Future Generation Computer Systems ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

User role identification based on social behavior and networking analysis for information dissemination

Xiaokang Zhou [a], Bo Wu [b], Qun Jin [c,*]

[a] Faculty of Data Science, Shiga University, Hikone, 522-8522, Japan
[b] Advanced Research Center for Human Sciences, Waseda University, Tokorozawa, 359-1192, Japan
[c] Faculty of Human Sciences, Waseda University, Tokorozawa, 359-1192, Japan

ARTICLE INFO

Article history:
Received 23 August 2016
Received in revised form
7 March 2017
Accepted 29 April 2017
Available online xxxx

Keywords:
User identification
User modeling
Social networking analysis
Social behavior
Information dissemination

ABSTRACT

Nowadays, along with the high development of emerging computational paradigms, more and more populations have been involved into the social revolution across various intelligent systems, which results in dynamic user connections associated with a variety of social behaviors. The associated users with different properties, who can be regarded as one kind of information resources, have become increasingly important, especially in social knowledge creation and human intelligence utilization processes. In this study, we concentrate on user role identification based on their social connections and influential behaviors, in order to facilitate information sharing and propagation in social networking environments. Following the construction of a dynamic user networking model, we propose a network-aware method to identify four kinds of special users, who may play an important role in information delivery among a group of users, or knowledge sharing between pairs of users. A set of attributes and measures is proposed and calculated to identify and represent these users based on the analysis of their influence-related social behaviors and dynamic connections. Experiments and evaluations are conducted to demonstrate the practicability and usefulness of the proposed method using Twitter data. Analysis results show the effectiveness of our approach in identifying the distinct features of four kinds of users from the user networking model. Comparison experiments indicate that the proposed identification method outperforms two other related works. Finally, a questionnaire-based evaluation demonstrates the accuracy and efficiency of the proposed method in terms of finding these users in a real social networking environment.

1. Introduction

With the rapid development of emerging computing paradigms, such as Ubiquitous Computing, Social Computing, and Mobile Computing, we have been continuously experiencing a fast change from all walks of our work, life, learning and entertainment. People are joining together to publish personal messages, share individual experience, and exchange their own opinions through online social networking services. As more and more people have been engaged into this social networking environment, a large number of user generated contents, which contain a variety of human experience and social knowledge, have spread widely in a higher speed than ever before [1]. Comparing with the traditional information dissemination which fully depends on the popularity of posted contents, the information flow propagated cross social networks mainly relies on interactions among individuals and groups associated with various social relationships. That is, the dissemination speed, scale, and controllability are increasingly influenced by the highly connected users. In particular, individual users with different background knowledge (e.g., cognition, interest, reputation, and etc.), are playing a significant role not only in shaping public opinions, but also in expanding access to diversified personal contents more efficiently. Thus, it becomes a challenge but essential issue to dynamically identify types or roles of individuals, who may help to deliver human intelligence and socialized knowledge to the person in need via user relationship chains during information sharing, exchanging, and propagating processes.

Generally, a social network [2] can be viewed as the complex network of social relationships and interactions among a series of associated people with different ideas, interests and perspectives. Since individuals have become an important information source,

* Corresponding author. Fax: +81 4 2947 6906.
E-mail address: jin@waseda.jp (Q. Jin).

http://dx.doi.org/10.1016/j.future.2017.04.043

recommending users can help people find the truly related information, and further motivate them to interact with each other more frequently. More importantly, discovering and identifying popular users or top influencers (e.g., finding strong friends [3] and significant friends [4]), who may have high social connectivity or strong influence on a large number of other related users, may provide the computational collective intelligence for a sensible decision making. Many computational intelligence methods have been developed to model or identify real world problems, and enable adaptive applications in complex environments. Specifically, challenges in information dissemination across social networks can be summarized as: How to understand social dynamics among social connections to design an efficient information dissemination mechanism; How to integrate various social factors (such as social influence and interactional behaviors), to maximize the extent of information dissemination; How to effectively identify and utilized the popular/important users to build bridges across individuals and communities during information delivery and knowledge sharing processes. Therefore, to benefit a target user from a well-structured user networking, we try to find a computational and intelligent way to leverage both network structure and social factors, and seek the associated users who can provide helpful information or personal experience in a certain social context.

In our previous study, we have built a Dynamically Socialized User Networking model to discover and represent implicit and explicit user relationships in the SNS (Social Networking Service) site [5]. Moreover, we have analyzed and built the dynamic user profiling to describe multi-dimensional user features and properties [6]. In this study, to facilitate the human intelligence utilization in the social environment, we concentrate on user identifications in terms of their different roles among a group of associated users during information dissemination processes. In details, following a social networking model to describe and analyze dynamic connections among users, we define and identify four kinds of special users to support the information dissemination process in different social contexts. A set of attributes and measures are then defined and calculated based on the analysis of social behaviors and user connections. Specifically, the major contributions of this paper are summarized as follows.

- A network-aware approach to identifying four kinds of special users in a multi-viewed way, which can facilitate information delivery among a group of users, or knowledge sharing between pairs of users respectively.
- A computational method with a set of measures to quantitatively describe and calculate distinct user features in accordance with their dynamic connections and influential behaviors, when they deliver, exchange, and share personal information among a group of people.
- A design of a questionnaire-based evaluation to demonstrate the dynamic finding of these users comparing with the manual method in a real social networking environment, which reveals useful insights in understanding the importance of user roles during the social information dissemination process.

The remainder of this paper is organized as follows. We give a brief overview on related issues and works in Section 2. In Section 3, after introducing several basic concepts, such as a user networking model and definitions of influence-related behaviors, we propose and define four kinds of special users, and further address a set of corresponding measures for their identifications respectively. The experiment and evaluation results using the Twitter data are demonstrated in Section 4. We conclude this study and give our promising perspectives regarding future research in Section 5.

2. Related work

Three issues related to this study are addressed in this section. Specifically, the analysis of social networking, methods of user modeling and recommendation, and issues of information propagation and dissemination, are discussed respectively.

2.1. Social networking analysis

The importance and popularity of social networking have attracted more and more researches to develop models and applications based on social behavior analysis. A variety of types of social behaviors have been focused to analyze users' behavioral features in social networking environments. For instance, Kelly et al. [7] examined people's location behaviors, and proposed a human mobility modeling framework to calculate several human behavior features, which aimed to build predictability vectors to describe an individual's behaviors. Zhao et al. [8] built a predictive model to analyze social relationships with different strengths, and studied the correlation between social popularity based on online and offline social behaviors. Pan and Wang [9] defined two kinds of user behaviors in Sina Weibo, and further utilized the sequential pattern algorithm to mine the social behavior pattern. Socievole et al. [10] proposed a degree centrality based method to analyze the mobility data with Facebook friendships collected from multi-layer social networks, which may help to better understand a user's social behaviors. As for the utilization of analysis results from social behaviors, Agreste et al. [11] analyzed online activities in a social platform named aNobii, and built three kinds of profiles for each user. Han and Xu [12] defined three types of social behavior features in microblog, and proposed a pair-user-based classification mechanism to improve the performance of friend recommendation. Adali and Golbeck [13] defined and analyzed a set of behavioral measures based on the intensity and the number of social interactions among users. The Twitter data were then utilized to demonstrate how these features could be used for the personality prediction. Ruan et al. [14] studied users' social behavioral features from their social activities on online social networks, and built social behavioral profiles which could be applied to distinguish different OSN users and detect compromised accounts. Comparing with these existing researches, the focus of this study is on analyzing social influence-related behaviors. We try to extract and represent the behavioral user features based on their social influence. The analysis results are further utilized for special user identifications, which can facilitate the information dissemination in social networking environments.

2.2. User modeling and recommendation

Analyzing and modeling of different properties of individual users have drawn a fair amount of works, not only in the system performance improvement, but also for the social application development. Lin et al. [15] analyzed Fan Pages on the Facebook, and developed a data mining method to find opinion leaders, which may be used to improve the design marketing plan in companies. Xu et al. [16] addressed an identification method to mine related users using the Weibo data, in order to support the public opinion monitoring in the social networking environment. Xie and Li [17] considered both heterophily and homophily issues in friend recommendations, which aimed to enhance the user experience in Twitter-like SNS sites. Typically, two basic aspects, the friend recommendation and special user finding, have become major researches in recent years. Fan et al. [18] proposed a tree-structure based data mining algorithm to find a group of friends in social networks. Wang et al. [19] employed identified topics and social networks to recommend groups of users based on the calculation

Please cite this article in press as: X. Zhou, et al., User role identification based on social behavior and networking analysis for information dissemination, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.04.043

of different user factors. Tang et al. [20] addressed a friend recommendation approach which taken four aspects into account based on a micro-blog user model. On the other hand, Ye et al. [21] applied an attribute synthetic evaluation method to identify different users based on multiple attribute decision making, which aimed to benefit reuse and sharing of user profiles. Instead of recommending people with user contents, Jiang et al. [22] utilized geo-tags to discover top-k local users in the Twitter system. Srinivas and Velusamy [23] defined a so-called enhanced degree centrality measure to find the most influential objects or persons in social networks. Fang et al. [24] proposed a relational latent SVM model in which six types of user attributes were considered to achieve the better user attribute inference performance in social media. Comparing with these methods, in particular, this study concentrates on modeling four kinds of special users based on the analysis of their social behaviors and dynamic connections. A set of attributes and measures are defined and calculated to identify their roles during information delivery and knowledge sharing process.

2.3. Information propagation and dissemination

Information propagation and dissemination across social media have been hotly discussed for many years. Several social factors have been employed in both model constructing and diffusion measuring. Chou et al. [25] presented an analytical model which could be considered as the independent-cascade model or susceptible–infected model to analyze the dynamics of information dissemination in social networks. Zhang et al. [26] built a so-called "following" link cascade model which could help to learn the diffusion strength in different triadic structures during the information diffusion process. Considering the time duration, Feng et al. [27] defined the information diffusion process with a direct measure, which aimed to analyze the diffusion efficiency in real online social networks. In addition to the consideration of users' relationships, for instance, Chen and Hu [28] presented a socialization information dissemination mode based on the interpretation of several aspects of users' relationships, in which users themselves were increasingly taken into account as an important role when analyzing the social information dissemination process. Han et al. [29] designed a targeted immunization policy with a distributed protocol. They proposed a heuristic algorithm to select influential mobile users, and promote the information dissemination in mobile social networks. Wen et al. [30] investigated and verified the importance of popular users in online social networks, which demonstrated several counter-intuitive results. Tang et al. [31] built an interest-based dissemination model, in order to choose users who could promote the information dissemination across social media. Qi et al. [32] measured users' secondary abilities from their neighborhoods' aspects based on the behavioral analysis, which could differentiate users during the social information dissemination process. Similarly, in this study, we analyze and extract user features in terms of their interpersonal influence and temporal connections. Moreover, a computational method is proposed to find these users who may either promote information diffusion among a group of people, or stimulate information sharing between pairs of individuals.

2.4. Summary

Research works have pointed out the necessity to take advantage of the social data, not only in the design of systems and models, but also in the practical recommendation and prediction services. In particular, the analysis and modeling of social behaviors among a series of connected individuals have become increasingly important for social information sharing and recommendation. In this study, comparing with other traditional methods, we emphasize the importance of individual users, and focus on identifying their special roles in facilitation of social information dissemination. Specifically, the dynamic user connections and influential behaviors are integrated together to describe and analyze the behavioral user features in different social contexts, which can efficiently improve information dissemination in social networking environments.

3. Social analysis for user identification

In this section, we first introduce the basic user networking model, and concepts to describe a user's influence-related behaviors. Based on these, we define and propose four kinds of special users with their corresponding measures, which can help to promote information delivery among a group of users, or knowledge sharing between two connected users.

3.1. Basic model and concepts

3.1.1. Model for user networking analysis

The sociological theory of homophily [33] told us that people tend to build connections with others who are similar to them, or often perform similar actions. The social influence network theory [34] also indicates the importance of a network of interpersonal influence, which plays a foundational role in modifying people's attitudes and opinions when they interact with each other during information dissemination processes. Accordingly, a well-structured user networking model is basic and necessary to link users with similar features together, and further find out those significant users who may contribute more on information delivery and social knowledge sharing.

Therefore, the basic model, named as DSUN (Dynamically Socialized User Networking) [5], is utilized to analyze and represent dynamic user connections. The definition and expression are addressed as follows.

$$G_{DSUN}(U, E, UC_T) \tag{1}$$

$U = \{u_1, u_2, u_3, \ldots, u_n\}$: A non-empty set of vertexes in the network model. Each u_i indicates a unique user.

In details, $u_i = (ID_i, H_i, A_i)$, in which ID_i indicates a unique user ID of each user; H_i indicates a user's time-varying intentions, such as his/her current interests or needs during time period T; and A_i indicates a set of user attributes that can be used to identify his/her special role during information dissemination processes. In a word, all the individual-related information will be extracted and stored in the corresponding vertex.

$E = \{e_{ij} = \langle u_i, UC_T, u_j\rangle |$ if a relationship exists between u_i and $u_j\}$: A collection of edges that connect the vertexes in U.

In particular, in the DSUN model, vertex u_i is the user on the head of edge e_{ij}, who may provide useful information relating to user u_j's current requirement as a potential benefactor. On the other hand, vertex u_j is the user on the tail of edge e_{ij}, who may obtain valuable information from user u_i during T as a beneficiary. Accordingly, in this connected user networking, the edge $\overrightarrow{u_iu_j}$, extending from u_i to u_j, not only points out the direction of helpful information delivery through them, but also describe the potential benefit between a pair of users.

$UC_T = \{UC_{T_{ij}} |$ if $\exists e_{ij} \in E\}$: A multi-tuple of measures appended on the corresponding edge which can be used to describe and quantify various types of user relationships.

Specifically, each measure with a corresponding weight is defined to calculate the strength of a specific correlation between user u_i and u_j during T.

3.1.2. Concepts for influence behavior analysis

Online social networks provide a variety of online features, such as the social content, relation, and behavior features, during user communications. Specifically, the generated corresponding social behaviors (e.g., sending messages, building connections, and interacting with friends), may have different effects on each user, and shape different types of users during information dissemination in social networking environments.

Social influence [35], which is defined as "change in an individual's thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group", has a long history in research field of sociology and psychology [36,37]. It has been widely utilized as a measure to specify how people influence thoughts, attitudes, and behaviors of others in direct or indirect ways, which reflects the reputation or significance of one person within the whole networked people or to a specific one. More precisely, the social impact theory [38] explains that the social impact/forces is/are diffused across a group of target people who may have sort of associations in terms of their works and communications. In some specific situations, a few of people may become extremely important in determining the impact degree.

Thus, we focus on description and analysis of a special type of social behaviors in terms of their influence among a group of associated users, which can be defined and categorized as the follows.

Influencing Behavior ($IgB_{\overrightarrow{ij}}$): A set of influence behaviors of user u_i, which indicates that user u_i influences user u_j. It can also be considered as a kind of behavior that indicates whether or not user u_i gives personal information to user u_j.

Influenced Behavior ($IdB_{\overrightarrow{ij}}$): A set of information behaviors of user u_j, which indicates whether or not user u_j has been influenced by user u_i. It can also be considered as a kind of behavior that indicates whether or not user u_j has received the personal information from user u_i or been in favor of his/her thought.

Accordingly, among a variety of social behaviors that users conducted to seek and utilize information across social media, the influence-related behaviors are specified to extract and analyze influence-based user features arising out of the interactional behaviors among a group of connected users during information dissemination in social networking environments.

3.2. Attributes and measures for user identification

In addition to several attributes defined in our previous study [6] to present some basic descriptions of one user's profile, in this paper, we go further to analyze and describe the behavior-related user features. In particular, four kinds of special users, are identified to support information delivery and knowledge sharing among a series of associated users.

3.2.1. User role identification among a group of users

Generally, users are sharing and exchanging their personal information and knowledge through complicated interactions. During this process, some users who usually prefer to share their personal ideas, attitudes, or opinions, will gradually become the origin, and constantly promote interactions among a group of people in terms of some specific topics In another aspect, other users who always follow these positive users, and tend to derive the information relating to their current interests or needs, will be gradually influenced by them. Consequently, to model such an information dissemination process, it is important and necessary to define and capture these features of users who appear to be the center in term of a group of people.

Hub user: A hub user is defined as the user who continuously shares and delivers information through his/her information behaviors to the extent of influencing others. Other users can benefit directly or indirectly, so as to result in a high reputation in regard to a group of individuals within specific limits.

Specifically, the diffusion of information through the DSUN model from one user to a group of users is employed to identify the so-called hub user It can be viewed as a collective and global measure of worthiness based on the influence scope of a certain group of individuals within information dissemination processes. Inspired by the studies presented in [39,40], the diffusion attribute used to identify the hub user can be defined as follows.

Diffusion attribute: Given a specific user u_i, the diffusion attribute indicates the density of the influence scope, caused by his/her information behaviors. The higher the density, the greater the number of individuals who may derive helpful information over an extensive range would be.

$$DA_i = \sum_{k \in IB_i} AvgD_k * \log IdU_k \tag{2}$$

where IB_i denotes a set of information behaviors of user u_i, $AvgD_k$ the average influence depth of a specific information behavior conducted by user u_i, and IdU_k the number of users who have been influenced by this behavior.

Furthermore, it is noted that the issue of promotion also plays a crucial role in information dissemination processes. The modeling and identification of promotion users can help promote referrals and ratings of information through interactional user behaviors in a certain community, which will finally benefit the sharing and exchanging of human intelligence in social networking environments.

Promotion user: A promotion user is defined as the user who can tremendously increase and promote sharing and delivering of the information that disseminates via him/her It means a large fraction of information will get the higher referrals if this kind of user is willing to deliver them through his/her information behaviors.

Specifically, the promotion attribute is defined as follows, to identify this kind of promotion users.

Promotion attribute: Given a specific user u_i, the promotion attribute indicates the change of referrals of information after he/she has delivered them It also shows the power of influence regarding to a topical scope of users.

$$PA_i = \sum_{k \in IdB_i} \sum_{n=1}^{MaxD_k} \frac{IdU_{kn}}{n} \tag{3}$$

where IdB_i denotes a set of influenced behaviors of user u_i, $MaxD_k$ the maximum influence depth of a specific influenced behavior conducted by user u_i and IdU_{kn} the number of users who have been influenced by this behavior of user u_i in the nth-depth ($n = 1, 2, 3, \ldots, MaxD_k$).

3.2.2. User role identification between users

On the other hand, personally, users may not concern who influence the most of users, or who benefit the most for the information dissemination. What they are really concerned may be who can provide more valuable information related to their own requirements. In order to deal with this situation, contrasting with the hub user and promotion user, who are globally visible to the whole body of users in the networking we further identify another two kinds of users to assist personalized information sharing within pairs of users.

Contribution user: A contribution user is the better benefactor u_i among the users linked to a target user u_j, who can better support user u_j with more relevant and valuable information, or transfer the beneficial influence to him/her through social behaviors.

Contribution attribute: Given a pair of users: $\langle u_i, u_j \rangle$, the contribution attribute is the value that reflects the contribution or importance of user u_i to user u_j, which can be quantified as follows.

$$CA_{ij} = \frac{BIC_{ij}}{\sum_{k=1}^{|Out_i|} BIC_{ik}} + \frac{BIC_{ij}}{\sum_{l=1}^{|In_j|} BIC_{ij}} \quad (4)$$

where BIC_{ij} denotes the beneficial-influence-based correlation between a pair of connected users $\langle u_i, u_j \rangle$ in a constructed DSUN model. It considers both of their similarity-based and influence-based relationships, and indicates the beneficial influence from u_i to u_j. Out_i denotes a group of users who linked from user u_i according to u_i's out-degree, In_j a group of users who linked to user u_j according to u_j's in-degree.

Reference user: A reference user is the most similar user u_i among users linked to a target user u_j, who can understand and complement with each other better, and share the similar information or experience in a reciprocal way.

Reference attribute: Given a pair of users: $\langle u_i, u_j \rangle$, with a user set U, $U = \{u_1, u_2, \ldots, u_n\}$, who have influence on them, the reference attribute indicates the similarity between these two users according to their influence-related behaviors with those users in U, which can be quantified as follows.

$$RA_{ij} = \frac{\sum_{n=1}^{|U|} \left(\left| IdB_{\overrightarrow{nj}} \right| \times \left| IdB_{\overrightarrow{nj}} \right| \right)}{\sqrt{\sum_{n=1}^{|U|} \left| IdB_{\overrightarrow{nj}} \right|^2} \times \sqrt{\sum_{n=1}^{|U|} \left| IdB_{\overrightarrow{nj}} \right|^2}} \quad (5)$$

where $\left| IdB_{\overrightarrow{nj}} \right|$ and $\left| IdB_{\overrightarrow{nj}} \right|$ denote the number of influenced behaviors from other users to user u_i and u_j respectively. $|U|$ denotes the number of users in a constructed DSUN model. The measure of Cosine similarity is used to calculate the value of reference attribute.

4. Experiment and analysis

In this section, experiments and evaluations are conducted in identifying four kinds of special users (i.e., hub user and promotion user, contribution user and reference user), in order to demonstrate the practicability and usefulness of our proposed methods.

4.1. Data set

As mentioned above, Twitter, one of the famous social media networking system, has been employed to collect the data set for our experimental analysis. We collected the data by crawling the contents generated from users in a Twitter list named "Awesomesocial" with their followees and followers. We selected this list because comparing with other languages, Tweets posted in this popular Twitter list were predominantly written in English. This is important for us to analyze and capture user features from the collected data in a computational way. In addition, we give two statements. First, after filtering the invalid users (e.g., advertiser), as well as the Internet slang words in the contents, we assume that there are no deviant users in our data set, which means all users have conducted their social behaviors normally and generated no spam or irrelevant actions. Second, due to the upper limit of each user's available collecting time for the officially provided Twitter API which we used to collect the data, we assume that we collect adequate data for each user in our experiment even if his/her posts exceed over the limitation.

The collecting period was from April 2 2014 to August 15 2014, in which the entire time period has been dynamically divided into several time slices according to the prevailing topic-based trends. It is noted that the collected data were filtered according to two rules: (i) Users who are eligible for experiments should continuously keep a high activeness in the entire experiment period (e.g., posting at least one tweet per two days); (ii) Tweets which are qualified for experiments cannot contain any Internet slang words. Finally, 4455 users with 463 644 tweets were used to conduct the experiments, together with 12 generated time slices as follows.

T1: April 2–April 8,
T2: April 9–April 25,
T3: April 26–April 28,
T4: April 29–May 1,
T5: May 2–May 10,
T6: May 11–May 12,
T7: May 13–May 16,
T8: May 17–May 28,
T9: May 29–June 14,
T10: June 15–July 28,
T11: July 29–August 9, and
T12: August 10–August 15.

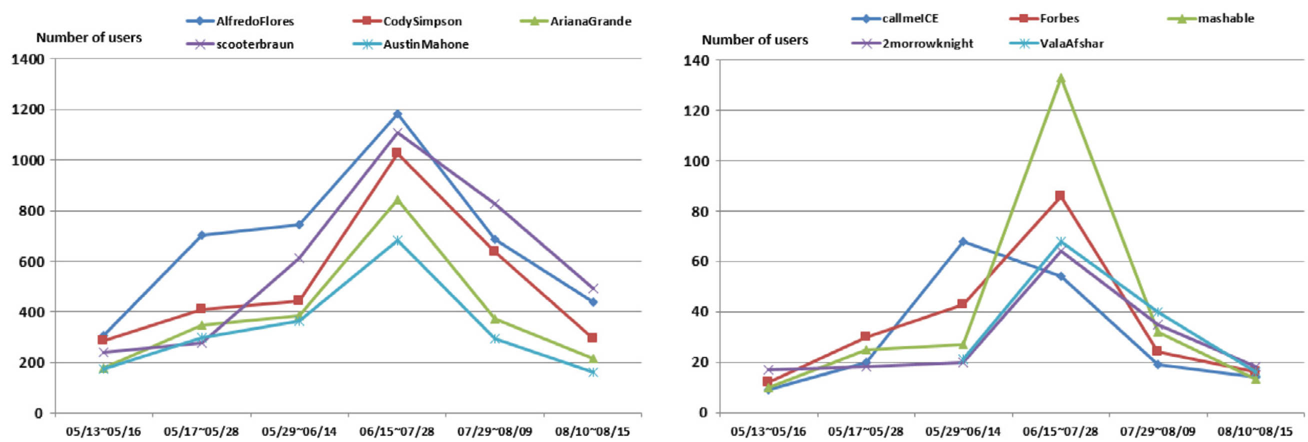
To model and analyze social behaviors relating to information dissemination in this data set from Twitter, posting messages, retweeting contents and mentioning other users are regarded as information behaviors. Social tags in the posted contents are utilized to identify these two kinds of influence-related behaviors. In particular, "@name" is taken as the influencing behavior, and "RT@name" the influenced behavior.

4.2. Analysis of user identification

We identify the top-ranked hub and promotion users in different time slices (say *T7*–*T12*) using Eqs. (1) and (3) respectively. To demonstrate the dynamics of the finding of these two kinds of special users, the numbers of other users who have been influenced by them in each time slice are involved to illustrate the time-varying changes. The comparison results of five top-ranked hub users and promotion users are shown in Fig. 1. We give our observations and discussions for the analysis of hub and promotion users as follows.

(1) Generally, it seems that both hub users and promotion users are sensitive to the length of time interval. The longer the time slice (such as *T10*) is, the more people will be influenced by these users.
(2) Obviously, hub users always influence more people than promotion users. In most cases, hub users are easy to be identified than promotion users because of the obvious gaps in terms of the number of influenced users. As shown in Fig. 1(b), distinguishing each promotion user will become extremely difficult in some special time slices (such as *T7* and *T12*).
(3) As shown in Fig. 1(a), only one hub user changes the rank, while other hub users are continuously keeping their ranks even though the numbers of influenced users are dynamically changed. On the contrary, as shown in Fig. 1(b), promotion users keep alternating their ranks along with the changes of topics in different time slices. These results indicate that the hub user usually tends to influence a certain group of users, while the promotion user seems to be more topic-sensitive.

To support the personalized information sharing process, we calculate the contribution and reference attributes based on Eqs. (4) and (5) respectively, and identify the corresponding contribution and reference users for a given specific user. Table 1 shows the best five contribution and reference users for a specific user.



(a) Hub users. (b) Promotion users.

Fig. 1. Changing in number of users influenced by hub users and promotion users.

Table 1
Results of top 5 contribution users and reference users for a specific user.

Contribution user	CA'_{ij}	InD_i	$OutD_i$
maximummiley	0.51	1276	479
thecinemafan	0.49	3150	1549
tatul92	0.45	33	9
srone82	0.32	2549	275
terapiaoral	0.18	186	49
Reference user	RA'_{ij}	InD_i	$OutD_i$
zarahlee91	0.38	151	42
rossandcompany	0.29	113	40
wpkofficial	0.28	134	10
zbleumoon	0.27	93	25
pamuk58	0.26	200	48

It is noted that CA'_{ij} and RA'_{ij} in Table 1 are the normalized contribution attribute CA_{ij} and reference attribute RA_{ij}, calculated by the equations given as follows.

$$CA'_{ij} = \frac{CA_{ij}}{\sqrt{\sum_{i=1}^{N} CA_{ij}^2}} \tag{6}$$

$$RA'_{ij} = \frac{RA_{ij}}{\sqrt{\sum_{i=1}^{N} RA_{ij}^2}} \tag{7}$$

where N indicates the number of identified contribution/reference users.

In addition, the in-degree InD_i, which is the number of head ends adjacent to u_i, and out-degree $OutD_i$, which is the number of tail ends adjacent to u_i, are calculated for comparisons in the DSUN model. We give our observations and discussions for contribution and reference users to a specific user as follows.

(1) Generally, most of the contribution users hold relatively high out-degree. Since the out-degree of each vertex in DSUN model can indicate the popularity of users in regard to a group of other people, this result demonstrates that the identified contribution user can have an effective influence range in terms of their posted helpful information. Moreover, this kind of beneficial influence will be particularly useful in providing individualized support. In this case (see Table 1), comparing with the most popular user (i.e., *thecinemafan*) who held the highest out-degree, the user (i.e., *maximummiley*) with the highest contribution attribute can contribute better to the target user. This result highlights the necessity of finding contribution users for personalized information delivery.

(2) Typically, reference users almost hold relatively lower in-degree and out-degree, comparing with contribution users. This is because the reference attribute is calculated to identify the users who may result in high similarities in terms of their influenced behaviors. The identified users can share each other with more relevant information according to their similar requirements. These results demonstrate that our method can contribute to finding users who can complement with each other, and share the similar experience based on indirect influence in social networking environments.

4.3. Comparison and evaluation

We compare the proposed approach with existing methods, and conduct a questionnaire-based evaluation to demonstrate the applicability and effectiveness of our model and method.

4.3.1. Comparison with two other methods

Two other methods which also took the social influence into account when finding influential users in information dissemination processes, are considered to make comparisons with our method.

First, we compare the identification of hub users with an AdHeat-model-based user identification method [41], which also consider the social influence among a group of users. This AdHeat-model-based method utilized an influence-based propagation model and identified influential users from a social networking graph using the HITS (Hypertext Induced Topic Selection) algorithm. It is noted that to have a better comparison, we employed the DSUN model instead of the HITS, to construct the social graph matrix which would be used in the method in [41].

We extract the top-ranked identified users in each method (i.e., the hub user, and AdHeat-model-based user in [41]), and analyze the number of people who have been influenced by them, to demonstrate their different outcomes. The ranking of top-10 users in terms of the number of their influenced people are shown in Table 2 respectively. Fig. 2 shows the comparison results of top-20 users according to the influenced users within a group of people. It is noted that the vertical axes on the left indicates the number of influenced people for the histogram of top-20 users in each method, while the vertical axes on the right indicates the grand totals of influenced people for the two methods as shown in the curves.

We give the observations and discussions based on comparison results shown in Table 2 and Fig. 2 as follows.

Table 2
Top-10 users based on influenced people.

Hub user		AdHeat user	
AlfredoFlores	1181	gomzhoran	1021
scooterbraun	1110	JustinsOurHeart	998
CodySimpson	1027	**scooterbraun**	801
ArianaGrande	842	StephLauren	768
AustinMahone	684	slobotski	685
KennyHamilton	640	**CodySimpson**	683
mashable	570	19julia98	562
pattiemallette	490	**KennyHamilton**	500
selenagomez	481	**selenagomez**	471
Trans1110	452	DaniielaaS	127

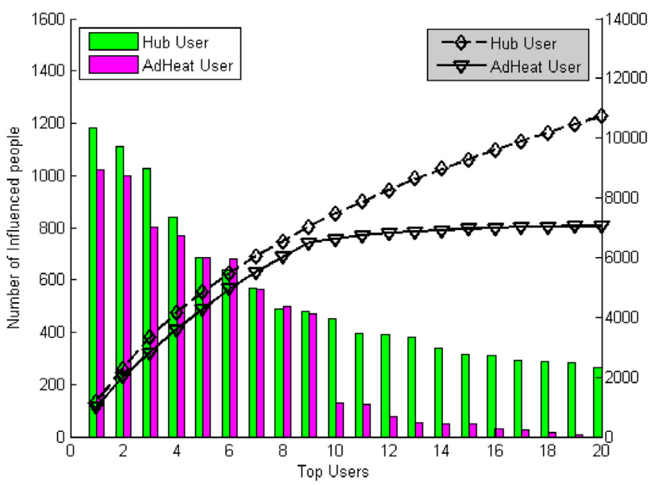


Fig. 2. Comparisons based on influenced users.

(1) Both of the proposed method and AdHeat-model-based method can find a number of users with high influence among a group of people. In particular, four same users are identified in the two lists of top-10 users, which indicates the accuracy and consistency of these two methods. Furthermore, the identification of hub users has a better result not only in the ranking of the four same users, but also in the numbers of their influenced users. This demonstrates the advantage of our proposed method in identifying such kind of important user.

(2) On the whole, the AdHeatmodelbased method performs less effectively than the proposed method. Although the number of people influenced by the top-ranked users in AdHeatmodel-based method can approach those by the identified hub users, the results become worse from the 8th ranked users. The reasons can be summarized as: First, users identified with high influence in the AdHeat method strongly depend on the high node degree in a constructed graph model. Thus, it would be able to identify a few of influential users, but will become an inefficient way to compute the coverage of social influence, which means it is less efficient in finding more influential users across social networks. Second, the influence calculated in the AdHeat model is limited between two adjacent nodes due to the matrix computation process. Thus, it is not suitable to study the social influence diffused among a group of networked people, and results will become worse when dealing with a sparse data situation.

Second, we compare identifications of contribution and reference users with a social-activity-based user identification method [42], which also consider the social influence between two connected users. This social-activity-based method was a probability-based method which calculated the so-called activation of users and further identified active or inactive users based on the analysis of interactive activities.

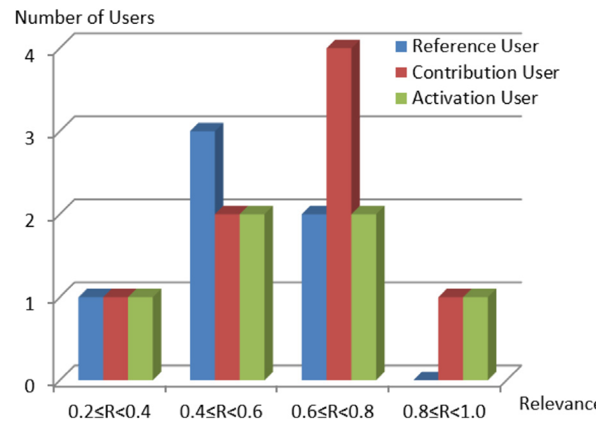


Fig. 3. Comparisons based on the influence-based relevance.

Table 3
Evaluation of three kinds of users.

	Reference user	Contribution user	Activation user
MAE	0.14	0.11	0.23
Hit rate	0.6	0.8	0.6

We evaluate identifications of contribution and reference users, and the activation user in [42] according to the influence-based relevance between these identified users and the target user. Specifically, the set of users who have influenced them is used to calculate the influence-based relevance which can be quantified as follows.

$$Rel_i = \frac{\sum_{j=1}^{K} IF_{ji} * w_{jt}}{\sum_{j=1}^{K} IF_{ji}} \tag{8}$$

where u_t indicates a given target user, and u_i one kind of identified user (i.e., contribution user, reference user, or activation user). u_j belongs to the top-K users who have influenced the identified user u_i. IF_{ji} indicates the Influence Frequency of u_j to u_i. $w_{jt} = \frac{IF_{jt}}{\sqrt{\sum_{j=1}^{K'} IF_{jt}^2}}$ indicates the weight according to the Influence Frequency of u_j to u_t, if u_j belongs to the top-K' users who have influenced the target user u_t.

Given a specific user as the target user in the time slice *T10*, we identify the top-10 contribution users, reference users, and activation users respectively, and set $K = 10$, and $K' = 5$, to calculate the influence-based relevance of these identified users. The results are shown in Fig. 3.

Furthermore, the MAE (Mean Absolute Error) method [43] with the ranking of influence-based relevance is employed to evaluate the efficiency of our proposed methods. Given a target user, the MAE of each kind of identified users is calculated as follows.

$$MAE = \frac{\sum_{i=1}^{|U_{iden}|} \frac{1-Rel_i}{Rank_i}}{|U_{iden}|} \tag{9}$$

where $|U_{iden}|$ indicates the number of one kind of identified users, and $Rank_i$ the corresponding rank-based weight of the users in each type according to the influence-based relevance to the target user.

We make the statistic of non-zero value users calculated in Eq. (6), to demonstrate the hit rate of finding each kind of identified users. The evaluation results of MAE and hit rate are shown in Table 3.

The observations and discussions based on Fig. 3 and Table 3 are given as follows.

Please cite this article in press as: X. Zhou, et al., User role identification based on social behavior and networking analysis for information dissemination, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.04.043

Table 4
Evaluation for finding of hub and promotion users.

Hub user		Promotion user	
Discovered (by algorithm)	Observed (by questionnaire)	Discovered (by algorithm)	Observed (by questionnaire)
AlfredoFlores	**mashable**	**mashable**	**callmeICE**
scooterbraun	**ArianaGrande**	Ogilvy	Daniel Andriyanov
CodySimpson	HE FOLLOWED ME	**Forbes**	HIGH VOLTAGE 500 kV
ArianaGrande	iaia	ValaAfshar	**mashable**
mashable	ZouisILoveYou4ever^^	CrewBiebsUK	Alejandra Morales
pattiemallette	Seun	2morrowknight	**Forbes**
KennyHamilton	HAPPY BDAY FREDO ILY	JBieberFC1	Lady Gaga
AustinMahone	A.anwar	Ms	CharsWings
selenagomez	TeamAngelAlquimia	**callmeICE**	Kristina
Trans1110	MadDaddyFirrzie	britneyspears	Dalvin lion king

(1) As shown in Fig. 3, relevance results of social-activity-based method appear to be relatively even. Our method in identifying contribution users has a better performance which results in more people with the higher relevance. This is because the proposed method considers more on the influential correlation between each pair of users when calculating the contribution from one to another. On the other hand, most of the relevance results of reference users seem to distribute in the intermediate interval. This is thought to be the finding of reference users considers more on the similarity of indirect influence among users.

(2) Based on the evaluation results in Table 3, it is observed that both contribution and reference users perform more efficiently than activation users identified in the social-activity-based method. In particular, the contribution user performs the lowest result according to the MAE estimation. This indicates the advantage of the proposed method in recommending more useful users in the individualized information sharing process.

4.3.2. *Questionnaire-based evaluation for special user finding*

To demonstrate the effectiveness of our method in identifying the special user roles (e.g., the hub user and promotion user), a questionnaire-based evaluation was conducted.

First, we used our proposed method to identify the lists of top-10 hub users and promotion users in the time slice *T10* respectively. We randomly selected five days in which most of these users kept active. Then, for each category of users, e.g., the hub user, we searched the related contents, for instance, the contents posted by the users in the top-10 list or the contents containing their names, from our data set. Once a suitable content is found, we selected 100 contents before and after it respectively, and recorded these contents as one test set. We continuously conducted this process every one hour until ten test sets were found. We discarded those test sets that do not contain enough related users to obtain five test sets for this evaluation experiment.

Second, we asked ten subjects to conduct the evaluation and fill in the questionnaire. Each subject was asked to browse all these ten test sets (five sets for hub user and five sets for promotion user), and then find one hub user or promotion user for each set. In each questionnaire, (1) we explain the definition of hub user (or promotion user), (2) for each test set, we provide a list of no more than 20 candidate users, and require the subjects to find the target user from them, (3) we ask the subjects to finish the choose for each list within 5 min.

Finally, we made the statistic based on these results and ranked a questionnaire-based list of hub/promotion users. Comparison results of both hub and promotion users are shown in Table 4. The observation and discussion for evaluations of finding hub and promotion users are given as follows.

(1) The statistics result shows that the observation experiment can find two out of ten hub users who have been identified by the algorithm. This not only demonstrates the accuracy of the proposed method in identifying these special users, but also indicates the difficulty to find them in an observation way. Besides, it is noted that this evaluation is more sensitive to the active users since we randomly selected test days and generated test sets following the timeline. This explains why the finding of hub users concentrates intensively on a few of certain users, and the number of observed hub users is less than those of observed promotion users.

(2) Considering both the number and corresponding ranks of these identified users, it seems that to identify the hub user is easier than to identify the promotion user through directly browsing the contents. Moreover, according to our monitoring records, the average time to find the promotion user is longer than to find the hub user. This is because according to our definitions, the promotion user is not easy to be observed manually, or it needs to be identified by observing a mass of data. Thus, the result highlights the importance to utilize the proposed method to discover such kind of useful user automatically.

4.4. *Discussions*

Generally, the hub and promotion users are identified to facilitate social information delivery among a group of associated people, while the contribution and reference users are identified to provide the personalized support for information sharing among pairs of users. The experiment results demonstrate the applicability of the DSUN model in discovering these four kinds of users in SNS sites.

For the personal aspect, a set of ranked contribution and reference users, can be selected to benefit a specific user. More precisely, the identified contribution user is enabled to directly provide more valuable information from more relevant social resources. The identified reference user is capable to share the personal experience in a reciprocal way, which can be viewed as an indirect way to help to approach the purpose progressively. Accordingly, all these can efficiently promote the information sharing process between two users, and further involve more other people into the social interactive communication and collaboration.

As for information delivery to a group of users along with social influence, Fig. 4 shows part of users in a certain group which is constructed by one hub user indicated in the center of the graph. Among these users, information related to a hotly discussed topic or issue of public concerns will originate from the hub user in the center, and then deliver one by one through users associated by the directed edges. Furthermore, promotion users indicated in the graph will promote this information dissemination process and assist to deliver the suitable information to more related users.

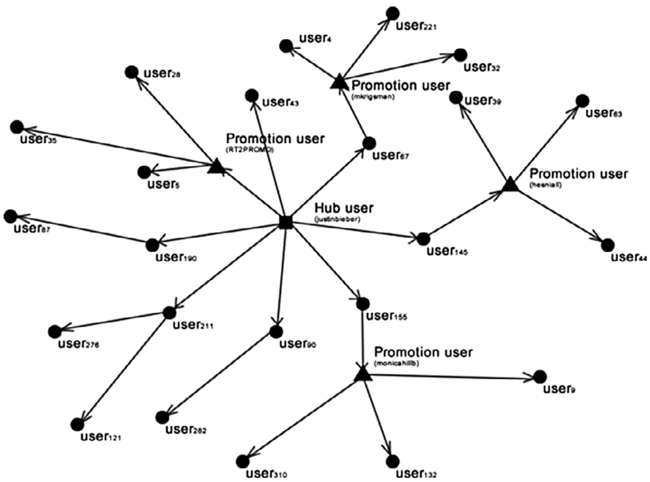


Fig. 4. Image of information dissemination through hub and promotion users.

In particular, the user-generated information, delivered from the hub user, refined by other users, and propagated by promotion users during the diffusing process, can be integrated with a great degree of collective intelligence of the whole community. Following this way, the derived ideas and opinions promoted by the hub user and promotion user will be shared with a range of abilities and knowledge that would not reside in a single individual. More people will thus be attracted and engaged into this networked social group with more diversified information, which could extremely boost the information dissemination in a socialized way.

5. Conclusion

In this study, we have proposed a computational analysis approach to model and identify different user roles within a group of associated people during social information dissemination processes. Models and mechanisms have been developed to describe, analyze, and identify four kinds of special users based on their dynamic connections and influence-related behaviors, which can facilitate information delivery and knowledge sharing in social environments.

First, we introduced a social networking model to represent the dynamic user connections, and basic concepts to describe the influence-related social behaviors. Based on these, we defined and proposed four kinds of special users, who may promote information delivery among a group of users, or assist knowledge sharing between pairs of associated users. A set of user attributes with the corresponding measures was then defined and calculated for the identification of their different roles. Experiment results using Twitter data demonstrated performances of identifying four newly defined users, namely the hub user, the promotion user, the contribution user, and the reference user, during social information dissemination processes. Comparisons with two other methods and the questionnaire-based evaluation proved that our method can efficiently help people find these important users, to support information dissemination processes in social networking environments.

In the future studies, we will go further to elaborate the proposed user attributes and measures, which may be used to enrich the dynamic user profiling in social networking environments. Mechanisms will be improved in a prototype system for user recommendations. More evaluations and experiments will be conducted to improve the proposed methods with better recommendation results in more complex situations.

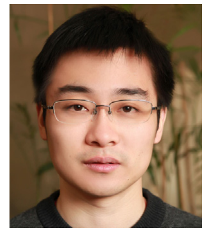
Acknowledgment

The work has been partially supported by 2015 and 2016 Waseda University Grants for Special Research Projects No. 2015B-381 and 2016B-233.

References

[1] Li-Chun Hsu, Kai-Yu Wang, Wen-Hai Chih, Kuan-Yu Lin, Investigating the ripple effect in virtual communities: An example of Facebook Fan Pages, Comput. Hum. Behav. 51 (Part A) (2015) 483–494.
[2] P.J. Carrington, J. Scott, S. Wasserman (Eds.), Models and Methods in Social Network Analysis, Cambridge University Press, 2005.
[3] J.J. Cameron, C.K.-S. Leung, S.K. Tanbeer, Finding strong groups of friends among friends in social networks, in: Proc. 9th IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC, Dec. 2011, pp. 824–831.
[4] S.K. Tanbeer, Fan Jiang, C.K.-S. Leung, R.K. MacKinnon, I.J.M. Medina, Finding groups of friends who are significant across multiple domains in social networks, in: Proc. 5th International Conference on Computational Aspects of Social Networks, CASoN, Aug. 2013, pp. 21–26.
[5] X. Zhou, Q. Jin, User correlation discovery and dynamical profiling based on social streams, in: Proc. 2012 International Conference on Active Media Technology, (AMT2012), in: Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence, Springer, Macao, China, 2012, pp. 53–62.
[6] X. Zhou, W. Wang, Q. Jin, Multi-dimensional attributes and measures for dynamical user profiling in social networking environments, Multimedia Tools Appl. 74 (14) (2015) 5015–5028.
[7] D. Kelly, B. Smyth, B. Caulfield, Uncovering measurements of social and demographic behavior from smartphone location data, IEEE Trans. Hum.-Mach. Syst. 43 (2) (2013) 188–198.
[8] H. Zhao, H. Zhou, C. Yuan, Y. Huang, J. Chen, Social discovery: Exploring the correlation among three-dimensional social relationships, IEEE Trans. Comput. Soc. Syst. 2 (3) (2015) 77–87.
[9] Y. Pan, Z. Wang, Data collection and modeling of personal social behaviors, in: Proc. 2015 International Conference on Service Science, ICSS, Weihai, May 2015, pp. 164–168.
[10] A. Socievole, A. Caputo, S. Marano, Multi-layer sociality in opportunistic networks: An extensive analysis of online and offline node social behaviors, in: Proc. 2015 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS, Chicago, Jul. 2015, pp. 1–8.
[11] S. Agreste, P. De Meo, E. Ferrara, S. Piccolo, A. Provetti, Analysis of a heterogeneous social network of humans and cultural objects, IEEE Trans. Syst. Man Cybern.: Syst. 45 (4) (2015) 559–570.
[12] S. Han, Y. Xu, Friend recommendation of microblog in classification framework: Using multiple social behavior features, in: Proc. 2014 International Conference on Behavior, Economic and Social Computing, BESC, Shanghai, Oct. 2014, pp. 1–6.
[13] S. Adali, J. Golbeck, Predicting personality with social behavior, in: Proc. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, Istanbul, Aug. 2012, pp. 302–309.
[14] X. Ruan, Z. Wu, H. Wang, S. Jajodia, Profiling online social behaviors for compromised account detection, IEEE Trans. Inf. Forensics Secur. 11 (1) (2016) 176–187.
[15] Kuan-Cheng Lin, Shih-Hung Wu, Liang-Pu Chen, Ping-Che Yang, Finding the key users in facebook fan pages via a clustering approach, in: Proc. 2015 IEEE International Conference on Information Reuse and Integration, IRI, Aug. 2015, pp. 556–561.
[16] Lin Xu, Xin Xu, Chenglong He, Shiqiang Zong, Identifying related users in Weibo platform, in: Proc. 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD, Aug. 2015, pp.1412–1417.
[17] Jin Xie, Xing Li, Make best use of social networks via more valuable friend recommendations, in: Proc. 2nd International Conference on Consumer Electronics, Communications and Networks, CECNet, Apr. 2012, pp. 1112–1115.
[18] Fan Jiang, C.K.-S. Leung, S.K. Tanbeer, Finding popular friends in social networks, in: Proc. 2nd International Conference on Cloud and Green Computing, CGC, Nov. 2012, pp. 501–508.
[19] Jing Wang, Zhijing Liu, Hui Zhao, Group recommendation using topic identification in social networks, in: Proc. 6th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC, Aug. 2014, pp. 355–358.
[20] Fan Tang, Bofeng Zhang, Jianxing Zheng, Yajun Gu, Friend recommendation based on the similarity of micro-blog user model, in: Proc. Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, Aug. 2013, pp. 2200–2204.
[21] Ye Na, Zhao Yinliang, Dong Lili, Bian Genqing, Enjie Liu, G.J. Clapworthy, User identification based on multiple attribute decision making in social networks, Commun. China 10 (12) (2013) 37–49.
[22] Jinling Jiang, Hua Lu, Bin Yang, Bin Cui, Finding top-k local users in geo-tagged social media data, in: Proc. 31st IEEE International Conference on Data Engineering, ICDE, Apr. 2015, pp. 267–278.
[23] A. Srinivas, R. Leela Velusamy, Identification of influential nodes from social networks based on Enhanced Degree Centrality Measure, in: Proc. 2015 IEEE International Advance Computing Conference, IACC, Jun. 2015, pp. 1179–1184.

[24] Quan Fang, Jitao Sang, Changsheng Xu, M.S. Hossain, Relational user attribute inference in social media, IEEE Trans. Multimedia 17 (7) (2015) 1031–1044.
[25] Yu-Feng Chou, Hsin-Heng Huang, Ray-Guang Cheng, Modeling information dissemination in generalized social networks, IEEE Commun. Lett. 17 (7) (2013) 1356–1359.
[26] Jing Zhang, Zhanpeng Fang, Wei Chen, Jie Tang, Diffusion of following links in microblogging networks, IEEE Trans. Knowl. Data Eng. 27 (8) (2015) 2093–2106.
[27] Yuhui Feng, Bo Bai, Wei Chen, Information diffusion efficiency in online social networks, in: Proc. 2015 IEEE International Conference on Digital Signal Processing, DSP, Jul. 2015, pp. 1138–1142.
[28] Shiyin Chen, Jiming Hu, Mechanism of information dissemination under socialization networking environment, in: Proc. 2013 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI, Jul. 2013, pp. 108–112.
[29] Bo Han, Jian Li, A. Srinivasan, Your friends have more friends than you do: Identifying influential mobile users through random-walk sampling, IEEE/ACM Trans. Netw. 22 (5) (2014) 1389–1400.
[30] Sheng Wen, Jiaojiao Jiang, Yang Xiang, Shui Yu, W. Zhou, Are the popular users always important for information dissemination in online social networks? IEEE Netw. 28 (5) (2014) 64–67.
[31] Mingsheng Tang, Xinjun Mao, Shuqiang Yang, Haibin Zhu, Policy evaluation and analysis of choosing whom to tweet information on social media, in: Proc. 2014 IEEE International Conference on Systems, Man and Cybernetics, SMC, Oct. 2014, pp. 45–50.
[32] Chao Qi, Hongchang Chen, Chenglong He, Yan Yu, Role analysis of social network users based on neighborhood dissemination capability, in: Proc. 5th IEEE International Conference on Software Engineering and Service Science, ICSESS, June 2014, pp. 1084–1087.
[33] M. Mcpherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. 27 (1) (2001) 415–444.
[34] N. Friedkin, E. Johnsen, Social influence networks and opinion change, Adv. Group Process. 16 (1999) 1–29.
[35] Rashotte, Lisa Slattery, Social influence, in: George Ritzer, J. Michael Ryan (Eds.), The Blackwell Encyclopedia of Sociology, Vol. IX, Blackwell Publishing, Oxford, 2007, pp. 4426–4429.
[36] B. Latane, The psychology of social impact, Amer. Psychol. 36 (1981) 343–356.
[37] Ronald Fischeer, Christin-Melanie Vauclair, Social influence and power, in: C.V. Torres, E.R. Neiva (Eds.), Psicologia Social: Principais Temas e Vertentes, Artmed, 2011, pp. 153–170.
[38] J.M. Jackson, Social impact theory: A social forces model of influence, in: B. Mullen, G.R. Goethals (Eds.), Theories of Group Behavior, Springer Verlag, New York, 1987, pp. 111–124.
[39] Palakorn Achananuparp, Ee-Peng Lim, Jing Jiang, Tuan-Anh Hoang, Who is retweeting the tweeters? modeling, originating, and promoting behaviors in the twitter network, ACM Trans. Manage. Inf. Syst. 3 (3) (2012).
[40] Jaewon Yang, Jure Leskovec, Modeling information diffusion in implicit networks, in: Proc. 2010 IEEE International Conference on Data Mining, (ICDM'10), IEEE Computer Society, Washington, DC, USA, 2010, pp. 599–608.
[41] Hongji Bao, Edward Y. Chang, AdHeat: an influence-based diffusion model for propagating hints to match ads, in Proc. 19th International Conference on World Wide Web, WWW'10, 2010, pp. 71–80.
[42] M. Doo, L. Liu, Probabilistic diffusion of social influence with incentives, IEEE Trans. Serv. Comput. 7 (3) (2014) 387–400.
[43] Ian H. Witten, Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, third ed., Morgan Kaufmann Publishers, San Francisco, CA, 2011.



Xiaokang Zhou received the Ph.D. degree in human sciences from Waseda University, Japan, in 2014. From 2012 to 2015, he was a research associate with the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Japan. He is currently a lecturer with the Faculty of Data Science, Shiga University, Japan. He has been engaged in the interdisciplinary research works in the fields of computer science and engineering, information systems, and human informatics. His current research interests include ubiquitous and social computing, data mining and analytics, behavior and cognitive informatics, human–computer interaction, and information seeking and recommendation. He is a member of IEEE, IEEE CS, and ACM, USA, and IPSJ, and JSAI, Japan.

Bo Wu received the M.S. degree in information science from Jilin University, China, in 2011, and the Ph.D. degree in human sciences from Waseda University, Japan, in 2015. He is currently working as a postdoctoral researcher with the Data Science Laboratory, Kansai University, Japan. His current research interests include information search and recommendation, data mining, and social network analysis. He is a member of IPSJ, Japan.

Qun Jin is currently a tenured full professor of the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Japan. He has been engaged extensively in research works in the fields of computer science, information systems, and social and human informatics. He seeks to exploit the rich interdependence between theory and practice in his work with interdisciplinary and integrated approaches. His recent research interests cover human-centric ubiquitous computing, human–computer interaction, behavior and cognitive informatics, big data, personal analytics and individual modeling, cyber-enabled applications in healthcare, cyber security, and computing for well-being. He is a member of IEEE, IEEE CS, and ACM, USA, IEICE, IPSJ, and JSAI, Japan, and CCF, China.