Analysis of User Network and Correlation for Community Discovery Based on Topic-Aware Similarity and Behavioral Influence

Xiaokang Zhou, *Member, IEEE*, Bo Wu, and Qun Jin, *Member, IEEE*

***Abstract*—While social computing related research has focused mostly on how to provide users with more precise and direct information, or on recommending new search methods to find requested information rapidly, the authors believe that network users themselves could be viewed as an important social resource. This study concentrates on analyzing potential and dynamic user correlations, based on topic-aware similarity and behavioral influence, which may help us to discover communities in social networking sites. The dynamically socialized user networking (DSUN) model is extended and refined to represent implicit and explicit user relationships in terms of topic-aware features and social behaviors. A set of measures is defined to describe and quantify interuser correlations, relating to social behaviors. Three types of ties are proposed to describe and discover communities according to influence-based user relationships. Results of the experiment with Twitter data are used to show the discovery of three types of communities, based on the presented model. Comparison with six different schemes and two existing methods demonstrates that the proposed method is effective in discovering influence-based communities. Finally, the scenario-based simulation of collective decision-making processes demonstrates the practicability of the proposed model and method in social interactive systems.**

***Index Terms*—Community discovery, social behavior analysis, social influence, social network analysis, user correlation.**

I. Introduction

MODELING of user network structures, which aims at discovering patterns in user associations with other users over time [1], has become an increasingly favored research topic for both research and government institutions. Maintaining and refining patterns of diversified social relationships are essential not only for the enlargement of social circles to promote information sharing and knowledge creation, but also for the enhancement of social capital to access current and accurate information, relating to various needs in different situations. A variety of network structure based applications, including analyzing the strength of user relations according to mutual friends [2] and predicting the user's personality trait [3], have emerged in recent years and have been proven to improve user experience within a community, with a common purpose [4]. In this study, a new user networking model, different from other graph models, is developed, based on the calculation of the similarity of topic-aware features and the influence from their interactional behaviors, to discover multiple types of user communities.

The high accessibility of the social networking service (SNS) has led to a new form of cooperative and pervasive data collection and analysis on a personal, local, or global scale. It is now possible to collect and share data, including more information from local environments, in terms of an individual's behavioral habits and daily routines [4]. Therefore, the user-generated data, associated with social behaviors, could be viewed as a valuable social resource for constructing dynamic user networking, not only to motivate collaborations among individuals, but also to facilitate information propagation across communities. Compared with most algorithms that consider user networks as static graphs, the key challenges to dynamic social network analysis are as follows: How to track and identify temporal changes and evolving features of social networks in different time frames over time; how to characterize the nature of dynamically changed user networks within a certain social context, when user associations and related information are involved on a large scale; and how to efficiently measure the strength of dynamic networks, which not only represent the hidden structure among various associated linkages, but also reveal value-added rich information in meaningfully structured communities.

To address these challenges, a model of dynamic user networking is constructed in this study, and the potential user correlation is analyzed from the user-generated data with social behaviors. Based on the authors' previous work [5], significant extensions are made with respect to the following aspects.

i) Characterizing two specific relationships for which the analysis of social behaviors is newly included to construct user networks in social environments.
ii) Measuring the strength of dynamic user correlations, in terms of social influence, along with interactive activities.
iii) Identifying influence-based communities according to dynamic and potential user correlations.

Manuscript received April 22, 2017; accepted June 10, 2017. Date of publication August 29, 2017; date of current version November 13, 2018. This work was supported in part by the 2015 and 2016 Waseda University Grants for Special Research Projects Nos. 2015B-381 and 2016B-233, and in part by the 2016–2018 Masaru Ibuka Foundation Human Sciences Research Project on Oriental Medicine. This paper was recommended by Associate Editor Dr. Bin Guo. *(Corresponding author: Qun Jin.)*

X. Zhou is with the Faculty of Data Science, Shiga University, Hikone 522-8522, Japan (e-mail: zhou@biwako.shiga-u.ac.jp).

B. Wu is with the Advanced Research Center for Human Sciences, Waseda University, Tokorozawa 359-1164, Japan (e-mail: wubo@ruri.waseda.jp).

Q. Jin is with the Faculty of Human Sciences, Waseda University, Tokorozawa 359-1164, Japan (e-mail: jin@waseda.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2017.2725341

In particular, the focus is on identifying two special kinds of social relationships to represent the dynamics among a group of connected users: *similarity based* and *influence based*. Both topic-aware similarity and behavioral influence are seamlessly incorporated to refine dynamic user correlations. Specifically, this study focuses on proposing the following.

i) A generic user networking model, named dynamically socialized user networking (DSUN) that represents a user's topic-aware features and interactional behaviors in a unified way, to identify and describe implicit and explicit user relationships in a certain social context.
ii) A set of measures to analyze, refine, and quantify the social behavior-related user correlations among a group of people, based on topic-aware similarity and behavioral influence.
iii) Three algorithms to extract and describe three types of ties, considering influence-based user relationships, which enable the discovery of different kinds of communities that assist in information dissemination and knowledge sharing.

The remainder of this paper is organized as follows. Section II presents an overview of related works. Section III introduces the basic structure of the DSUN model and discusses the identification and description of two special relationships within the model. Section IV defines a set of measures to describe, analyze, and quantify user correlations, and describes the algorithms to extract different types of ties for community discovery. Section V presents the evaluation and comparison of the proposed method in influence-based community discovery. Finally, Section VI concludes this study with a brief presentation of the authors' perspectives regarding future research.

II. Related Work

A. Social Data Utilization and Influence Analysis

The widespread utilization of social data is becoming increasingly important for the web application improvement, including user feature identification (e.g., computing credit, trust, or identity scores) [6], information diffusion [7], and recommendation systems [8]. For instance, Huang *et al.* [9] proposed a social-network mining method to analyze the user-group-level similarity from heterogeneous social networks. The findings can improve trust prediction between users. Fire *et al.* [10] introduced the "friend" and "community" features in social media to identify missing links, which can improve link prediction in social networks. Similarly, in this study, social contextual information, such as social networks and social tags, is extracted from the user-generated contents along with information behaviors. Moreover, the social-networking-aware features are analyzed among a group of users to recommend suitable communities.

Among these works, which relate to the use of social data, the analysis of social influence has been drawing great attention in recent years. Research based on the modeling of social influence, ranging from identification of influential user [11] to detection of influence diffusion [12], has been conducted. They focused on two kinds of influence, namely topic based and activity based. Tang *et al*. [13] proposed topical affinity propagation to model and identify topic-level social influence across a large social network. Achananuparp *et al.* [14] examined user behaviors, especially the retweeting activities among Twitter users, to model the behaviors relevant to information propagation. They used the model for event detection in Twitter. In this study, the focus is on analyzing and measuring behavioral influence, hidden in social activities. Furthermore, the so-called beneficial influence, which is sensitive to a user's time-varying interests (or needs), is newly defined and measured to refine influence-based user correlations.

B. User Relationship Analysis and Community Discovery

The focus of recent research has been more on the analysis of user correlations, which can result in diversified community discovery and individualized friend or group recommendation. Yu *et al.* [15] presented a method that suggests suitable social groups that considers both the similarities and the relationships of the groups. Wilson *et al.* [16] proposed interaction graphs to quantify user interactions in Facebook to determine whether social links are valid indicators of real user interaction. A group-aware system was developed to support how to organize group activities in the real world, in which a concept model was built to characterize activity-related features of online and offline communities [17]. Aiello *et al.* [18] studied the presence of homophily, in terms of correlations between social and topical features in three online social network systems, to predict the existence of social links based on similarities among user profiles. Lin *et al.* [19] utilized various social contexts and interactions to enable community structure extraction and support community discovery. Zheng *et al*. [20] measured the similarity of user relationships in terms of location histories, and recommended a particular type of group of social media "friends" based on a community's geographical information. Zhang *et al.* [21] presented a unified framework that combines the author-topic model with the user friendship network analysis for community discovery in online social networking sites. A reference framework was proposed to build human-in-the-loop system for large-scale sensing and computing, in which features were characterized to utilize community intelligence in mobile crowd sensing environments [22]. Paliouras [23] focused on discovering user communities from the logs of user activities across the social network, which can be used to model user interests and personalize web applications. Bu *et al*. [24] proposed and extended an autonomy-oriented community mining method based on the structural similarity of dynamic networks.

The existing works focus on discovering a diverse range of social media communities for various purposes, in which the following three major features are considered:

i) the semantics-based features, such as topics, interests, and locations;
ii) the activity-based features, such as individual behaviors and interactions; and
iii) the relationship-based features, such as friendships in Facebook.

Existing research on these features is compared with the method proposed in this study, in terms of the following four

TABLE I
COMPARISON OF RESEARCH WORKS ON COMMUNITY DISCOVERY ACCORDING TO FOUR FACTORS

Research	Semantics	Relationship	Activity	Influence
Yu [15], Zheng [20]	√			
Zhang [21], Bu [24]	√	√		
Wilson [16], Guo [17]		√	√	
Aiello [18], Guo [22]	√		√	
Lin [19], Paliouras [23]			√	
DSUN model	√	√	√	√

factors: semantics, relationship, activity, and influence; and the results are shown in Table I. In this paper, in addition to a comprehensive consideration of topic-based and activity-based similarities, a new feature, namely social influence, is incorporated into the DSUN model to quantify the strength of dynamic user correlations.

III. DSUN MODEL

A. Basic Model Description

Following the definitions given in [5], the basic DSUN model, constructed on the basis of analysis of social streams [25], is extended and refined in this paper to discover and represent potential and dynamic user relationships. The definition is expressed as

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

$U = \{u_1, u_2, \ldots, u_n\}$ is a nonempty set of vertexes in the network model, in which each u_i indicates a unique user. Specifically, $u_i = (ID_i, H_i, A_i)$, in which ID_i indicates the user ID to identify a unique user; H_i indicates the *n*-dimensional interests (or needs) of user u_i during time period *T*; and A_i denotes a set of user attributes, which can enrich the user profiling.

$E = \{e_{ij} =< u_i, UC_T, u_j > |$ if a relationship exists between u_i and $u_j\}$ is a collection of edges that connect the vertexes in *U*. Specifically, vertex u_i is the user on the head of edge e_{ij}, which indicates the potential benefactor who may provide useful information relating to user u_j's current requirements, whereas vertex u_j is the user on the tail of edge e_{ij}, which indicates the beneficiary who may obtain some valuable information from user u_i during T. Consequently, in this connected user networking, the edge $\overrightarrow{u_i u_j}$, extending from u_i to u_j, not only indicates the potential benefit between a pair of users, but also illustrates the direction of helpful information delivery between them.

$UC_T = \{UC_{T_{ij}} |$ if $\exists\ e_{ij} \in E\}$ is a multituple appended on the corresponding edge, which includes a set of measures to describe and quantify various types of user relationships. Each measure is defined to calculate the strength of a specific correlation between user u_i and u_j during T.

According to the sociological theory of homophily [26], individuals are inclined to construct connections with people having some similar features (e.g., interests and needs). Such a connection may be particularly strong when one person can provide help to the other. The stronger the connection between two individuals, the more similar their features could be [27]. Therefore, in the DSUN model, the time-varying user similarity and social influence are taken into account in describing two special kinds of relationships among a group of users, defined as follows.

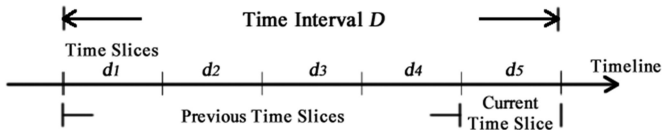


Fig. 1. Illustration of a dynamic division of time slices.

Similarity-based relationship: This describes a kind of implicit relationship, specifically, one that can be identified by extracting time-varying user features and calculating topic-aware similarities. The similarity of features between a pair of users may lead to a stronger connection between them.

Influence-based relationship: This describes a kind of explicit relationship, specifically, one that can be identified by analyzing the user's interactional influence behaviors. For a pair of users, frequent mutual exchange of influence behavior may result in a stronger connection between them.

The similarity-based relationship, which is used to determine whether a pair of users can be connected, is generally considered as the basic relation in the DSUN model. It can also be viewed as a prerequisite to influence-based relationship. Stronger similarity-based relationships between two users may result in stronger influence between them. However, the influence-based relationship, which can be employed to detect useful information sharing among related users, is considered as an essential relation in the DSUN model to illustrate the social perspective of information dissemination. Thus, stronger influence-based relationships between a pair of users may lead to higher similarities between them.

B. Similarity-Based Relationship Analysis

For this study, the user-generated contents in social networking sites (e.g., Twitter) are analyzed, and a user's time-varying interests are extracted to represent topic-aware features. Then, the relevant information is organized to calculate the similarities and to identify the similarity-based relationships among them.

1) Extracting Topic-Aware User Features

Two types of interest, namely *transilient* and *durative*, are introduced to describe topic-aware user features. The transilient interest (H_T) describes a kind of time-evolving interest that changes during some periods of time or is triggered by some hot topics or events. The durative interest (H_D) describes a kind of consistent or inherent interest that is continuously held over a long period of time.

To quantify and distinguish between the transilient and durative interests for a specific user u_i, as shown in Fig. 1, given a time interval D with several dynamically divided time slices $d_t\ (j = 1, 2, 3, 4, 5)$, the transilient interest is extracted from the current time slice d_5, while the previous time slices, d_1, d_2, d_3, and d_4, are treated as references. The durative interest is extracted considering the entire time interval D. Each d_t will hold the same durative interest.

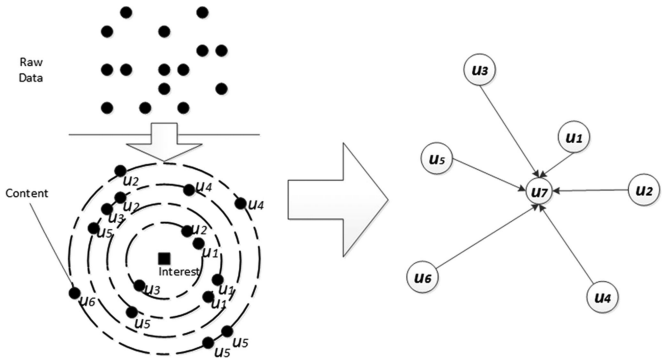


Fig. 2. Conceptual image of building topic-aware similarity-based relationships.

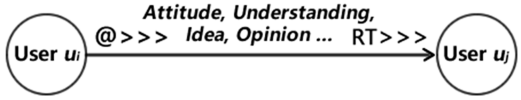


Fig. 3. Illustration of an influence-based relationship.

The TF-IDF (Term Frequency-Inverse Document Frequency) based method is developed to calculate the frequency-based transilient interest $H_{T_{rt}}$ for a keyword k_r ($r = 1, 2, 3, \ldots, |K|$) in a time slice d_t ($t = 1, 2, 3, \ldots, |D|$). $|K|$ is the number of keywords and $|D|$ is the total number of d_t in D. It is expressed by

$$H_{T_{rt}} = \frac{|D| * F_{rt}}{S_i * \sum_{i=1}^{|K|} F_{rt}} \tag{2}$$

D denotes the entire time interval and d_t is a time slice, $D = \{d_t\}$. F_{rt} denotes the frequency of a specific keyword k_i in a time slice d_t. $\sum_{i=1}^{|K|} F_{rt}$ denotes the sum of the frequencies of all the keywords in d_t. $S_i = |\{j \mid \text{if } \exists k_i \in K_j\}|$ denotes the number of time slices in which the keyword k_i is included in K_j, and K_j is the keyword set in d_t.

On the other hand, Eq. (3) is employed to calculate the frequency-based durative interest, H_D:

$$H_D = \frac{F_i}{\sum_{i=1}^{|K|} F_i} \tag{3}$$

where F_i denotes the frequency of the specific keyword k_i over the entire time interval D, and $\sum_{i=1}^{|K|} F_i$ is the sum of the frequencies of all the keywords.

Finally, depending on the values of the dimension coefficient, the top-*n* scored keywords are selected as the specific user interests to represent topic-aware features. Based on the incorporation of these kinds of user interests, the proposed model can timely capture the user's current features in dynamic social contexts referring to the time-varying changes during time slices d_{j-2}, d_{j-1}, etc.

2) Identifying Similarity-Based Relationships

The associative ripple (AR) [25], which was proposed to organize raw social stream data into meaningful content according to a specific user's current interests, is employed to facilitate the calculation of topic-aware similarities. As shown in Fig. 2, in one time slice, divided by the entire time interval, the AR consists of a series of circles, which represent the strength of relations with regard to one specific interest, ranging from inside to outside. The selected data, distributed among different circles, are clustered to the specific interest at the center, according to the strength of relevance. The closer the data to the center, the more relevant the information may be. The AR can be viewed as an integration of a series of ranked data, relating to a specific topic in terms of the user interest. Thus, for a specific user u_i with one of his/her ARs, other users who have provided more relevant data to this AR, will be considered as more related or useful users to him/her. Given a specific AR_i of u_i, this relation or usefulness is quantified in the following:

$$TSR_{ij} = \frac{\sum_{n=1}^{N} w_{c_n} * |\{j \mid \text{if } u_j \in U_{c_n}\}|}{\sum_{n=1}^{N} w_{c_n} * |U_{c_n}|} \tag{4}$$

where $|U_{c_n}|$ denotes the total number of contents provided by all related users and clustered on the *n*-level circle ($n = 1, 2, 3, \ldots, N$) in user u_i's AR, and *N* is the number of circles. $|\{j \mid \text{if } u_j \in U_{c_n}\}|$ denotes the total number of contents provided by user u_j to that circle. w_{c_n} is a frequency-based weight assigned to each circle, which indicates the relevance between *n*-level circle and the center.

As an example, seven users, $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$, are considered as an input user group, to demonstrate how the proposed model works. Given a generated AR from u_7 with a set of user-related data, distributed in four circles (see Fig. 2), it is assumed that the weight to each circle is assigned from inside to outside as $w_{c_1} = 4$, $w_{c_2} = 3$, $w_{c_3} = 2$, and $w_{c_4} = 1$. In each circle, the distributions of data, which was indexed to users from U, are described, thus: $U_{c_1} = \{u_1, u_2, u_3\}$, $U_{c_2} = \{u_1, u_1, u_5\}$, $U_{c_3} = \{u_2, u_3, u_4, u_5\}$, and $U_{c_4} = \{u_2, u_4, u_5, u_5, u_6\}$. For instance, the similarity-based relationship between users u_5 and u_7 (see in the center) can be calculated using (5) as follows:

$$TSR_{75} = \frac{4*0+3*1+2*1+1*2}{4*3+3*3+2*4+1*5} = 0.206.$$

This kind of similarity-based relationship, constructed from the analysis of AR, can be calculated accordingly to describe dynamic user connections, along with topic-aware changes among a group of people within a specific time period.

C. Influence-Based Relationship Analysis

1) Definition of Influence Behavior

Wilson defined human information behavior as "the totality of human behavior in relation to sources and channels of information" [28], which can be viewed as an important resource to analyze contextual human interactions with information. Moreover, Friedkin explained how the process of social influence could clarify interpersonal coordination and agreements among actors in a network of interpersonal influence [29]. Thus, it can be considered that users can influence each other through interactions of similar features in communication, work, and social activities. In particular, social network influence can be considered as a process of information transmission and opinion formation [30], which involves repeated information transfer

among interacting individuals across social networks [31]. Therefore, influence behaviors can be analyzed with a computational approach to gain insight hidden in social interactions.

Consequently, the authors focus on the influence-based relationships arising out of interactional information behaviors among a group of connected users in social networking sites. In particular, among the various social behaviors that users exhibit to seek and utilize information across social media, *influencing behavior* and *influenced behavior* are specified for use in analyzing the influence-based relationships.

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

Influenced behavior($\boldsymbol{IdB}_{\overrightarrow{ij}}$) is a set of influence behaviors of user u_j, which indicates whether or not user u_j has been influenced by user u_i. It can be considered as a kind of behavior that indicates whether or not user u_j has received personal information from user u_i.

2) Analyzing Influence-Based Relationships

The social tags in the posted contents are utilized to efficiently identify these two kinds of social behaviors in social networkin g sites. For instance, in Twitter, the social tag "@name" can be utilized to identify the influencing behavior. It means user u_i tends to build a connection or deliver information that may be related to user u_j, when u_i mentions "@u_j" in the tweets. On the other hand, the social tag "RT @name" can be utilized to identify the influenced behavior. It means user u_j has selected and received personal opinions of u_i, when u_j mentions "RT @ u_i" in the tweets.

As illustrated in Fig. 3, the directed edge $\overrightarrow{u_i u_j}$, which is built based on the similarity calculation from user u_i to user u_j, can be viewed as a bridge for information transmission. In this way, the personal attitudes, understandings, ideas, and opinions of u_i are delivered to u_j. Thus, the influencing behavior can be viewed as a force that promotes transmission of information, whereas the influenced behavior can be viewed as the receiving behavior, which indicates that the information has already been transmitted. In addition, the influenced behavior can also be viewed as a driving force for beneficial information propagation. Consequently, these influence-related behaviors, associated with topic-aware user similarities, can describe the dynamic information flow according to the interactive connections between pairs of users.

Therefore, given a pair of users $\overrightarrow{u_i u_j}$ in the DSUN model, the influencing relationship based on influencing behaviors from user u_i to user u_j $RIgB_{\overrightarrow{ij}}$ can be calculated by

$$RIgB_{\overrightarrow{ij}} = \frac{\left|IgB_{\overrightarrow{ij}}\right|}{|IgB_i|} \tag{5}$$

where $|IgB_i|$ denotes the total number of influencing behaviors conducted by user u_i, and $|IgB_{\overrightarrow{ij}}|$ is the total number of influencing behaviors from user u_i to user u_j during a given time period T.

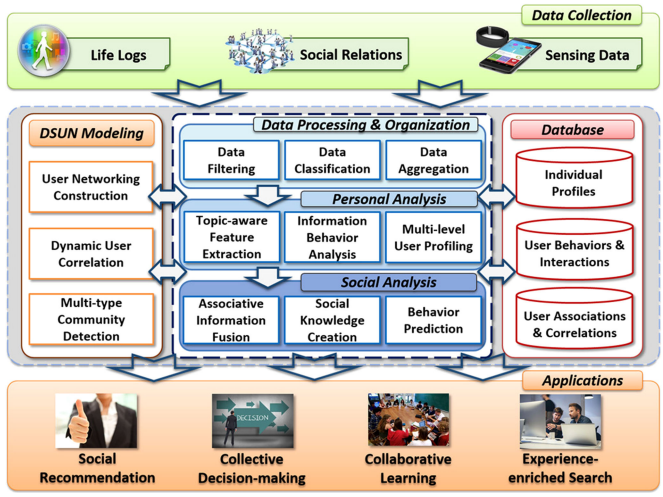


Fig. 4. A conceptual framework for generic application development.

Likewise, given a pair of users $\overrightarrow{u_i u_j}$ in the DSUN model, the influenced relationship based on influenced behaviors from user u_i to user u_j $RIdB_{\overrightarrow{ij}}$ can be expressed by

$$RIdB_{\overrightarrow{ij}} = \frac{\left|InB_{\overrightarrow{ij}}\right|}{|InB_i|} + \frac{\left|IdB_{\overrightarrow{ij}}\right|}{|IdB_j|} \tag{6}$$

where $|InB_i|$ denotes the total number of influence behaviors conducted by user u_i, which have influenced other users during a given time period T, and $|InB_{\overrightarrow{ij}}|$ denotes the total number of influence behaviors conducted by user u_i, which have influenced user u_j during T. $|IdB_j|$ denotes the total number of influenced behaviors conducted by user u_j, and $|IdB_{\overrightarrow{ij}}|$ is the total number of influenced behaviors conducted by user u_j who is influenced by user u_i during T.

D. Generic Application Framework

A conceptual framework of generic applications is presented in Fig. 4, to demonstrate the use of the DSUN model in facilitating the development of human–machine applications in social networking sites.

The framework, as shown in Fig. 4, consists of six major components. In detail, the DSUN modeling module is the core module to build dynamic social networks, to analyze potential user correlations, and to discover multiple types of communities from various relationships. The data collection module is utilized to collect heterogeneous data from multilevel sources (e.g., life logs, social relations, and sensing data), associated with a variety of human activities. The data processing and organization module is used for formulation, analysis, and organization of the collected raw data. The functions of this module include data filtering, classification, and aggregation. Based on these, the personal analysis module is used to characterize the individual user personalities, such as extracting topic-aware features and analyzing influence and interaction behaviors, to achieve multidimensional individual profiling. Moreover, the social analysis module is used to leverage values within the identified

communities, which can benefit information fusion, knowledge creation, behavior prediction, etc. All the generated model and analysis results, such as individual profiles and user correlations, will be saved in the corresponding database. Finally, the application module provides users with a variety of personalized support and intelligent services (such as market navigation, social activity suggestion, and shopping recommendation).

IV. User Correlation Analysis and Community Discovery

A. Measures for User Correlation Analysis

The dynamic correlations between a pair of users are described from three aspects: similarity, influence, and interaction. Thus, the basic correlation can be defined as a three tuple:

$$UC_T = \varphi(TSC,\ BIC,\ DIC) \tag{7}$$

where the topic-aware similarity correlation (TSC), which is employed to describe the basic relationship in the DSUN model, is considered as the prerequisite correlation between a pair of users;

The behavioral influence correlation (BIC), which is calculated based on the influenced behaviors among users in the DSUN model, is employed to describe the influenced correlation between a pair of users (it can also be used to analyze the information or knowledge delivery among a group of users); and

The direct interaction correlation (DIC), which is calculated based on the influencing behaviors among users in the DSUN model, is employed to describe the interactional correlation between a pair of users.

Based on these, two influence-based measures are proposed to calculate social correlations as follows.

Interactional-Influence-based Correlation (IInFC): Given a pair of users u_i and u_j in the DSUN model, the IInFC indicates the interactional influence between u_i and u_j, which considers the direct influencing behaviors between them.

Considering that the users may influence each other during their interaction process, the influencing behaviors are taken into account to calculate this kind of direct interactions between them. The more the influencing behaviors conducted with the counterpart, the higher the interactional correlation would be. Thus, a specific $IInFC_{ij}$ is represented as

$$IInFC_{ij} = \frac{\min\left(\left|IgB_{\vec{ij}}\right|, \left|IgB_{\vec{ji}}\right|\right)}{\left|\left|IgB_{\vec{ij}}\right| - \left|IgB_{\vec{ji}}\right|\right| + 1} \tag{8}$$

Beneficial-Influence-based Correlation (BInFC): Given a pair of users u_i and u_j in the DSUN model, the BInFC indicates the beneficial influence from u_i to u_j, which considers both topic-aware similarity and behavioral influence between them.

In this situation, the availability of the influence-based relationship depends on the construction of the similarity-based relationship. For a pair of users, u_i and u_j, if a beneficial influence from u_i to u_j exists, then these two users are expected to display a similarity in terms of their interests or requirements, so that when u_j is influenced by u_i, the information transferred from u_i can be useful to support the requirement of u_j. Therefore, a specific $BInFC_{ij}$ can be quantified as

$$BInFC_{ij} = TSR_{ji} * RIdB_{\vec{ij}}. \tag{9}$$

Consequently, these specifically identified correlations, which consider the integration of topic-aware user similarities and behavioral influences, can represent the dynamics among a group of networked people, not only in describing the time-varying user associations and interactive behaviors, but also in revealing the potential influence delivery within a dynamic social context.

B. Algorithms for Community Discovery

Using the DSUN model, the authors defined and developed the concept of tie [5] to discover and describe various types of identifiable communities.

Tie: Tie describes a group of users, who undergo confluence following different types of rules, and connect under a certain relationship among them. Users can communicate and benefit from each other through interactional behaviors in a collaborative way within different ties.

Considering the user correlation and user attribute, two basic kinds of ties are kept in focus, namely user-correlation-based and hub-user-based. The user-correlation-based tie can be further categorized into two sub-types, as shown below.

Strong-correlation-based tie: This tie (abbreviated, hereafter, as strong tie) is constructed based on a user's direct interactional behavior.

The algorithm used to discover and construct the strong tie is shown in Fig. 5. Specifically, the strong tie is constructed in accordance with the IInFC between each user, and the discovery process is similar to the breadth-first search process.

Weak-correlation-based tie: This tie (abbreviated, hereafter, as weak tie) is constructed based on a user's indirect influenced behavior from others.

The algorithm used to discover and construct the weak tie is shown in Fig. 6. Specifically, the weak tie is constructed in accordance with the BInFC among users, and the discovery process is similar to the clustering process.

Generally, during information dissemination, some users will become the origin of information dissemination within a community, whereas other users who tend to follow them to derive useful information will be influenced gradually by them. Therefore, such kinds of communities are described and discovered, based on the identification of the hub user, who can be expressed as follows

Hub user: A hub user is one who shares and delivers information continuously 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 with regard to a group of individuals within specific limits.

Following the studies presented in [14] and [32], the diffusion attribute used to identify the hub user can be expressed as follows.

Input: The user set $U = \{u_1, u_2, \ldots, u_n\}$ in the *DSUN* model,
The strong tie *threshold* δ_{st}
Output: The strong tie set $C_{st} = \{C_{st_1}, C_{st_2}, \ldots, C_{st_m}\}$

Create a new C_{st} ;
Create a new queue Q ;
$n = |U|$;
for i =1 to n
  while $U \neq \emptyset$ **do**
    Fetch u_i from U ;
    Create a new C_{st_x} ;
    $U = U - \{u_i\}$;
    for user $u_j \in U$ who is a neighbor of u_i
      $IInFC_{ij} = \frac{\min(|IgB_{\vec{ij}}|, |IgB_{\vec{ji}}|)}{\left||IgB_{\vec{ij}}| - |IgB_{\vec{ji}}|\right|+1}$ *(Eq. (8))* ;
      if $IInFC_{ij} > \delta_{st}$
        Push u_j into Q ;
        $C_{st_x} = C_{st_x} \cup \{< u_i, u_j >\}$;
        $U = U - \{u_j\}$;
      end if
    end for
    while $Q \neq \emptyset$ **do**
      Pop u_l from Q ;
      for user $u_k \in U$ who is a neighbor of u_l
        $IInFC_{lk} = \frac{\min(|IgB_{\vec{lk}}|, |IgB_{\vec{kl}}|)}{\left||IgB_{\vec{lk}}| - |IgB_{\vec{kl}}|\right|+1}$ *(Eq. (8))* ;
        if $IInFC_{ij} > \delta_{st}$
          Push u_k into Q ;
          $C_{st_x} = C_{st_x} \cup \{< u_l, u_k >\}$;
          $U = U - \{u_k\}$;
        end if
      end for
    end while
    $C_{st} = C_{st} \cup \{C_{st_x}\}$;
  end while
end for
Return $C_{st} = \{C_{st_1}, C_{st_2}, \ldots, C_{st_m}\}$;

Fig. 5. Algorithm for generating a strong-correlation-based tie.

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.

Accordingly, the calculation of diffusion attribute for hub user is expressed as

$$DIF_i = \sum_{j \in IB_i} AvgD_j * \log IdU_j \tag{10}$$

where IB_i denotes the set of information behaviors of user u_i; $AvgD_j$ is the average influence depth of an information behavior; and IdU_j is the number of users influenced by this behavior.

Hub-user-based tie: This tie is constructed based on identifying the hub users and other users who have been influenced by them directly or indirectly, and who, together with the hub users, encompass the whole set of users.

The algorithm used to discover and construct the hub-user-based tie is shown in Fig. 7. This tie is constructed in accordance with the analysis of hub users and the influenced behaviors among the related users.

Input: The user set $U = \{u_1, u_2, \ldots, u_n\}$ in the *DSUN* model
The weak tie *threshold* δ_{wt}
Output: The weak tie set $C_{wt} = \{C_{wt_1}, C_{wt_2}, \ldots, C_{wt_m}\}$

Create a new C_{wt} ;
$n = |U|$;
for i =1 to n
  for j =1 to n
    $BInFC_{ij} = TSR_{ji} * RIdB_{\vec{ij}}$ *(Eq. (9))* ;
  end for
  $v_i = \{BInFC_{1i}, BInFC_{2i}, \ldots, BInFC_{ni}\}$;
end for
$V = \{v_1, v_2, \ldots, v_n\}$;
if $U \neq \emptyset$
  Fetch u_1 from U ;
  $V = V - \{v_1\}$;
  $C_{wt} = C_{wt} \cup \{\{u_1\}\}$;
end if
if $V \neq \emptyset$
  for j =2 to n
    Calculate similarity between v_j and each element of C_{wt};
    if similarity $< \delta_{wt}$
      $C_{wt} = C_{wt} \cup \{\{u_j\}\}$;
    else
      Add u_j into the most similar element in C_{wt} ;
    end if
  end for
end if
Return $C_{wt} = \{C_{wt_1}, C_{wt_2}, \ldots, C_{wt_m}\}$

Fig. 6. Algorithm for generating a weak-correlation-based tie.

Input: The user set $U = \{u_1, u_2, \ldots, u_n\}$ in the *DSUN* model
Output: The hub-user-based tie set $C_{hu} = \{C_{hu_1}, C_{hu_2}, \ldots, C_{hu_m}\}$

Create a new C_{hu} ;
$n = |U|$;
for i =1 to n
  $DIF_i = \sum_{j \in IB_i} AvgD_j * \log IdU_j$ *(Eq. (10))* ;
end for
Descend all the users in U according to DIF_i ;
Set the ordered result to U';
while $U' \neq \emptyset$ **do**
  Create a new C_{hu_x} ;
  Fetch the top user u_i from U';
  $n = |U'|$;
  for j =1 to n
    if u_j influenced by u_i
      $C_{hu_x} = C_{hu_x} \cup \{u_j\}$;
    end if
  end for
  $C_{hu_x} = C_{hu_x} \cup \{u_i\}$;
  $U' = U' - C_{hu_x}$;
  $C_{hu} = C_{hu} \cup \{C_{hu_x}\}$;
end while
Return $C_{hu} = \{C_{hu_1}, C_{hu_2}, \ldots, C_{hu_m}\}$

Fig. 7. Algorithm for generating a hub-user-based tie.

V. Experiment and Analysis

A. Dataset

Twitter data were collected by crawling contents generated by users in a Twitter list, named "Awesomesocial," with their followees and followers. This list was selected for two reasons: 1) It was one of the most popular Twitter lists. 2) The tweets posted in this list were written predominantly in English, which is convenient for data analysis. The data were collected from April 2, 2014 to August 15, 2014. The whole time period was

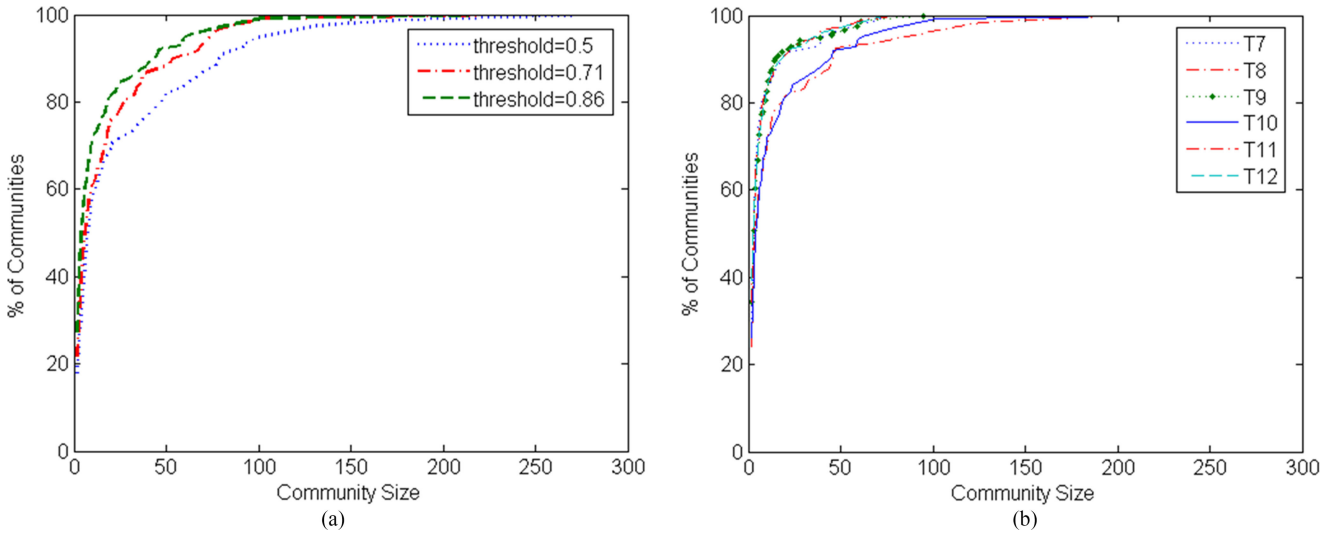


Fig. 8. Community discovery using weak ties. (a) Community discovery under different thresholds. (b) Community discovery in different time slices.

divided into several time slices, depending on the prevailing topic-based trends. Keywords were extracted from the posted contents and changes of their ranks were detected to determine the changes in trends. Finally, 4455 users with 463 644 tweets were used to conduct the experiments, in terms of the 12 time slices generated: *T1*: April 2–8, *T2*: April 9–25, *T3*: April 26–28, *T4*: April 29–May 1, *T5*: May 2–10, *T6*: May 11–12, *T7*: May 13 –16, *T8*: May 17–28, *T9*: May 29–June 14, *T10*: June 15–July 28, *T11*: July 29–August 9, *T12*: August 10–15. The DSUN model was constructed using the first six time slices as the reference periods. For constructing the model, the tweets posted by the users were used to analyze the similarity-based relationships, whereas social tags (e.g., @, RT) in these tweets were extracted to calculate the influence-based relationships.

B. Results of the Experiment on Community Discovery

To discuss the community discovery, based on the three types of ties from the DSUN model, the weak tie was taken as an example to demonstrate the dynamic changes in different time periods. First, the results of discovered communities were evaluated under different thresholds. For the algorithm shown in Fig. 5, the thresholds were set as $\delta_{wt} = 0.5$, 0.71, and 0.86, respectively, to test the results of weak ties in the time slice *T10*. After running the algorithm, the number of discovered communities and that of the members in each community were recorded. The distribution of community size, in terms of the number of communities, is shown in Fig. 8(a). Obviously, a high threshold leads to a fine-grained community discovery. Therefore, to identify more lightweight communities, using the proposed algorithm, the dynamic changes of weak ties were demonstrated in terms of the community size in different time slices, by setting the threshold as $\delta_{wt} = 0.86$. The results are shown in Fig. 8(b).

Approximately 80% of the communities contain less than 40 users. Typically, the number of users in the largest community of any time slice is around 100, and among them, only two communities contain more than 200 users. These statistics indicate that a weak tie is more prevalent in small-scale populations, thereby implying that a small-scale population is better suited for promoting information sharing.

Moreover, the distributions in time slices *T8* and *T10* are quite different from those in other slices, which show only slight changes. This result implies that the influence-based community is topic sensitive, especially when some big events happen. For instance, two big events, namely "European Parliament Elections" and "World Cup" happened in time slices *T8* and *T10*, after investigating the original data.

Likewise, a comparative study of the results of the three different ties in the longest time slice, *T10,* was carried out. Notably, for this experiment, the threshold was set as $\delta_{st} = 0.5$ to generate a strong tie. The two changes, one according to the number of communities, and the other according to the number of users, were considered while analyzing the results. The distribution results of the discovered communities in terms of the strong tie, weak tie, and hub-user-based tie are shown in Fig. 9(a) and those of the users in Fig. 9(b).

Fig. 9(a) shows that most of the strong ties are observed in communities whose size is quite small. Nearly 70% of the strong ties are users in pairs. These results indicate that for this dataset, one user tends to interact directly only with a few other users. The weak tie does not result in very big-sized communities either. Most of the communities contain fewer than 100 users. For the hub-user-based tie, although most of the communities are of normal size, there are several big-sized communities. The biggest community contains over 600 users. Since a hub-user-based tie facilitates information dissemination, the bigger the size of community, the better the information sharing will be.

For the strong tie, the distribution of users along with the community size shows polarization [see Fig. 9(b)], which means that most of the users either interact with another user in pairs, or are involved in a large interactional group. Only about 20% of the users are involved in this kind of community. The results from this dataset indicate that direct interactions are generated

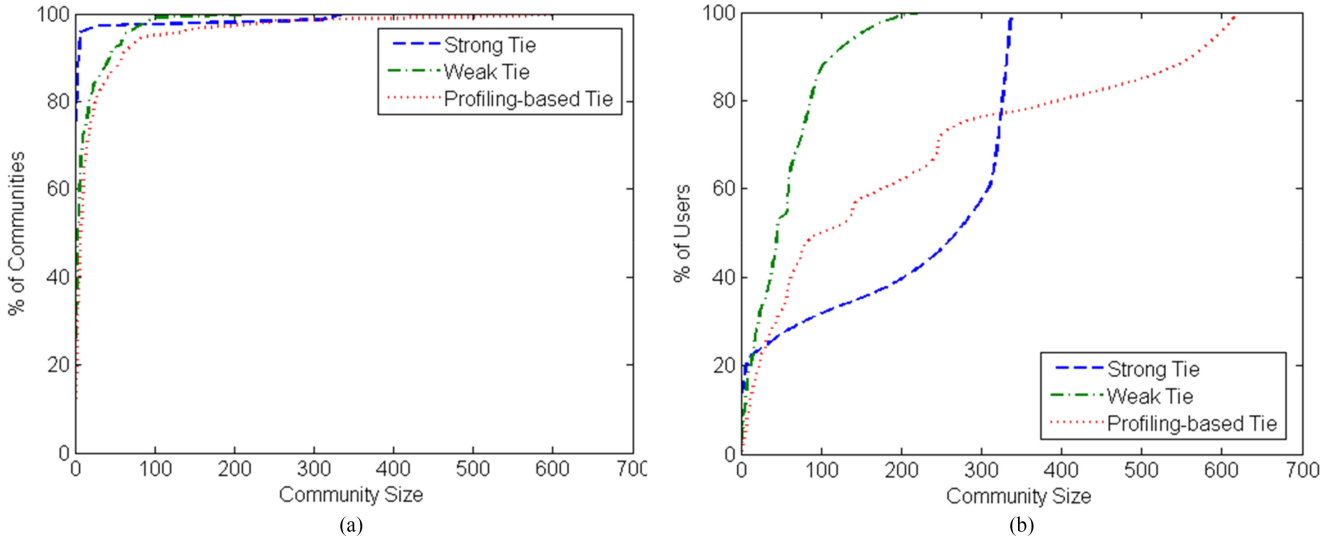


Fig. 9. Distribution of the three types of ties. (a) Distribution of communities. (b) Distribution of users.

within a small group of users, mainly comprising a series of user pairs, in most cases. As for the weak tie, nearly 80% of the users belong to communities of fewer than 100 users. The biggest one in this time slice contains 236 users. This result proves that the proposed method can effectively assign users to several groups of suitable size, with the factor of social influence considered. Compared with the distributions of the former two ties, the hub-user-based tie results show a relatively even distribution of users. These results indicate that the size of the hub-user-based tie fluctuates significantly, depending on the hub user. Unlike the correlation-based tie, whose users are centralized into small-sized communities, half of the users in the hub-user-based tie tend to be centralized into bigger-sized communities.

C. Comparison and Evaluation

1) Evaluation of Community Discovery

To examine the contributions of different factors considered for the proposed method, the community discovery from the weak tie was evaluated using the following schemes.

i) The topic-based community discovery method (TbC): Calculates the similarity TSR_{ji} solely to construct the community.

ii) The behavior-based community discovery method (BbC): Calculates the behavioral similarity solely to construct the community.

iii) The topic- and behavior-based community discovery method (TBbC): Calculates similarities of both topic and behavior, without their existing relationships (e.g., follower and followee), to construct the community.

iv) The topic-, behavior-, and relationship-based community discovery method (TBRbC): Considers the above-mentioned three factors together, to construct the community.

v) The behavioral influence-based community discovery method (BIbC): Calculates the influenced behaviors among users, without considering their topic-based similarities, which directly computes $RIdB_{\vec{ij}}$, to discover the community.

vi) The DSUN model-based community discovery method (DSUNbC): The proposed method in this paper.

All the foregoing methods were evaluated according to top-*k* accuracy-based metrics. For discovering the influence-based community, each discovered community was associated with a group of influencing users, according to the proposed method. Therefore, the extracted interests of members in each of the generated communities were matched with the top-*k* ranked topics, extracted from the corresponding groups of the influencing users. The top-*k* accuracy is quantified as

$$\text{Accuracy} = \frac{\sum_{C_j \in C_{\text{set}}} \sum_{i=1}^{|C_j|} \varphi\left(\text{Match}\left(h_i|_{C_j}, T_{\text{set}}\right) \le k\right)}{\sum_{C_j \in C_{\text{set}}} |C_j|} \tag{11}$$

where C_{set} denotes the set of generated communities; h_i is the interest of each member in the corresponding community C_j; $|C_j|$ is the number of element in C_j; and T_{set} is the set of topics from the corresponding groups of the influencing users. $\text{Match}(h_i|_{C_j}, T_{\text{set}})$ will return the matched position according to the ranking of topics, and $\varphi(\text{Match}(h_i|_{C_j}, T_{\text{set}}) \le k)$ will return 1 if the condition is satisfied; otherwise, it will return 0.

Specifically, the top-3, top-5, and top-10 accuracy-based evaluations were compared in terms of the longest time slice *T10*. The results are shown in Fig. 10(a). Furthermore, the top-10 accuracy was selected for the evaluation in terms of time slices $T7-T12$, and the results are shown in Fig. 10(b).

In discovering the weak-tie-based community, the method TbC, considering only the topic-based similarity, performed least effectively among all the methods used. This was attributed to the sparsity of data collected from Twitter. Extracting time-varying topics or interests is extremely difficult in a time interval, which is either too short or too long. This underlines the need

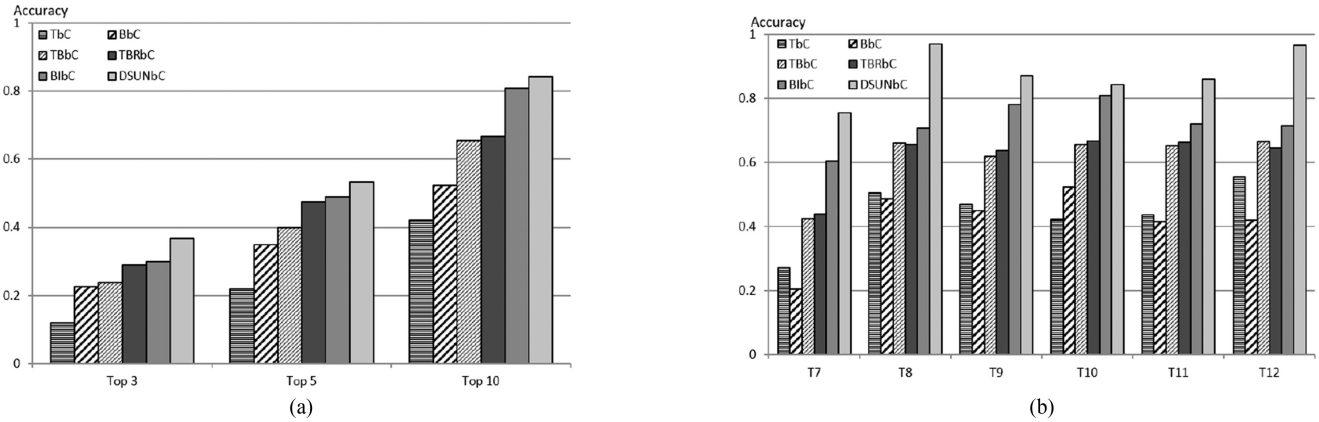


Fig. 10. Comparison of different schemes. (a) Results based on top-*k* accuracy for different schemes. (b) Results based on top-10 accuracy in different time slices.

to consider other factors for improving the results under similar experiment conditions.

The behavior analysis based method could achieve higher accuracy with an increase in time period, although the method BbC that considered only behavior similarity could achieve only lower accuracy. In contrast, the behavioral influence-based method BIbC achieved a better result. In particular, the result is almost the same as that obtained by using the DSUN model-based method DSUNbC for the longest time period. This result indicates the importance of analyzing the influence among the interactional users.

The TBbC and TBRbC methods gave almost the same results for the six time periods. First, this is because the existing relationship in Twitter is not quite as important as that in other SNS environments, such as Facebook. Second, this result indicates that the existing relationship works only as a feeble factor in constructing such kind of influence-based community.

The result of the DSUN model based method DSUNbC looks stable for all the time periods, with an average accuracy of 87.7%. In addition, the results for *T8* and *T12* are remarkably good (approximately 96%).

2) Comparison With Two Related Works

Most of the existing methods focus on discovering communities for a variety of purposes. Thus, to demonstrate the effectiveness of the proposed method, the following two methods, which consider both topic and influence issues, are chosen for comparison.

i) Latent Dirichlet allocation (LDA) based community discovery [33]: This is a probabilistic method that obtains the topic-based community by running an LDA model. For a better comparison to discover the influence-based community, the "RT" and "RT by" were employed, instead of the "follower" and "followee" relationships considered in the method used in [33].

ii) Influencer based community discovery [34]: This is a method that identifies a set of influencing users as the core users in a pre-extracted topic based community. Note that in this method, users who are influenced by the core users would constitute the influence-based community for comparison with the proposed method.

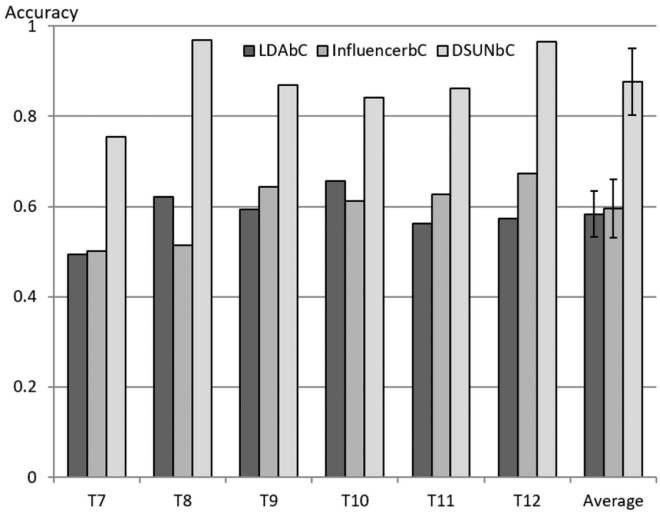


Fig. 11. Comparison of the proposed method with two related works.

The results of comparison based on top-10 accuracy are shown in Fig. 11. Among the three methods including the proposed one, the performance of LDA-based method is least effective. Although the accuracy of the results is over 60% (in *T8* and *T10*), sporadically, it is only 49% in a shorter time slice. This result can be explained in two aspects. First, the LDA-based method was designed to discover topic-sensitive communities. Although its sensitivity to influence was improved in this comparison experiment, it still failed in discovering such kind of influence-based communities. Second, the data space becomes extremely sparse in some short time slices, a situation in which the LDA model is not sufficiently suitable to deal with. That is why the results improved when obvious topics were identified in some specific time slices, but worsened in some short time slices. This finding underlines the need for a behavioral analysis solution to overcome the problem.

Fig. 11 also shows the average accuracy with the corresponding standard deviation for the three methods. Overall, the influencer-based method shows better performance than the LDA-based method because the former identifies the influencers after extracting topic-based communities. However, preidenti-

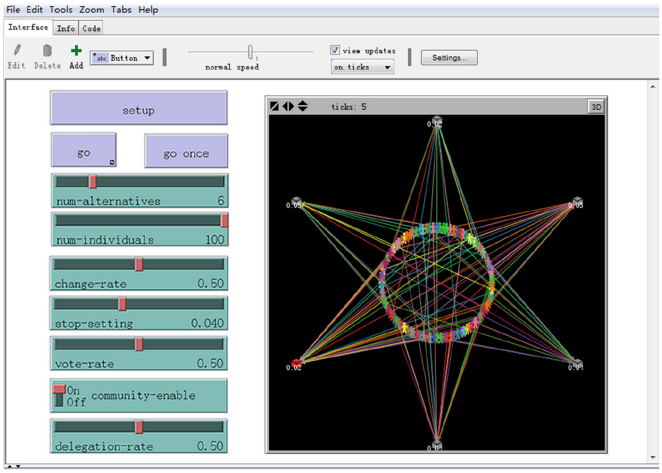


Fig. 12. Snapshot of a decision-making scenario.

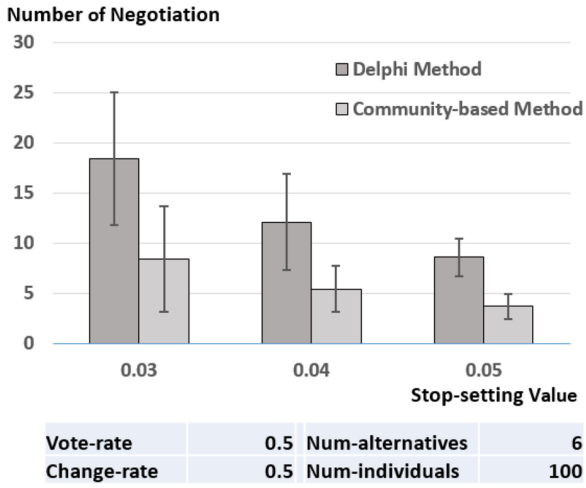


Fig. 13. Comparison with Delphi method.

fied influencers may not be sufficient for dynamic discovery of communities. Because the influencer-based method utilizes only the influence information among users to form the "core groups," resulting in loss of information of similarity among users while delivering helpful information, the proposed method in this study performs better in terms of beneficial influence analysis.

D. Application-Scenario-Based Simulation

The proposed model was applied in collective decision-making processes [35] to evaluate its operability in practical application systems. NetLogo [36], an integrated agent-based programmable modeling environment, is widely used to simulate applications in interactive systems. A NetLogo-based application was developed to simulate how the identified influence-based communities can be used to support collective decision-making processes. This application is capable of exemplifying collective decision-making from multiple persons to multiple alternatives, with controllable vote rate and change rate, in voting–negotiation–consensus processes. A snapshot of the simulation is shown in Fig. 12.

The Delphi method [37], an established management approach to building consensus without any need for face-to-face meetings, was employed as the baseline method for this simulation. The discoveries of weak tie and hub-user-based tie were utilized to improve the Delphi-based decision-making process in two aspects. First, the factor of hub user was characterized as an additional weight to influence the voting process. That is, given a group of users who are involved in a specific decision-making process, each individual identified as a hub user will hold a higher voting weight. The application system will then assign them different additional voting weights, from high to low, according to their corresponding calculated attribute values in terms of hub users. Second, delegation [38] will be done based on the influence among people from an identified weak tie, to accelerate the negotiation process, and finally, reach a consensus. That is, a delegation rate is added while using the community-based method, which means that an individual can delegate to another person, who has similar influence within a community, discovered from weak ties during negotiation.

From the collected dataset, 100 active users were randomly identified to perform collective decision-making processes in the application simulation. From these participants, the communities were identified using the weak tie and hub-user-based tie, and the number of negotiations of the community-based method was compared with that of the Delphi method. The variance in the voting results by all the involved users for each alternative was calculated as the threshold to evaluate the stop-setting value [39]. For a comparative study, both the vote rate and change rate were set to 0.5, the number of alternatives to six, and the stop-setting value to 0.03, 0.04, and 0.05, respectively, as shown in Fig. 13, which shows the average number of negotiation for ten times with the corresponding standard deviation. The results demonstrate that the community-based method can help achieve a consensus with a fewer number of negotiations, as compared with the Delphi method. This result establishes the effectiveness of utilizing influence-based communities, identified by the proposed method, in facilitating collective decision-making processes.

E. Discussion

Generally, in a social networking model, the user-correlation-based ties are constructed according to structural features, in terms of dynamic and potential user correlations, associated with direct or indirect influence. On the other hand, the hub-user-based tie is constructed based on specific user attributes with regard to the scope of the influence among a group of connected users.

Based on the results of the experiment presented above, it is found that communities discovered from the strong tie can be used to facilitate collaborative works, along with better interactions and communications among users. Users involved in strong ties can have better discussions, which can assist them in social information seeking or sharing in interactive application systems. However, detecting or forming such kind of big-sized community is not easy in some situations because most people tend to conduct direct interactions with only a few others.

On the other hand, communities discovered from the weak tie can promote individual information acquisition and personal experience sharing among a group of people who may not have direct interactions with each other. Comparative studies have proven that the proposed method can accurately identify such kind of influence-based community because it considers behavior analysis in the social networking model. The application-scenario-based simulation demonstrates that this method can be used to effectively improve the voting and negotiation efficiency in collective decision-making processes.

In addition, communities discovered from the hub-user-based tie can be useful in promoting influence-based information delivery and knowledge creation in a certain social context. Typically, the discovery process of hub-user-based ties relies mainly on user selections. Different orders of choosing hub users will lead to different results of generated communities. With the proposed algorithms, the hub users are to be selected based on their high to low diffusion attributes, to involve more active users in the interactional process. Results of the experiment indicate that such kind of identified community holds a larger number of individuals than the other two kinds of communities do, which is advantageous for exchanging information in social interactive systems.

VI. Conclusion

In this study, a unified approach was proposed to model, analyze, and quantify interuser correlations, based on the user-generated data and social behaviors from social networking sites. Mechanisms were developed to discover communities among a group of users, which could facilitate information seeking and knowledge sharing in the same networking environments.

The main findings of this study are summarized as follows.

First, the DSUN model for dynamic user networking was extended and refined in accordance with two special relationships: similarity-based to represent implicit relationships based on topic-aware features, and influence-based to represent explicit relationships based on interactional behaviors.

Second, a set of measures was proposed to describe and quantify the interuser correlations, relating to social behaviors. Algorithms were developed to analyze and identify three types of ties for community discovery. Results on community discovery, using Twitter data, were given to demonstrate that three types of communities, namely the strong tie, weak tie, and hub-user-based tie, could be dynamically discovered to satisfy different requirements of users.

Comparative studies, including the evaluation of six different schemes and comparison with two related works, were detailed to demonstrate the effectiveness and accuracy of the proposed model and method for discovering influence-based community.

Finally, an application-scenario-based simulation, carried out in a collective decision-making system, was provided to show the practicability of the proposed method in improving the social collaborative efficiency.

In the future studies, the authors will focus on in-depth understanding of individual and collective intelligence of users and communities. For further evaluation of the proposed method, the authors plan to conduct more experiments to optimize the coefficients in equations, to better adjust the weights of each factor, and adapt them for more complex situations.

Acknowledgment

The authors are also deeply grateful to Dr. R. Gray, Emeritus Professor of Waseda University, for his invaluable editorial comments. For the first author, part of this work was conducted for his Ph.D. thesis at Waseda University.

References

[1] Y. Liu, J. Venkatanathan, J. Goncalves, E. Karapanos, and V. Kostakos, "Modeling what friendship patterns on Facebook reveal about personality and social capital," *ACM Trans. Comput.–Human Interact.*, vol. 21, no. 3, 2014, Art. no. 17.

[2] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 211–220.

[3] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, "Friends don't lie: Inferring personality traits from social network structure," in *Proc. ACM UbiComp*, 2012, pp. 321–330.

[4] K. Shilton, "Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection," *Commun. ACM*, vol. 52, no. 11, pp. 48–53, 2009.

[5] X. Zhou and Q. Jin, "User correlation discovery and dynamical profiling based on social streams," in *Proc. 2012 Int. Conf. Active Media Technol.*, Macao, China, 2012, pp. 53–62.

[6] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," *ACM Trans. Knowl. Discovery Data*, vol. 5, no. 1, 2010, Art. no. 6.

[7] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Detecting changes in information diffusion patterns over social networks," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 3, 2013, Art. no. 55.

[8] H. Ma, T. C. Zhou, M. R. Lyu, and I. King, "Improving recommender systems by incorporating social contextual information," *ACM Trans. Inf. Syst.*, vol. 29, no. 2, 2011, Art. no. 9.

[9] J. Huang, F. Nie, H. Huang, Y. C. Tu, and Y. Lei, "Social trust prediction using heterogeneous networks," *ACM Trans. Knowl. Discovery Data*, vol. 7, no. 4, 2013, Art. no. 17.

[10] M. Fire, L. Tenenboim-Chekina, R. Puzis, O. Lesser, L. Rokach, and Y. Elovici, "Computationally efficient link prediction in a variety of social networks," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, 2014, Art. no. 10.

[11] X. Tang and C. C. Yang, "Ranking user influence in healthcare social media," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 4, 2012, Art. no. 73.

[12] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Trans. Knowl. Discovery Data*, vol. 5, no. 4, 2012, Art. no. 21.

[13] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Paris, France, 2009, pp. 807–816.

[14] P. Achananuparp, E. Lim, J. Jiang, and T. Hoang, "Who is retweeting the tweeters? Modeling, originating, and promoting behaviors in the twitter network," *ACM Trans. Manage. Inf. Syst.*, vol. 3, no. 3, 2012, Art. no. 13.

[15] J. Yu, X. Jin, J. Han, and J. Luo, "Collection-based sparse label propagation and its application on social group suggestion from photos," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, 2011, Art. no. 12.

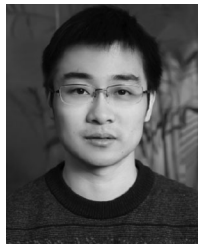
[16] C. Wilson, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "Beyond social graphs: User interactions in online social networks and their implications," *ACM Trans. Web*, vol. 6, no. 4, 2012, Art. no. 17.

[17] B. Guo, Z. Yu, L. Chen, X. Zhou, and X Ma, "MobiGroup: Enabling lifecycle support to social activity organization and suggestion with mobile crowd sensing," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 390–402, 2016.

[18] L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer, "Friendship prediction and homophily in social media," *ACM Trans. Web*, vol. 6, no. 2, 2012, Art. no. 9.

[19] Y. Lin, J. Sun, H. Sundaram, A. Kelliher, P. Castro, and R. Konuru, "Community discovery via metagraph factorization," *ACM Trans. Knowl. Discovery Data*, vol. 5, no. 3, 2011, Art. no. 17.

[20] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W. Ma, "Recommending friends and locations based on individual location history," *ACM Trans. Web*, vol. 5, no. 1, 2011, Art. no. 5.
[21] Z. Zhang, Q. Li, D. Zeng, and H. Gao, "User community discovery from multi-relational networks," *Decis. Support Syst.*, vol. 54, no. 2, pp. 870–879, 2013.
[22] B. Guo, Z. Wang, Z. Yu, Y. Wang, Neil Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Comput. Surveys*, vol. 48, no. 1, pp. 1–31, 2015.
[23] G. Paliouras, "Discovery of web user communities and their role in personalization," *User Model. User-Adapt. Interact.*, vol. 22, no. 1/2, pp. 151–175, 2012.
[24] Z. Bu, Z. Wu, J. Cao, and Y. Jiang, "Local community mining on distributed and dynamic networks from a multiagent perspective," *IEEE Trans. Cybern.*, vol. 46, no. 4, pp. 986–999, Apr. 2016.
[25] X. Zhou, N. Y. Yen, Q. Jin, and T. K. Shih, "Enriching user search experience by mining social streams with heuristic stones and associative ripples," *Multimedia Tools Appl.*, vol. 63, no. 1, pp. 129–144, 2013.
[26] M. Mcpherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.
[27] M. Granovetter, "The strength of weak ties: A network theory revisited," *Sociol. Theory*, vol. 1, pp. 201–233, 1983.
[28] T. D. Wilson, "Human information behavior," *Inf. Sci.*, vol. 3, no. 2, pp. 49–56, 2000.
[29] N. E. Friedkin, *A Structural Theory of Social Influence*, Cambridge, U.K.: Cambridge Univ. Press, 1998.
[30] P. M. DeMarzo, D. Vayanos, and J. Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *Q. J. Econ.*, vol. 118, no. 3, pp. 909–968, 2003.
[31] M. O. Jackson and B. Golub, "Naive learning in social networks: Convergence, influence and wisdom of crowds," FEEM Working paper no. 64.2007, 2007.
[32] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," in *Proc. 2010 IEEE Int. Conf. Data Mining*, Washington, DC, USA, 2010, pp. 599–608.
[33] G. Zhao, M. L. Lee, W. Hsu, W. Chen, and H. Hu, "Community-Based user recommendation in uni-directional social networks," in *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, San Francisco, CA, USA, 2013, pp. 189–198.
[34] M. Kardara, G. Papadakis, T. Papaoikonomou, K. Tserpes, and T. Varvarigou, "Influence patterns in topic communities of social media," in *Proc. 2nd Int. Conf. Web Intell., Mining Semantics,* Craiova, Romania, 2012, Art. no. 10.
[35] M. A. Rodriguez *et al.*, "Smartocracy: Social networks for collective decision making," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, Washington, DC, USA, 2007, pp. 90–100.
[36] U. Wilensky, and W. Rand, *An Introduction to Agent-Based Modeling: Modeling Natural, Social and Engineered Complex Systems With NetLogo*. Cambridge, MA, USA: MIT Press, 2015.
[37] M. R. Geist, "Using the Delphi method to engage stakeholders: A comparison of two studies," *Eval, Program Plan.*, vol. 33, pp. 147–154, 2010.
[38] S. Alonso, I. J. Perez, F. J. Cabrerizo, and E. Herrera-Viedma, "A consensus reaching model for web 2.0 communities," in *Modeling Decisions for Artificial Intelligence*, vol. 5861, V. Torra, Y. Narukawa, and M. Inuiguchi, Eds. New York, NY, USA: Springer, 2009, pp. 247–258.
[39] J. P. Clarys, J. Tresignie, A. Scafoglieri, and E. Cattrysse, "Monitoring and classifying evidence-based workload for profiling manual handling occupations," in *Proc. IEEM 2011*, pp. 377–381, 2011.



Xiaokang Zhou (M'12) received the Ph.D. degree in human sciences from Waseda University, Tokorozawa, Japan, in 2014.

From 2012 to 2015, he was a Research Associate in the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University. Since 2016, he has been a Lecturer in the Faculty of Data Science, Shiga University, Hikone Japan. He has been engaged in interdisciplinary research works in the fields of computer science and engineering, information systems, and human informatics. His research interests include ubiquitous and social computing, data mining and analytics, behavior and cognitive informatics, and human–computer interaction.

Dr. Zhou is a member of the IEEE CS, and ACM, USA, IPSJ, Japan, and JSAI, Japan.

Bo Wu received the M.S. degree in information science from Jilin University, Changchun, China, in 2011, and the Ph.D. degree in human sciences from Waseda University, Tokorozawa, Japan, in 2015.

He is a Postdoctoral Researcher in the Data Science Laboratory, Kansai University, Suita, Japan. His research interests include information search and recommendation, data mining, and social network analysis.

Dr. Wu is a member of the IPSJ, Japan.

Qun Jin (M'95) received the B.S. Degree in Control Theory and Engineering from Zhejiang University, China, in 1982, the M.S. Degree in Computer Science from Hangzhou Institute of Electronic Engineering and the Fifteenth Research Institute of Ministry of Electronic Industry, China, in 1984, and the Ph.D. Degree in Electrical Engineering and Computer Science from Nihon University, Japan, in 1992.

He is a Tenured Full Professor in the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Tokorozawa, Japan. He was with the Department of Computer Science, Hangzhou Institute of Electronic Engineering, China, from December 1984 to March 1989, the Systems Research Center, INES Corporation, Japan, from April 1992 to March 1995, the Department of Information Science and Intelligent Systems, Faculty of Engineering, Tokushima University, Japan, from April 1995 to March 1999, and the School of Computer Science and Engineering, the University of Aizu, Japan, from April 1999 to March 2003. 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.

Dr. Jin is a member of the IEEE CS, and ACM, USA, IEICE, Japan, IPSJ, Japan, and JSAI, Japan, and CCF, China.