Hindawi
Complexity
Volume 2017, Article ID 4783159, 16 pages
https://doi.org/10.1155/2017/4783159

Research Article

Mining Community-Level Influence in Microblogging Network: A Case Study on Sina Weibo

Yufei Liu,[1] Dechang Pi,[1,2] and Lin Cui[1]

[1]*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China*
[2]*Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, Jiangsu 211106, China*

Correspondence should be addressed to Dechang Pi; dc.pi@nuaa.edu.cn

Received 7 June 2017; Accepted 12 November 2017; Published 4 December 2017

Academic Editor: Jia Wu

Social influence analysis is important for many social network applications, including recommendation and cybersecurity analysis. We observe that the influence of community including multiple users outweighs the individual influence. Existing models focus on the individual influence analysis, but few studies estimate the community influence that is ubiquitous in online social network. A major challenge lies in that researchers need to take into account many factors, such as user influence, social trust, and user relationship, to model community-level influence. In this paper, aiming to assess the community-level influence effectively and accurately, we formulate the problem of modeling community influence and construct a community-level influence analysis model. It first eliminates the zombie fans and then calculates the user influence. Next, it calculates the user final influence by combining the user influence and the willingness of diffusing theme information. Finally, it evaluates the community influence by comprehensively studying the user final influence, social trust, and relationship tightness between intrausers of communities. To handle real-world applications, we propose a community-level influence analysis algorithm called CIAA. Empirical studies on a real-world dataset from Sina Weibo demonstrate the superiority of the proposed model.

1. Introduction

Community-level influence analysis is an emerging problem, which can be used in many filed, for example, recommendation system [1, 2], public opinion prediction [3], and cybersecurity analysis [4]. There are many researchers who are interested in analyzing the social influence in social networks [5], but rarely assessing the influence in community level. With the rapid spread of online social networks, such as Twitter, Facebook, and Sina Weibo, large amounts of data with the real world are produced, which provide support for the social influence analysis.

How to establish an effective model for analyzing community-level influence has become an important research for online social network. Community-level influence is greater than individual-level influence, but few researchers have studied community influence. The existing studies establish various social influence analysis models [6, 7], but they just study the influence in the individual level and mostly ignore the existence of a common influence pattern from a community that includes multiple nodes. A large number of achievements have been obtained on individual-level influence, but most of the studies are based on static statistics method [8–11], link analysis algorithms [12–14], or probabilistic models [15–17]. These studies do not consider whether the user is willing to receive or diffuse information or what the role of social trust between users is or do not remove zombie fans. However, these factors are very important for analyzing the social influence. Meanwhile, the existing works about community-level influence focus on the influence strength between communities and ignore the problem of analyzing the community-level influence. For example, Belák et al. [18] calculated the community-level influence by only averaging influence of all users in a community.

An important observation is that zombie fans have no contribution to the social influence, and the willingness of users to diffuse information has a certain effect on the accuracy of calculating social influence, and social trust plays an important role in social influence. The trust degree of user A to user B determines the influence of user B on user A. The more the user A trusts user B, the more influence the user B has on the user A. Because user influence is the basis of the community influence, a little carelessness on the former will lead to errors on the later.

Aiming to assess the community-level influence effectively and accurately, we construct a community-level influence analysis model that can assess community influence. Based on our model, a community-level influence analysis algorithm (short for CIAA) is proposed, which can assess the community influence more effectively and accurately. The main idea of our model is as follows. First, we eliminate the interference of zombie fans on the social influence to make the results more accurate. Then, in the process of calculating user influence, we consider the social trust and use the random walk method to calculate the user influence. In evaluating the user's theme information, the user mean willingness is calculated by exploring the content related to the user's theme information. We combine these two factors (the user influence and the user willingness to diffuse theme information) to calculate the user final influence. Finally, the community-level influence is calculated by comprehensively studying the user final influence, the social trust, and relationship tightness between intrausers of communities. Experiments are conducted on a real-world dataset crawled from Sina Weibo. Comparing with the state-of-the-art algorithm (the averaging user influence algorithm [18]), the results show that our model is more effective and accurate to evaluate the community-level influence.

The contributions of this paper can be summarized as follows. (1) We formulate the problem of analyzing the community-level influence and design a community-level influence analysis model. (2) CIAA, a community-level influence analysis algorithm based on our model, is proposed, which is effective and reliable to evaluate the community influence of microbloggers from Sina Weibo. (3) We conduct extensive experiments to assess the performance of the proposed model. Experimental results on the real-world dataset demonstrate the superiority of the proposed CIAA.

The rest of the paper is organized as follows. In Section 2, we summarize the related works. In Section 3, we propose the community-level influence analysis model and give an example to illustrate its working principle, and the CIAA is proposed. In Section 4, we conduct experiments on the real-world dataset crawled from Sina Weibo and then analyze the performance of the proposed approach. Finally, we state the conclusion and future work in Section 5.

2. Related Works

Since Katz and Lazarsfeld [19] found that social influence plays an important role in social life and decision-making in the 1950s, researchers in computer field have spare no effort to study the relevant problems. It is found that the popular users play an important role in adopting innovation, social public opinion propagation and guidance, group behavior formation and development [5], and so on.

There are a great deal of research efforts to measure individual-level influence [20, 21], typically, the "opinion leaders." Existing methods can be categorized into three types: the network structure based methods, the user behavior based methods, and the mutual information based methods. The network structure based methods are degree centrality [22], closeness centrality [23], betweenness centrality [24], eigenvector centrality [25], Katz centrality [26], PageRank [27], and clustering coefficient [28]. We know that node degree essentially means the connection between a node and its neighbors. The method based on node degree can intuitively express this meaning, and its computational cost is smaller than other methods [29]. These methods are widely used in measuring the users' influence in the social network. However, the methods based on node degree only reflect the connection between the users and their neighbors and cannot measure the users' influence in the entire social network for the local influence of users. For example, based on the community scale-sensitive maxdegree, Hao et al. [30] proposed an influential users discovering approach called CSSM when placing advertisements. CSSM uses the degree centrality and neighbor's degree to evaluate node's (microbloggers) influence. However, the algorithm does not consider the contribution of microblogs to user influence. Comparing with the methods based on the degree, the method based on the shortest path (closeness centrality and betweenness centrality) can measure the individual-level influence in the entire social network. Nevertheless, its computational complexity is higher than the degree centrality method. For example, based on text mining and social network analysis, Bodendorf and Kaiser [31] proposed an approach to detect opinion leaders in directed graph of user communication relationship. It can predict tendency of network opinion leaders via closeness centrality and betweenness centrality. Moreover, measuring the individual-level influence by the shortest path is an ideal status, and it is difficult to achieve in the real-world application scenarios. Besides, the methods based on random walk only consider the structure characteristics of the node while ignoring the behavior characteristics. For example, Xiang et al. [32] provided an understanding of PageRank and authority from an influence propagation perspective by performing random walks. However, they did not consider the personal attributes to understanding of PageRank as well as the relationship between PageRank and social influence analysis. Zhu et al. [33] proposed a novel information diffusion model called CTMC-ICM, which introduces the continuous-time Markov Chain theory into the Independent Cascade Model. Based on the model, they proposed a new ranking metric called SpreadRank. Based on continuous-time Markov process, Li et al. [34] proposed a dynamic information propagation model called IDM-CTMP to predict the influence dynamics of social network users. IDM-CTMP defined two other dynamic influence metrics and could predict the spreading coverage of a user within a given time period. Zhou et al. [35] established new upper bounds to significantly reduce the number of Monte-Carlo simulations in greedy-based algorithms, especially at the initial step. Based on the bound, they proposed a new upper bound based lazy forward algorithm for discovering the top-k influential nodes in social networks.

The aforementioned models focus only on assessing the social influence of single individuals. However, a small number of works attempt to build models on the community influence analysis. Qi et al. [36] applied degree centrality, closeness centrality, and betweenness centrality to groups and classes as well as individuals. Latora and Marchiori [37] put forward a group information centrality to measure the importance of node sets. Mehmood et al. [38] exploited information diffusion records to calculate the influence strength between different communities. Although these works preliminarily study the community-level influence, none of them focuses on how to measure a community's influence. Belák et al. [18] assessed the community-level influence according to the average of the all users' influence in the same community. Because the distribution of the users' influence is uneven in different communities, average based method is inequitable to bigger communities, while summation based method is inequitable to smaller ones. At present, community-level influence analysis is still a challenging problem.

3. Proposed Methodology

We construct our model and implement the corresponding algorithm in this section. First, we give the related definitions in Section 3.1. Then, we propose the community-level influence analysis model for microbloggers. Next, we describe the working principle of our model via an example in Section 3.2. Finally, the community-level influence analysis algorithm is proposed in Section 3.3.

3.1. Related Definitions and Community-Level Influence Analysis Model

3.1.1. Related Definitions. Social networks and communities are described as follows: a typical social network can be represented as a bipartite graph $G = \{V, E\}$, V is a set of nodes (users) in a social network, and E is a set of edges used to describe the relationships between nodes. A community can be represented as a subgraph of a social network: that is, $C = \{CV, CE\}$; $CV \subseteq V$ is a set of users in a community. $CE \subseteq E$ is a set of relationships between users within a community. A node is defined as a user within the community if he/she belongs to the community; otherwise, he/she is defined as a user outside the community. The set of users outside the community is written as UOC. Modeling and calculating the community influence of C_i are the basis of our work, and the objective function of our model is as follows:

$$\mathrm{CI}\left(C_i\right) = f\left(G, C_i\right). \tag{1}$$

$\mathrm{CI}(C_i)$ denotes the community influence of the community C_i, and the function $f(G, C_i)$ indicates that the assessment method is based on G and C_i. There are two entities (i.e., users and communities) which can produce influence. To study the community-level influence, we give the related definitions as follows.

Definition 1.

Trust. A node in a social network has a certain trust degree in other nodes according to its past contact with other nodes or the reputation of other nodes [39, 40]. According to the different sources of trust, we divide the trust into direct trust and indirect trust.

(1) Direct Trust (DT). Assume that the node v is the entry node of the node u, indicating that there is contact between u and v. According to the previous contacts and the reputation of u, v will have direct trust on u.

(2) Indirect Trust (IT). Assume that the node u is the reachable node of the node v; v will have indirect trust on u because the reputation of u can be transmitted to v.

Users not only have mutual trust, but also mutually influence each other. According to the different sources of influence, this paper divides the influence into direct influence and indirect influence.

Definition 2.

(1) Direct Influence (DI). Assume that the node v is the entry node of the node u; u will have an influence on v: that is, u produces direct influence on v.

(2) Indirect Influence (II). Assume that the node u is a reachable node of the node v; u will have an influence on v through transmission layer by layer: that is, u produces indirect influence on v.

In order to assess the overall influence of u on v, we define the user combined influence.

Definition 3.

User Combined Influence (UCI). Because v has direct trust or indirect trust to u, and u has direct influence or indirect influence on v, we comprehensively combine the four factors to calculate the combined influence of u on v.

Definition 4.

(1) User Influence (UI). User influence refers to the influence of individual on other users.

(2) Community Influence (CI). Community influence is the overall influence of the community, which is formed by the UI of all the users in the community and the community's self-factors.

Definition 5.

Mean Willingness to Diffuse Theme Information (MW). In communities, some users receiving the theme information may not diffuse it, some users prefer to post their own blog, and some users prefer to forward others' blog. We assess the community influence by taking into account the diffusion of information between users. MW represents a user' willingness to diffuse the information of a blog. The theme information of the user u is stored in the set $T(u) = \{t_{u1}, t_{u2}, \ldots, t_{uj}, \ldots\}$, where t_{uj} represents the user's jth theme information. If t_{uj} is diffused in a social network, a path map g_{uj} is formed to describe the propagation path. We store the path graphs formed by $T(u)$ in the set $g(u) = \{g_{u1}, g_{u2}, \ldots, g_{uj}, \ldots\}$.

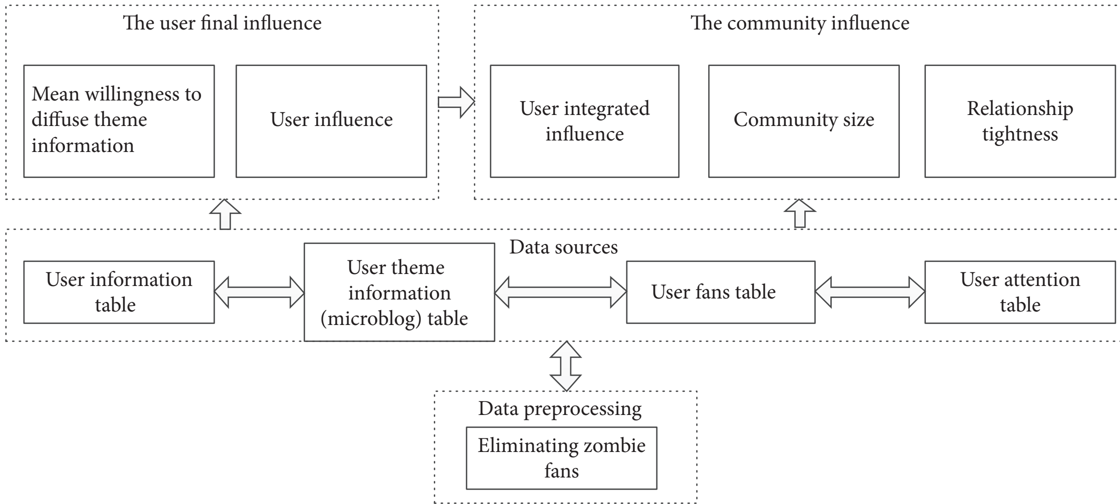

Figure 1: The framework of the proposed model.

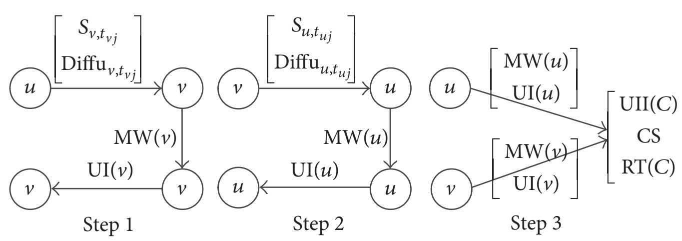

Figure 2: The working steps of the community-level influence analysis model.

3.1.2. Model Framework. Our model consists of four modules: data preprocessing module, data source module, the user final influence module, and the community influence module. Figure 1 shows our model framework.

Data preprocessing module is used to eliminate zombie fans. We judge the zombie fans from the behavior dimension and time dimension. Behavior dimension is based on the amount of theme information posted by the user and the fans' influence of the user. Time dimension is based on the user login frequency and the frequency of diffusing theme information. Finally, the data preprocessing results are stored to the data source.

Data source module is responsible for providing the relevant data needed for influence analysis. We establish the user information table, the microblog table, the user fans information table, and the user attention table to access the user's relevant information efficiently.

The user final influence module first calculates the mean willingness to diffuse theme information for each user in a community and then calculates the user's influence. Next, it combines these two results to get the user final influence.

The community influence module first calculates the community size, the tightness of user relationship, and the user-integrated influence in the community and then evaluates the community influence by integrating the three factors.

3.2. Working Principle. In this subsection, we introduce the working principle of each module in the model framework in detail. We assume that u and v are two users in community C. After performing data preprocessing, Figure 2 shows the working principle, where the mathematical notations will be described in the following subsections in detail.

The working principle can be described as the following steps.

Step 1. Calculate the Diffu_v and S_v of v. Then calculate the $\mathrm{MW}(v)$ of v. Finally, calculate $\mathrm{UI}(v)$ of v.

Step 2. According to Step 1, calculate the $\mathrm{MW}(v)$ and $\mathrm{UI}(v)$ of u.

Step 3. Integrate MW and UI to calculate the $\mathrm{UII}(C)$. Then calculate CS and $\mathrm{RT}(C)$. Finally, combine the three factors to calculate the community influence.

3.2.1. Data Preprocessing. In microblogging networks, some users of ulterior motives or business purpose lead to producing the zombie fans. According to the definition in [41], zombie fans are the users who are fake fans generated and maintained mostly for economic purpose. Zombie fans certainly interfere in analyzing the social influence. A small number of empirical researches have been conducted on recognizing zombie fans [41–43]. The existing studies were mostly subject to the Twitter platform.

Presently, researchers generally detect the zombie fans based on the amount of attention, the number of fans,

(1) **Input**: V, E, LF, DAF, NUI, NAU, NUF
(2) **Output**: $G = (V, E)$
(3) Select the users who are the last 10% of the login frequency and whose login time interval is greater than 7 days, into the set LF
(4) Put the users with the top 10% of the diffusing advertisement frequency into the set DAF
(5) Select the users who are the last 10% of the number of user' theme information into the set NUI
(6) Put the users with the top 10% of the attention users into the set NAU
(7) Put the users with the number of fans between 10–200 into the set NUF
(8) ZF = LF ∩ DAF ∩ NUI ∩ NAU ∩ NUF
(9) Update $V = V - \mathrm{ZF}$ and $E = E - E_{\mathrm{ZF}}$
(10) **return** V, E

Algorithm 1: Eliminating zombie fans.

original and forward information frequencies, and other basic attributes. With the ever-changing escalation of zombie fans, zombie fans will produce more features [44]. The existing feature-based methods to eliminate zombies may gradually fail. We observe that because zombie fans are occasionally managed via software program or a few people behind, zombie fans often rarely speak, even seldom log in, or no longer are used; and their behaviors can be vastly different with ordinary users in profile information and contents. Moreover, no matter how the features of zombie fans change, they can be split into time dimension and behavior dimension. Thus, it is reasonable to recognize zombie fans from the time dimension and behavior dimension, and it is more able to adapt to the needs of detecting zombie fans in microblogging networks.

According to expert knowledge criteria [45], in the time dimension, we assess zombie fans from the user login frequency and the diffusing advertisement frequency. Thus, time dimension includes login frequency (LF) and diffusing advertisement frequency (DAF). Login frequency refers to the number of logins in a period. The lower the frequency of login is, the higher the probability of the user becoming zombie fans is. The login frequency is calculated as follows:

$$\mathrm{LF} = \frac{\Delta t.\mathrm{LoginNumber}}{\Delta t}, \tag{2}$$

where LoginNumber indicates the number of logins. The higher the diffusing advertisement frequency is, the higher the probability of the user becoming zombie fans is. The diffusing advertisement frequency is calculated as follows:

$$\mathrm{DAF} = \frac{\Delta t.\mathrm{NumberOfDiffusingAdvertisement}}{\Delta t}, \tag{3}$$

where NumberOfDiffusingAdertisement indicates the number of diffusing advertisement frequencies.

For the same reason, in the behavior dimension, we assess zombie fans from the amount of user theme information and the individual influence of the user's fans. Thus, we take into account the number of user theme information (NUI), the number of attention users (NAU), and the number of user's fans (NUF).

To ensure that the criteria of the parameters are reliable, the corresponding criteria are obtained by prior knowledge, expert knowledge, or experimental trial. For example, we select the users who are the last 10% of the login frequency and whose login time interval is greater than 7 days into the set LF. To reduce the amount of calculation, we filter all users in a microblogging network. If a user has a certified user in his/her fans, the user is not considered a zombie fan. If a user does not have a certified user in his/her fans, the details to eliminate zombie fans can be described in Algorithm 1.

As we can see that, unlike the classification and pattern recognition, the proposed method to eliminating zombie fans does not require labeled data and training model. It is effective and easy to use in practice.

3.2.2. The User Final Influence. The traditional models are simple, not taking into account the degree of social trust between users and the user's willingness to diffuse theme information. However, the two factors are important to the user final influence. In this paper, the user final influence is calculated by integrating the MW and UI. Because the influence of a user on other users is related to the user's willingness to exert his/her influence, the bigger the value of MW, the greater the probability of the user diffusing a theme information. UFI is calculated as follows:

$$\mathrm{UFI}(u) = \mathrm{MW}(u) \times \mathrm{UI}(u). \tag{4}$$

Mean Willingness to Diffuse Theme Information. The higher frequency of diffusing theme information means a higher user influence, because more users will know the user. Therefore, MW reflects the probability that a user has high-impact in a microblogging network. The parameter $S_{v,T_{uj}}$ indicates the state of receiving theme information for the user v as follows:

$$S_{v,t_{uj}} = \begin{cases} 0, & \text{The user has never received the theme information} \\ 1, & \text{The user has received the theme information.} \end{cases} \tag{5}$$

The initial value of $S_{v,T_{uj}}$ is set to 0. Meanwhile, to know the result of v diffusing the theme information t_{uj}, we observe

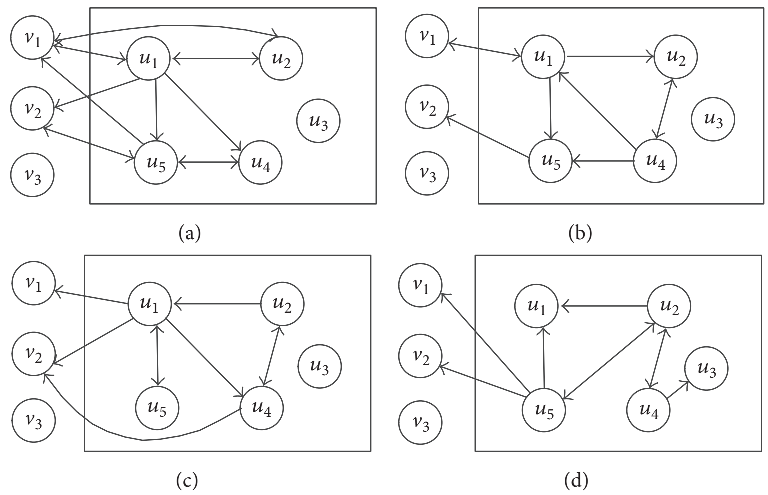

FIGURE 3: An example of calculating MW: there are five users inside a community, that is, u_1, u_2, u_3, u_4, and u_5. There are three users outside the community, that is, v_1, v_2, and v_3. (a) shows the relationship between these users. (b) shows the diffusion of theme information from u_1. (c) also shows the diffusion of theme information from u_1. (d) shows the diffusion of theme information from u_2.

g_{uj}. The parameter $\mathrm{Diffu}_{v,t_{uj}}$ indicates whether v diffuses the theme information that he/she received.

$$\mathrm{Diffu}_{v,t_{uj}} = \begin{cases} 0, & \text{outdegree} \le 0 \\ 1, & \text{others.} \end{cases} \tag{6}$$

When the outdegree of v is greater than 0, it indicates that v has already diffused the theme information; otherwise, v has never diffused the theme information. The number of users receiving theme information is written as NRTI and the number of users diffusing theme information is written as NDTI.

$$\begin{aligned} \mathrm{NRTI} &= \sum_{u\in(V-\{v\})} \sum_{t_{uj}\in T(u)} S_{v,t_{uj}}, \\ \mathrm{NDTI}_v &= \sum_{u\in(V-\{v\})} \sum_{t_{uj}\in T(u)} \mathrm{Diffu}_{v,t_{uj}}. \end{aligned} \tag{7}$$

MW is calculated as

$$\begin{aligned} &\mathrm{MW}(v) \\ &= \frac{\theta \times (\mathrm{NDTI}_v/\mathrm{NRTI}_v) + (1-\theta) \times \sum_{u\in\mathrm{In}(v)} \mathrm{MW}(u) \times w(u)}{\mathrm{num}_S} \\ &\quad + \frac{|\mathrm{NP}(v)|}{\mathrm{num}_S}, \end{aligned} \tag{8}$$

where $w(u) = 1/\mathrm{outdegree}(u)$. $\mathrm{MW}(v)$ is the MW of v. $\theta \in [0, 1]$ is the weight. $\mathrm{NP}(v)$ represents the total number of theme information posts by v. $\mathrm{In}(v)$ is the set of indegree nodes of v. $w(u)$ represents the weight of the user u, which is determined by his/her outdegree. num_S is the total number of g_{uj}. The initial value of $\mathrm{MW}(v)$ is set as 1. We give an example for calculating MW in Figure 3.

Assume that the MW of all users initially are 1, $\theta = 0.6$, and then calculate the MW as follows.

(1) $MW(u_1)$. From Figures 3(b)–3(d), we have $\mathrm{num}_s = 3$. For u_1, he/she posts two-theme information, which forms two theme information graphs in Figures 3(b) and 3(c). Thus, we get the set $T(u_1)$ ($|T(u_1)| = 2$). From Figure 3(d), $\mathrm{NRTI}_{u_1} = 1$, $\mathrm{NDTI} = 0$, because the outdegree of node u_1 is 0, and u_1 forms its one theme information graph. The $\mathrm{MW}(u_1)$ is calculated as follows:

$$\begin{aligned} A(u_1) &= \{u_2, u_5\}, \\ B(u_1) &= \varnothing; \\ w(u_2) &= \frac{1}{2}, \\ w(u_5) &= \frac{1}{4}, \\ \mathrm{MW}(u_1) & \\ &= \frac{0.6 \times (0/1) + 0.4 \times (1 \times (1/2) + 1 \times (1/4))}{3} + \frac{2}{3} \\ &= \frac{23}{30}. \end{aligned} \tag{9}$$

(2) $MW(u_2)$. Similar to the calculation of $\mathrm{MW}(u_2)$, we have the set $T(u_2)$, $|T(u_2)| = 1$. From Figures 3(b) and 3(c), we have $\mathrm{NDTI}_{u_2} = 1$, $\mathrm{NRTI}_{u_2} = 2$. $\mathrm{MW}(u_2)$ is calculated as follows:

$$\begin{aligned} A(u_2) &= \{u_1, u_4\}, \\ B(u_2) &= \varnothing; \\ w(u_2) &= 1, \\ w(u_4) &= \frac{1}{3}, \\ \mathrm{MW}(u_2) &= \frac{0.6 \times (1/2) + 0.4 \times (1 \times 1 + 1 \times (1/3))}{3} \\ &\quad + \frac{1}{3} = \frac{1}{18}. \end{aligned} \tag{10}$$

Similarly, for u_3, u_4, and u_5, we have

$$\begin{aligned}
\mathrm{NDTI}_{u_3} &= 0+0+0=0,\\
\mathrm{NRTI}_{u_3} &= 0+0+1=1,\\
\mathrm{MW}\left(u_3\right) &= \frac{0.6\times 0+0.4\times 0}{3}+0=0,\\
\mathrm{NDTI}_{u_4} &= 1+1+1=3,\\
\mathrm{NRTI}_{u_4} &= 1+1+1=3,\\
A\left(u_4\right) &= \left\{u_2\right\},\\
B\left(u_4\right) &= \varnothing,\\
w\left(u_2\right) &= 1,\\
A\left(u_4\right) &= \left\{u_1,u_2\right\},\\
B\left(u_4\right) &= \left\{v_2\right\},\\
w\left(u_1\right) &= \frac{1}{3},\\
w\left(u_2\right) &= \frac{1}{2},\\
w\left(v_2\right) &= 1,\\
A\left(u_4\right) &= \left\{u_2\right\},\\
B\left(u_4\right) &= \varnothing,\\
w\left(u_2\right) &= \frac{1}{3},\\
\mathrm{MW}\left(u_4\right) &= \frac{0.6\times(3/3)+0.4\times(1\times 1+1\times(1/3)+1\times(1/2)+1\times 1+1\times(1/3))}{3}+0=\frac{28}{65},\\
\mathrm{NDTI}_{u_5} &= 0+1+1=2,\\
\mathrm{NRTI}_{u_5} &= 1+1+1=3,\\
A\left(u_5\right) &= \left\{u_1\right\},\\
B\left(u_5\right) &= \varnothing,\\
w\left(u_2\right) &= \frac{1}{3},\\
w\left(u_4\right) &= \frac{1}{3},\\
A\left(u_5\right) &= \left\{u_1\right\},\\
B\left(u_5\right) &= \varnothing,\\
w\left(u_1\right) &= \frac{1}{3},\\
A\left(u_5\right) &= \left\{u_2\right\},\\
B\left(u_5\right) &= \left\{v_2\right\},
\end{aligned}$$

$$w(u_2) = \frac{1}{3},$$
$$w(v_2) = 1,$$
$$\mathrm{MW}(u_5) = \frac{0.6 \times (2/3) + 0.4 \times (1 \times (1/3) + 1 \times (1/3) + 1 \times (1/3) + 1 \times (1/3) + 1 \times 1)}{3} + 0 = \frac{4}{9}. \tag{11}$$

The User Influence. There are mutual impact and mutual trust between users. Social trust plays an important role in calculating the user influence. She/he is impacted by others including users inside and outside the community.

(1) Calculating Direct Trust and Direct Influence. If v is an entry node of u, then v will have direct trust on u.

$$\mathrm{DT}_{vu} = \frac{\mathrm{RU}(u)}{\mathrm{outdegree}(v)},$$
$$\mathrm{RU}(u) = \frac{\sum_{w\in \mathrm{In}(u)} \mathrm{RU}(w)}{\mathrm{indegree}(u)}, \tag{12}$$

where DT_{vu} is the direct trust of v on u. $\mathrm{RU}(u)$ is the reputation of user u. $\mathrm{In}(u)$ is the set of entry nodes of u, and $\mathrm{RU}(u \leftarrow w)$ is the reputation of the entry neighbor w of u. The value of $\mathrm{RU}(u)$ depends on the average reputation of all u's entry neighbors. For each node, we give the initial direct trust value 0.1. In Figure 3(a), we calculate the direct trust on u_1 from other nodes as follows:

$$\mathrm{RU}(u_1) = \frac{0.1 + 0.1 + 0.1 + 0.1}{4 + 1} = 0.08,$$
$$\mathrm{In}(u_1) = \{u_2, u_4, u_5, v_1\},$$
$$\mathrm{DT}_{u_2,u_1} = \frac{0.08}{2} = 0.04,$$
$$\mathrm{DT}_{u_3,u_1} = \frac{0.08}{0} \ (\text{written as } 0),$$
$$\mathrm{DT}_{u_4,u_1} = \frac{0.08}{2} = 0.04, \tag{13}$$
$$\mathrm{DT}_{u_5,u_1} = \frac{0.08}{4} = 0.02,$$
$$\mathrm{DT}_{v_1,u_1} = \frac{0.08}{2} = 0.04,$$
$$\mathrm{DT}_{v_2,u_1} = \frac{0.08}{1} = 0.08,$$
$$\mathrm{DT}_{v_3,u_1} = \frac{0.08}{0} \ (\text{written as } 0).$$

u has a direct influence on v as follows:

$$\mathrm{DI}_{uv} = \frac{I(u \leftarrow v)}{\mathrm{outdegree}(v)},$$
$$W(u \longleftarrow v) = \frac{|\mathrm{theme}(v,u)|}{\mathrm{NRTI}_v}, \tag{14}$$

where DI_{uv} is the direct influence of u on v. $I(u \leftarrow v)$ is the degree of interest of v to u. $|\mathrm{theme}(v,u)|$ is the amount of the theme information from u in the receiving theme information of v.

In Figure 3, we calculate the direct influence on u_1 produced by other users as follows:

$$I(u_1 \longleftarrow u_2) = \frac{2}{2} = 1,$$
$$I(u_1 \longleftarrow u_3) = \frac{0}{1} = 0,$$
$$I(u_1 \longleftarrow u_4) = \frac{2}{3} = 0.667,$$
$$I(u_1 \longleftarrow u_5) = \frac{2}{3} = 0.667, \tag{15}$$
$$I(u_1 \longleftarrow v_1) = \frac{2}{3} = 0.667,$$
$$I(u_1 \longleftarrow v_2) = \frac{2}{3} = 0.667,$$
$$I(u_1 \longleftarrow u_3) = \frac{2}{0} \ (\text{written as } 0).$$

In Figure 3(a), we have

$$\mathrm{DI}_{u_1u_2} = \frac{1}{2} = 0.5,$$
$$\mathrm{DI}_{u_1u_3} = \frac{0}{0} \text{ is } 0,$$
$$\mathrm{DI}_{u_1u_4} = \frac{0.667}{2} = 0.334,$$
$$\mathrm{DI}_{u_1u_5} = \frac{0.667}{5} = 0.133, \tag{16}$$
$$\mathrm{DI}_{u_1v_1} = \frac{0.667}{2} = 0.334,$$
$$\mathrm{DI}_{u_1v_2} = \frac{1}{1} = 1,$$
$$\mathrm{DI}_{u_1v_3} = \frac{0}{0} \ (\text{written as } 0).$$

(2) Indirect Trust and Indirect Influence. If u is the reachable node of v, then v will have indirect trust on u as follows:

$$\mathrm{IT}_{vu} = \frac{\mathrm{RU}(u)}{\min_{vu}}. \tag{17}$$

IT_{vu} is v's indirect trust on u. $\min_{vu}$ is the length of the shortest path from v to u.

In Figure 3(a), we calculate the indirect trust on u_1 gained from other nodes as follows:

$$\begin{aligned}
\mathrm{IT}_{u_2u_1} &= \frac{0.08}{1} = 0.08,\\
\mathrm{IT}_{u_3u_1} &= \frac{0.08}{0}\ \text{(written as 0)},\\
\mathrm{IT}_{u_4u_1} &= \frac{0.08}{1} = 0.08,\\
\mathrm{IT}_{u_5u_1} &= \frac{0.08}{1} = 0.08,\\
\mathrm{IT}_{v_1u_1} &= \frac{0.08}{1} = 0.08,\\
\mathrm{IT}_{v_2u_1} &= \frac{0.08}{2} = 0.04,\\
\mathrm{IT}_{v_3u_1} &= \frac{0.08}{0}\ \text{(written as 0)}.
\end{aligned} \tag{18}$$

u has an indirect influence on v as follows:

$$\begin{aligned}
\mathrm{II}_{uv} &= \frac{I(u \leftarrow v)}{\min_{vu}},\\
I(u \longleftarrow v) &= \frac{|\mathrm{theme}(v,u)|}{\mathrm{NRTI}_v}.
\end{aligned} \tag{19}$$

In Figure 3(a), we calculate the indirect influence of other nodes on u_1 as follows. The calculation of I is the same as the above formula.

$$\begin{aligned}
\mathrm{II}_{u_1u_2} &= \frac{1}{1} = 1,\\
\mathrm{II}_{u_1u_3} &= \frac{0}{0}\ \text{(written as 0)},\\
\mathrm{II}_{u_1u_4} &= \frac{0.667}{1} = 0.667,\\
\mathrm{II}_{u_1u_5} &= \frac{0.667}{1} = 0.667,\\
\mathrm{II}_{u_1v_1} &= \frac{0.667}{1} = 0.667,\\
\mathrm{II}_{u_1v_2} &= \frac{1}{2} = 0.5,\\
\mathrm{II}_{u_1v_3} &= \frac{0}{0}\ \text{(written as 0)}.
\end{aligned} \tag{20}$$

(3) User Combined Influence. Assuming that v can reach u through a path, we introduce the factor λ ($\lambda \in [0,1]$).

If v is the entry node of u, the combined influence of u on v is

$$\mathrm{UCI}_{uv} = \lambda \mathrm{DI}_{uv} + (1-\lambda)\,\mathrm{DT}_{vu}. \tag{21}$$

If v is not an entry node of node u, but u is a reachable node of v, the combined influence is

$$\mathrm{UCI}_{uv} = \lambda \mathrm{II}_{uv} + (1-\lambda)\,\mathrm{IT}_{vu}. \tag{22}$$

Assume $\lambda = 0.3$. In Figure 3, we calculate the combined influence of other nodes on u_1 as follows.

u_2 is the entry node of u_1; then we have $\mathrm{UCI}_{u_1u_2} = 0.3 \times 0.5 + 0.7 \times 0.04 = 0.178$.

u_4 is the entry node u_1; then we have $\mathrm{UCI}_{u_1u_4} = 0.3 \times 0.334 + 0.7 \times 0.04 = 0.1282$.

u_5 is the entry node of u_1; then we have $\mathrm{UCI}_{u_1u_5} = 0.3 \times 0.133 + 0.7 \times 0.02 = 0.0539$.

v_1 is the entry node of u_1; then we have $\mathrm{UCI}_{u_1v_1} = 0.3 \times 0.334 + 0.7 \times 0.04 = 0.1282$.

v_2 is the reachable node of u_1; then we have $\mathrm{UCI}_{u_1v_2} = 0.3 \times 0.5 + 0.7 \times 0.04 = 0.178$.

(4) User Influence. User influence is got by combining all users' influence:

$$\mathrm{UI}(u) = \frac{\sum_{v\in \mathrm{SUCP}(u)} \mathrm{UCI}_{uv}}{|\mathrm{SUCP}(u)|}, \tag{23}$$

where SUCP represents a set of users that can reach u through a certain path. For example, in Figure 3, the user influence of u_1 is calculated as follows:

$$\begin{aligned}
&\mathrm{UI}(u_1)\\
&= \frac{\mathrm{UCI}_{u_1u_2} + \mathrm{UCI}_{u_1u_4} + \mathrm{UCI}_{u_1u_5} + \mathrm{UCI}_{u_1v_1} + \mathrm{UCI}_{u_1v_2}}{5}\\
&= 0.133.
\end{aligned} \tag{24}$$

When we get $\mathrm{MW}(u_1)$ and $\mathrm{UI}(u_1)$, the user final influence can be calculated according to (4).

3.2.3. Community Influence. The community influence is composed of the users' interaction inside and outside the community. In this paper, we consider it from three factors, that is, the user-integrated influence, the community size, and the degree of relationship tightness among users inside the community.

User-integrated influence (UII) is integrated from the final influence of all users within the community.

$$\mathrm{UII}(C_i) = \sum_{u\in CV(u)} \mathrm{UFI}(u), \tag{25}$$

where $\mathrm{UII}(C_i)$ is UII of the community C_i. $CV(u)$ is the set of users inside community C_i.

The community size (CS) is important to the calculation of the community-level influence. The larger the number of users in a community is, the greater the influence of the community becomes. The formula is as follows:

$$\mathrm{CS}(C_i) = \frac{|CV(C_i)|}{\max(V)}, \tag{26}$$

where $|CV(C_i)|$ represents the number of users in a community and $\max(V)$ represents the total number of users in the social network.

Input: $G = \{V, E\}$; C; $T(u)$; $g(u)$; UII = 0; τ; ρ; RT = 0
Output: community influence
(1) **for** $i = 0$ to $|V|$ **do**
(2) MW(i)
(3) UI(i)
(4) **end for**
(5) **for** $j = 0$ to $|CV|$ **do**
(6) $\mathrm{UII}(j) = \mathrm{MW}(j) \times \mathrm{UI}(j) + \mathrm{UII}(j)$
(7) **end for**
(8) CS(C)
(9) **for** $i = 0$ to $|CV|$ **do**
(10) $\mathrm{RT}(C_i) = \dfrac{\sum_{u \in CV(C_i)} (\mathrm{outdegree}(u) + \mathrm{indegree}(u))}{CV(C_i)}$
(11) **end for**
(12) $\mathrm{CI}(C_i) = \tau \times \mathrm{UII}(C_i) + \rho \times \mathrm{CS} + (1 - \tau - \rho) \times \mathrm{RT}(C_i)$
(13) **return** $\mathrm{CI}(C_i)$

Algorithm 2: Community-level influence analysis algorithm (CIAA).

The degree of relationship tightness (RT) represents the degree of closeness between users inside a community. We describe it from the user's outdegree and indegree as follows:

$$\mathrm{RT}(C_i) = \frac{\sum_{u \in CV(C_i)} (\mathrm{outdegree}(u) + \mathrm{indegree}(u))}{CV(C_i)}. \quad (27)$$

Therefore, we calculate the CI as follows:

$$\mathrm{CI}(C_i) = \tau \times \mathrm{UII}(C_i) + \rho \times \mathrm{CS} + (1 - \tau - \rho) \times \mathrm{RT}(C_i), \quad (28)$$

where τ and ρ ($\tau, \rho \in [0, 1]$) are used to distinguish the importance of different factors.

3.3. The Proposed Algorithm. According to the above description, we propose a community-level influence analysis algorithm, called CIAA, in a pseudo-code format in Algorithm 2. It can be seen from the algorithm that the total time complexity is $O(n)$. This means that our algorithm can be applied on large-scale social dataset.

4. Experiments

We conduct experiments to validate the effectiveness of the proposed approach on a real-world microblogging network. In this section, we describe the experimental setup followed by the discussion of experiment results.

4.1. Dataset. The real-world dataset in this paper is crawled from Sina Weibo by Weibo crawler. Similar to a hybrid of Twitter and Facebook, Sina Weibo is one of the most popular sites in China. It has more than 33% of the Internet users in China, and its market penetration is equivalent to that of Twitter in the United States. As released by the Sina Weibo, as of June 2016, the active users from different social and cultural backgrounds have reached 282 million monthly and 86.8 million daily. Moreover, there are nearly 100 million new microblogs every day. They promote and disseminate views and attitudes on business, culture, education, and so forth. The crawled data includes 20,151,129 microblogs, 932,578,467 comments, and 9,218 users. In this paper, we collected more than 1000 users from the crawled dataset and divided the related information into Tables 1, 2, 3, and 4 for data sources according to our model framework. They are stored in txt-formatted files.

Table 1: Data structure and description of the user information.

Features	Description
UserID	User' ID
IsVIP	Authenticated user
FansNum	Number of fans
AttenNum	Number of attention users
ThemeAmo	Amount of theme information
Tag	User' label
Time	Login time

Table 2: Data structure and description of the user theme information (microblogs).

Features	Description
ThemeID	Theme information ID
ThemeFromID	Source ID of theme information
ProNum	Number of processes
ThemeClass	Theme information class
PTime	Post time of theme information

Table 3: Data structure and description of the user fans.

Features	Description
UserID	User' ID
FansID	Fans' ID

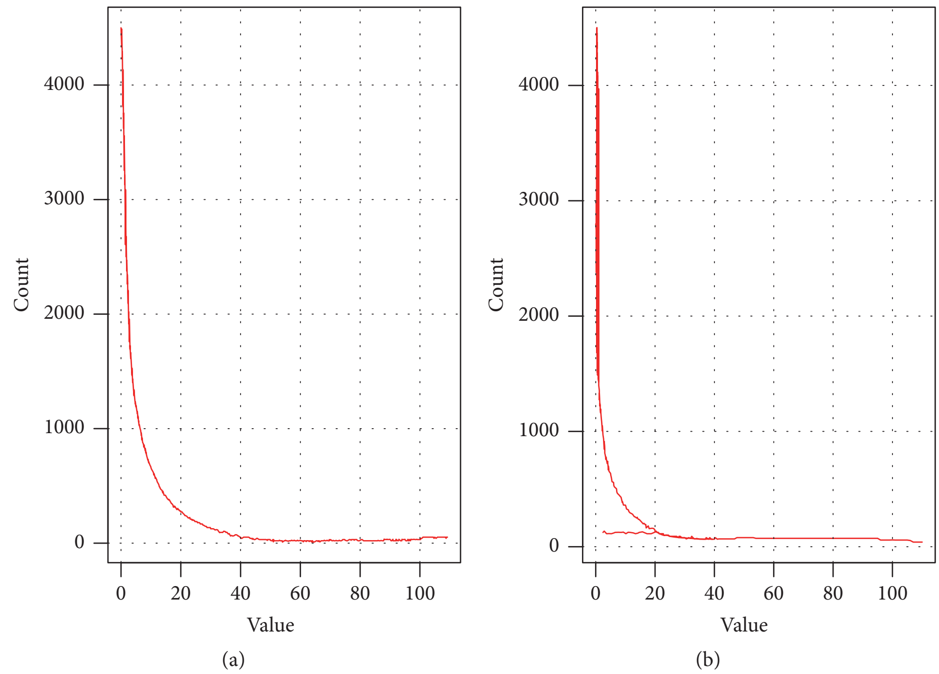

Figure 4: (a) is the outdegree distribution and (b) is the degree distribution.

Table 4: Data structure and description of the user attention.

Features	Description
UserID	User' ID
AttenID	User-attended ID

Table 5: Parameters for experiments.

Symbol	Description	Value
V	The total number of nodes	1127
CV	The total number of nodes in the community	20
λ	Parameter	0.3
θ	Parameter	0.5
τ	Parameter	0.5
ρ	Parameter	0.3

4.2. Experimental Setting. All experiments are conducted on a PC with Intel Core i5 processor, 8 GB RAM. According to prior knowledge, we set the parameters of the experiments as Table 5.

4.3. Results

4.3.1. Community Structure Analysis. In order to mine and study the characteristic of community, we plot the outdegree distribution and degree distribution of users in community. In a directed social network, the indegree of nodes is the number of fans of the user. The outdegree of nodes is the amount of the user's attention. Figure 4 shows the outdegree and degree distribution of data sources.

As shown in Figure 4, the outdegree distribution and the degree distribution of Sina Weibo dataset follow the power-law distribution, which indicates that the social network composed of the dataset is a scale-free network.

4.3.2. Eliminating Zombie Fans. In order to improve the accuracy of our model, we remove zombie fans. According to the eliminating zombie fans method in Algorithm 1, we finally remove 12 zombie fans, as shown in Table 6.

As shown in Table 6, the three sets are NUI, NAU, and NUF. The little black boxes in Table 6 represent the shared users of three sets, and they are the same as the shared users from time dimension and behavior dimension. Therefore, the shared users will be removed. We compare the user final influence without the zombie fans with the user final influence with the zombie fans, as shown in Table 7.

From Table 7, the result of the comparison shows that the accuracy of the UFI with zombie fans for the actual user ranking is only 60%. It is concluded that the elimination of zombie fans is very important for the accuracy of the user final influence.

4.3.3. Accuracy Analysis of the User Final Influence. We calculate the user final influence of users in community, but we compare the top ten users for simplicity. The top 10 user final influences and their related information are shown in Table 8.

According to the UFI ranking in Table 8, we find that these users are authenticated user. It is concluded that the authenticated users are more influential in microblogging networks. There are two reasons for this phenomenon. First, the majority of well-known users are authenticated users, and the influence of well-known users is larger than the user

TABLE 6: Three user sets for eliminating zombie fans. The boxes represent zombie fans.

	NUI			NAU			NUF	
511 * * * *843	320 * * * *657	226 * * * *535	257 * * * *813	122 * * * *644	384 * * * *495	348 * * * *495	214 * * * *635	522 * * * *846
267 * * * *275	377 * * * *140	506 * * * *228	508 * * * *382	203 * * * *473	569 * * * *865	569 * * * *865	514 * * * *515	565 * * * *964
209 * * * *054	314 * * * *751	551 * * * *783	$\boxed{535 * * * *588}$	540 * * * *732	512 * * * *879	512 * * * *879	314 * * * *302	553 * * * *291
$\boxed{535 * * * *588}$	260 * * * *165	$\boxed{564 * * * *561}$	540 * * * *495	272 * * * *407	345 * * * *320	345 * * * *820	560 * * * *696	550 * * * *598
569 * * * *524	299 * * * *713	326 * * * *401	541 * * * *396	532 * * * *553	553 * * * *237	553 * * * *237	362 * * * *483	557 * * * *097
519 * * * *908	255 * * * *954	546 * * * *117	236 * * * *681	$\boxed{508 * * * *496}$	241 * * * *385	241 * * * *885	169 * * * *032	528 * * * *140
174 * * * *367	295 * * * *285	366 * * * *383	124 * * * *474	567 * * * *764	538 * * * *374	538 * * * *874	568 * * * *540	551 * * * *812
176 * * * *904	312 * * * *963	140 * * * *523	541 * * * *048	514 * * * *452	237 * * * *312	237 * * * *812	293 * * * *367	295 * * * *820
381 * * * *512	312 * * * *885	357 * * * *742	$\boxed{365 * * * *215}$	561 * * * *240	267 * * * *275	267 * * * *275	512 * * * *708	549 * * * *817
522 * * * *989	275 * * * *525	547 * * * *573	557 * * * *157	219 * * * *090	516 * * * *282	516 * * * *382	531 * * * *888	108 * * * *870
180 * * * *713	272 * * * *524	558 * * * *440	562 * * * *840	554 * * * *983	216 * * * *527	$\boxed{535 * * * *588}$	540 * * * *397	563 * * * *989
$\boxed{508 * * * *496}$	393 * * * *610	520 * * * *974	295 * * * *781	519 * * * *173	395 * * * *398	395 * * * *898	$\boxed{508 * * * *496}$	560 * * * *564
267 * * * *724	325 * * * *361	564 * * * *326	$\boxed{217 * * * *423}$	395 * * * *459	531 * * * *874	531 * * * *874	514 * * * *924	320 * * * *232
194 * * * *451	299 * * * *433	291 * * * *885	155 * * * *451	240 * * * *653	531 * * * *985	531 * * * *885	503 * * * *355	553 * * * *123
519 * * * *020	$\boxed{398 * * * *168}$	564 * * * *548	535 * * * *748	$\boxed{398 * * * *168}$	518 * * * *654	518 * * * *554	$\boxed{217 * * * *423}$	$\boxed{365 * * * *215}$
213 * * * *014	526 * * * *623	564 * * * *703	563 * * * *796	569 * * * *999	540 * * * *388	540 * * * *888	368 * * * *450	565 * * * *147
299 * * * *593	$\boxed{297 * * * *117}$	551 * * * *728	523 * * * *767	308 * * * *265	393 * * * *530	393 * * * *530	241 * * * *965	561 * * * *032
$\boxed{365 * * * *215}$	506 * * * *354	269 * * * *324	516 * * * *694	553 * * * *815	$\boxed{107 * * * *161}$	260 * * * *887	301 * * * *065	524 * * * *860
263 * * * *023	327 * * * *315	377 * * * *804	562 * * * *886	315 * * * *642	553 * * * *284	553 * * * *284	546 * * * *749	315 * * * *640
505 * * * *471	184 * * * *620	349 * * * *961	286 * * * *383	199 * * * *843	282 * * * *601	282 * * * *501	$\boxed{398 * * * *168}$	530 * * * *776
281 * * * *650	293 * * * *863	$\boxed{387 * * * *165}$	537 * * * *642	$\boxed{564 * * * *561}$	$\boxed{387 * * * *165}$	506 * * * *834	175 * * * *475	558 * * * *546
249 * * * *881	530 * * * *172	202 * * * *075	266 * * * *792	531 * * * *022	558 * * * *740	558 * * * *740	559 * * * *740	557 * * * *957
$\boxed{217 * * * *423}$	206 * * * *147	561 * * * *896	564 * * * *344	563 * * * *288	381 * * * *565	381 * * * *565	559 * * * *435	$\boxed{565 * * * *036}$
393 * * * *557	227 * * * *201	562 * * * *656	554 * * * *847	190 * * * *733	377 * * * *522	377 * * * *522	521 * * * *073	564 * * * *950
367 * * * *587	324 * * * *272	282 * * * *244	181 * * * *912	190 * * * *415	532 * * * *773	532 * * * *773	$\boxed{564 * * * *561}$	535 * * * *470
354 * * * *437	246 * * * *555	524 * * * *753	550 * * * *247	163 * * * *152	326 * * * *463	326 * * * *463	561 * * * *058	558 * * * *005
202 * * * *713	$\boxed{107 * * * *161}$	524 * * * *189	558 * * * *343	567 * * * *057	183 * * * *325	183 * * * *825	533 * * * *829	527 * * * *830
140 * * * *971	259 * * * *422	546 * * * *882	$\boxed{562 * * * *957}$	548 * * * *952	$\boxed{297 * * * *117}$	$\boxed{107 * * * *161}$	536 * * * *141	537 * * * *866
206 * * * *863	395 * * * *128	554 * * * *705	558 * * * *610	562 * * * *816	215 * * * *573	215 * * * *673	384 * * * *830	528 * * * *914
240 * * * *727	371 * * * *200	508 * * * *954	219 * * * *403	186 * * * *260	373 * * * *905	373 * * * *905	207 * * * *025	$\boxed{297 * * * *117}$
292 * * * *683	177 * * * *177	$\boxed{565 * * * *036}$	356 * * * *633	532 * * * *335	331 * * * *172	331 * * * *172	361 * * * *345	535 * * * *483
289 * * * *077	321 * * * *383	548 * * * *304	557 * * * *693	329 * * * *831	372 * * * *172	372 * * * *172	561 * * * *310	539 * * * *709
378 * * * *432	299 * * * *217	376 * * * *382	363 * * * *234	$\boxed{558 * * * *008}$	385 * * * *668	385 * * * *668	$\boxed{562 * * * *957}$	562 * * * *106
280 * * * *733	315 * * * *540	557 * * * *957	559 * * * *028	327 * * * *271	564 * * * *754	$\boxed{558 * * * *008}$	528 * * * *672	558 * * * *843
386 * * * *371	$\boxed{562 * * * *957}$	293 * * * *987	551 * * * *896	554 * * * *403	375 * * * *410	375 * * * *410	$\boxed{387 * * * *165}$	316 * * * *442
219 * * * *655	346 * * * *220	$\boxed{558 * * * *008}$	185 * * * *423	362 * * * *913	$\boxed{565 * * * *036}$	569 * * * *628	173 * * * *242	560 * * * *121
166 * * * *754	562 * * * *363	352 * * * *153	122 * * * *644	292 * * * *807	387 * * * *841	387 * * * *841	393 * * * *428	288 * * * *500
248 * * * *174	521 * * * *857	561 * * * *406	531 * * * *740	558 * * * *488	538 * * * *261	538 * * * *261	565 * * * *531	549 * * * *206
246 * * * *354	257 * * * *813		531 * * * *866	557 * * * *762		531 * * * *866	558 * * * *569	

Table 7: Comparison of the user final influence.

User ID	UFI without zombie fans	UFI with zombie fans	The actual rankings
263 ∗ ∗ ∗ ∗023	1	3	1
511 ∗ ∗ ∗ ∗843	2	2	2
519 ∗ ∗ ∗ ∗020	3	1	3
508 ∗ ∗ ∗ ∗496	4	4	4
550 ∗ ∗ ∗ ∗598	5	5	5
267 ∗ ∗ ∗ ∗724	6	6	6
365 ∗ ∗ ∗ ∗215	8	8	7
299 ∗ ∗ ∗ ∗593	7	7	8
522 ∗ ∗ ∗ ∗989	9	9	9
194 ∗ ∗ ∗ ∗451	10	10	10

Table 8: Top 10 user information of the UFI.

UFI ranking	User ID	Number of fans	Number of blogs	Authenticated or not
1	263∗∗∗∗023	128	1515	1
2	511∗∗∗∗843	282	1282	1
3	519∗∗∗∗020	66	101	1
4	508∗∗∗∗496	261	5471	1
5	550∗∗∗∗598	14	22	1
6	267∗∗∗∗724	823	1452	1
7	299∗∗∗∗593	158	109	1
8	365∗∗∗∗215	177	945	1
9	522∗∗∗∗989	13	29	1
10	194∗∗∗∗451	69	11	1

average influence. Second, the authenticated user's identity is transparent, which makes the user have higher social trust. Table 8 also shows that the user final influence needs to be considered from the quality of the user fans, the number of user microblogs, and user authentication.

Table 9 and Figure 5 show the comparison between the UFI method and the microblog-fans ranking algorithm. Table 9 shows the UFI method ranking and the corresponding ranking via microblog-fans ranking algorithm. Figure 5 shows the overall ranking order via the microblog-fans ranking algorithm.

It can be seen from Table 9 and Figure 5 that the UFI ranking is almost completely different from the microblog-fans ranking. Overall, according to the UFI method, the number of microblogs and fans of the top users must reach a certain quantity to support individual influence. Thus, the number of microblogs and fans is a factor of measuring influence in UFI method. However, social trust between users can help improve individual influence in the UFI method.

The user final influence is an experimental evaluation of the user, and there is no existing dataset with its comparison. We can only refer to the ranking of the user influence from some affiliations. Based on the ranking of user influence provided by Sina Weibo official, we verify the calculation

Table 9: Comparison of UFI method with microblog-fans ranking algorithm.

UFI ranking	User ID	Number of fans	Number of blogs	Microblog-fans ranking
1	263∗∗∗∗023	128	1515	3
2	511∗∗∗∗843	282	1282	4
3	519∗∗∗∗020	66	101	8
4	508∗∗∗∗496	261	5471	1
5	550∗∗∗∗598	14	22	6
6	267∗∗∗∗724	823	1452	2
7	299∗∗∗∗593	158	109	7
8	365∗∗∗∗215	177	945	5
9	522∗∗∗∗989	13	29	10
10	194∗∗∗∗451	69	11	9

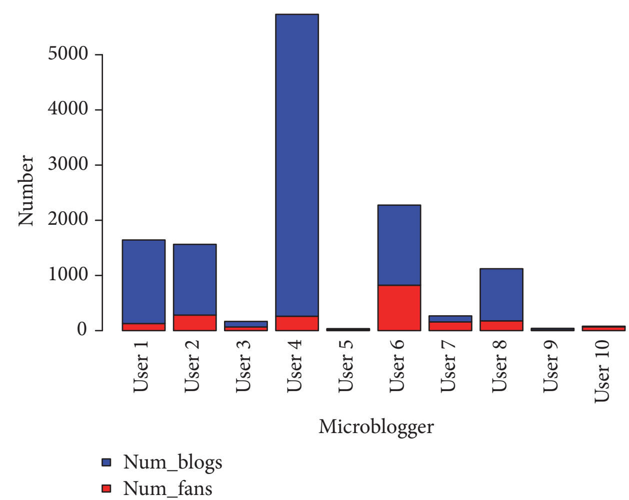

Figure 5: The overall ranking via the microblog-fans ranking algorithm.

method proposed in this paper. We compare the results of the proposed method with the official ranking to verify the correctness of the user final influence. Because each microblogging platform has its own influence calculation method, we cannot numerically compare the results, but we compare the results from the relative position, that is, ranking. If the influence rankings of the two methods are in the similar order, we consider the results of the influence analysis to be similar. The comparison of the users ranking by Sina Weibo officially and UFI method is shown in Table 10.

In Table 10, the user final influence calculation method and the user actual ranking are mainly the same but having the user pair of 299 ∗ ∗ ∗ ∗593 and 365 ∗ ∗ ∗ ∗215. That is because user influence ranking by Sina Weibo emphasizes the number of microblogs and fans, and the number of microblogs and fans of user 299∗∗∗∗593 and user 365∗∗∗∗215 is largely different. However, the UFI method considers the factors of influence more reasonably.

Table 10: Comparison of user actual ranking with UFI ranking.

User ID	The actual ranking	UFI value	UFI ranking
263 ∗ ∗ ∗ ∗023	1	1.0000	1
511 ∗ ∗ ∗ ∗843	2	0.0384	2
519 ∗ ∗ ∗ ∗020	3	0.0215	3
508 ∗ ∗ ∗ ∗496	4	0.0107	4
550 ∗ ∗ ∗ ∗598	5	0.0099	5
267 ∗ ∗ ∗ ∗724	6	0.00726	6
299 ∗ ∗ ∗ ∗593	8	0.0028	7
365 ∗ ∗ ∗ ∗215	7	0.0021	8
522 ∗ ∗ ∗ ∗989	9	0.0019	9
194 ∗ ∗ ∗ ∗451	10	0.0016	10

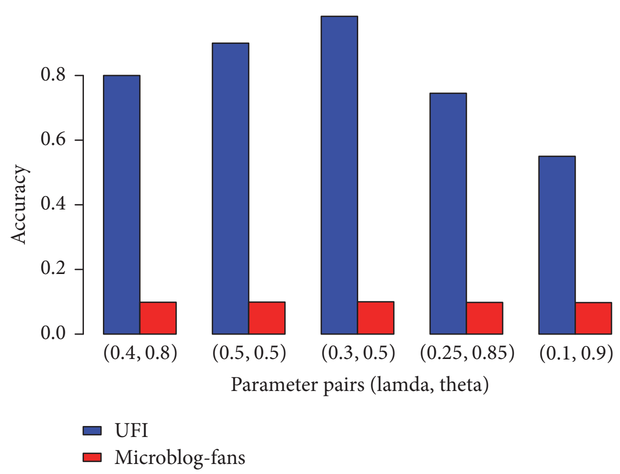

Figure 6: Comparison of accuracy of two methods with different λ and θ.

Considering the results of Sina Weibo official as the standard, the accuracy of UFI method will change with different λ and θ, as shown in Figure 6.

From Figure 6, it can be seen that the UFI method accuracy changes with the different λ and θ. When $\lambda = 0.3$, $\theta = 0.5$, UFI method has the highest accuracy. Therefore, the parameter pair $(0.3, 0.5)$ is used for other experiments. We also find that the UFI method is more accurate than the microblog-fans ranking algorithm. Moreover, this experiment indicates the importance of the user willingness to diffusing theme information in the accuracy of the user influence.

4.3.4. Accuracy Analysis of CIAA. Because the existing studies of community influence are few, we compare the proposed algorithm CIAA with the averaging user influence algorithm (AI). We set different parameters pair τ and ρ for comparing the two algorithms. Then, we can calculate the corresponding community influence, as shown in Figure 7.

Figure 7 shows that the results of the CIAA are changing with the different parameter values. When $\tau = 0.5$ and $\rho = 0.2$, the results of the two algorithms are closest. That is because the AI algorithm is mainly the weighted average of the user influence, and the CIAA is the integration of the user-integrated influence, the community size, and the degree of relationship tightness among users inside the community. The greater the proportion of the user final influence, the closer the results of the two algorithms. Therefore, the proposed algorithm outperforms the state-of-the-art baseline algorithm.

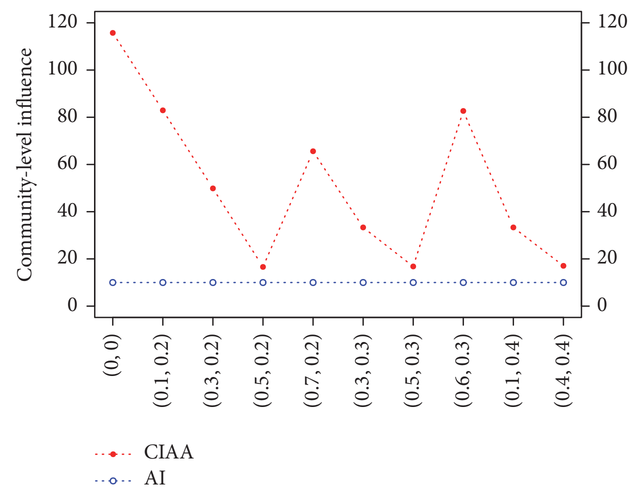

Figure 7: The community-level influence by two measuring algorithms with different (τ, ρ) pairs.

5. Conclusion

In this paper, we studied the emerging problem on how to model community-level influence. Online social networks, especially microblogging networks, are more and more important in our daily life. Previous works can effectively cope with the individual influence in microblogging network, but they rarely evaluate the social influence in community level, which outweighs the individual influence. We defined the related concepts for the community-level influence and constructed a model that combined the user influence, social trust, and relationship tightness of intrausers in a community to reveal the community-level influence appropriately. We proposed the algorithm CIAA to cope with the real-world applications. We conducted empirical studies on a real-world microblogging crawled from Sina Weibo, where the CIAA outperformed the state-of-the-art baseline algorithm. To the best of our knowledge, the proposed approach has a significant effect on community influence in microblogging network. The highlights of this paper can be summarized as follows: (1) formulating the problem of analyzing community-level influence and designing a community-level influence analysis model; (2) proposing community-level influence analysis algorithm called CIAA, to cope with real-world microblogging applications; and (3) extensively demonstrating the superiority of the proposed method. In the future work, we plan to extend the proposed method to assess the community influence in dynamic online social network.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants U1433116 and 61702355, in part by the Fundamental Research Funds for the Central Universities under Grant NP2017208, and in part by the Funding of Jiangsu Innovation Program for Graduate Education under Grants KYLX15_0324 and KYLX15_0321.

References

[1] L. Yao, Q. Z. Sheng, A. H. H. Ngu, J. Yu, and A. Segev, "Unified collaborative and content-based web service recommendation," *IEEE Transactions on Services Computing*, vol. 8, no. 3, pp. 453–466, 2015.

[2] J. Wu, S. Pan, X. Zhu, C. Zhang, and X. Wu, "Multi-instance Learning with Discriminative Bag Mapping," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–16, 2018.

[3] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2016.

[4] T. Cruz, L. Rosa, J. Proenca et al., "A Cybersecurity Detection Framework for Supervisory Control and Data Acquisition Systems," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2236–2246, 2016.

[5] D. Kim, D. Hyeon, J. Oh, W.-S. Han, and H. Yu, "Influence maximization based on reachability sketches in dynamic graphs," *Information Sciences*, vol. 394-395, pp. 217–231, 2017.

[6] G. Wang, W. Jiang, J. Wu, and Z. Xiong, "Fine-grained feature-based social influence evaluation in online social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 9, pp. 2286–2296, 2014.

[7] J. M. Hofman, A. Sharma, and D. J. Watts, "Prediction and explanation in social systems," *Science*, vol. 355, no. 6324, pp. 486–488, 2017.

[8] S. Myers and J. Leskovec, "The bursty dynamics of the twitter information network," in *Proceedings of the 23rd International Conference on World Wide Web, WWW 2014*, pp. 913–923, Republic of Korea, April 2014.

[9] Y. Liu, Q. Li, X. Tang, N. Ma, and R. Tian, "Superedge prediction: What opinions will be mined based on an opinion supernetwork model?" *Decision Support Systems*, vol. 64, pp. 118–129, 2014.

[10] R. De Caux, C. Smith, D. Kniveton, R. Black, and A. Philippides, "Dynamic, small-world social network generation through local agent interactions," *Complexity*, vol. 19, no. 6, pp. 44–53, 2014.

[11] J. Wu, S. Pan, X. Zhu, C. Zhang, and P. S. Yu, "Multiple structure-view learning for graph classification," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–16, 2017.

[12] L. Zhu, D. Guo, J. Yin, G. V. Steeg, and A. Galstyan, "Scalable temporal latent space inference for link prediction in dynamic social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2765–2777, 2016.

[13] S.-Y. Tan, J. Wu, L. Lü, M.-J. Li, and X. Lu, "Efficient network disintegration under incomplete information: The comic effect of link prediction," *Scientific Reports*, vol. 6, Article ID 22916, 2016.

[14] J. Wu, S. Pan, X. Zhu, C. Zhang, and X. Wu, "Positive and unlabeled multi-graph learning," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 818–829, 2017.

[15] A. Almaatouq, L. Radaelli, A. Pentland, and E. Shmueli, "Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change," *PLoS ONE*, vol. 11, no. 3, Article ID e0151588, 2016.

[16] Q. Fang, J. Sang, C. Xu, and Y. Rui, "Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 796–812, 2014.

[17] J. Wu, S. Pan, X. Zhu, and Z. Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[18] V. Belák, S. Lam, and C. Hayes, "Towards maximising cross-community information diffusion," in *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, pp. 171–178, Turkey, August 2012.

[19] P. F. Lazarsfeld, *Personal Influence: The Part Played by People in the Flow of Mass Communications*, Transaction Publishers, New York, NY, USA, 2006.

[20] C. Dong, Y. Zhao, and Q. Zhang, "Assessing the influence of an individual event in complex fault spreading network based on dynamic uncertain causality graph," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1615–1630, 2016.

[21] N. Ma and Y. Liu, "SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1357–1368, 2014.

[22] X. Tang, J. Wang, J. Zhong, and Y. Pan, "Predicting essential proteins based on weighted degree centrality," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 11, no. 2, pp. 407–418, 2014.

[23] M. K. Tarkowski, P. Szczepa, T. Rahwan, T. P. Michalak, and M. Wooldridge, "Closeness centrality for networks with overlapping community structure," in *Proceedings of the in Thirtieth AAAI Conference on Artificial Intelligence*, pp. 622–629, Phoenix, Ariz, USA, 2016.

[24] N. Kourtellis, G. De Francisci Morales, and F. Bonchi, "Scalable Online Betweenness Centrality in Evolving Graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2494–2506, 2015.

[25] G. Lohmann, D. S. Margulies, A. Horstmann et al., "Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain," *PLoS ONE*, vol. 5, no. 4, Article ID e10232, 2010.

[26] P. Grindrod and D. J. Higham, "A matrix iteration for dynamic network summaries," *SIAM Review*, vol. 55, no. 1, pp. 118–128, 2013.

[27] D. F. Gleich, "PageRank beyond the web," *SIAM Review*, vol. 57, no. 3, pp. 321–363, 2015.

[28] J. Wang, M. Li, H. Wang, and Y. Pan, "Identification of essential proteins based on edge clustering coefficient," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 4, pp. 1070–1080, 2012.

[29] O. Sporns, "Contributions and challenges for network models in cognitive neuroscience," *Nature Neuroscience*, vol. 17, pp. 652–660, 2014.

[30] F. Hao, M. Chen, C. Zhu, and M. Guizani, "Discovering influential users in micro-blog marketing with influence maximization mechanism," in *Proceedings of the 2012 IEEE Global Communications Conference, GLOBECOM 2012*, pp. 470–474, USA, December 2012.

[31] F. Bodendorf and C. Kaiser, "Detecting opinion leaders and trends in online social networks," in *Proceedings of the 2nd ACM Workshop on Social Web Search and Mining, SWSM'09, Co-located with the 18th ACM International Conference on Information and Knowledge Management, CIKM 2009*, pp. 65–68, China, November 2009.

[32] B. Xiang, Q. Liu, E. Chen, H. Xiong, Y. Zheng, and Y. Yang, "PageRank with priors: An influence propagation perspective," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence, IJCAI'13*, pp. 2740–2746, Beijing, China, 2013.

[33] T. Zhu, B. Wang, B. Wu, and C. Zhu, "Maximizing the spread of influence ranking in social networks," *Information Sciences*, vol. 278, pp. 535–544, 2014.

[34] J. Li, W. Peng, T. Li, T. Sun, Q. Li, and J. Xu, "Social network user influence sense-making and dynamics prediction," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5115–5124, 2014.

[35] C. Zhou, P. Zhang, W. Zang, and L. Guo, "On the upper bounds of spread for greedy algorithms in social network influence maximization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2770–2783, 2015.

[36] X. Qi, E. Fuller, R. Luo, and C.-Q. Zhang, "A novel centrality method for weighted networks based on the Kirchhoff polynomial," *Pattern Recognition Letters*, vol. 58, pp. 51–60, 2015.

[37] V. Latora and M. Marchiori, "A measure of centrality based on network efficiency," *New Journal of Physics* , vol. 9, article 188, 2007.

[38] Y. Mehmood, N. Barbieri, F. Bonchi, and A. Ukkonen, "CSI: Community-level social influence analysis," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 8189, no. 2, pp. 48–63, 2013.

[39] C. S. E. Bale, N. J. Mccullen, T. J. Foxon, A. M. Rucklidge, and W. F. Gale, "Modeling diffusion of energy innovations on a heterogeneous social network and approaches to integration of real-world data," *Complexity*, vol. 19, no. 6, pp. 83–94, 2014.

[40] P. De Meo, E. Ferrara, D. Rosaci, and G. M. L. Sarné, "Trust and compactness in social network groups," *IEEE Transactions on Cybernetics*, vol. 45, no. 2, pp. 205–216, 2015.

[41] H. Liu, Y. Zhang, H. Lin, J. Wu, Z. Wu, and X. Zhang, "How many zombies around you?" in *Proceedings of the 13th IEEE International Conference on Data Mining, ICDM 2013*, pp. 1133–1138, USA, December 2013.

[42] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?" *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 811–824, 2012.

[43] Y. Liu, D. Pi, and L. Cui, "Metric Learning Combining With Boosting for User Distance Measure in Multiple Social Networks," *IEEE Access*, vol. 5, pp. 19342–19351, 2017.

[44] Q. Zhang, J. Wu, Q. Zhang, P. Zhang, G. Long, and C. Zhang, "Dual influence embedded social recommendation," *World Wide Web*, 2017.

[45] Q. Yan, L. Wu, and L. Zheng, "Social network based microblog user behavior analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 7, pp. 1712–1723, 2013.