users according to the features of information sharing in online SNS. Then, we define the states of users and state transition rules according to dynamics of infectious disease. Finally, we use dynamic evolution equations to describe the evolution process of information diffusion with multiple diffusion mechanisms. Besides, we simulate and analyze the density's evolutions of susceptible nodes, public disseminated nodes, privacy disseminated nodes and immunized nodes with time, respectively. Detailed simulations and analysis show that our proposed DMPS can describe the information diffusion accurately, and the information diffusion in DMPS agrees with the real online SNS. Furthermore, the fewer privacy setting policies of users, the faster information spreads and the densities of users with various privacy setting policies have the similar trends. Therefore, DMPS can be used to identify the underlying mechanism of information diffusion in online SNS and forecast the trend of it. In our future work, we intend to analyze the impact of different information's features in online SNS accurately.

ACKNOWLEDGEMENT

The authors thank the anonymous reviewers for their constructive comments, which have helped improve the quality of this paper.

FUNDING

This work was supported by National Natural Science Foundation of China (61303218, 61272457), National Science & Technology Major Projects (2012ZX03001009), Fundamental Research Foundations for the Central Universities of China (K5051301017) and the 111 Project (B08038).

REFERENCES

[1] Fatima, I., Fahim, M., Lee, Y. and Lee, S. (2013) Modm: multi-objective diffusion model for dynamic social networks using evolutionary algorithm. *J. Supercomput.*, **66**, 738–759.
[2] Sharifi, B., Inouye, D. and Kalita, J. (2014) Summarization of twitter microblogs. *Comput. J.*, **57**, 378–402.
[3] Mavromoustakis, C. and Karatza, H. (2011) Embedded socio-oriented model for end-to-end reliable stream schedules by using collaborative outsourcing in MP2P systems. *Comput. J.*, **54**, 1235–1247.
[4] Santonja, F., Villanueva, R., Jódar, L. and González-Parra, G. (2010) Mathematical modelling of social obesity epidemic in the region of valencia, spain. *Math. Comput. Model. Dyn. Syst.*, **16**, 23–34.
[5] Leskovec, J., Lang, K., Dasgupta, A. and Mahoney, M. (2009) Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. *Internet Math.*, **6**, 29–123.
[6] Nekovee, M., Moreno, Y., Bianconi, G. and Marsili, M. (2007) Theory of rumour spreading in complex social networks. *Phys. A: Stat. Mech. Appl.*, **374**, 457–470.
[7] Vázquez, A. and Weigt, M. (2003) Computational complexity arising from degree correlations in networks. *Phys. Rev. E*, **67**, 027101.
[8] Leskovec, J. (2012) *Stanford large network dataset collection*. http://snap.stanford.edu/data/egonets-Twitter.html. (accessed March 10, 2014).

[9] Kitsak, M., Gallos, L., Havlin, S., Liljeros, F., Muchnik, L., Stanley, E. and Makse, H. (2010) Identification of influential spreaders in complex networks. *Nat. Phys.*, **6**, 888–893.

[10] Xu, B. and Liu, L. (2010) Information Diffusion Through Online Social Networks. *IEEE Int. Conf. Emergency Management and Management Sciences(ICEMMS2010)*, Beijing, China, August, pp. 53–56. IEEE, Piscataway.

[11] Goldenberg, J., Libai, B. and Muller, E. (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Mark. Lett.*, **12**, 211–223.

[12] Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. and Hurst, M. (2007) Patterns of Cascading Behavior in Large Blog Graphs. *Proc. 7th SIAM Int. Conf. Data Mining*, Minneapolis, USA, April, pp. 551–556. Society for Industrial and Applied Mathematics Publications, Philadelphia.

[13] Bakshy, E., Hofman, J., Mason, W. and Watts, D. (2011) Everyone's an Influencer: Quantifying Influence on Twitter. *Proc. 4th ACM Int. Conf. Web Search and Data Mining*, Hong Kong, China, February, pp. 65–74. ACM, New York.

[14] Romero, D., Meeder, B. and Kleinberg, J. (2011) Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. *Proc. 20th Int. Conf. World wide web*, Hyderabad, India, March, pp. 695–704. ACM, New York.

[15] Saito, K., Kimura, M., Ohara, K. and Motoda, H. (2013) Detecting changes in information diffusion patterns over social networks. *ACM Trans. Intell. Syst. Technol. (TIST)*, **4**, 55: 1–23.

[16] Small, L. and Mason, O. (2013) Information diffusion on the iterated local transitivity model of online social networks. *Discrete Appl. Math.*, **161**, 1338–1344.

[17] Karnik, A., Saroop, A. and Borkar, V. (2013) On the diffusion of messages in on-line social networks. *Perform. Eval.*, **70**, 271–285.

[18] Budak, C., Agrawal, D. and Elabbadi, A. (2011) Structural trend analysis for online social networks. *Proc. VLDB Endowment*, **4**, 646–656.

[19] Cimini, G., Chen, D., Medo, M., Lü, L., Zhang, Y. and Zhou, T. (2012) Enhancing topology adaptation in information-sharing social networks. *Phys. Rev. E*, **85**, 046108.

[20] Stieglitz, S. and Dangxuan, L. (2013) Emotions and information diffusion in social media sentiment of microblogs and sharing behavior. *J. Manage. Inform. Syst.*, **29**, 217–248.

[21] Papagelis, M., Murdock, V. and Vanzwol, R. (2011) Individual Behavior and Social Influence in Online Social Systems. *Proc. 22nd ACM Conf. Hypertext and Hypermedia*, Eindhoven, The Netherlands, June, pp. 241–250. ACM, New York.

[22] Aloufi, S., Kim, H. and Elsaddik, A. (2012) A group trust metric for identifying people of trust in online social networks. *Expert Syst. Appl.*, **39**, 13173–13181.

[23] Borgeholthoefer, J., Rivero, A. and Moreno, Y. (2012) Locating privileged spreaders on an online social network. *Phys. Rev. E*, **85**, 066123.