

Research Article

A Topic Space Oriented User Group Discovering Scheme in Social Network: A Trust Chain Based Interest Measuring Perspective

Wang Dong,¹ Wang Peng,^{2,3} Li Meizi,⁴ and Zhang Bo⁴

¹Department of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China

²School of Management, Fudan University, Shanghai, China

³Shanghai Cloud Valley Development Co., Ltd., Shanghai, China

⁴College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

Correspondence should be addressed to Zhang Bo; sh.zhangbo@gmail.com

Received 29 September 2015; Revised 10 February 2016; Accepted 17 February 2016

Academic Editor: Wan Fokkink

Copyright © 2016 Wang Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Currently, user group has become an effective platform for information sharing and communicating among users in social network sites. In present work, we propose a single topic user group discovering scheme, which includes three phases: topic impact evaluation, interest degree measurement, and trust chain based discovering, to enable selecting influential topic and discovering users into a topic oriented group. Our main works include (1) an overview of proposed scheme and its related definitions; (2) topic space construction method based on topic relatedness clustering and its impact (influence degree and popularity degree) evaluation; (3) a trust chain model to take user relation network topological information into account with a strength classification perspective; (4) an interest degree (user explicit and implicit interest degree) evaluation method based on trust chain among users; and (5) a topic space oriented user group discovering method to group core users according to their explicit interest degrees and to predict ordinary users under implicit interest and user trust chain. Finally, experimental results are given to explain effectiveness and feasibility of our scheme.

1. Introduction

Currently, user group in social network site (SNS) has been garnering increased attention in fields of topic related opinion expression and information sharing [1]. Commonly, there is a set of related topics, which interest all members in the user group. Therefore, individual users who maintain high interests in the set of topics would join the group and interact with other members conveniently. By joining topic associated group, users can deliver their attitudes and discuss with other members and share other related information about the topic. In user group, group related information would be shared more rapidly. Thereby, user group impacts users more deeply than other organizations. Generally, a user group which is related to more influential topics would obtain more attentions and have a larger impact in social network. Thus,

how to discover influential user groups to attract more users is a significant problem for social network analysis.

(1) *Theoretical Background and Consideration.* Topic is the primary factor for user group. The topics, which users discuss and communicate around, should have close relations to make sure that all members in the group have most common interests. In addition, the topics must have large impacts in SNS to attract mass users' interests into the user group. From this point, it is indispensable to find out a set of topics that have close relations and large impacts in SNS.

Interest reflects the sense of concern and curiosity about the topics that have the power of attracting or holding users' attentions in SNS. Thereby, user's interest about specific topics is another significant factor to evaluate whether a user has probability to join the group or not. That means the more

he/she is interested in the topics, the more likely he/she is to be a member of the user group.

Additionally, information is propagated in a fission pattern based on a large-scale social relation network which is formed by users' relationships, such as friend relationship and followed relationship [2]. Such propagation pattern of users' relationships will also facilitate detecting and organizing user groups, since there would be more probabilities of maintaining similar interests among users who have closer relationships. That means user relationships in social network provide an important parameter to calculate and evaluate user group discovering.

Based on above consideration, the aim of our work is to find the most influential topics which are related to each other in SNS and then organize users who keep interests in these topics into groups to achieve information sharing through channeling close user relationships.

Currently, scholars cluster topics mainly through topic detection technology to construct network model for user grouping [3–5]. However, not only is user group organization based on users' behaviors, such as sending, forwarding, or accepting, but also it contains implicit effects of social relationship in SNS [6]. It is more probable that information would be shared among mutual trusted users. Therefore, the effectiveness of social relationships cannot be ignored in user clustering and information propagation. Many existing researches have already explored that relationship among people plays key influence on information propagation [7]. The interests of users, which cause the people to be clustered in user groups [8], can also transfer through their trust relationships. Thus, relationship is critical to discover user group accurately in SNS.

Trust reflects user's confidence or faith to others based on his past experiences or other factors. It has been used to measure the closeness of relationships among users and calculate related reliability in social network [9, 10]. Users can pursue their favorite items, news, and other related information about the topics and also be concerned with or even accept information which is related to them or sent by trustworthy persons in SNS [11]. In our consideration, user group is formed by the trustworthiness among users and essentially reflects their confidences towards a specific topic. Such point-to-point trust relationships would bring users together and form group under their common interests. Thereby, we can accurately discover qualified users to organize topic user group based on trust chain in SNS.

(2) *Main Contributions of Proposed Work.* In this paper, we propose a topic space oriented user group discovering scheme based on trust in social network, which is composed of three phases: topic space detection, interest evaluation, and user grouping based on trust chain. Firstly, we address an overview of our scheme and give related definitions, that is, graph model of social network, topic space, user interest, and trust chain. Secondly, we propose a detection method of core topic set (named topic space in this work) through topic impact evaluating and relatedness clustering. Thirdly, we present the user interest evaluation method including explicit and implicit interest degree. Then, we address a user

grouping method for discovering users based on trust chain in SNS. Finally, we perform experimental analysis to verify the effectiveness and feasibility of our method. The main contributions of this work include (1) putting forward a topic space construction method based on topic relatedness clustering and impact evaluation, including influence degree and popularity degree evaluation; (2) setting up a trust chain model by taking user relation network topological information into account with a strength classification perspective; (3) presenting a user interest degree evaluation method which involves explicit and implicit interest degree calculation based on trust chain prediction; (4) proposing a topic space oriented user group discovering method. By this method, core users who have large explicit interest in topic space are grouped according to explicit interest evaluation and ordinary users who have implicit interest are further estimated based on trust chain in social network.

2. Related Work

2.1. *Community Discovering in Network.* Community discovering in network environment is a traditional research area [12–15]. In many existing works, networks including SNS, P2P, or distributed system were characterized by graph theory. In graph theory, machines or users are regarded as a set of vertices and their relationships or communication interactions are described as a set of edges. On this basis, communities with features of small-world network [16] or scale-free network [17] can be defined as induced subgraphs of the network graphs. There are dense and tight links among vertices (nodes) in community, while their relationships are sparse and loose outside of the community [14]. Essentially, community discovering finds out that the vertices have relatively dense links according to the topological structures and graph features of the network. Many efforts have been made for network community discovering. Girvan and Newman proposed a method of detecting community, called GN algorithm, by using the property of community structure. In this structure, network nodes are joined together to form tightly knit groups; however, between groups there are only looser connections [12]. Besides, other methods such as k NN [18] and k -means clustering algorithm [19] are also widely used in discovering methods.

2.2. *Influence Evaluation in Social Network.* In the area of impact evaluation, a lot of researches about the maximized influence of social network have been done [20]. Chen and his colleagues propose a series of works about influence maximization such as greedy algorithm evaluation [21], influence diffusion dynamics and influence maximization [22], and scalable influence maximization under the linear threshold model [23]. In our consideration, relations among topics and their popularities are two significant aspects in influence evaluation. Since topics are not independent in SNS, there should be inherent relations for topics, and influential topics would be likely to link to other more influential topics. In addition, popularity is another explicit dataset showing the influence of a topic directly. Therefore, we take the above two aspects (relation among topics and their popularity degrees)

that are not addressed in traditional works into account to evaluate the influence of social network.

2.3. User Interest Analysis. User interest analysis has been used in many fields, such as online user clustering, recommender system, and service quality evaluation. Zeng et al. [24] proposed a user interest analysis method based on user activities on the web. Li et al. [8] addressed a method based on user interest popularity distribution in recommender system. Hegde et al. [25] presented an approach that automatically assigns tags to places, based on interest profiles and visits or check-ins of users at places. Most of these traditional studies are based on users' explicit data, such as behaviors, profiles, or other related data. That means only explicit interest was analyzed. However, many users keep their interest implicitly and did not express their interests explicitly in SNS. These users expressing their interests implicitly would be lost under those traditional measuring methods. Obviously, it is insufficient to discover users just based on their explicit interest. Implicit interest is another important criterion to find potential users. That is why we take it into account for user interest analysis in this paper.

2.4. Trust Computation. There have been a large number of researches on trust and reputation in the past decades [10, 26, 27]. Many methods, such as summation/average/iteration of past trust rating [26] and Bayesian model [27], have been proposed to optimize one or more aspects of trust computation performance. In addition, the weighted average of ratings method is a typical and widely used method in trust computation [28]. In this method, all trust ratings about the target object are aggregated and then a weighted average of the aggregation is calculated as trust value. In social networks research, there are also many works for trust computing. Javier Ortega and colleagues propose a method to compute a ranking of the users in a social network and propagate both positive and negative opinions of the users [29]. The opinions from each user about others can influence their global trust score. Qureshi et al. propose a decentralized framework and the related algorithms for trusted information exchange and social interaction among users based on the dynamicity aware graph relabeling system [30]. In [31], an extended Advogato trust metric is proposed to facilitate the identification of trustworthy users and diffuse a capacity of a target user throughout personal network. Golbeck proposed TidalTrust that gets trust in social networks using numeric trust values [32]. It utilized the shortest path based on the breadth-first search. Furthermore, TidalTrust can be used to retrieve accurate information from the highest trust adjacent nodes.

Different from traditional network community, topic user group is composed of users interested in the same topic. Members of topic user group might disperse in different locations in network and do not have tight and frequent interactions with each other in past. Therefore, there are the following considerations for topic user group discovering in our view. (1) User's interest degree is a significant factor for measuring whether a user should be detected in the user group. Those who maintain strong interests on the topic are

definitely core member of user group. (2) There are many users who keep high interests on topic and do not express their interests explicitly. These potential members should be recognized in the user group. (3) Interests may be transferred through users' relationships based on their trustworthiness. That is, if there is a pretty high level of trust between two users, they might have great possibility to keep similar interests on the same topic. For example, one of them would have his positive or negative effect on another through their trust relationship. Thereby, trust is a crucial linkage among users to share their interests in common and plays an important role in user group discovering.

3. Overview of Our Scheme

Topic space oriented user group (TUG) is organized by three phases: topic space detection, interest evaluation, and user grouping based on trust chain. The topic space gathers influential and interrelated topics that can attract peoples' attentions and public concerns. Specifically speaking, it is meaningless to detect and organize a user group about inessential and unremarkable topics. User interest reflects how interested he/she will be in the topic. That is, interest degree is the criterion for evaluating and grouping user into a TUG. User relationship model, called trust chain model in this paper, reflects the close degrees among users. Through trust chain, we can measure the probabilities of users' topic interest similarities and then group those with mutual trusts in the same TUG.

The overview of our scheme is shown in Figures 1(a)–1(e) as follows. (a) Topics are linked through their relations (black lines) in SNS. (b) Users are linked via their trust chains (blue lines) in SNS. (c) The impacts and relatedness degree of topics are evaluated according to indicators of topic rank and popularity. Then, influential and close related topics are clustered into topic space (marked in red in Figure 1(c)). (d) The interest degrees of users towards the topic are measured based on explicit or implicit interest. (e) Core users of TUG can be identified according to their explicit interests (the core users are marked in red with their interest degrees in green dash lines in Figure 1(d)). Furthermore, ordinary users of TUG are detected based on selected core users and their trust chains (ordinary users are marked in pale red in Figure 1(e)).

Correspondingly, we address the following definitions in this paper.

Firstly, we introduce the graph theory for modeling the social network formally.

Definition 1 (social network graph model). Social network graph model can be described as $SN = \langle V, E \rangle$, where $V = \{v_1, v_2, v_3, \dots\}$ is the nonempty set of vertexes which denote the users in SNS, while $E = \{e_1, e_2, \dots\}$ is the set of edges which denote user relationships among users.

Through Definition 1, we can describe the trustworthy relationship among users by vertexes (users) and edges (user relationships). That is, if a user v_i keeps a trust relationship with another user v_j , the trust can be described as $e(v_i, v_j) = \langle v_i, v_j \rangle$.

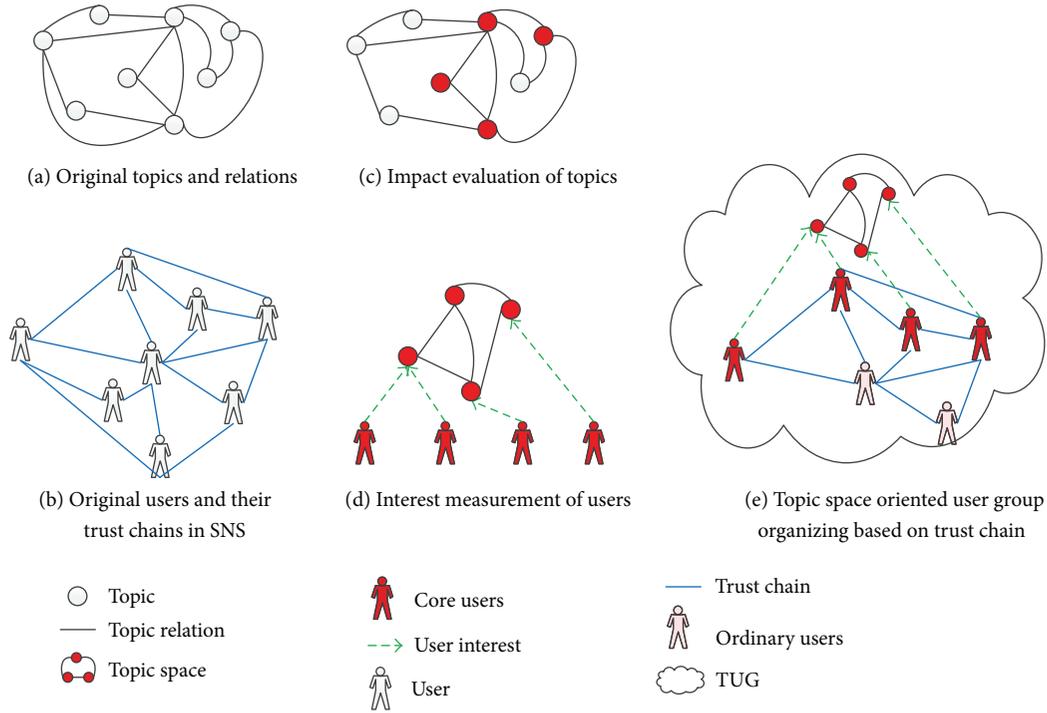


FIGURE 1: Overview of topic user group discovering.

Definition 2 (topic space). Topic space is a set of topics which have large impacts and close relations. Topic space can be defined as $TS = \langle \{\Theta_1, \Theta_2, \dots\}, \text{imp}(TS) \rangle$, where $\{\Theta_1, \Theta_2, \dots\}$ is the set of topics in topic space; Θ_i denotes a topic in the topic space and it can be described as $\Theta = \langle \text{content}, \text{impact} \rangle$, and $\text{imp}(TS)$ is the impact degree of whole topic space.

In the above definition, there are two elements for describing a topic: content and impact; content = $\langle \text{core}, \text{Parent}, \text{Subtopic} \rangle$ which contains core content of topic (core), its subtopic set (Subtopic = $\{st_1, st_2, \dots\}$), and its parent topic set (Parent = $\{pt_1, pt_2, \dots\}$), while impact $\in [0, 1]$ denotes the weight value of the topic influence. Meanwhile, $\text{imp}(TS)$ is the impact degree of the topic space, which is an integrated value by combining all the impact degrees of topics.

Definition 3 (interest degree). Interest degree of user reflects the quantified value of user's interest level about a specific topic or a topic space, which can be used to predict the probability of joining a topic group for users.

In this study, we give two kinds of interests for users as follows: explicit interest and implicit interest. In our consideration, the interests expressed by users' direct behaviors, such as judgments, browsing time, approving, and forwarding, are defined as explicit interest, while the potential feelings or opinions which have not been expressed by users are seen as implicit interest. Apparently, explicit interest can be evaluated directly through users' past behaviors, and the implicit interest can be extracted through users' relationships since users are linked through their relationships in social network and such relationships enable revealing the possible

implicit interests. That is, we can estimate implicit interest through their trustworthy relationships, which are regarded as trust chains in this paper. For example, if user A , who has no direct evidence to express his interest in a topic, keeps a very high trustworthy degree to his friend B who has a strong interest in the topic, we can predict that user A might have a certain interest in the topic. In this example, explicit interest is delivered through users' trust relationship and thus generates implicit interest, which is the underlying rationale for implicit prediction in this work. Correspondingly, users in social network have their interest degree of both topic and topic space.

Additionally, relationship is another significant entity connecting users in SNS. Consequently, it can be used for evaluating closeness degree among users, predicting implicit interest degree, and further organizing users to form groups in this work. To achieve that, we use the notion of trust to reveal the relatedness among users. That is because there would be more probabilities of users to share common interests and then join the same user group if they trust each other. In this work, the trust relationships, including direct relationships and indirect relationships among users, are defined formally as a conception, trust chain, as follows.

Definition 4 (trust chain). Trust chain is a model for describing the direct and indirect link among users. It reflects the trustworthy relationship and can be defined as $\Omega = (\text{Node}, \text{Atom}, \text{Combine}, \text{Class}, \text{Trust})$, where $\text{Node} \subseteq \text{SN} \cdot V$ denotes nonempty set of user nodes in trust chain and the user nodes can be divided into three roles as follows: source user nodes N_S , intermediate user nodes N_I , and target user nodes N_E ; $\text{Atom} \subseteq N \times N \subseteq \text{SN} \cdot E$ denotes the finite set

of atomic trust chain; $Combine \subseteq F \times F \cup F \parallel F$ denotes the combined trust chains which are composed of atomic trust chains and symbols \times and \parallel denote serial trust chain and parallel trust chain, respectively; $Class$ denotes the chain classification of trust chain; and $Trust : F \rightarrow [0, 1] \cup C \rightarrow [0, 1]$ denotes trust value of atomic trust chain or combined trust chain.

In addition, there are two categories for trust chain: by topological route composition and by strength of trust chain. Firstly, since there are different route compositions of indirect trust chain among users, we divide trust chain into four kinds: atom trust chain, serial trust chain, parallel trust chain, and combined trust chain. Meanwhile, to signify the strength of trust chain and define the constraints of trust chain strictly, we classify trust chain as strong and weak trust chain. Details of trust chain will be discussed later. Through Definition 4, the direct and indirect trust chains can be described formally according to the topological composition of the relationships among users.

Definition 5 (topic space oriented user group). In SNS, topic space oriented user group (TUG) can be defined as a 3-tuple $\Delta = (TS, U, I)$, where TS denotes the topic space; $U = \langle U_c, U_o \rangle$ is nonempty set of users in which U_c is the core user set and U_o is the ordinary user set; $I = \{I_{u1}(TS), I_{u2}(TS), \dots\}$ denotes the set of users' explicit or implicit interest degrees of topic space, respectively.

A TUG contains a topic space and a set of users who maintain strong interests to it. With the consideration that many users do not express their interests through explicit behaviors or evidences, their implicit interests can be estimated and evaluated through their trust relationships with others. Therefore, there are the following properties for TUG in this paper:

- (1) For user set U in Δ , $U_c \neq \emptyset$.
- (2) For $U_c \in U$ and $U_o \in U$, $U_c \cap U_o = \emptyset$.
- (3) For each $u \in U$, if s/he has an explicit interest ($EL_u(TS)$) to topic space TS, his/her interest value satisfies condition $EL_u(TS) \geq \eta$.
- (4) For each $u \in U$, if s/he has an implicit interest ($ML_u(TS)$) to topic space TS, s/he must satisfy the condition $(ML_u(TS) \geq \eta) \wedge (\exists u_c \in U_c \rightarrow T_{\Omega}(u, u_c) \geq 0)$.

For future facilitating of the reading, Table 1 presents the nomenclatures proposed in our paper.

4. Topic Space Construction and Impact Evaluation Method

We first address the method of structuring topic space for TUG discovering. As mentioned above, only influential and close related topics can be selected for organizing the topic space, and thus our method of topic space construction is based on evaluation of the relatedness degree and influence degree for topics in social network.

As defined in Definition 2, topic space is composed of a set of topics and each topic can be described as two elements: semantic (content) and impact level (impact). In our consideration, the evaluation of relatedness degree can be measured from semantic perspective, while the impact degree comprises two aspects as influence degree and popularity degree. Accordingly, there are the following three aspects for detecting topic space: relatedness clustering, influence evaluation, and popularity evaluation.

4.1. Relatedness Clustering for Topic Space. Relatedness clustering aims to find out topics which have close relations and then form a strong related topic set. That is, irrelevant topics should be removed from a topic space because they have few correlations with those topics in topic space and might contribute a little to attracting users' interests to join TUG. Hence, we here provide a method called relatedness clustering for topic space.

We propose a factor, denoted as relatedness degree, to describe the closeness of topics' relations. Assume that there are two topics, Θ_i and Θ_j , and their corresponding sample sets, which include topic related posts, comments, or other items, are $Z(\Theta_i)$ and $Z(\Theta_j)$, respectively. Then, the relatedness degree of two topics can be calculated by *Jaccard similarity* as follows:

$$re(\Theta_i, \Theta_j) = \frac{|Z(\Theta_i) \cap Z(\Theta_j)|}{|Z(\Theta_i) \cup Z(\Theta_j)|}. \quad (1)$$

It is noteworthy that the result of relatedness degree is impacted by the sample sets of topics. That means different sample sets would result in different relatedness degree. Therefore, we propose an iterative algorithm for stabilizing the relatedness degree as shown in Algorithm 1.

Based on Algorithm 1, we can get relatedness degree factor of every two topics for measuring their closeness relation. Further, we can detect close related topics and then cluster them based on the relatedness degree factor.

Assume that there is a set of candidate topics, $Can_ \Theta = \{can_ \Theta_1, can_ \Theta_2, \dots, can_ \Theta_n\}$, for discovering topic space. Then, the relatedness degree of every two topics can be calculated based on Algorithm 1. We can get a relatedness degree matrix as follows:

$$R(Can_ \Theta) = \begin{bmatrix} re(can_ \Theta_1, can_ \Theta_1) & re(can_ \Theta_1, can_ \Theta_2) & \cdots & re(can_ \Theta_1, can_ \Theta_n) \\ re(can_ \Theta_2, can_ \Theta_1) & \cdots & \cdots & \cdots \\ \cdots & re(can_ \Theta_i, can_ \Theta_j) & \cdots & \cdots \\ re(can_ \Theta_n, can_ \Theta_1) & \cdots & \cdots & re(can_ \Theta_n, can_ \Theta_n) \end{bmatrix}. \quad (2)$$

```

select  $Z_1(\Theta_i)$  and  $Z_1(\Theta_j)$ ;
get  $re(\Theta_i, \Theta_j)_1$ ;
 $re(\Theta_i, \Theta_j)_2 = 0$ ;
while  $|re(\Theta_i, \Theta_j)_2 - re(\Theta_i, \Theta_j)_1| > \rho$ 
{
  select  $Z_2(T_1)$  and  $Z_2(T_2)$  which satisfy condition  $Z_1(T_1) \cap Z_2(T_1) = \emptyset \wedge Z_1(T_2) \cap Z_2(T_2) = \emptyset$ ;
  get  $re(\Theta_i, \Theta_j)_2$ ;
   $re(\Theta_i, \Theta_j)_1 = \frac{re(\Theta_i, \Theta_j)_1 + re(\Theta_i, \Theta_j)_2}{2}$ ;
   $(Z_1(T_1) = Z_1(T_1) \cup Z_2(T_1)) \wedge (Z_1(T_2) = Z_1(T_2) \cup Z_2(T_2))$ ;
}
return  $re(\Theta_i, \Theta_j) = re(\Theta_i, \Theta_j)_1$ .

```

ALGORITHM 1: Relatedness degree stabilizing algorithm for topics.

TABLE 1: Nomenclature.

Category	Conception or term	Symbol	Description
Social network related terms	Social network graph model	$SN = \langle V, E \rangle$	Formal description of social network (Definition 1)
	Trust chain	$\Omega = (Node, Atom, Combine, Class, Trust)$	Trustworthy relationship among users (Definition 4)
	Topic space oriented user group	$\Delta = (TS, U, I)$	User group organized according to a topic space (Definition 5)
Topic related terms	Topic space	$TS = \langle \{\Theta_1, \Theta_2, \dots\}, imp(TS) \rangle$	A set of topics which have large impacts and close relations (Definition 2)
	User interest	$I = \{I_{u1}(TS), I_{u2}(TS), \dots\}$	The quantified value of user's interest level about a specific topic or a topic space (Definition 3)
Topic space related calculation	Relatedness clustering for topic space	$re(\Theta_i, \Theta_j)$	The closeness degree of topics
	Relatedness degree matrix	$R(Can_ \Theta)$	A matrix of relatedness degree for candidate topics in topic space clustering
	Topic influence rank (TIR) degree of link relation	$TIR_in(p(\Theta_i)_j)$	Influence degree of a topic based on its link relation
	TIR degree of hierarchy relation	$TIR_con(\Theta_i)$	Influence degree of a topic based on its hierarchy relation
	Influence degree value of a topic	$influence(\Theta_i)$	Influence degree of a topic based on TIR
	Popularity of topic	$popular(\Theta_i)$	Popularity level of a topic
Trust chain related calculation	Impact of the topic space	$impact(TS)$	Impact degree of the topic space
	Atomic trust chain	$atom_T(N_1, N_2)$	Trust degree of atomic trust chain
	Serial trust chain	$ser_T(N_S, N_E)$	Trust degree of serial trust chain
	Parallel trust chain	$par_T(N_S, N_E)$	Trust degree of parallel trust chain
	Combined trust chain	$com_T(N_S, N_E)$	Trust degree of combined trust chain
User interest related calculation	User influence	$ef(v)$	The influence degree of a user on his/her trust chain
	Explicit interest	$EI(v \Theta)$	The degree of explicit interests which can be measured through users' direct behaviors or other direct witness evidences
	Implicit interest	$MI(N_E \Theta)$	The degree of implicit interests which reflects the users' potential opinion toward topic

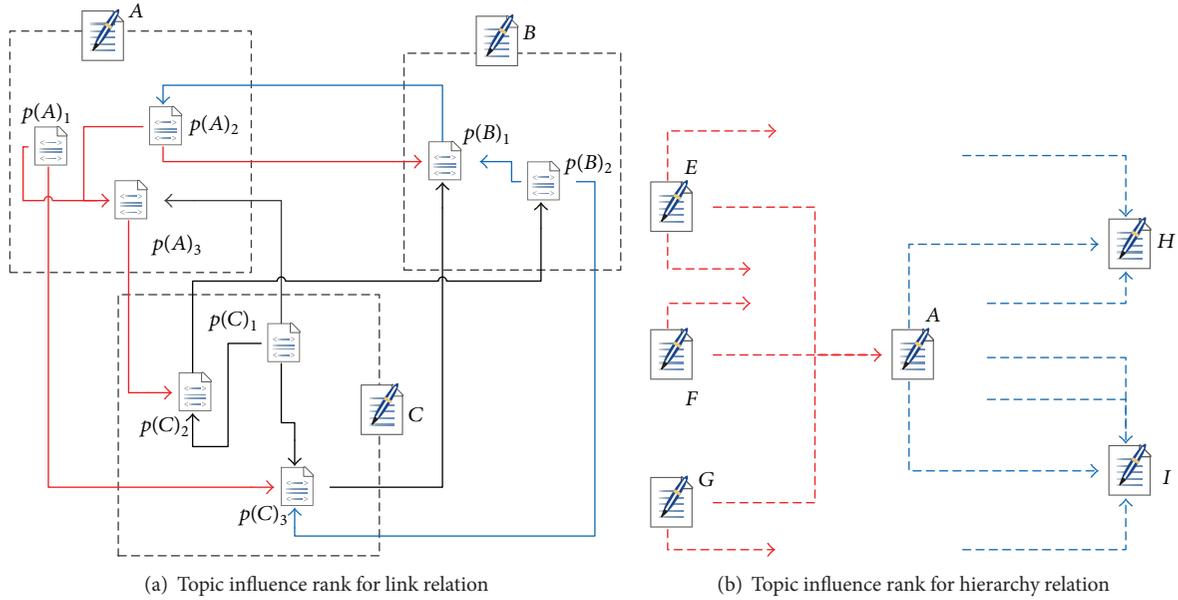


FIGURE 2: Example of topic influence rank evaluation for topics.

According to the relatedness degree matrix, we can see that the candidate topic with maximum value of sum of its column values would be the topic which has the highest relevance with all other candidate topics. Here, we denote the topic with maximum value of sum of its column values as *Topic Space Kernel*. Then, topic space can be clustered based on the *Topic Space Kernel*. We propose a relatedness clustering algorithm for topic space as shown in Algorithm 2.

In Algorithm 2, close related topics are discovered and thus clustered to form TS. The step of *stabilizing*($R(\text{Can_}\Theta)$) indicates utilizing Algorithm 1 to stabilize the relatedness degree values of topics. Also, $\beta \in [0, 1]$ is a threshold which is given in advance.

4.2. Influence Evaluation for Topic Space. According to the core content, parent, and subtopic sets, influence degree of a topic can be evaluated from the above parts. That means influence degree is an integrated value which is calculated based on importance of core content, parent, and subtopic. Here we address a method, called topic influence rank (TIR) method, which is similar to PageRank [33]. Assume that there are different topics in SNS; TIR method works by counting the number and quality of relationships to a topic to determine a rough estimate of how important the topic is. That is, more influential topics are likely to have more relationships with other topics. As shown in Figure 2, there are two kinds of relations between two topics. (1) The first kind is link relation (solid lines in Figure 2(a)). That is, there are semantic relationships among topics and the topics are linked through their inner links in topic oriented web pages. Here, the topic oriented web page means a page whose content is mainly about a specific topic. For example, a page including a text about topic of “music” can be seen as a “music oriented page.” Most topic oriented web pages are categorized manually.

```

Can_Θ = {can_Θ1, can_Θ2, ...}; TS = ∅;
get R(Can_Θ);
stabilizing(R(Can_Θ));
for all can_Θi ∈ Can_Θ do
{
  get sum(can_Θi) = ∑i re(can_Θi, can_Θj);
  if get max(sum(can_Θi)) then kernel = can_Θi;
}
Can_Θ ← Can_Θ - kernel; TS ← kernel;
for all can_Θj ∈ Can_Θ do
{
  if re(kernel, can_Θj) ≥ β
  then TS ← can_Θj, Can_Θ ← Can_Θ - can_Θj;
}
for all ti ∈ TS ∧ can_Θj ∈ Can_Θ do
{
  if re(ti, can_Θj) ≥ β
  then TS ← can_Θj, Can_Θ ← Can_Θ - can_Θj;
}
return TS;

```

ALGORITHM 2: Relatedness clustering algorithm for topic space.

(More specifically, there are the following types: (1) if a topic is included in the title of the text in a page, the page is marked as a topic oriented page; (2) if a topic is included in the keywords or label words of the text in a page, the page is marked as a topic oriented page; (3) the text in a page is marked as a topic oriented one through semantic analysis technologies (owing to length limitation of the paper, detail semantic analysis technologies are discussed in other works). All the data is prepared through the data preprocessing in

this work.) For example, topics of “pollution” and “disease” are in semantic relation since there are inner links among their pages. (2) The second type is hierarchy relation (dotted lines in Figure 2(b)). Such relation is also called parent-child relation. In our consideration, the subtopics or parent topics can bring their contributions to the topic which has semantic containment relations. For example, topics of “pollution” and “air pollution” are in semantic containment relation and the two topics contribute their impact degrees to each other.

We first propose the TIR calculation method of semantic incompatible relation. In this case, all topics have completely different semantics of core contents. Assume there is topic set as $\{\Theta_1, \Theta_2, \dots, \Theta_n\}$, and, for each topic Θ_i , it has a page set as $P(\Theta_i) = \{p(\Theta_i)_1, p(\Theta_i)_2, \dots\}$. Then, for each $p(\Theta_i)_j \in P(\Theta_i)$, it can calculate its TIR degree of link relation, $\text{TIR_in}(p(\Theta_i)_j)$, as follows:

$$\begin{aligned} \text{TIR_in}(p(\Theta_i)_j) &= \frac{\eta}{|\bigcup_{l=1}^n P(\Theta_l)|} \\ &+ (1 - \eta) \sum_{p_k \in \text{Linkin}(p(\Theta_i)_j)} \frac{\text{weight}(p_k) \times \text{TIR_in}(p_k)}{|\text{Linkout}(p_k)|}, \end{aligned} \quad (3)$$

$\text{weight}(p_k)$

$$= \begin{cases} 1 \times \text{TIR_in}(\Theta_K) & p_k \notin P(\Theta_i) \wedge p_k \in P(\Theta_K) \\ 0.8 \times \text{TIR_in}(\Theta_K) & p_k \in P(\Theta_i) \wedge p_k \in P(\Theta_K), \end{cases}$$

$$\text{TIR_in}(\Theta_i) = \frac{\sum_{j=1}^{|P(\Theta_i)|} \text{TIR_in}(p(\Theta_i)_j)}{|P(\Theta_i)|},$$

where $\text{Linkin}(p(\Theta_i)_j)$ is the set of topics that have a link to page $p(\Theta_i)_j$; $\text{Linkout}(p_k)$ denotes the set of pages that have links from page p_k ; and $|\text{Linkout}(p_k)|$ denotes the number of topics in set $\text{Linkout}(p_k)$. Similar to damping factor set in PageRank, we set a parameter η here to describe the probability of topic change. Additionally, we give a weight factor, $\text{weight}(p_k)$, for page $p_k \in \text{Link}(p(\Theta_i)_j)$ to distinguish the importance of pages as follows: (1) if the page p_k belongs to same topic with page $p(\Theta_i)_j$, which means the link from p_k to $p(\Theta_i)_j$ is an inner link, page p_k would contribute less influence than the page which has external link with $p(\Theta_i)_j$ and (2) a page belonging to a more influential topic would bring more contributions to the pages in its $\text{Linkin}()$ set. The example is shown as in Figure 2(a) and we can see that there are three topics (A , B , and C) and their pages and links. Then, $\text{TIR_in}(p(\Theta_i)_j)$ can be calculated iteratively and finally can be convergent to stable values (suppose that η is set as 0.2) as follows:

$$\begin{aligned} \text{TIR_in}(p(A)_3) &= \frac{0.2}{8} + (1 - 0.2) \\ &\cdot \left[\frac{0.8 \times \text{TIR_in}(A) \times \text{TIR_in}(p(A)_1)}{2} \right. \\ &\left. + \frac{0.8 \times \text{TIR_in}(A) \times \text{TIR_in}(p(A)_2)}{2} \right] \end{aligned}$$

$$\begin{aligned} &+ \frac{1 \times \text{TIR_in}(C) \times \text{TIR_in}(p(C)_1)}{3} \Big], \\ \text{TIR_in}(p(B)_1) &= \frac{0.2}{8} + (1 - 0.2) \\ &\cdot \left[\frac{1 \times \text{TIR_in}(A) \times \text{TIR_in}(p(A)_2)}{2} \right. \\ &+ \frac{0.8 \times \text{TIR_in}(B) \times \text{TIR_in}(p(B)_2)}{2} \\ &\left. + \frac{1 \times \text{TIR_in}(C) \times \text{TIR_in}(p(C)_2)}{1} \right], \\ \text{TIR_in}(p(C)_2) &= \frac{0.2}{8} + (1 - 0.2) \\ &\cdot \left[\frac{1 \times \text{TIR_in}(A) \times \text{TIR_in}(p(A)_3)}{1} \right. \\ &\left. + \frac{0.8 \times \text{TIR_in}(C) \times \text{TIR_in}(p(C)_1)}{3} \right]. \end{aligned} \quad (4)$$

In addition, we calculate TIR degree of hierarchy relation. Likewise, we assume that if a topic's subtopics or parents have higher TIR degrees, the topic would get a higher TIR degree. Let the subtopic set of a topic Θ_i be $\text{Subtopic}(\Theta_i) = \{st_1, st_2, \dots, st_K\}$ and let the parent set be $\text{Parent}(\Theta_i) = \{pt_1, pt_2, \dots, pt_M\}$. The TIR value of Θ_i for hierarchy relation, $\text{TIR_con}(\Theta_i)$, can be calculated as follows:

$$\begin{aligned} \text{TIR_con}(\Theta_i) &= \frac{1}{M} \sum_{pt_j \in \text{Parent}} \frac{\text{re}(pt_j, \Theta_i) \times \text{TIR_con}(pt_j)}{|\text{Subtopic}(pt_j)|} \\ &+ \frac{1}{K} \sum_{st_k \in \text{Subtopic}} \frac{\text{re}(st_j, \Theta_i) \times \text{TIR_con}(st_k)}{|\text{Parent}(st_k)|}, \end{aligned} \quad (5)$$

where $|\text{Subtopic}(pt_j)|$ is the number of subtopics of topic pt_j and $|\text{Parent}(st_k)|$ is the number of parent topics of topic st_k . Meanwhile, $\text{TIR_con}(\Theta_i)$ is impacted by the relatedness degree between topic and its subtopic or parent topic. The value of $\text{TIR_con}(\Theta_i)$ could be 0 while a topic has no parent or subtopic. For example, there are three parent topics (E , F , and G) and two subtopics (H and I) of topic A , and then value of $\text{TIR_con}(A)$ in Figure 2(b) can be calculated as follows:

$$\begin{aligned} \text{TIR_con}(A) &= \frac{1}{3} \left[\frac{\text{re}(E, A) \times \text{TIR_con}(E)}{3} \right. \\ &+ \frac{\text{re}(F, A) \times \text{TIR_con}(F)}{2} \\ &\left. + \frac{\text{re}(G, A) \times \text{TIR_con}(G)}{3} \right] \end{aligned}$$

$$+ \frac{1}{2} \left[\frac{\text{re}(H, A) \times \text{TIR_con}(H)}{2} + \frac{\text{re}(I, A) \times \text{TIR_con}(I)}{3} \right]. \quad (6)$$

In summary, influence degree value of a topic by TIR can be calculated based on the above two equations:

$$\text{influence}(\Theta_i) = \frac{1}{2} (\text{TIR_in}(\Theta_i) + \text{TIR_con}(\Theta_i)). \quad (7)$$

4.3. Popularity Evaluation for Topic Space. Popularity is another significant criterion for topic influence evaluation. In our consideration, the underlying assumption of popularity evaluation is that the more popular a topic is, the more influential it would be. Hence, we calculate the popularity of a topic based on its related data, including user number, propagated communities, average browsing time, and lasting time.

We first propose several types of topic related data for popularity evaluation in SNS as follows.

- (1) Number of followers: $\text{follow}(\Theta_i)$ denotes the number of users who follow the topic Θ_i .
- (2) Number of communities: $\text{community}(\Theta_i)$ denotes the number of communities in which the topic Θ_i is propagated.
- (3) Browsing time: $\text{browse}(\Theta_i)$ denotes the average length of time that users spend on topic Θ_i .
- (4) Lasting time: $\text{last}(\Theta_i)$ denotes the length of time that the topic Θ_i keeps hot in SNS.
- (5) Activity: a topic Θ_i is active in a time slice if and only if it is posted, followed, browsed, or propagated or wins other social behaviors in SNS. $\text{activity}(\Theta_i)$ denotes the activity level of topic Θ_i in each time slice.

All the above types of data are available through specific collection methods in SNS. In this paper, popularity evaluation is the average of five indicators produced by the above five types of topic related data. Let the total numbers of user and community in SNS be n and m , respectively. The maximum lengths of browsing time and lasting time of all topics in SNS are $|\text{max_browse}|$ and $|\text{max_last}|$, respectively. Then, the popularity level, $\text{popular}(\Theta_i)$, of topic Θ_i can be calculated as follows:

$$\text{popular}(\Theta_i) = \frac{1}{5} \left(\frac{\text{follow}(\Theta_i)}{n} + \frac{\text{community}(\Theta_i)}{m} + \frac{\text{browse}(\Theta_i)}{|\text{max_browse}|} + \frac{\text{last}(\Theta_i)}{|\text{max_last}|} + \text{activity}(\Theta_i) \right). \quad (8)$$

In the above equation, $\text{activity}(\Theta_i) = \bar{k}/k$, where k is the number of time slices in the life cycle of topic Θ_i and \bar{k} is the number of time slices in which Θ_i keeps active status.

Through the above two aspects of impact evaluation, that is, influence and popularity, we can get the total impact of a topic in SNS as follows:

$$\text{impact}(\Theta_i) = \frac{1}{2} [\text{influence}(\Theta_i) + \text{popular}(\Theta_i)]. \quad (9)$$

Furthermore, impact of whole topic space is measured based on its included topics. Let there be a topic space $\text{TS} = \{\Theta_1, \Theta_2, \dots, \Theta_n\}$, and the impact of each topic Θ_i is $\text{impact}(\Theta_i)$. Then, the impact of the topic space is a comprehensive evaluation of all topics in TS as follows:

$$\text{impact}(\text{TS}) = \frac{\sum_{i=1}^n \text{impact}(\Theta_i)}{n}. \quad (10)$$

5. Trust Chain Model and Its Computation Method

5.1. Trust Chain Model. Here, we propose the model of trust chains in detail based on their different network topologies and their trust value calculation methods. We divide trust chain into four types based on the topology.

(1) *Atomic Trust Chain.* A trust relationship between users is an atomic trust chain if and only if there is a direct link between two nodes and no intermediate node between them.

(2) *Serial Trust Chain.* A trust relationship between users is a serial trust chain if and only if there is a serial path from source node to target node and the path has the following properties: (1) for source node, its out-degree is 1 and in-degree is 0; (2) for the target node, its out-degree is 0 and in-degree is 1; (3) for each intermediate node, its out-degree is 1 and in-degree is 1.

(3) *Parallel Trust Chain.* A trust relationship between users is a parallel trust chain if and only if there are two or more trust paths from source node to target node and there is no intersection node among the paths, and the path has the following properties: (1) for source node, its out-degree is m ($m \geq 2$) and in-degree is 0; (2) for target node, its out-degree is 0 and in-degree is m ($m \geq 2$); (3) for each intermediate node, its out-degree is 1 and in-degree is 1.

(4) *Combined Trust Chain.* A trust relationship between users is a combined trust chain if and only if the trust chain is composed of the above three kinds of trust chain.

Furthermore, we here classify the trust chain between users into two types according to the mutual trust degrees as follows.

(1) *Strong Trust Chain.* A trust chain is a strong one if and only if there are two mutual accessible trust chains between

two users and the trust degrees of the both trust chains are all higher than a given threshold ($\vartheta_1 \in [0, 1]$).

(2) *Weak Trust Chain*. A trust chain is a weak one if and only if there is a trust chain higher than a given threshold ($(\vartheta_2 \in [0, 1]) \wedge (\vartheta_2 \leq \vartheta_1)$) from source node to target node and such trust chain is not a strong one.

In our definition, a strong trust chain reveals a mutual high trust relationship between two nodes, while a weak trust chain reflects a unidirectional trust degree or a bidirectional trust degree with a relative high value.

5.2. Computation of Trust Chain Model

5.2.1. Atomic Trust Chain. In atomic trust chain, there is no intermediate node between two nodes. Then, we can calculate the trust degree through their direct trustworthy interactions and their interest similarity. Let there be two nodes N_1, N_2 , and $f_i(N_1, N_2) \in [0, 1]$ denotes the trustworthy opinion which is expressed by N_1 to N_2 . Assume that there is a common topic set, $CT = \{\Theta_1, \Theta_2, \dots\}$, which denotes the set of topics of interest by both nodes. For each $CT \cdot \Theta_i \in CT$, the number of nodes that maintain explicit interest degrees to it is $\text{num}(CT \cdot \Theta_i)$, and the maximum number of nodes that maintain explicit interest degrees to all the topics in social network is max_top . Then, the degree of atomic trust chain from N_1 to N_2 can be calculated as follows:

$$\begin{aligned} \text{atom_}T(N_1, N_2) &= \text{trust_interact}(N_1, N_2) \\ &\quad \times \text{sim_interest}(N_1, N_2), \\ \text{trust_interact}(N_1, N_2) &= \frac{\sum_{i=1}^n (f_i(N_1, N_2))}{n}, \end{aligned}$$

$\text{ser_}T(N_S, N_E)$

$$= \frac{\left[T(N_S, N_I^1) \times \omega(N_S, N_I^1) + \sum_{k=1}^{\text{depth}(N_E)-1} \left(T(N_I^k, N_I^{k+1}) \times \omega(N_I^k, N_I^{k+1})^{k+1} \right) + \left(T(N_I^{\text{depth}(N_E)-1}, N_E) \times \omega(N_I^{\text{depth}(N_E)-1}, N_E)^{\text{depth}(N_E)} \right) \right]}{\text{depth}(N_E)}, \quad (12)$$

$$\omega(N_I^k, N_I^{k+1}) = \begin{cases} 1 & \text{atomic trust chain from } N_I^k \text{ to } N_I^{k+1} \text{ is a strong trust chain} \\ 0.9 & \text{atomic trust chain from } N_I^k \text{ to } N_I^{k+1} \text{ is a weak trust chain.} \end{cases}$$

Here function $\text{depth}(N_E)$ denotes the depth of serial trust chain; namely, $\text{depth}(N_E) = |\Omega \cdot N_I| + 1$. We can see that the deeper the depth of serial trust chain is, the weaker the trust value among users is. That means longer serial trust chain would be punished since the trust would be damped with the number of intermediate nodes increasing. In addition, we give a parameter $\omega(N_I^k, N_I^{k+1})$ for distinguishing the weights of strong and weak atomic trust chain parts in serial trust chain.

5.2.3. Parallel Trust Chain. In parallel trust chain, there are at least two serial paths without intersection from source node to target node. In addition, we here present a constraint for ensuring the reliability of parallel trust chain as follows.

$$\begin{aligned} \text{sim_interest}(N_1, N_2) &= \frac{\sum_{j=1}^{|\text{CT}|} (\text{EI}_{N_1}(\Theta_j) \times \text{EI}_{N_2}(\Theta_j))}{\sqrt{\sum_{j=1}^{|\text{CT}|} (\text{EI}_{N_1}(\Theta_j))^2} \times \sqrt{\sum_{j=1}^{|\text{CT}|} (\text{EI}_{N_2}(\Theta_j))^2}} \\ &\quad \cdot e^{-\sum_{j=1}^{|\text{CT}|} (\text{num}(\Theta_j) / \text{max_top}) / |\text{CT}|}, \end{aligned} \quad (11)$$

where $\text{trust_interact}(N_1, N_2)$ denotes the trust degree based on nodes' direct interactions and $\text{sim_interest}(N_1, N_2)$ is the similarity degree based on nodes' interests. Here, we use the factors, $\text{num}(CT \cdot \Theta_i)$ and max_top , to reveal that a topic which has the less number of explicit interested nodes would give more contributions to nodes' interest similarity calculation.

5.2.2. Serial Trust Chain. In serial trust chain, we give a constraint for its composition as follows.

Constraint 1. Each atomic trust chain part in serial trust chain must be a strong atomic trust chain or a weak atomic trust chain.

That is, an atomic trust chain part with a low trust degree would be excluded from the serial trust chain, and thus the indirect path from source node to target node cannot be considered as a serial trust chain.

Let there be source node N_S , target node N_E , and intermediate node N_I^k in serial trust chain and $\text{atom_}T(N_i, N_j)$ which denotes trust value of atomic trust chain part in the serial trust chain. Therefore, we can calculate the trust value of serial trust chain as follows:

Constraint 2. A serial path can be seen as a serial trust chain in a parallel trust chain if and only if it is a strong or weak serial trust chain.

From Constraint 2, we can divide the serial trust chains into two types as follows: (1) the strong or weak serial trust chains, called *active serial trust chains*, are taken into consideration in parallel trust chain evaluation; and (2) the serial trust chains with low trust degrees, called *inactive serial trust chains*, would be excluded in trust degree calculation of parallel trust chain. However, the number of inactive trust degree serial trust chains, in which all the atomic trust chain parts are strong or weak ones, is also used in parallel trust

chain evaluation since their low trust degrees also reflect the untrustworthy perspective of parallel trust chain.

Let there be m_1 ($m_1 \geq 2$) serial trust chains in the parallel trust chain from source node N_S to target node N_E . Assume that there are m_2 ($m_2 \leq m_1$) active serial trust chains in parallel trust chain. C_l ($0 \leq l \leq m_2$) denotes each serial trust chain, and $\text{ser}_T(C_l)$ represents trust degree value of serial trust chain C_l . Then, the trust degree of parallel trust chain from N_S to N_E , $\text{par}_T(N_S, N_E)$, can be calculated as follows:

$$\begin{aligned} \text{par}_T(N_S, N_E) &= \frac{\sum_{l=1}^{m_2} [\text{ser}_T(C_l) \times \omega(C_l)]}{m_2} \left(\frac{m_2}{m_1} \right)^{(m_1-m_2)/m_2}, \quad (13) \\ \omega(C_l) &= \begin{cases} 1 & C_l \text{ is a strong serial trust chain} \\ 0.9 & C_l \text{ is a weak serial trust chain.} \end{cases} \end{aligned}$$

In the above equation, trust degrees of parallel trust chain are calculated as follows: (1) trust degrees of active serial trust chains are calculated by weighted average method ($\omega(C_l)$ is the weight of each active serial trust chain) and (2) number of inactive serial trust chains degrees are considered as negative aspects and thus are used to weaken the trust degree of parallel trust chain by exponential weighting as in (33). We can see that the higher ratio of effective serial trust chains in parallel trust chain and the lower ratio of inactive serial trust chain to active serial trust chain imply a higher exponential weighting.

5.2.4. Combined Trust Chain. Combined trust chain includes crossing paths which are above three kinds of trust chains. We introduce an iterative optimizing approach for combined trust chain, called IOA, by including strong or weak trust chain part and excluding other parts. The constraint of the proposed approach is as follows.

Constraint 3. There are the following four rules for IOA.

- (i) Local trustworthy rule (LTR): for each atomic path in combined trust chain, it can be seen as an active atomic trust chain for combined trust chain if and only if it is a strong or weak trust chain. That is, those atomic paths which are not strong or weak trust chains from nodes to their neighbors can be ignored in the combined trust chain, and thus their successor paths are ignored due to the breakage occurring.
- (ii) Serial trustworthy rule (STR): for a serial path from N_S to N_E , it can be seen as active trust chain if and only if it satisfies Constraint 1.
- (iii) Serial merging rule (SMR): if there is a combined trust chain from N_S 's indirect neighbor nodes to N_E , it would be merged as a serial one iteratively.
- (iv) Parallel calculating rule (PCR): if there are two or more direct neighbor nodes which have serial trust chains from N_S to N_E , the combined trust chain

can be reconstructed as parallel trust chain with its neighbors iteratively if and only if the reconstructed parallel trust chain satisfies Constraint 2.

An example of our scheme is shown in Figures 3(a)–3(e). In Figures 3(a) and 3(b), we can see the parallel paths and the inverse paths from source node S to target node T , while the intermediate nodes are B , C , and D . Assume that the thresholds, ϑ_1 and ϑ_2 , of strong trust chain and weak trust chain are set as 0.6 and 0.5, respectively. Then, we can get the strong atomic trust chains ($B \rightarrow T$, $S \rightarrow C$, $C \rightarrow T$) and weak trust chains ($S \rightarrow B$, $D \rightarrow T$), while the path $S \rightarrow D$ is an ignorable path due to its low trust degree. Then, we can calculate the trust degrees of serial paths ($S \rightarrow B \rightarrow T$, $S \rightarrow C \rightarrow T$) based on (32) and Constraint 1 as follows:

$$\begin{aligned} \text{ser}_T(S, T)_{S \rightarrow B \rightarrow T} &= \frac{[T(S, B) \times \omega(S, B) + (T(B, T) \times \omega(B, T)^2)]}{2} \\ &= \frac{1 \times 0.9 + 0.9 \times 1^2}{2} = 0.9, \quad (14) \\ \text{ser}_T(S, T)_{S \rightarrow C \rightarrow T} &= \frac{[T(S, C) \times \omega(S, C) + (T(C, T) \times \omega(C, T)^2)]}{2} \\ &= \frac{0.8 \times 1 + 1 \times 1^2}{2} = 0.9. \end{aligned}$$

Likewise, we can get the active serial trust chains as $S \rightarrow C \rightarrow T$ (strong trust chain) and $S \rightarrow B \rightarrow T$ (weak trust chain) according to inverse serial trust chain calculation and Constraint 1 in Figure 3(b), while the serial path of $S \rightarrow D \rightarrow T$ is an inactive path. Then further, we can get the trust degree of the parallel path in Figure 3(a) through (33) as follows:

$$\begin{aligned} \text{par}_T(S, T) &= \frac{\sum_{l=1}^{m_2} [\text{ser}_T(C_l) \times \omega(C_l)]}{m_2} \left(\frac{m_2}{m_1} \right)^{(m_1-m_2)/m_2} \quad (15) \\ &= \frac{[0.9 \times 0.9 + 0.9 \times 1]}{2} \left(\frac{2}{3} \right)^{(3-2)/2} \approx 0.7. \end{aligned}$$

In Figure 3(c), we can see a combined path which includes joint nodes from source node S to target node T . According to LTR in Constraint 3, the atomic paths, $S \rightarrow B$, $S \rightarrow C$, $D \rightarrow F$, and $G \rightarrow K$, are seen as ignorable paths, and thus we can get the paths for calculating trust as in Figure 3(d). Further, the parallel parts of path from S to I and path from E to T can be eliminated by using SMR to generate serial trust chains, as in Figure 3(e). Based on PCR

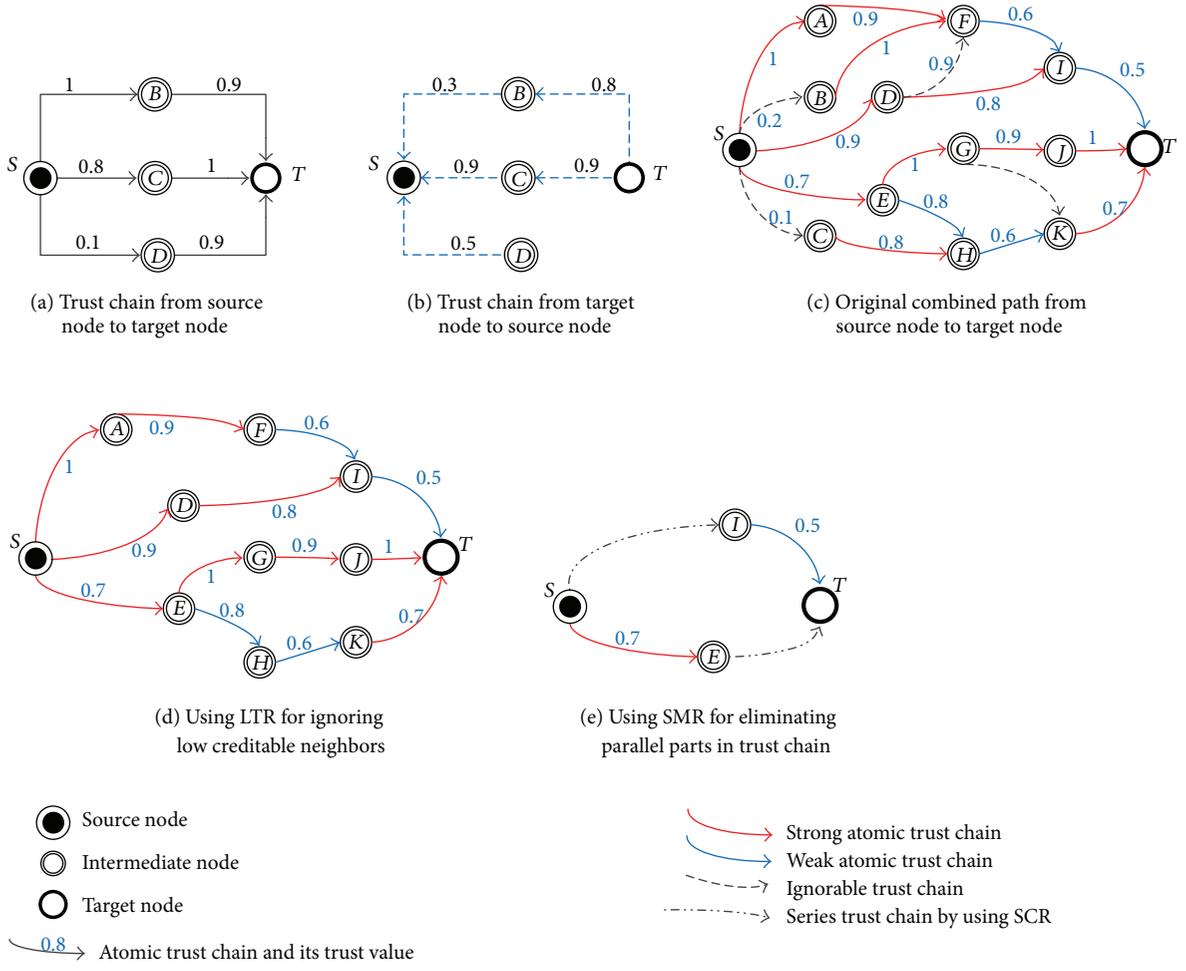


FIGURE 3: An example for trust chain calculation.

in Constraint 3, the trust degree of combined trust chain in Figure 3(e) can be calculated as follows:

$$\begin{aligned}
 S \rightarrow A \rightarrow F \rightarrow I: \quad \text{ser}_T(S, I) &= \frac{[T(S, A) \times \omega(S, A) + T(A, F) \times \omega(A, F)^2 + T(F, I) \times \omega(F, I)^3]}{3} \\
 &= \frac{1 + 0.9 + 0.6 \times 0.9^3}{3} \\
 &\approx 0.779,
 \end{aligned} \tag{16}$$

$$S \rightarrow D \rightarrow I: \quad \text{ser}_T(S, I) = \frac{[T(S, D) \times \omega(S, D) + T(D, I) \times \omega(D, I)^2]}{2} = \frac{0.9 \times 1 + 0.8 \times 1^2}{2} = 0.85, \tag{17}$$

$$\begin{aligned}
 E \rightarrow G \rightarrow J \rightarrow T: \quad \text{ser}_T(E, T) &= \frac{[T(E, G) \times \omega(E, G) + T(G, J) \times \omega(G, J)^2 + T(J, T) \times \omega(J, T)^3]}{3} = \frac{1 + 0.9 + 1}{3} \\
 &\approx 0.967,
 \end{aligned} \tag{18}$$

$$\begin{aligned}
E \rightarrow H \rightarrow K \rightarrow T: \text{ ser_}T(E, T) &= \frac{[T(E, H) \times \varpi(E, H) + T(H, K) \times \varpi(H, K)^2 + T(K, T) \times \varpi(K, T)^3]}{3} \\
&= \frac{0.8 \times 0.9 + 0.6 \times 0.9 + 0.7 \times 1^3}{3} \\
&\approx 0.653,
\end{aligned} \tag{19}$$

$$S \rightarrow I: \text{ par_}T(S, I) = \frac{[0.779 \times 0.9 + 0.85 \times 1]}{2} \left(\frac{2}{2}\right)^{(2-2)/2} \approx 0.776, \tag{20}$$

$$E \rightarrow T: \text{ par_}T(E, T) = \frac{[0.967 \times 1 + 0.653 \times 0.9]}{2} \left(\frac{2}{2}\right)^{(2-2)/2} \approx 0.777, \tag{21}$$

$$S \rightarrow I \rightarrow T: \text{ ser_}T(S, T) = \frac{[T(S, I) \times \varpi(S, I) + T(I, T) \times \varpi(I, T)^2]}{2} = \frac{0.776 \times 0.9 + 0.8 \times 0.9^2}{2} = 0.672, \tag{22}$$

$$S \rightarrow E \rightarrow T: \text{ ser_}T(S, T) = \frac{[T(S, E) \times \varpi(S, E) + T(E, T) \times \varpi(E, T)^2]}{2} = \frac{0.7 \times 1 + 0.777 \times 0.9^2}{2} = 0.665, \tag{23}$$

$$\text{com_}T(S, T) = \frac{[0.672 \times 0.9 + 0.665 \times 0.9]}{2} \left(\frac{2}{2}\right)^{(2-2)/2} \approx 0.602.$$

5.3. *User Influence Evaluation through Trust Chain.* Each user has his/her influence in SNS. Here, we can evaluate user influence by measuring the users who maintain high trustworthiness in the trust chain. The more the users who trust source user in the trust chains, the more influential the source user.

For a user v , let there be a trust chain, $\Omega(v)$, in which user v is the source node. We denote that the valid impacted user, iu_i , is the node which satisfies the following condition in trust chain $\Omega(v)$:

$$\begin{aligned}
& (iu_i \in \Omega(v) \cdot N) \\
& \wedge \left(T(v, iu_i) \geq \frac{\sum_{v_j \in \Omega(v) \cdot N} T(v, v_j)}{|\Omega(v) \cdot N|} \right). \tag{24}
\end{aligned}$$

Assume that there are m trust chains, $\Omega(v)_k$, in which user v is the source node, and the valid impacted user set in each trust chain is denoted as $\Omega(v)_k \cdot IU$. Then, the user influence can be evaluated as follows:

$$\begin{aligned}
\text{ef}(v) &= \frac{1}{2m} \\
& \cdot \sum_{k=1}^m \left[\left(1 - \frac{1}{|\Omega(v)_k \cdot IU|} \right)^{(1-1/|\Omega(v)_k \cdot N - \Omega(v)_k \cdot IU|)} \right. \\
& \cdot \left(\frac{|\Omega(v)_k \cdot IU|}{|\Omega(v)_k \cdot N|} \right) \\
& \left. + \left(\frac{\sum_{iu_i \in \Omega(v)_k \cdot IU} \text{ef}(iu_i)^{(1-T(v, iu_i))}}{|\Omega(v)_k \cdot IU|} \right) \right]. \tag{25}
\end{aligned}$$

6. The Calculation Method of Interest Degree

6.1. *User Explicit Interest Degree Evaluation towards Topic.* Explicit interest degree can be measured through users' direct behaviors or other direct witness evidences. In this study, we denote these direct items for evaluating explicit interest degree as *interest evidences*. For example, the behaviors of a node, such as "forwarding," "approving," "following," and "comments," can be seen as *interest evidences*. In our work, we have the following considerations for user explicit interest degree evaluation towards topic. (1) Explicit interest is measured by the level of each interest evidence and its weight. This means explicit interest is aggregated by users' past interest evidences, and, meanwhile, interest evidences have different impacts in interest aggregation. (2) Explicit interest is impacted by the impact degree of the topic. That is, a more influential topic would attract more users to browse it. From this point, we consider that if a user shows his/her explicit interest towards a topic with lower influence, he might have more interests in the topic.

Then we address the calculation method of explicit interest degree towards a topic. Assume that user v has his/her past different kinds of interest evidences as D_1, D_2, \dots, D_m , and $P(D_j)$ denotes the appearance probability of D_j in all kinds of interest evidences. Meanwhile, for a topic Θ , assume that user v did n times of interest evidences which were recorded as a set $\Phi = \{\delta_1, \delta_2, \dots, \delta_i, \dots\}$ and $P(\Theta | D_j)$ denotes the probability of interest evidence category D_j appearing when the user faced the topic Θ . The weight of each D_j is $\text{right}(D_j)$ ($\text{right}(D_j) \in [0, 1]$). Then, user node v would have an explicit interest degree about the topic Θ when it appears in next future as

$$\text{EI}(v | \Theta) = \begin{cases} 0 & n = 0 \\ \left[\frac{1}{n} \times \sum_{i=1}^n \left(\text{right}_{\delta_i \in D_j}(D_j) \times P_v(D_j | \Theta) \right) \right] \times \lambda^{\text{impact}(\Theta)} & n \geq 1. \end{cases} \quad (26)$$

In the above equation, $P_v(D_j | \Theta)$ denotes the probability of appearing interest evidence D_j for topic Θ and it can be calculated as

$$P_v(D_j | \Theta) = \frac{P(D_j) \times P(\Theta | D_j)}{\sum_{j=1}^m (P(D_j) \times P(\Theta | D_j))}. \quad (27)$$

Then, we give the calculation method of weight $\text{right}(D_j)$. In our view, there are inherent relations among interest evidences: many interest evidences appeared simultaneously and serially. For instance, interest evidence of long time browsing would be likely to lead to interest evidences of “approving” or “adding to the favorites list.” Then, their impacts are related to each other, and these relations can be enhanced in their increasing appearance. From this view, our underlying principle of interest evidence weight calculation is similar to the method of PageRank as more important interest evidences are likely to be related to other more important trust evidences. Here, interest evidence y caused by another interest evidence x is conveniently written as a link $x \rightarrow y$. From this, we can calculate the weight of trust evidence as follows.

Let there be an interest evidence set as $\text{ID} = \{D_1, D_2, \dots, D_M\}$, and the set of interest evidences that can link to D_i is denoted as $\text{Link}(D_i) = \{D_j | \exists(D_j \rightarrow D_i) \wedge (i \neq j)\}$. Then, the equation of calculating link-weight of interest evidence D_i is as follows:

$$\text{right}(D_i) = \frac{1 - p(D_i)}{|\text{ID}|} + p(D_i) \sum_{D_j \in \text{Link}(D_i)} \frac{\text{right}(D_j)}{L(D_j)}, \quad (28)$$

where $p(D_i)$ is the probability that D_i occurred in users' past behaviors and $L(D_j)$ is the number of interest evidences which link out of D_j .

Moreover, user's interest is a dynamic feeling and keeps changing with the time-passing. Here, we propose a dynamic predicting method based on *aging algorithm* for describing the interest changing. Suppose the original explicit interest degree calculated at the end of time quantum t_0 is $\text{EI}(v | \Theta)_0$. Now, suppose the next value of explicit interest is changed as $\text{EI}(v | \Theta)_1$ at the end of next time quantum t_1 . Then, we can consider that the explicit interest degree of topic Θ is an integrated value of the past two values and, then, renew the

dynamic predicting value by taking a weighted sum of these two numbers; that is,

$$\text{EI}(v | \Theta)_1 = \alpha \times \text{EI}(v | \Theta)_0 + (1 - \alpha) \times \text{Change}_{I_v}(\Theta)_1. \quad (29)$$

By analogy, the dynamic estimating value of explicit interest degree can be calculated as

$$\text{EI}(v | \Theta)_n = \alpha \times \text{EI}(v | \Theta)_{n-1} + (1 - \alpha) \times \text{Change}_{I_v}(\Theta)_n. \quad (30)$$

In this work, we have the following constraint for interest degree.

Constraint 4. A user has an explicit interest degree toward a topic if and only if his explicit interest degree value is larger than a given threshold τ .

6.2. Implicit Interest Degree of Topic. Implicit interest degree reflects a kind of users' potential opinion toward topic. In our consideration, our proposed implicit interest evaluation mainly focuses on solving the problem of interest prediction without direct evidences or data. Thus, in our proposed work, there are two main considerations for implicit interest evaluation. (1) Implicit interest manifests a likelihood of a user's potential interest while he/she has not shown any explicit interest or direct evidence before, such as a new register. Therefore, the only direct evidence for predicting the interest of such user is his/her relationships with others. Of course it is impossible that all the relationships can be used to reflect and predict his/her potential interest. Consequently, we in this work use the relationship with “high trustworthy degree” for implicit interest evaluation. That is because the relationship with high trust worthy degree is established based on users' past interactions and experience, which can reflect the similarity between users in a higher level. (2) Trustworthy relationship based implicit interest has a performance of relatively low computational overhead. The reason is that by evaluating the implicit interest through trust relationship the proposed method only relies on the data of trustworthy relationships and their degrees among users rather than other detail data of users. That is, for each evaluation, it only needs to inquire the trust relationships and degrees among users, which avoids querying detail records, whether or not it has to do with the implicit interest evaluation, and storing such data.

In this study, there are two factors for implicit interest evaluation—trust and user similarity. Trust degree reflects the

probability of user's approving or against attitude through his trust relationships. Therefore, users would tend to be influenced by those who they trust. On the other hand, user similarity reveals the common interests between users. It means users would be easily influenced by those who have similar interests with them. For example, a user might have potential interests to a topic because the topic is his/her best friend's favorite topic and most of the time the user has common interests with his/her friend to other topics. Thereby, we give an estimating method for evaluating the implicit user interest degree according to the user trust chain model and their interest similarity. This approach comprises two steps: firstly, we calculate the value, called deliverable implicit interest degree, one-to-one from a source node to target node based on their trust chain, explicit interest degree, and user interest similarity; then, we integrate deliverable implicit interest degrees to target node from all source nodes which have trust chains with target node. In addition, there is a following constraint for implicit interest degree of topic.

Constraint 5. The explicit interest degree of a source node can be used for implicit interest evaluation if and only if there is a strong or weak trust chain from target node to the source node.

Let there be a trust chain $\Omega(N_E, N_S)$ from target node to source node at time quantum t_0 , and the original explicit interest degree of N_S toward a topic Θ is $EI_{N_S}(\Theta)_0$. Suppose that the original trust value of $\Omega(N_E, N_S)$ is $T(N_E, N_S)_0$ and then the original deliverable implicit interest degree from N_S to N_E is $MI(N_E | \Theta)_0^{N_S}$ which can be calculated as

$$\begin{aligned} MI(N_E | \Theta)_0^{N_S} &= EI_{N_S}(\Theta)_0 \times T(N_E, N_S)_0 \\ &\quad \times \text{sim_interest}(N_E, N_S) \times \omega(\Omega(N_E, N_S)), \\ \omega(\Omega(N_E, N_S)) &= \begin{cases} 1 & \Omega(N_E, N_S) \text{ is a strong trust link} \\ 0.9 & \Omega(N_E, N_S) \text{ is a weak trust link,} \end{cases} \end{aligned} \quad (31)$$

where $\text{sim_interest}(N_E, N_S)$ denotes the user interest similarity between N_S and N_E about all their other common explicit interested topics, and $\omega(\Omega(N_E, N_S))$ is the weight for distinguishing the impact of strong and weak trust chain.

Now suppose the trust value of $\Omega(N_E, N_S)$ and explicit interest degree of N_S are measured to be $T(N_S, N_E)_1$ and $EI_{N_S}(\Theta)_1$ at the end of next time quantum t_1 . Then, we can renew the estimation value of implicit interest degree by taking a weighted sum of the new values as

$$\begin{aligned} MI(N_E | \Theta)_1^{N_S} &= \alpha \times MI(N_E | \Theta)_0^{N_S} + (1 - \alpha) \\ &\quad \times (EI_{N_S}(\Theta)_1 \times T(N_E, N_S)_1 \\ &\quad \times \text{sim_interest}(N_E, N_S) \times \omega(\Omega(N_E, N_S))). \end{aligned} \quad (32)$$

By analogy, we can get the equation for deliverable implicit interest degree after n quantum as

$$\begin{aligned} MI(N_E | \Theta)_n^{N_S} &= \alpha \times MI(N_E | \Theta)_{n-1}^{N_S} + (1 - \alpha) \\ &\quad \times (EI_{N_S}(\Theta)_n \times T(N_E, N_S)_n \\ &\quad \times \text{sim_interest}(N_E, N_S) \times \omega(\Omega(N_E, N_S))). \end{aligned} \quad (33)$$

After finishing the first step of deliverable implicit interest degree estimation, we can calculate the overall implicit interest degree from all the source nodes to target node. Let there be a set of source nodes, $NS = \{N_S^1, N_S^2, \dots\}$ which satisfy Constraint 4, and each node in NS has trust chain from target node N_E . Then the overall implicit interest degree can be calculated as

$$\begin{aligned} MI(N_E | \Theta) &= \frac{\sum_{N_S^i \in NS} [MI(N_E | \Theta)_{n-1}^{N_S^i} \times T(N_S^i, N_E)]}{\sum_{N_S^i \in NS} T(N_S^i, N_E)}. \end{aligned} \quad (34)$$

6.3. User Interest Degree of Topic Space. User interest degree of topic space is comprised of user interest degrees of topics which are included in topic space. Here, we use weighted average method for calculating interest degree of topic space.

Assume that there is topic space $TS = \{\Theta_1, \Theta_2, \dots, \Theta_n\}$ and user v has explicit interest degree (or implicit interest degree) of topic Θ_i as $EI(v | \Theta_i)$ (or $MI(v | \Theta_i)$). Then, user interest degree of topic space can be calculated as

$$I(v | TS) = \frac{\sum_{i=1}^n [\text{impact}(\Theta_i) \times EI(v | \Theta_i)]}{\sum_{i=1}^n \text{impact}(\Theta_i)}. \quad (35)$$

7. TUG Discovering Algorithm Based on Trust Chain

In this work, the basic idea of single topic user group organization is as follows: firstly, influential topic is selected based on impact evaluation; secondly, core users, who have strong interests in the specific topic, should be discovered; then, ordinary users, who keep certain explicit or implicit interests in the topic, should be organized through trust chains. Correspondingly, we set two sets, U_C and U_O , for recording the above two types of users.

Based on the above consideration, we first propose the influential topic discovering algorithm as shown in Algorithm 3.

Here, the topic space, which has impact values larger than average of the whole topic space set $Top = \{TS_1, TS_2, \dots\}$ and where also the numbers of follower users are not equal to 0, would be selected as influential topics. Then, for each topic space $TS_i \in Top$, we propose the algorithm of core user discovering for TUG to discover core users from a candidate set $Candidate \subseteq SN \cdot V$ as shown in Algorithm 4.

```

Top = {TS1, TS2, ...}, Selected = ∅;
for TSi ∈ Top do
  get impact(TSi);
  ave =  $\frac{\sum \text{impact}(TS_i)}{|Top|}$ ;
for TSi ∈ Top do
  if (follower(TSi) = 0) ∨ (impact(TSi) ≤ ave) then Top ← Top - TSi;
return Top;

```

ALGORITHM 3: Influential topic space evaluation algorithm.

```

TS = {Θ1, Θ2, ...}, Candidate = {v1, v2, ...}, Δ · UC = ∅;
while Candidate ≠ ∅ do
{
  for vi ∈ Candidate do
  {
    get ef(vi);
    if (ef(vi) ≥ σ) ∨ (EI(vi | TS) ≥ μ)
    then (Δ · UC ← Δ · UC ∪ {vi}) ∧ (Candidate ← Candidate - {vi});
  }
  for vj ∈ {vj | vj ∈ Ω(vi) · N ∧ vi ∈ Δ · UC ∧ vj ∉ Δ · UC} do
  {
    if (ef(vj) ≥ σ) ∨ (EI(vj | TS) ≥ μ)
    then (Δ · UC ← Δ · UC ∪ {vj}) ∧ (Candidate ← Candidate - {vj})
    else Candidate ← Candidate - {vj};
  }
}
return Δ · UC;

```

ALGORITHM 4: Core user discovering algorithm for TUG.

In Algorithm 4, we set two thresholds, σ and μ , for selecting qualified core users. The thresholds can be calculated as

$$\sigma = \frac{\sum_{v_i \in \text{Candidate}} \text{ef}(v_i)}{|\text{Candidate}|}, \quad (36)$$

$$\mu = \frac{\sum_{v_i \in \text{Candidate}} \text{EI}(v_i | \text{TS})}{|\text{Candidate}|}.$$

After discovering the set of core users, ordinary users of TUG can be found based on their trust chains and explicit (or implicit) interest degree. Then, we propose the ordinary user discovering algorithm for TUG as shown in Algorithm 5.

8. Experiment and Analysis

In this section, we give examinations to explain the performances of our proposed scheme. In our scenario of examinations, the data comes from a real social network platform, *Tencent microblog*, which is very popular in China and the dataset is collected manually. Our data included about 3,750 IDs (some IDs located in two or more communities) and more than 457,000 records, including posts, comments, and users' behaviors (browsing, approving, forwarding, and others). The topology of collected dataset is from the users'

real relationships and the average out-degree of a node is about 9. We select eight kinds of topic, that is, education, entertainment, sporting, technique, financial, food, touring, and history, for user group discovering experiment. Details about initial data setting are shown in Table 2. Additionally, we also get about 780 robot nodes by our previous data processing method in the dataset and we call them invalid nodes, which only follow real users, forward posts, or hit approving automatically and should be excluded in user grouping.

8.1. Examination for Topic Space Construction and Impact Evaluation

8.1.1. Performance of Topic Space Relatedness Clustering Method. In this examination, we aim to reveal the performance of topic space relatedness clustering method. Here, the candidate topic sets for space clustering are selected randomly from eight kinds of topics. For comparison, we give four groups as follows: (1) topic grouping based on K -means method (KM), which classifies topics into specified number groups, (2) topic grouping based on cosine similarity (CS), which selects feature words for feature vectors in calculation by TF-IDF method, (3) topic grouping based on k NN method (k NN), and (4) proposed relatedness clustering method (RC).

```

TS = {Θ1, Θ2, ...}, Δ · UO = ∅;
for vi ∈ {vi | vi ∈ SN · V ∧ vi ∉ Δ · UC ∧ ∃Ω(vi, vj) ∧ vj ∈ Δ · UC} do
{
  if T(vi, vj) ≥ η then
    if (EI(vi | TS) ≥ μ/2) ∨ (MI(vi | TS) ≥ μ/2)
      then Δ · UO ← Δ · UO ∪ vi;
}
for vi ∈ Δ · UO do
{
  p = 0;
  while p ≠ n do
  {
    if (EI(vi | TS)p ≥ μ/2) ∨ (MI(vi | TS)p ≥ μ/2)
      then p = p + 1;
    else UO ← UO - {vi}, p = n;
  }
}
return Δ · UO;

```

ALGORITHM 5: Ordinary user discovering algorithm for TUG.

TABLE 2: Characteristics of five communities in examination prototype.

Community	Education	Financial	Sporting	Entertainment	Technique	Food	Touring	History
Number of IDs	705	832	887	1,051	605	852	719	594
Number of posts	4,974	5,046	8,147	10,419	6,930	3,095	5,394	2,943
Number of comments	30,519	68,215	104,535	94,749	49,053	28,451	42,067	21,596

To verify the performances of comparing groups, we set nonnoise data and 20% and 30% noise data sets, respectively, where the noise data sets include the topics that did not belong to any of the above eight kinds. The results are shown as in Figures 4(a)–4(c). We can see that the performance of our proposed method outperforms other three groups a little.

And further, we analyze the impact of two thresholds, ρ and β , in Algorithms 1 and 2. As shown in Figure 4(e), the average accuracy of relatedness between topics based on Algorithm 1 is decreased with the increasing of value setting of threshold ρ . The reason is that a higher value of ρ would allow Algorithm 1 ending with less stable relatedness values between topics. In addition, we can see that the average accuracy is lower while the value of β is set too low or too high. We consider that a too low value of β leads to unrelated topics being grouped in topic space while a too high value of β leads to appropriate topics being excluded from topic space. This test validates that the thresholds around 0.1 and 0.7 are often a reasonable compromise for ρ and β .

8.1.2. Performance of Topic Space Impact Evaluation. In this examination, we reveal the performance of the proposed topic space impact evaluation method. We give about 240 topic spaces which are grouped by proposed space clustering method for evaluating their impact degrees. All topics in topic spaces are contained in the initial data set. We set six impact evaluation groups in different methods for comparison as follows: only link relation based impact calculation (LR), only hierarchy relation based impact calculation (HR), only proposed influence degree based impact calculation

(ID), only proposed popularity degree based impact calculation (PD), linear threshold model based method (LT), and our proposed integrated impact degree method (TP). And then, we record the average precisions of six groups as in Figure 5(a). Here, the optimized results of topic impact evaluation are set manually in advance, and if the impact evaluation results of six groups are larger or smaller than optimized results within 0.1, the evaluation results are seen as accurate. We can see that LT method and our proposed TP method get similar performances, which both outperform other methods. Meanwhile, we can see that the precision of TP keeps increasing with the increasing of topic number in topic space as in Figure 5(b). That means that as the topic space includes more topics, the data for impact evaluation would be enriched more, which results in better performance in impact evaluation.

8.2. Performance Evaluation of Proposed Trust Chain Calculation. In this examination, we get the accuracy of predicting trustworthiness among users through invalid node detection. In this test, we use the trust degrees between real users and invalid users for testing the accuracy of invalid node detection. For a real user, if he/she has a strong or weak trust chain with an invalid node, the trust relationship between them is wrong. For comparison, we introduce the following groups: atomic trust chain (ATC), serial trust chain (STC), parallel trust chain (PTC), hybrid trust chain (HTC), EigenTrust method (ET) [28], weighted average trust rating (WA) [30], and ultimate trust rating (UT) [31]. Figure 6(a) reveals that the UT method gets the best performance, while

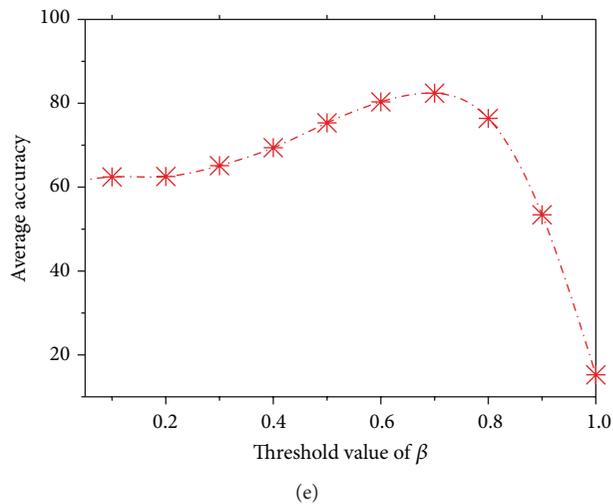
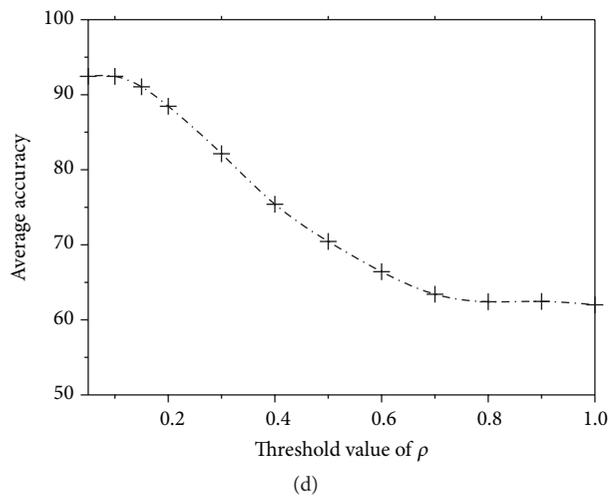
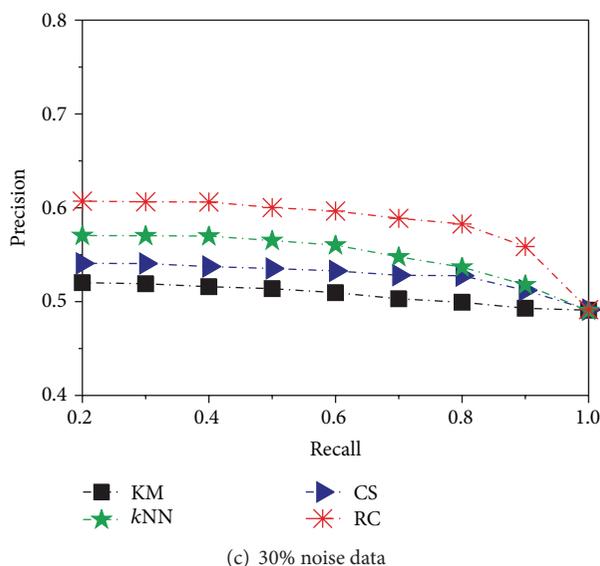
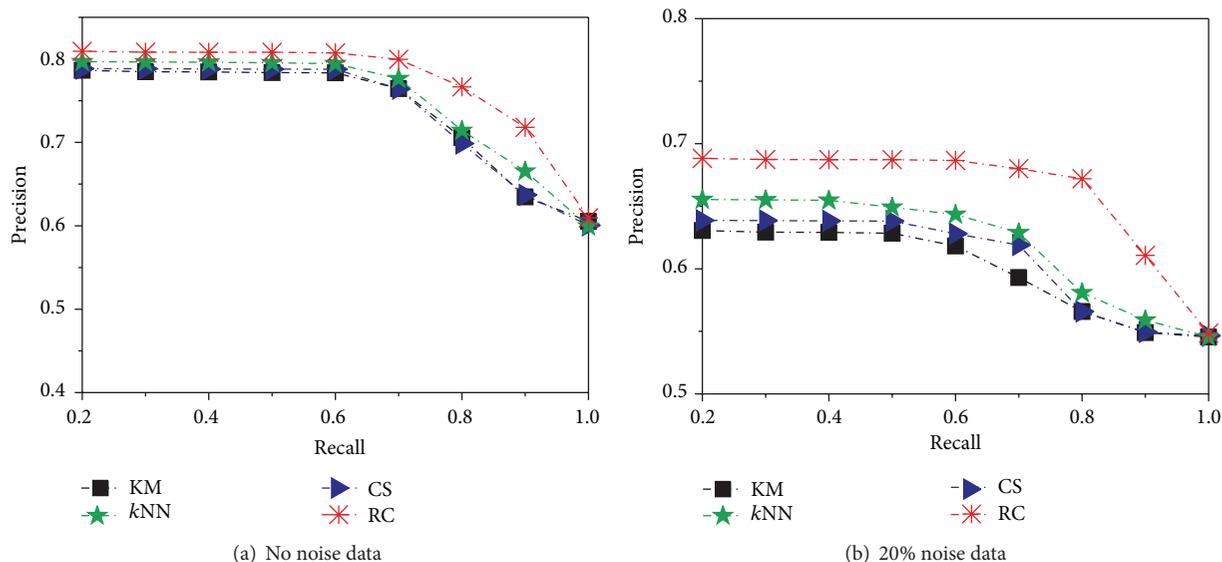


FIGURE 4: Performance evaluation of topic space relatedness clustering method.

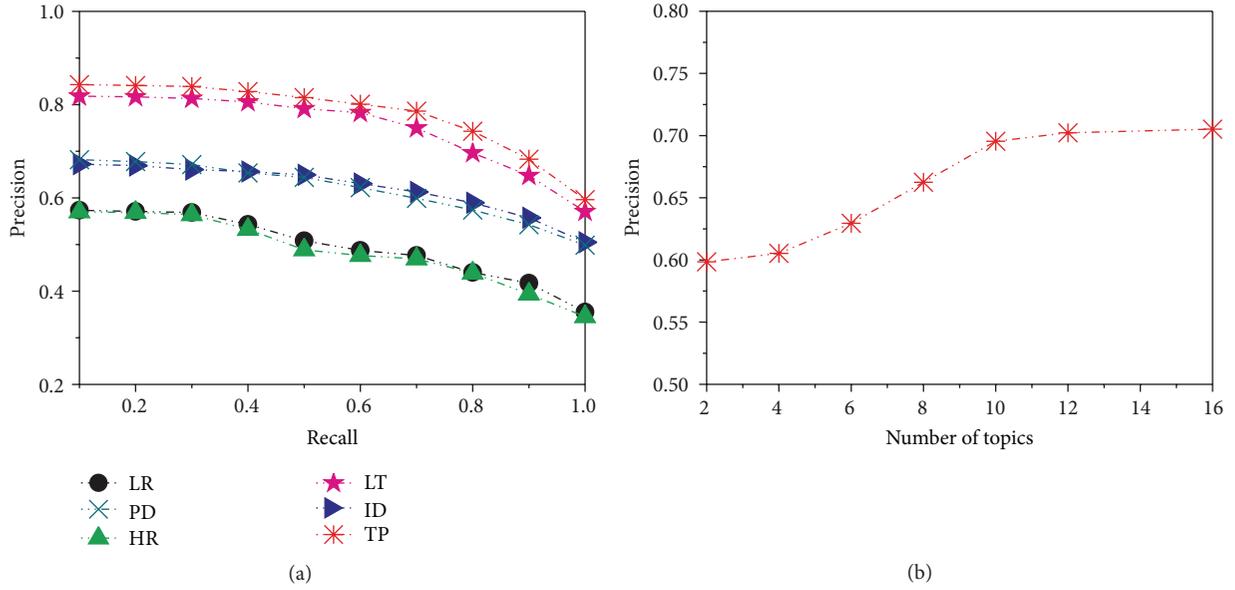


FIGURE 5: Effects of topic space impact evaluation.

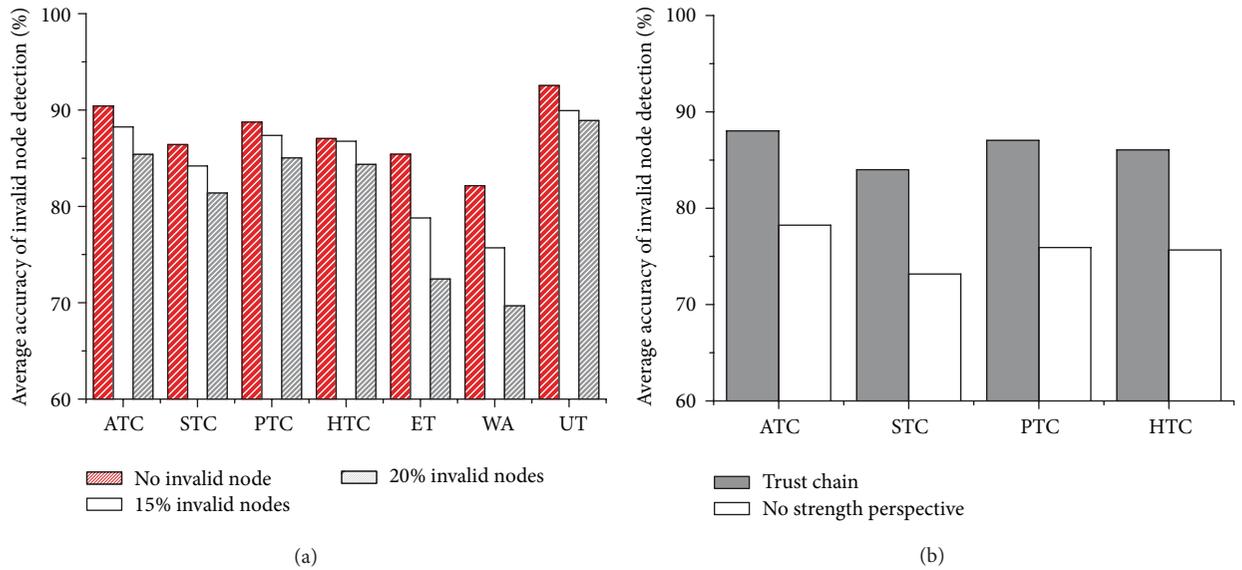


FIGURE 6: Performance evaluation of proposed trust chain calculation.

our trust chain methods including ATC, STC, PTC, and HTC get similar performances and are better than methods of ET and WA. We consider that the reasons are as follows: (1) the ultimate trust method provides trust calculation by its dynamic adjusting of factor by large computational costs, which results in all users maintaining trustworthy knowledge about others for detecting malicious interactions, and (2) trust chain methods can provide more comprehensive evaluation for relationships among users through proposed factors under relative low costs. In addition, we compare the performance of proposed trust chains (ATC, STC, PTC, and HTC) with strength perspective (strong and weak trust chain) and trust chains without strength constraints. The results are shown in Figure 6(b), and we can see that the performances of

proposed trust chains are obviously higher than the methods without strength constraints. With respect to value setting of thresholds ϑ_1, ϑ_2 , we have that values around 0.7 and 0.5 are reasonable compromise according to our empirical testing.

8.3. Examination of User Interest Degree Measurement. In this example, we examine the performance of proposed interest degree for topic space. Firstly, we test the effects of explicit interest degree calculation. We calculate users' explicit interest toward selected topic spaces and then compare the explicit interest quality of the following groups: view-time based interest (VT), weighted average of interest evidence based interest degree (WA), the explicit interest without dynamic predicting (ND), and the proposed explicit interest

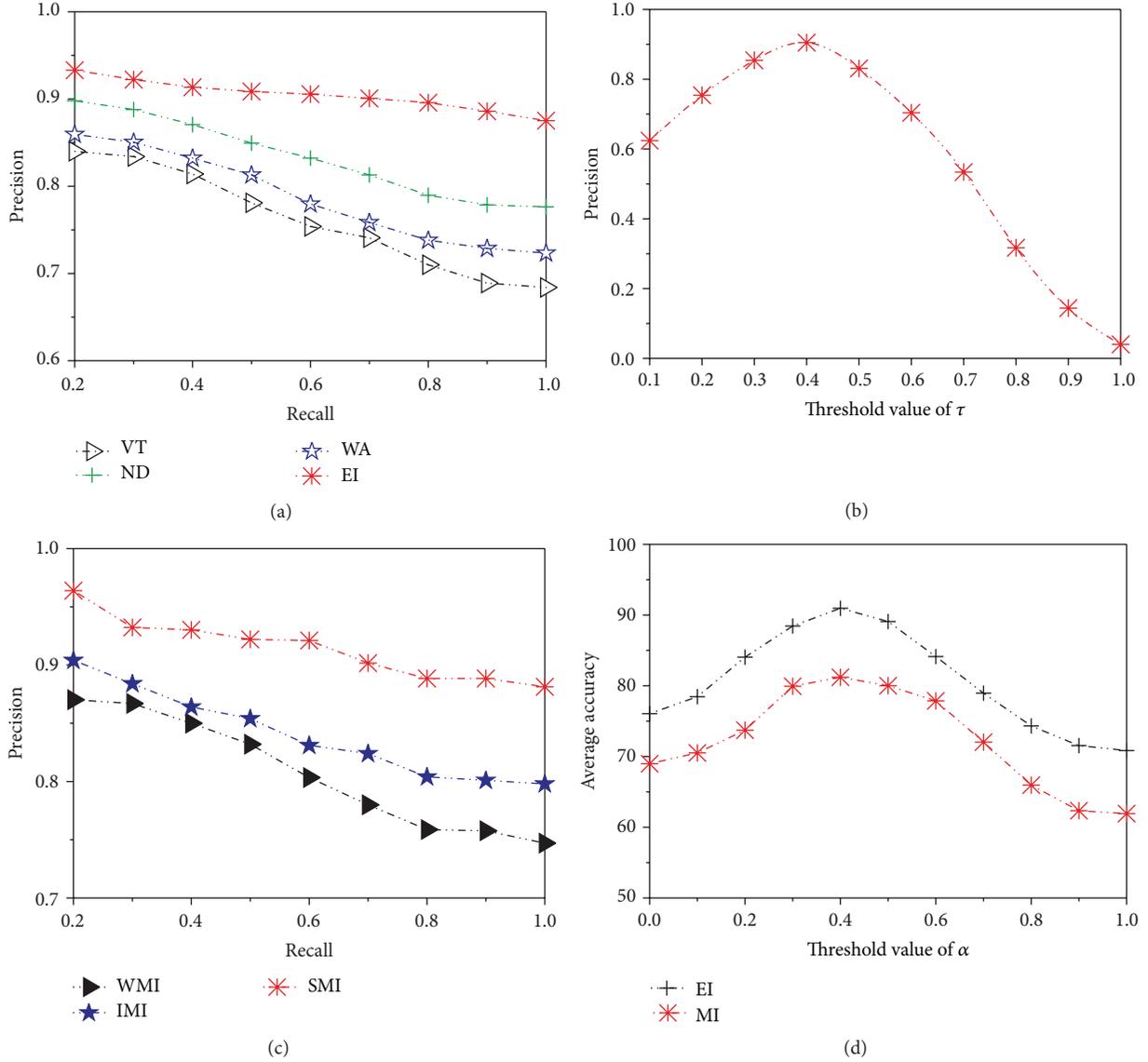


FIGURE 7: Examinations of user interest degree calculation.

calculation in this paper (EI). As shown in Figure 7(a), EI method outperforms VT, WA, and ND, with an improvement of 28%, 21%, and 13%, respectively. We consider that the reasons are as follows: (1) interest evidences are impacted by appearing probability and the weights of link and (2) dynamic predicting method is used for measuring latest interest. In addition, we can see the impact of threshold τ in Constraint 4 and results are shown in Figure 7(b). By empirical analysis, the threshold can be set as 0.4.

Then, we test the performance of implicit interest prediction for topic space. Figure 7(c) shows the performance of different implicit interest calculation methods, that is, strong trust chain based implicit interest (SMI), weak trust chain based implicit interest (WMI), and integrated implicit interest based on both strong trust chain and weak trust chain (IMI). We can see that the performance of SMI is best since the strong trust chain implies a mutual trustworthy relationship which brings a higher precision of implicit

interest prediction. The total average accuracy of implicit interest prediction based on trust chain is about 80.9%.

Finally, we verify the impacts of parameter α for interest degree calculation. We test the different values of α for explicit interest degree (EI) and implicit interest degree (MI), respectively. In Figure 7(d), we can see that the average accuracy is lower while α is too low or too high. On the basis of test validation, a threshold around 0.4 is a reasonable compromise.

8.4. Topic Space Oriented User Group Discovering Analyzing. In this examination, we verify the performance of our user group discovering method with trust chain based interest measuring (TI). We compare our proposed method with three methods: k nearest neighbor method (k NN), only explicit interest based user grouping method (EI), and relationship closeness (edge density) based user grouping method (RC). Figure 8 compares the user group discovering

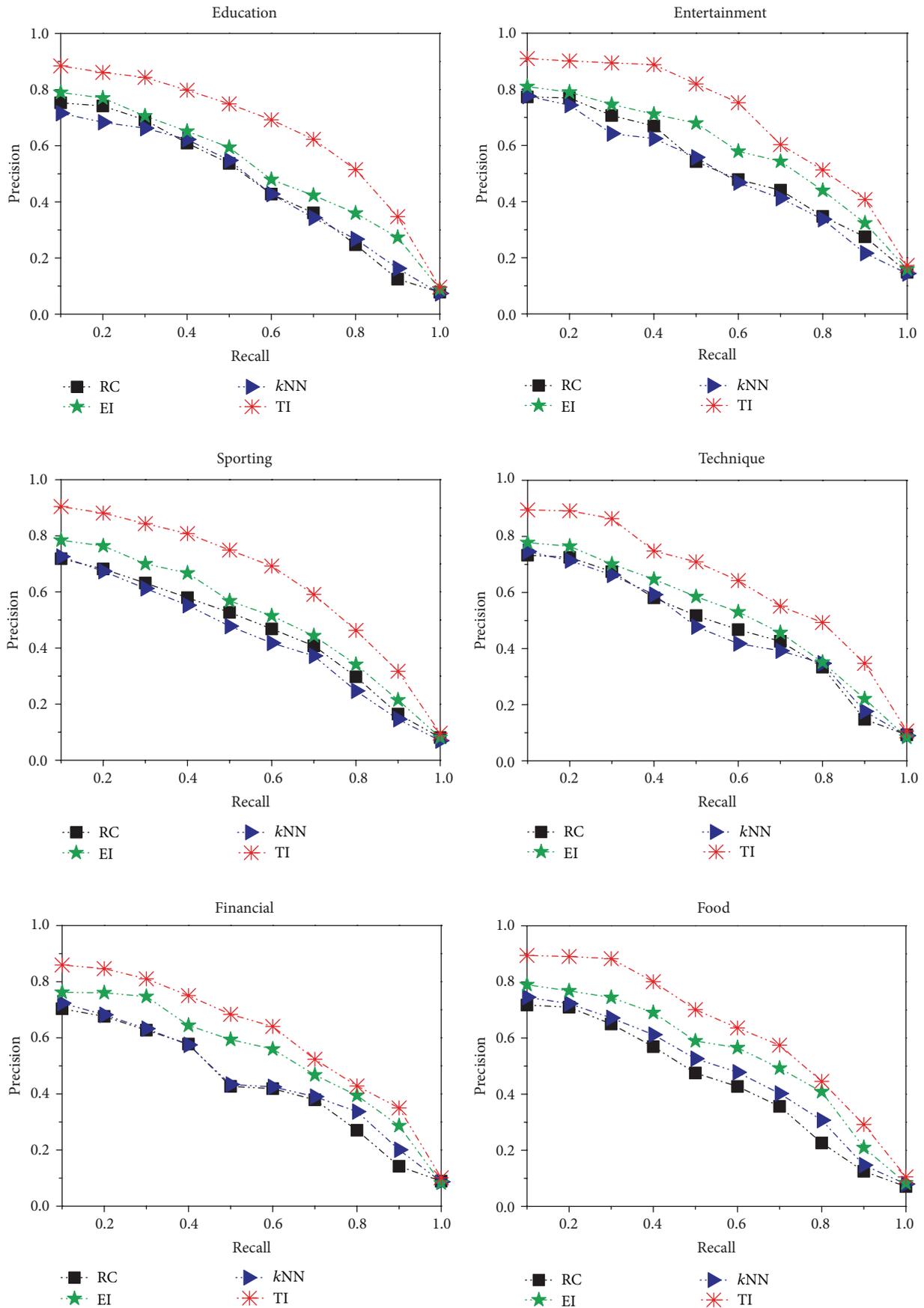


FIGURE 8: Continued.

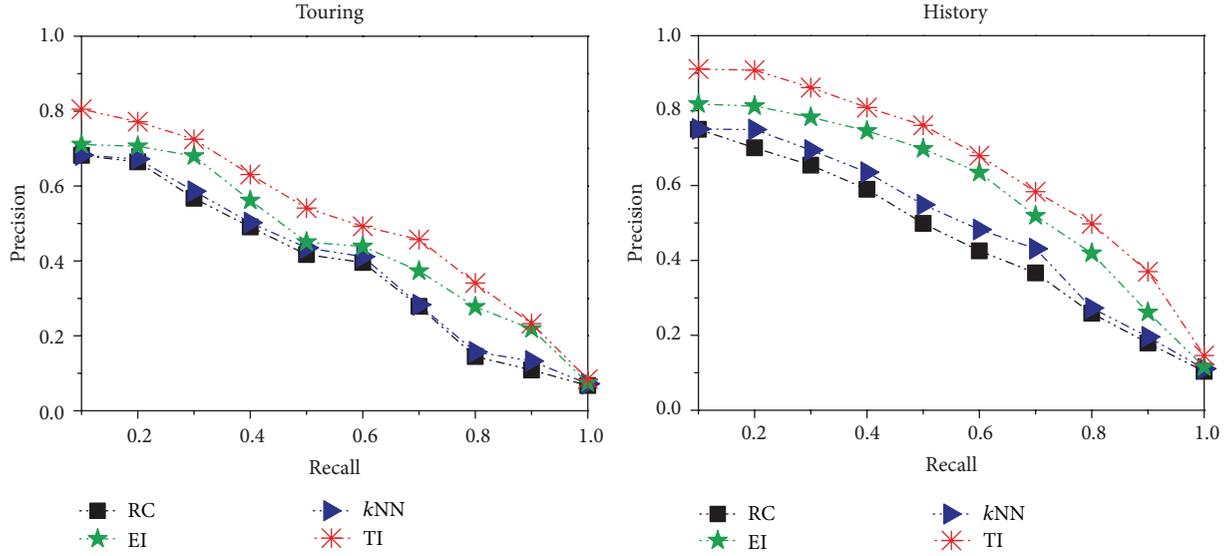


FIGURE 8: User group discovering comparison of proposed method TI with k NN, EI, and RC.

quality of TI, k NN, EI, and RC. As shown in Figure 8, k NN, EI, and RC obtain lower performance than TI in all eight kinds of topics, respectively. We consider that the reasons are as follows: (1) TI is an interest sensitive approach which adopts both explicit and implicit interests as a key criterion for user grouping, while k NN and RC do not take interest into account and EI only considers the explicit interest; (2) in TI, users are grouped within a close related way, strong or weak trust chain, for manifesting their close correlations, which decreases the error rate in ordinary user discovering.

Furthermore, we examine the effectiveness of user group discovering through online topic recommendation quality. In this examination, we recommend manually preprocessed topics, including both close related and irrelevant topic space, and then record the quality of recommendation. In the recommendation, the above mentioned topics are recommended to users who are in four kinds of user groups that are organized by methods of TI, with k NN, EI, and RC. Then, if a user performs behaviors of hitting, reading, forwarding, or posting judgments on his/her received topic, the recommendation is recorded as an accurate one. The precision of a recommendation to a user group is calculated as the average of all members in the user groups. There are six user groups corresponding to each kind of organizing method, respectively, and about 120 times of recommendation were launched in our experiment. We record the average precision of all kinds of user groups and the results are shown in Figure 9. We can see that recommendation quality of user group discovered through our proposed method is better than other three methods. That means close related topics are accepted and irrelevant topics are excluded by users in user group based on TI, which implies higher accuracy and better efficiency than user groups through k NN, EI, and RC.

9. Conclusion

For users in SNS, joining a user group to facilitate their communication is very common and inevitable. Commonly, the user groups are organized based on a set of topics having close internal correlations. However, most existing researches focus on user clustering based on their relation closeness degree and common explicit interests, while few efforts have been paid on users' interest interaction and expanding. Establishing user group based on influential topics has its practical significance. First, our work provides a valuable guideline on describing the topic oriented user group and computation methodology of user group organized formally and specifically for users. Each user can inquire about details of topic space, including influence degree, members of group, and estimating user interests for his/her further purpose of information communication. In practice, a machine driven mechanism can achieve higher efficiency than manual methods to reduce the overload of user group discovering and organizing. In addition, our proposed formal definitions and conceptions are quite appropriate for machine reading and understanding the calculation methods. Secondly, application of the influence and popularity helps us aggregating all aspects to get a more comprehensive impact degree about topics and topic spaces. This is because the proposed influence degree of topic reflects the structural closeness directly which also shows the integrated impacts about topic more or less, while proposed popularity degree reflects how influential a topic is from direct data through impartial views. Thereby, our proposed method can provide sound usability for influential topic space discovering. Thirdly, the proposed trust chain model shows acquiring direct and indirect trust among users and acquiring strength of indirect trust (strong or weak trust chains) for users. Meanwhile, the topological information of trust chain is fully considered in this work. That is, we calculate trust degrees of trust chain in different route compositions corresponding to the trustworthy

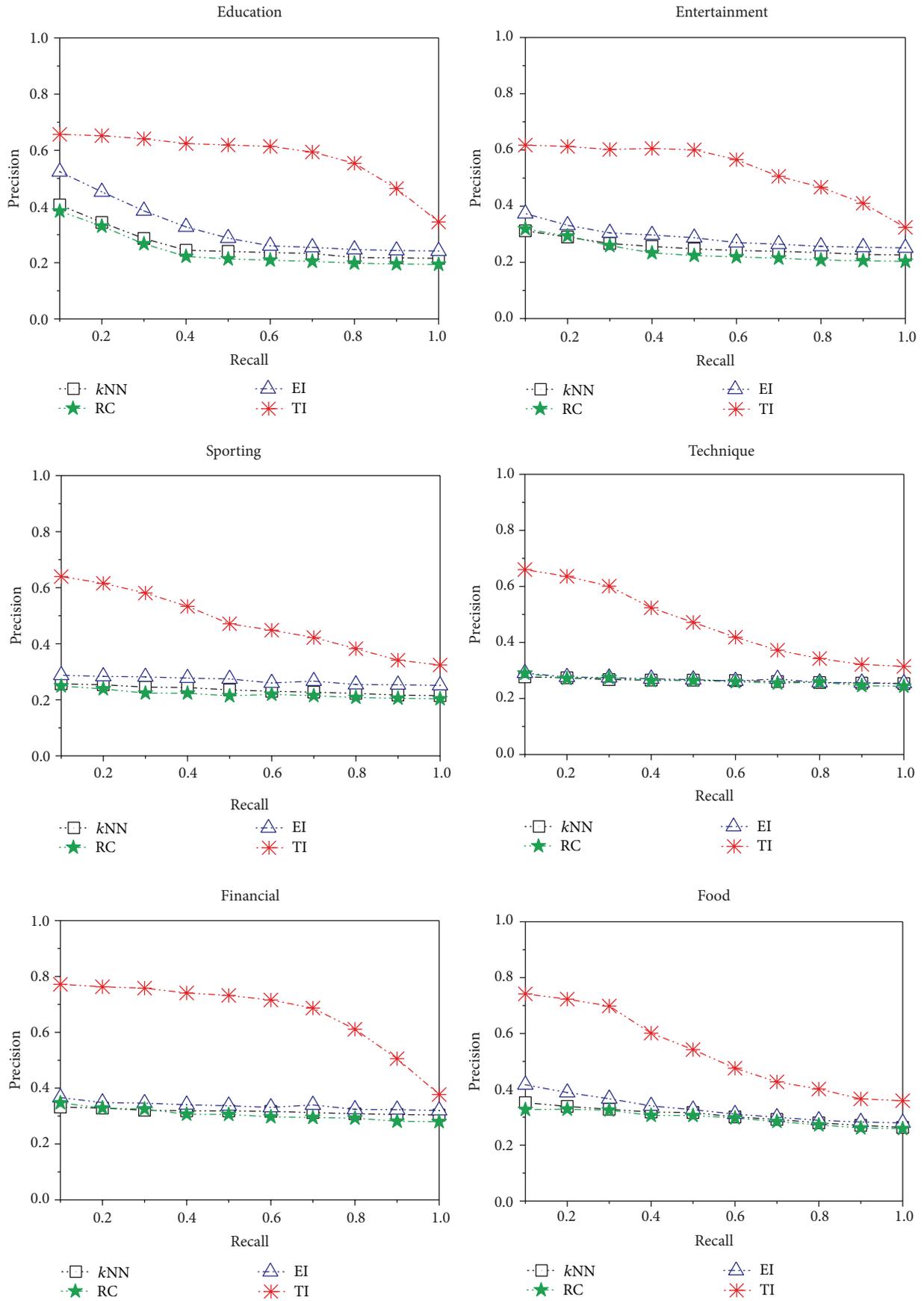


FIGURE 9: Continued.

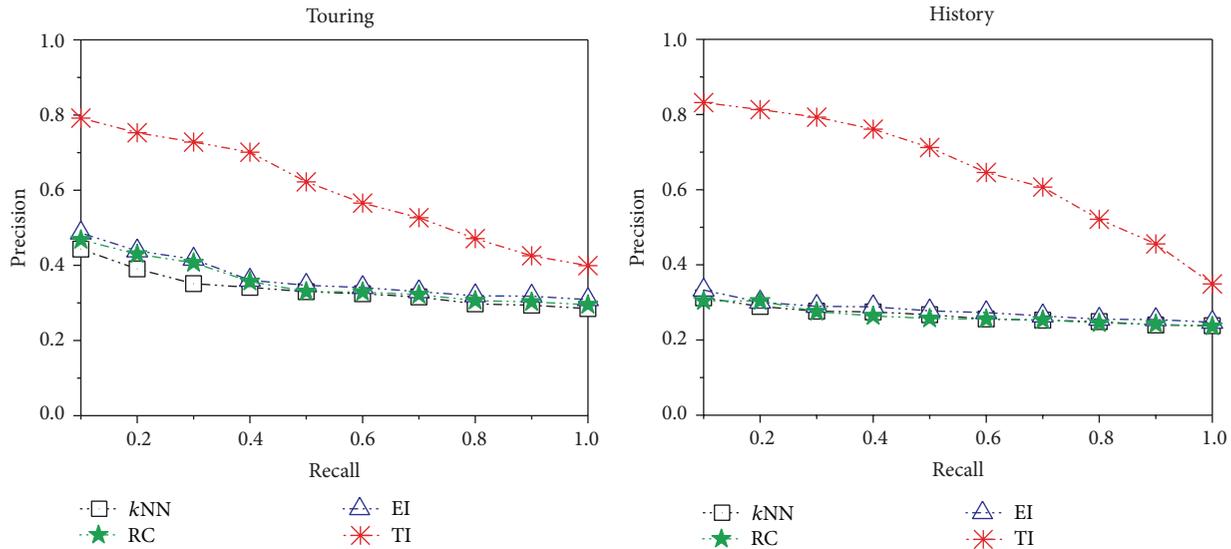


FIGURE 9: Effects of topic recommendation quality.

manners among people in real world; for example, trust is damped with the depth of route increasing and trust is enhanced when multiple direct neighbors link to the same target simultaneously from parallel paths. Furthermore, the proposed interest evaluation gives a novel method of calculating both explicit and implicit interest degrees, which may facilitate the automatic user preference analysis.

In our view, user group in SNS is an important community for users' daily discussing, information sharing, and recommendation under similar interests. Commonly, the user groups are organized based on a set of topics having close internal correlations, which means that close related topics may attract more interests of users to share their information. In this study, our main purpose is to find a method for gathering influential and close related topics and predicting all probable interested users (including explicit interested and implicit interested users) to form user groups. To do so, we propose a topic space discovering scheme with a trust relationship based interest measuring perspective, which contains four aspects: influential topic space construction through topic relatedness clustering and impact evaluation, user trust chain evaluation based on SNS topological information, explicit and implicit interest estimation based on user trust chain, and user group discovering. Through empirical examinations based on our collected dataset manually, results and evaluations show the efficiency and feasibility of our proposed scheme. Our future work will be in study of user group organization with the concern of overlapped structure.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

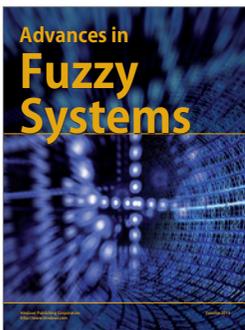
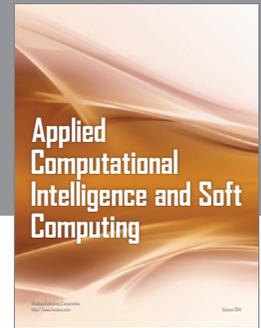
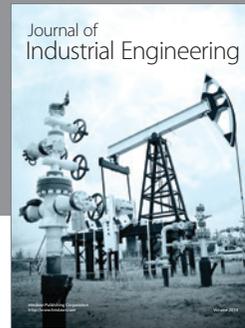
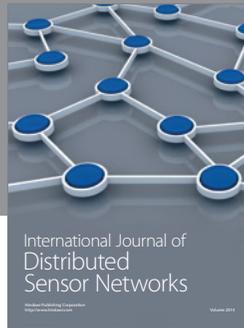
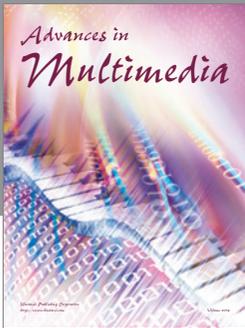
This work is funded by National Basic Research Program of China (2014CB340404), National Natural Science Foundation of China (61572326, 61103069), Innovation Program

of Shanghai Municipal Education Commission (13YZ052, ZZyy14003), China Postdoctoral Science Foundation funded project (2014M551334), and Program of Shanghai Institute of Technology (YJ2014-06).

References

- [1] K. Yang, X. Cheng, L. Hu, and J. Zhang, "Mobile social networks: state-of-the-art and a new vision," *International Journal of Communication Systems*, vol. 25, no. 10, pp. 1245–1259, 2012.
- [2] L. Backstrom, D. Huttenlocher, J. Kleinberg et al., "Group formation in large social networks: membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*, pp. 44–54, Philadelphia, Pa, USA, August 2006.
- [3] B. Gérard, D. Miliotis, and P. Jacquet, "One day in twitter: topic detection via joint complexity," in *Proceedings of the Snow Challenge, 2nd Workshop on Social News on the Web (WWW '14)*, Seoul, South Korea, April 2014.
- [4] D. Ramage, S. Dumais, and D. Liebling, "Characterizing microblogs with topic models," in *Proceedings of the 4th International Conference on Weblogs and Social Media*, Washington, DC, USA, May 2010.
- [5] N. B. Ellison, J. Vitak, R. Gray, and C. Lampe, "Cultivating social resources on social network sites: facebook relationship maintenance behaviors and their role in social capital processes," *Journal of Computer-Mediated Communication*, vol. 19, no. 4, pp. 855–870, 2014.
- [6] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 497–506, Paris, France, July 2009.
- [7] C. T. Li, T. T. Kuo, C. T. Ho, S. Hong, W. Lin, and S. Lin, "Modeling and evaluating information propagation in a microblogging social network," *Social Network Analysis and Mining*, vol. 3, no. 3, pp. 341–357, 2013.

- [8] D. Li, Q. Lv, X. Xie et al., "Interest-based real-time content recommendation in online social communities," *Knowledge-Based Systems*, vol. 28, pp. 1–12, 2012.
- [9] Q. Wang, J. Wang, J. Yu, M. Yu, and Y. Zhang, "Trust-aware query routing in P2P social networks," *International Journal of Communication Systems*, vol. 25, no. 10, pp. 1260–1280, 2012.
- [10] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.
- [11] B. Danah, G. Scott, and L. Gilad, "Tweet, tweet, retweet: conversational aspects of retweeting on twitter," in *Proceedings of the 43rd IEEE Hawaii International Conference on System Sciences (HICSS '10)*, pp. 1–10, Honolulu, Hawaii, USA, January 2010.
- [12] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [13] L. Gauvin, A. Panisson, and C. Cattuto, "Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach," *PLoS ONE*, vol. 9, no. 1, Article ID e86028, 2014.
- [14] W.-Y. Gan, N. He, D.-Y. Li, and J.-M. Wang, "Community discovery method in networks based on topological potential," *Journal of Software*, vol. 20, no. 8, pp. 2241–2254, 2009.
- [15] L. Yang, J. Xin-Sheng, and L. Cai-Xia, "Optimizing community detection using the pre-processing of edge weighted based on random walk in networks," *Journal of Electronics & Information Technology*, vol. 35, no. 10, pp. 2335–2340, 2013.
- [16] F. Zaidi, "Small world networks and clustered small world networks with random connectivity," *Social Network Analysis and Mining*, vol. 3, no. 1, pp. 51–63, 2013.
- [17] F. L. Pinheiro, M. D. Santos, F. C. Santos, and J. M. Pacheco, "Origin of peer influence in social networks," *Physical Review Letters*, vol. 112, no. 9, Article ID 098702, 2014.
- [18] G. Zhu, Z. Zhou, F. Han, and Z. Ying, "The domain classification algorithm based on KNN in Micro-blog," in *Proceedings of the 4th IEEE International Conference on Software Engineering and Service Science (ICSESS '13)*, pp. 188–192, IEEE, Beijing, China, May 2013.
- [19] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Expert Systems with Applications*, vol. 40, no. 1, pp. 200–210, 2013.
- [20] S. K. Pal, S. Kundu, and C. A. Murthy, "Centrality measures, upper bound, and influence maximization in large scale directed social networks," *Fundamenta Informaticae*, vol. 130, no. 3, pp. 317–342, 2014.
- [21] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 199–207, ACM, July 2009.
- [22] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang, "Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM '13)*, pp. 657–666, ACM, Rome, Italy, February 2013.
- [23] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM '10)*, pp. 88–97, IEEE, Sydney, Australia, December 2010.
- [24] J. Zeng, S. Zhang, and C. Wu, "A framework for WWW user activity analysis based on user interest," *Knowledge-Based Systems*, vol. 21, no. 8, pp. 905–910, 2008.
- [25] V. Hegde, J. X. Parreira, and M. Hauswirth, "Semantic tagging of places based on user interest profiles from online social networks," in *Advances in Information Retrieval*, pp. 218–229, Springer, Berlin, Germany, 2013.
- [26] S. D. Kamvar and M. T. Schlosser, "EigenRep: reputation management in P2P networks," in *Proceedings of the 12th International World Wide Web Conference (WWW '03)*, pp. 123–134, Budapest, Hungary, May 2003.
- [27] K. Thirunarayan, P. Anantharam, C. Henson, and A. Sheth, "Comparative trust management with applications: Bayesian approaches emphasis," *Future Generation Computer Systems*, vol. 31, no. 1, pp. 182–199, 2014.
- [28] A. A. Selçuk, E. Uzun, and M. R. Pariente, "A Reputation-based trust management system for P2P networks," *International Journal of Network Security*, vol. 6, no. 3, pp. 235–245, 2008.
- [29] F. Javier Ortega, J. A. Troyano, F. L. Cruz, C. G. Vallejo, and F. Enríquez, "Propagation of trust and distrust for the detection of trolls in a social network," *Computer Networks*, vol. 56, no. 12, pp. 2884–2895, 2012.
- [30] B. Qureshi, G. Min, and D. Kouvatsos, "Trusted information exchange in peer-to-peer mobile social networks," *Concurrency Computation Practice & Experience*, vol. 24, no. 17, pp. 2055–2068, 2012.
- [31] S. Al-Oufi, H.-N. Kim, and A. El Saddik, "A group trust metric for identifying people of trust in online social networks," *Expert Systems with Applications*, vol. 39, no. 18, pp. 13173–13181, 2012.
- [32] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *ACM Transactions on the Web*, vol. 3, no. 4, pp. 1–33, 2009.
- [33] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: bringing order to the web," Tech. Rep. 1999-66, Stanford University, 1999.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

