Online Information Review

Computational modeling of Weibo user influence based on information interactive network

Dingguo Yu Nan Chen Xu Ran

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/1468-4527.htm

Computational modeling of Weibo user influence based on information interactive network

Dingguo Yu
School of New Media, Zhejiang University of Media and Communications, Hangzhou, China
Nan Chen
Qianjiang College, Hangzhou Normal University, Hangzhou, China, and
Xu Ran
Center for Internet and Society, Zhejiang University of Media and Communications, Hangzhou, China

Received 21 December 2015
Revised 10 May 2016
Accepted 12 May 2016

Abstract
Purpose – With the development and application of mobile internet access, social media represented by Weibo, WeChat, etc. has become the main channel for information release and sharing. High-impact users in social networks are key factors stimulating the large-scale propagation of information within social networks. User influence is usually related to the user's attention rate, activity level, and message content. The paper aims to discuss these issues.
Design/methodology/approach – In this paper, the authors focused on Sina Weibo users, centered on users' behavior and interactive information, and formulated a weighted interactive information network model, then present a novel computational model for Weibo user influence, which combined multiple indexes such as the user's attention rate, activity level, and message content influence, etc., the model incorporated the time dimension, through the calculation of users' attribute influence and interactive influence, to comprehensively measure the user influence of Sina Weibo users.
Findings – Compared with other models, the model reflected the dynamics and timeliness of the user influence in a more accurate way. Extensive experiments are conducted on the real-world data set, and the results validate the performance of the approach, and demonstrate the effectiveness of the dynamics and timeliness. Due to the similarity in platform architecture and user behavior between Sina Weibo and Twitter, the calculation model is also applicable to Twitter.
Originality/value – This paper presents a novel computational model for Weibo user influence, which combined multiple indexes such as the user's attention rate, activity level, and message content influence, etc.

Keywords Social media, Sina Weibo, Opinion leader, User influence
Paper type Research paper

1. Introduction
In recent years, the popularity of social media such as Twitter, Facebook, and Sina Weibo has been booming, and these sites have become the prime platform for the publishing and sharing of information, as well as serving as important forums for the production

The authors truly thank the anonymous reviewers for their pertinent comments. This work was supported by the Philosophy and Social Science Planning Project of Zhejiang province, China (15NDJC135YB), the public welfare Technology Application Research Project of Zhejiang province, China (2016C33196), and the Science and Technology Project of state administration of press, publication, radio, film and television of China (201309).

1468-4527
DOI 10.1108/OIR-12-2015-0391

and dissemination of hot topics. With the increasing popularity of mobile social media, users do not even need a computer to surf online. They are able to utilize mobile internet services, via the smart mobile terminal, to send and receive information anytime, anywhere. This development not only overcomes the communication barriers set by time and location, but also enhances the immediacy of these interactions. Social media is able to link its users through its "following," "follower," and other functions. Users share news with their fans by forwarding posts, while fans can share the same posts with even more users by forwarding the same message. In Twitter there are 31 percent forwarding microblogs (Asur *et al.*, 2011), while in Sina Weibo the proportion of forwarding microblogs reaches 47.8 percent (Wang *et al.*, 2012). The sharing and propagation of network information has been greatly changed by the structure of user relations in social media. The forwarding behavior has become an important form of information transmission in Weibo, and also an important way for users to share and acquire information. Research shows (Donghao and Wenbao, 2014) that information propagation in social media, compared with information propagation in traditional media, appears to be large-scale, multi-mode, real-time, and rapid. The impact of social media on society and national public security continues to expand. The Egyptian Color Revolution is a good example of the great impact social media can have on society.

Weibo user influence measures the impact of each user on overall information propagation of microblogging platform. Greater user influence implies a higher attention rate, and also greater influence on network information dissemination. The presence and forwarding behavior of high-impact users are key factors in stimulating continuous information spreading and forming larger scale information propagation (Yi, 2011). Based on user influence measurements, a reasonable ranking of users within a microblogging system can be determined, which provides the theoretical basis for extended applications of Weibo, such as public opinion monitoring, network marketing, etc.

Weibo user influence is mainly determined by the user's attention rate, activity level, and microblog content influence. Currently, the investigation of Weibo user influence mostly focuses on the statistical analysis of number of fans, number of posted and forwarded posts, and number of comment received (Wang *et al.*, 2015). For example, the Sina Weibo platform uses the number of fans to indicate the user influence. While we agree that the number of fans can reflect user influence to a certain extent, we believe this measurement more accurately reflects the user's popularity, rather than influence. One complicating factor is the existence of "zombie fans." The presence of "zombie fans" may inaccurately inflate a user's total number of fans, resulting in a misleading attention rate. In addition, researches based on Twitter showed a weak correlation between the number of fans and user influence in information propagation (Cha *et al.*, 2010). Therefore, having a large number of fans is a necessary condition of user influence, rather than a sufficient condition. In fact, user influence is highly correlated to user activity and microblog content influence, whereas users with a high-attention rate are merely potential high-influence users. If a user with a high-attention rate is not active, that user's potential influence is not actuated. A higher quality microblog post also gains a stronger attraction for the user. A higher authoritative user in the field attracts more people who are interested in the message to transmit the post to even more users. Therefore when measuring user influence, more emphasis needs to be given to the user's ability to influence its fans with regards to information spreading, rather than only considering the user's ability to spread information to a large audience (Wang *et al.*, 2015; Romero *et al.*, 2011). This paper considered existing information on

Weibo users' attention rate and activity level, and constructed an information interactive network based on the Weibo user relations network, calculated Weibo content influence, and finally built a comprehensive user influence computational model comprising multiple indexes, such as users' attention rate, activity level, microblog impact, and other factors.

The main contributions of this paper are summarized as following:

(1) Focus on users' behavior and interactive information, we formulate the user relation network, Weibo element and information interaction relations in Sina Weibo system, and build the weighted information interactive network model based on the user relation network.

(2) According to the information interactive network model, we present a novel computational model for Weibo user influence, which combined multiple indexes such as the user's attention rate, activity level, and microblog content influence, etc. Compared with other models, our model reflected the dynamics and timeliness of the user influence in a more accurate way.

(3) We develop crawler and collect the real-world data set on Sina Weibo, and perform extensive experiments to evaluate our proposed model. The results validate the performance of our approach, and demonstrate the effectiveness of the dynamics and timeliness of user influence.

The rest of the paper is organized as follows. Section 2 displays some works about user influence computing approaches in social media. Section 3 introduces the definition of information interactive network. Section 4 details the proposed novel computational model for Weibo user influence, including direct influence computation and indirect influence computation. Section 5 presents the experimental design and results analysis. Finally, we conclude the study work in Section 6.

2. Related work

In social networks, users establish inter-user connections by utilizing the "following" function in order to construct the social network topology. Users exchange information by publishing, forwarding, and commenting on posts. The social network topology structure reflects the characteristics of user influence from the perspective of network topology. User behaviors in the social network, in addition to inter-user interactive information, can be used to reflect the emergence and evolution of user influence (Xindong *et al.*, 2014). Therefore, current research on social media user influence mainly focus on network topology, user behaviors, and interactive information. Using these parameters, network structure analysis can to a great extent reflect the user's network influence and also measure node user influence. For these reasons, network structure analysis is currently the most mainstream research method. In addition, the social network topological structure also forms the basis of other metrics research based on user behavior and interactive information.

Network topology structure based measurement methods mainly consist of methods that are node degree based, shortest path based, Hypertext Induced Topic Search (HITS) algorithm based, and PageRank algorithm based extension methods. In social media, a user's number of fans and number of microblog comment received are direct indicators of the user's influence. Node degree-based metrics indexes mainly include in-degree, out-degree, and degree centrality. The in-degree can assess the impact of the current node on its neighbor nodes, while the out-degree can assess the impact of neighbor nodes on the current node. The degree centrality (Freeman, 2012) can assess the average impact of the current node on its neighbor nodes. The social network shortest path based method focuses on closeness centrality (Newman, 2005) and between centrality (Newman, 2003). The closeness centrality can be used to measure the indirect impact of the current node on other nodes: a greater value indicates shorter distances between the current user and other users, and also a faster impact speed of the user on other users. The between centrality measures the importance of the node location within the network structure: a greater value indicates a greater amount of information flow through the node when information travels in the network, and also a greater node influence in information propagation.

The HITS (Kleinberg, 1999) algorithm was initially used in search engines. This algorithm was used to evaluate the importance of a webpage according to the centrality and authority of the webpage. Similarly, we can use the HITS algorithm in social networks to evaluate node influence via the centrality and authority of this node. Romero *et al.* (2011) proposed the Influence Passivity algorithm, based on the HITS algorithm, to measure the node user influence on Twitter.

The PageRank algorithm calculates node ranking in the directed graph. Assuming we treat the propagation of user influence as a random walk, the metrics can also be used to represent the user influence scale. Researchers have proposed a variety of social media user influence measurement models based on the PageRank algorithm. For example, Weng *et al.* (2010) proposed the TwitterRank algorithm, which calculated the influence of the user on each topic according to Twitter's following structure and user interest similarity. Xianhui *et al.* (2015) developed the TopicLeaderRank algorithm, which computed the user influence based on a constructed user behavior network containing Weibo users' content attributes and social attributes. In addition, there are also the forum reply emotional polarity based LeaderRank algorithm (Yu *et al.*, 2010), the microblog novelty and inter-microblog linking relation-based InfluenceRank algorithm (Song *et al.*, 2007), the user comments relation-based Microblog-Rank algorithm (Lin *et al.*, 2013), and others. Pal and Counts (2011) computed individual spreading influence, forwarding influence and mentioning influence based on individual number of fans, post, reply, forwarding and mentioning from Twitter's database, respectively.

Li *et al.* (2012) relied on Weibo historical messages and social interaction records to construct historical opinions and opinion influence using a statistical learning process, and developed an opinion influence model in terms of discussion topic level. Zhaoyun *et al.* (2013) considered time, microblog content, and network relations, and investigated an influence calculation method in Weibo. Aggarwal *et al.* (2011) presented a stochastic information flow model to identify the representative authority node in Twitter. Ver Steeg and Galstyan (2012) used the transfer entropy theory to describe the information flow among users in order to identify influential Twitter links. Goyal *et al.* (2010) adopted network log information in order to calculate the influence of the user and user behavior.

The network topology structure-based metric method conducts user influence analysis and modeling from network structure instead of a large amount of interactive information, which has these features including simple modeling and easy to use and expand, it is advantageous in studying scenario with a big-scale social network or limited interactive information. However, its sampled network topologies are static, like a snapshot of the social network structure, so it is incapable of representing the timeliness of the connection relation, as it treats links established ten years ago the same as it does links created a second ago. It is also unable to describe the intimacy of the connection, as it treats links receiving a single notice and links between two friends in heated conversation equally. This means that all users who have been connected are treated as having the same influence on one another. The above example clearly contradicts the actual situation. The root cause of these defects is that the network topology method simply emphasizes the existence of the relationship, while ignoring the timeliness and intimacy of the relationship. It does not make full use of the data related to the user behavior and interactive information, leading to a deviation of the metric method from the actual situation (Xindong *et al.*, 2014).

In this paper, in order to center the focus on users' behavior and interactive information in the Weibo network, we constructed a directed weighted information interactive network, incorporated time dimensions, calculated node attribute influence and interaction influence, and finally built a Weibo user influence computing model.

3. Network modeling

In this section, we will analyze the system architecture and user behavior in Sina Weib platform, define the user relation network as well as the Weibo dissemination relations, Weibo forwarding relations and so on in Weibo system, then formulate the weighted information interactive network model of Weibo system.

3.1 User relation network

In Weibo systems, users are classified as nodes, while the inter-user following relation is referred to as an edge. An edge begins at the original user and ends at the followed object. In light of this, the entire Weibo system can be viewed as a huge network. In this network, the influence of users with millions of fans differs from the influence of ordinary users, which means each user node is ascribed its own weight value. In addition, the connection of a single received notice also differs from the connection in a heated discussion between two friends in inter-user linkage intimacy. The propagation outcomes of user influence also differ between these scenarios, which means each edge has a distinct direction and weight. In short, the Weibo system user relation network can be defined as a directional weighted network:

Definition 1. The Weibo systems user relation network can be sketched as a directional weighted graph $UR = (V, E, P, W)$, wherein V denotes the user node set, E denotes the inter-user directional edge set, P represents the node weight set, and W represents the directional edge weight set. $\forall v_i, v_j \in V$, if $\exists e_{ij} = < v_i, v_j > \in E$, i.e., there is an edge starting from node v_i to node v_j, then it means that user v_i follows user v_j. In other words, user v_i is user v_j's fan/follower. $p_i \in P$ denotes the weight of node v_i, $w_{ij} \in W$ represents the weight of the edge e_{ij}.

3.2 Information interactive network

The unique network topology structure, information push, and forwarding mode of the Weibo system make its information dissemination different from the traditional linear transmission and network transmission mode, presenting a fissional transmission mode. This fissional transmission greatly broadens the span and speed of information transmission. In a user relation network, a greater node in-degree implies a user has more fans and a larger span of information transmission. However, the Weibo system network topology structure can only prove Weibo users' potential information dissemination ability, instead of the actual information dissemination ability and user influence. If a user with millions of fans seldomly posts or forwards microblogs, then his information transmission ability and user influence are minor, despite his large number of fans. Or if the user just posts or forwards microblogs with outdated content, due to the lack of novelty it will be difficult to arouse other users' resonance or motivate his fans to forward or comment on these messages. In this case, the user information dissemination ability and user influence are also diminished.

In summary, the Weibo system network topology structure determines users' potential information dissemination ability. A user's activity level and posted or forwarded microblogging content directly determines the user's influence, and the user's activity level is directly related to the user's microblog posting or forwarding behavior. This paper begins with the investigation of users' microblog posting or forwarding behavior and the subsequent Weibo content influence, and then constructs the Weibo information interactive network, and finally builds a user influence computational model. Next we will define the information interactive network.

In the microblog system, the main actions users take involving microblog messages include: posting, forwarding, actively pushing (@username), commenting, etc. Information release and forwarding are carried out in the user relation network. Information propagates in the opposite direction of edges within the relation network. The active information push can be toward a user's fans or other users. Information can be added as a comment under an original Weibo post. The comment transmission range is the same as the original Weibo post transmission range. Figure 1 shows the microblog message propagation network diagram.

In Figure 1, user v_1 posts microblog m_{11}. The microblog message travels along the opposite direction of the user's following relation. User v_1's fans v_2 and v_3 both receive

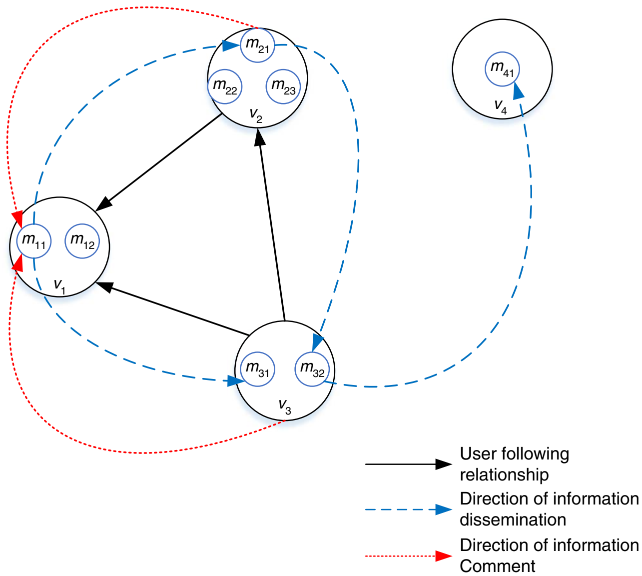

Figure 1. Microblog information transmission relation network

the microblog. v_2 and v_3 both add their comments under the microblog post. User v_2 forwards microblog m_{21}, v_2's fan v_3 receives the message, user v_3 directly pushes microblog m_{32} (@v_4) to user v_4. Analyzing the microblog information transmission network, so we give the following definition for the microblog system information interactive network:

Definition 2. Weibo elements: microblog system elements can be defined as an eight-element group: $m = (ID, Content, Author, Type, Time, FN, CN, PN)$, wherein m is a post in Sina Weibo, ID is the serial number of m, $Content$ is the content of m, $Author$ is the author of m, $Type$ is the type of m (e.g. original posts, forwarded posts), $Time$ is the post timestamp of m, FN is the number to be forwarded of m, CN is the number to be commented of m, and PN is the received "like" count of m.

Definition 3. Weibo dissemination relations: let $UR = (V, E, P, W)$ as the Weibo system user relation network. User v_i posts (including forwards) microblog m. The dissemination relation of m can be defined as $sr = (m, <v_i, v_j>)$, wherein $<v_i, v_j>$ is the directional edge, indicating the information transmission direction. Microblog m is passed on from v_i to v_j, so either user v_j is a fan of user v_i, i.e. $<v_j, v_i> \in E$, or v_j is a direct push object of v_i's microblog m, i.e. m's content contains content like "@v_j."

Definition 4. Weibo forwarding relations: Weibo forwarding relations in the Weibo system can be defined as $fr = <m_i, m_j>$, where m_j is a microblog forwarded from m_i, the direction of the microblog forwarding relation is pointing to m_j originating from m_i.

Definition 5. Weibo commenting relations: Weibo commenting relations in the Weibo system can be defined as $cr = <v_j, m_i>$, which means that user v_j comments on microblog m_i, the direction of the microblog commenting relations is pointing from v_j to m_i.

Definition 6. Information interactive network: the user relation network in Weibo systems is represented as $UR = (V, E, P, W)$, so the Weibo system information interactive network can be defined as $MI = (M, U, SR, FR, CR, WM)$, wherein M is the collection of Weibo element m in the Weibo system; U is the collection of Weibo users who post, forward, and comment on the microblog, $U \subseteq V$; SR is the collection of Weibo dissemination relations; FR is the collection of Weibo forwarding relations; CR is the collection of Weibo commenting relations; WM is the collection of the influence weight of Weibo elements. $\forall m_i \in M$, $Author(m_i) \in U$ denotes the author of microblog m_i, $wm(m_i) \in WM$ denotes the influence weight of microblog m_i's content.

$FU(m_i)$ denotes the collection of users who forwarded microblog m_i:

$$FU(m_i) = \{Author(m_j) \mid \exists m_j \in M, <m_i, m_j> \in FR,\}$$

$CU(m_i)$ denotes the collection of users who commented on microblog m_i:

$$CU(m_i) = \{v_j \mid \exists v_j \in U, <v_j, m_i> \in CR,\}$$

4. User influence model

Weibo user influence refers to the fact that Weibo users, through network social behavior, e.g. posting and forwarding microblogs, or commenting on and liking others' Weibo content within the microblog system, influence other people's thoughts, feelings, attitudes, and subsequent influence. We divided Weibo user influence into direct and indirect influence. Direct influence refers to the direct social impact of the user on its fans, while indirect influence refers to the social impact of the user on other users via its fans, as shown in the following formula:

$$IF = \rho \cdot IF_{dir} + (1-\rho) \cdot IF_{indir} \quad (0 \leqslant \rho \leqslant 1) \tag{1}$$

where IF is the Weibo user influence, IF_{dir} is the direct influence, IF_{indir} is the indirect influence, and parameter ρ and $(1-\rho)$ are the respective proportion of direct influence and indirect influence. Suppose $\rho = 1$, then the direct influence accounts for all the user influence and no indirect influence is present.

4.1 Direct influence calculation

The user's direct influence is mainly determined by the user's basic attributes and the user's posted (including forwarded) Weibo content. For user's basic attributes, we can get the user social information including real-name authentication symbol, following number and follower number, and the user behavior information including total microblog count, total forwarding count and total comment count from user profile page. For each microblog the user posted, we can get the statistical value including total forwarded count, total commented count and total like count from microblog information page. So, the assessment indexes of the user's direct influence are listed in Table I.

Hence, the user's direct influence can be calculated by the following formula:

$$IF_{dir} = \alpha IF_{dir\_user} + (1-\alpha) IF_{dir\_mb} \tag{2}$$

where IF_{dir_user} is the direct influence determined by the user's basic attribute, which can be calculated by Formula (3), IF_{dir_mb} is the direct influence decided by the user's Weibo content, and parameter α and $(1-\alpha)$ are the respective proportion of the user's basic attribute and the user's Weibo content:

$$IF_{dir\_user} = \beta_0 V + \beta_1 N_1 + \beta_2 N_2 + \beta_3 N_3 + \beta_4 N_4 + \beta_5 N_5 \tag{3}$$

where V denotes the user's real-name authentication mark, N_1 denotes the user's following number, N_2 denotes the user's follower number, N_3 denotes the user's total

Table I.
Assessment indexes of user's direct influence

Attributes	Detailed indexes
User basic attribute	Real-name authentication
	Following number
	Follower number
	Total microblog count
	Total forwarding count
	Total comment count
Webo content attribute	Total forwarded count
	Total commented count
	Total like count

number of microblog posts, N_4 denotes the user's total number of forwarded posts, N_5 denotes the user's total number of comments given, and the parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the respective proportion of the user's real-name authentication mark, user's following number, user's follower number, user's total number of microblog posts, user's total number of forwarded posts and user's total number of comments given, $\beta_0+\beta_1+\beta_2+\beta_3+\beta_4+\beta_5=1$.

The user influence decided by the user's Weibo content is composed of the content influence of each microblog posted by the user. The content influence of each microblog post is relevant to how the microblog is forwarded, commented and liked, such as the number of forwards and also who forwarded the microblog. So the influence of microblog m_j can be calculated by the following formula:

$$wm(m_i) = \gamma_1 PN + \gamma_2 \sum_{v_j \in FU(m_i)} w(v_j) + \gamma_3 \sum_{v_k \in CU(m_i)} w(v_k) \tag{4}$$

where PN represents the like count of microblog m_i, $FU(m_i)$ represents the collection of users who forwarded microblog m_i, $CU(m_i)$ represents the collection of users who commented on microblog m_i, and $w(v_j)$ represents the weight of user v_j. The parameters γ_1, γ_2, γ_3 are the weight percentages of likes, forwards and comments, respectively, $\gamma_1+\gamma_2+\gamma_3=1$.

The content attribute for the user's posted microblog factors in a timeline, namely, the difference between the influence of a microblog posted one year ago and the influence of a microblog posted one day ago. To reflect the activity level of user behavior and the timeliness of the Weibo content, we introduced the concepts of the time window and decay factor:

Definition 7. Time window (also known as time frame): the time window is defined as a time interval (window size equals to the length of time interval) used to depict the time span of a user posting, forwarding, or commenting on a particular microblog. The window sequentially slides forward as time advances, i.e. when the current window ends, the next window starts.

The length of the time window and the number of windows can be determined according to specific application scenarios. For simplicity, we used positive integers to represent the time windows; a greater integer implies a window closer to the current time. Specifically, we used 1, 2, … n to denote the first, second, … to the nth time window, among which the nth time window is the current time window.

In addition, the timeliness of Weibo content and more accurate calculation of Weibo content influence demand a differentiation among the user Weibo posted in different time windows. A common practice is to assign different weights to the calculated results of posts in different time windows, proportional to the distance from the current time window. A window closer to the current time window would be assigned a larger weight, while a time window with a farther distance from the current time would have a smaller weight associated with it. We also introduced a decay function. Making use of the constraints of the decay function, we achieved a more robust and reasonable outcome than with weight allocation alone. This is true because the weight allocation method is a subjective and experience-based decision, whereas the decay function has constraints for internal parameter selection, so the function value is easy to control.

Also the decay rate theorectically has more operational flexibility. The definition of the decay function is given below:

Definition 8. Decay function: the amplitude discount function of the content influence of the Weibo posted in the kth time window relative to the Weibo posted in the current time window (the nth window) is called the decay function, which can be expressed as $f(k) = \phi^{n-k}$, $0 < \phi < 1$, $1 \leqslant k \leqslant n$.

For the defined decay function f, there is a decay factor (function value) corresponding to each time window. For instance, the corresponding decay factor for time window W^k is ϕ^{n-k}. So, the decay factor of the current window is 1, which means there is no attenuation, whereas the decay factor for the first time window relative to the current window is ϕ^{n-1}, which implies the largest attenuation of amplitude.

From the above analysis, we can conclude that the direct influence determined by a user's Weibo content attribute IF_{dir_mb} can be calculated using the following formula:

$$IF_{dir\_mb} = \sum_{k=1}^{n} f(k) \sum_{m_i.Time \in W^k} wm(m_i) \tag{5}$$

where W^k represents time window k, $m_i.Time \in W^k$ denotes that microblog m_i is posted or forwarded in time window k.

4.2 Indirect influence calculation

A user's indirect influence is the influence of the user on other people through its fans and direct push objects (via @) in Weibo. For simplicity, this paper only discusses the user indirect influence passed through its fans.

In calculating the user's indirect influence, we referred to the TopicLeaderRank algorithm, which is based on the iterative voting ideas of the PageRank algorithm and expand the PageRank voting strategy (Xianhui *et al.*, 2015). The PageRank algorithm is based on the idea of linked voting, i.e. edge e_{ij} represents a vote of node v_i to node v_j. Each node will distribute its own PageRank weight evenly to the oriented node during voting.

On the basis of the Weibo systems user relationship network, the user's indirect influence is calculated following the idea of network topology-based voting. In the user relationship network, a directed edge indicates a poll, the user distributes its own influence to the objects it follows by a certain proportion, i.e., the user's indirect influence is composed of the cumulative influence of its fans. Different from the traditional PageRank algorithm, the node no longer treats all the linked nodes equality during its voting. Not only should the edge weights be considered, but the load weights should also be considered. Edge weight represents the degree of interaction between users. A larger weight implies a more intimate interaction between users, and also a greater inter-user influence. User node weight refers to the user's influence. Therefore, in voting the node needs to distribute its own influence based on the edge weight.

Based on the analysis above, the indirect influence of a user is composed of his fans' influence, and the fans' influence is divided into direct influence and indirect influence, while indirect influence and with his fans about. So, we can only use an iterative way to calculate the indirect influence of users, let vectors π_{IF} and $\pi_{IF_{indir}}$ denote the influence and indirect influence of all users, the iterative formula for calculation of the user's indirect influence can be expressed in the following equation:

$$\pi_{IF_{indir}}^{(k+1)T} = (1-c)e^T + c\pi_{IF}^{(k)T} V \tag{6}$$

where $\pi_{IF_{indir}}^{(k+1)T}$ is the updated indirect influence vector of all nodes, $\pi_{IF}^{(k)T}$ is the current influence vector of all nodes, $\pi_{IF}^{(k)T} = \pi_{IF_{dir}}^{(k)T} + \pi_{IF_{indir}}^{(k)}{}^{T}$, substitute into Formula (6). Formula (7) can be obtained; the parameter c is the damping coefficient, which is used to solve the sparseness problem of graph and ensure the convergence of algorithm; e^T is the unit row vector; V is the voting matrix, wherein element v_{ij} denotes the voting weight computed using Formula (8):

$$\pi_{IF_{indir}}^{(k+1)T} = (1-c)e^T + c\left(\pi_{IF_{\mathrm{dir}}}^{(k)} + \pi_{IF_{\mathrm{indir}}}^{(k)}\right)^T V \tag{7}$$

$$v_{ij} = \begin{cases} 0, & e_{ij} \notin E \\ \frac{p_i w_{ij}}{\sum_{e_{ik} \in E} w_{ik}}, & e_{ij} \in E \end{cases} \tag{8}$$

wherein p_i stands for the weight of node v_i, which is replaced in this paper by the user influence determined by the user's basic attribute IF_{dir_user} and calculated by Formula (3). w_{ij} stands for the weight of edge e_{ij} and is calculated using the following equation:

$$w_{ij} = \eta N_4 + (1-\eta) N_5 \tag{9}$$

wherein N_4 is the number of microblogs that node user v_i forwards from node user v_j, N_5 is the number of commentaries that node user v_i comments on the microblogs of node user v_j, and the parameter η and $(1-\eta)$ are the respective proportion of N_4 and N_5.

5. Experiments

In this section, we will report the experiments conducted on the real data set which crawled from Sina Weibo by our own developed crawlers, and to validate the performance of our approach.

5.1 Experiment data

In this paper, we developed a specific crawler with the help of Weibo API for collecting Sina Weibo data. The crawler configured 200 relatively high-network influence Weibo users as seed users. Based on these seed users, the system automatically collected information on some of these users' fans and Weibo direct push objects, added them into the user database to become new Weibo users, and finally collected the Weibo articles posted or forwarded and comments by all the Weibo users in the system. As of July 30, 2015, the system had collected a total of 23.1 million pieces of microblog information from 3,360 users. Among these microblogs there were about 11.06 million forwarded Weibo articles and about 3.56 billion comments. Based on this data, we deployed the proposed Weibo user influence computational model and calculated the influence of the Weibo users in the system.

5.2 Experiment setup and results analysis

On the one hand, in order to verify the dynamics and timeliness of user influence, and determine the user with high number of fans or high number of microblogs whether has a high influence. On the other hand, in order to comprehensively evaluate the effectiveness of our proposed method, we design this experiment as follows. First, we computed the direct influence for all users by using the users' attributes information and the statistical value for each microblog. Second, using the user's list of followees and followers to build the user relation network, then calculate the user indirect influence though ten iterations. Finally, computing user influence via the direct influence and indirect influence. All the algorithms are run on the real Sina Weibo data set described above using threefold validation, and the results reported in this section are the average values.

Table II lists the top 50 ranked Sina Weibo users based on influence, it can be observed from Table II that a larger fan count does not necessarily lead to greater user influence, for example, user "姚晨" had the greatest number of fans, but her user influence only ranked 22th. Likewise, a larger microblog count did not necessarily lead to greater user influence either, for example, user "侯宁" occupied first place in the microblog count ranking, yet the same user's associated influence is not high.

Moreover, the model incorporated the time dimension when calculating user influence. The model considered the timeliness of the user's Weibo content influence in a way that better reflected the timeliness and dynamics of the user's influence. In Figure 2, the influence-changing curve of Kai-fu Lee (李开复) *et al.* over the two months (from May 1, 2015 to June 30, 2015) is plotted.

6. Conclusions

Weibo user influence mainly relates to the user's attention rate, activity level, and Weibo content influence. Traditional static analysis based on data such as user fan count, microblog count, and comment count, etc., and the metric study based on network topology does not properly describe the user's activity level and microblog content influence, etc. Based on existing studies, we constructed the information interactive network model from the user relation network, and built a user influence computational model which combined multiple indexes such as the user's attention rate, activity level, and microblog content influence, etc. The real-world data from Sina Weibo verified that our model reflected the dynamics and timeliness of the user influence in a more accurate way. Our experiment results further illustrated that fan count and microblog count are the most important factors in affecting user influence, but these are not the sole determining factors of user influence. Additionally, due to the culture difference between China and western countries, there are some differences in message content between Sina Weibo and Twitter. But the user influence calculation method our proposed is equally applicable to Twitter platform, because both system architecture and user behavior are similar in Sina Weibo and Twitter, and the calculation method only relates to the influence of message content does not involve the specific message content.

Since the iterative calculation method is adopted in computing indirect influence, the computational efficiency is low when the user scale is large. In future work, we will further improve the model, to improve the efficiency of calculating indirect influence. Moreover, develop a complete user influence calculation software for Sina Weibo, for effective monitoring and analysis of social media opinion.

Nickname	Number of follower	Number of followee	Number of Weibo	Influence value	Influence ranking
任志强	33,139,844	196	84,459	316.961242	1
李开复	50,077,407	515	14,276	296.760611	2
王思聪	14,085,252	206	1,313	285.836069	3
薛蛮子	11,067,856	2,078	94,641	208.413862	4
陈里	25,256,365	2,636	21,136	182.959833	5
胡锡进	4,377,386	188	4,235	169.903007	6
董藩	370,281	532	26,445	166.987958	7
周小平同志	549,103	1,504	157	158.352268	8
戴旭	490,393	3,144	7,234	137.289961	9
评论员杨禹	5,039,752	1,445	2,449	123.256069	10
郭德纲	62,455,140	66	1,450	114.553592	11
侯宁	*1,479,150*	*1,805*	*148,930*	*110.906823*	*12*
但斌	8,868,639	1,249	108,477	109.692709	13
孟非	33,981,917	18	3	107.822192	14
韩寒	41,920,758	1,040	398	106.51315	15
杨澜	34,875,375	156	6,444	101.853064	16
苏芩	46,533,086	959	3,927	96.887776	17
袁裕来律师	20,333,393	558	52,697	93.436502	18
谢娜	69,866,384	665	8,042	92.024898	19
何炅	67,371,627	610	7,131	89.193744	20
罗永浩	11,823,435	1,905	25,348	88.901864	21
姚晨	*78,318,421*	*428*	*7,993*	*85.899286*	*22*
思想聚焦	10,779,003	1,521	46,961	85.463468	23
老榕	1,060,736	1,468	89,641	83.512057	24
王石	21,021,645	327	7,756	82.886052	25
袁岳	3,993,085	2,990	49,836	81.743995	26
徐昕	29,432,810	770	5,920	80.670278	27
马云	19,451,323	1	105	80.386883	28
雷军	12,315,938	956	5,050	79.432485	29
谷大白话	5,802,593	1,181	15,632	78.647789	30
陈坤	77,908,508	514	4,860	78.636992	31
赵薇	72,191,335	424	4,069	77.181463	32
angelababy	62,477,211	656	2,107	76.712447	33
延参法师	38,599,060	1,321	26,367	76.661526	34
杨幂	51,722,048	606	3,064	76.363485	35
李易峰	22,059,975	138	1,964	75.550746	36
周立波	36,211,839	25	1,023	75.4979	37
留几手	10,188,187	238	1,578	74.787838	38
陈彤	6,983,696	2,976	34,221	74.61107	39
邓超	31,917,600	123	506	73.444529	40
潘石屹	17,232,590	117	22,097	73.117811	41
黄健翔	17,814,974	309	21,265	70.758606	42
巴曙松	9,228,947	1,962	25,571	70.499229	43
刘同	18,073,098	1,236	8,078	70.430449	44
司马平邦	480,809	2,740	53,686	69.745841	45
顾剑	1,563,028	1,045	39,049	68.268736	46
学习粉丝团	2,820,445	3,061	3,318	68.209568	47
董路	7,247,110	7	41,533	68.103966	48
朱学东	3,026,289	1,879	41,194	68.057412	49
姜岚昕	4,275,455	2,003	16,713	67.914833	50

Table II. The top 50 ranked Sina Weibo users based on influence

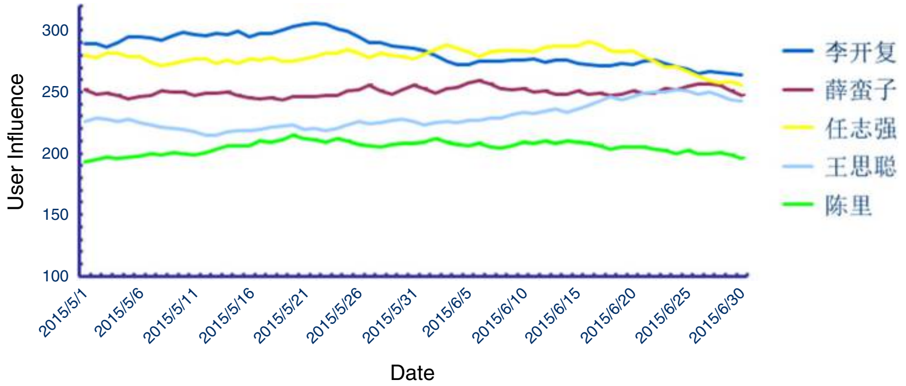

Figure 2.
The influence-changing curve of Kai-fu Lee (李开复) *et al.* over the two months

References

Aggarwal, C.C., Khan, A. and Yan, X. (2011), "On flow authority discovery in social networks", *Proceedings of the Eleventh Siam International Conference on Data Mining, SDM*, pp. 522-533.

Asur, S., Huberman, B.A., Szabo, G. and Wang, C. (2011), "Trends in social media: persistence and decay", *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*.

Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, P.K. (2010), "Measuring user influence in Twitter: the million follower fallacy", *Proceedings of the International Conference on Weblogs and Social Media, ICWSM*, pp. 10-17.

Donghao, Z. and Wenbao, H. (2014), "DiffRank: a novel algorithm for information diffusion detection in social networks", *Chinese Journal of Computers*, Vol. 37 No. 4, pp. 884-893 (in Chinese).

Freeman, L.C. (2012), "Centrality in social networks conceptual clarification", *Social Networks*, Vol. 1 No. 3, pp. 215-239.

Goyal, A., Bonchi, F. and Lakshmanan, L.V.S. (2010), "Learning influence probabilities in social networks", *Proceedings of the International Conference on Web Search and Web Data Mining, WSDM, February*, pp. 241-250.

Kleinberg, J.M. (1999), "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, Vol. 46 No. 5, pp. 604-632.

Li, D., Shuai, X., Sun, G., Tang, J., Ding, Y. and Luo, Z. (2012), "Mining topic-level opinion influence in microblog", *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM, October*, pp. 1562-1566.

Lin, Y., Li, H., Liu, X. and Fan, S. (2013), "Hot topic propagation model and opinion leader identifying model in microblog network", *Abstract and Applied Analysis*, No. 6, pp. 1-13.

Newman, M.E.J. (2003), "The structure and function of complex networks", *SIAM Review*, Vol. 45 No. 2, pp. 40-45.

Newman, M.E.J. (2005), "A measure of betweenness centrality based on random walks", *Social Networks*, Vol. 27 No. 1, pp. 39-54.

Pal, A. and Counts, S. (2011), "Identifying topical authorities in microblogs", *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, ACM, February*, pp. 45-54.

Romero, D.M., Galuba, W., Asur, S. and Huberman, B.A. (2011), "Influence and passivity in social media", *Machine Learning and Knowledge Discovery in Databases*, Springer, Berlin and Heidelberg, pp. 18-33.

Song, X., Chi, Y., Hino, K. and Tseng, B. (2007), "Identifying opinion leaders in the blogosphere", *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management. ACM, November*, pp. 971-974.

Ver Steeg, G. and Galstyan, A. (2012), "Information transfer in social media", *Proceedings of the 21st International Conference on World Wide Web, ACM, April*, pp. 509-518.

Wang, C., Guan, X., Qin, T. and Li, W. (2012), "Who are active? An in-depth measurement on user activity characteristics in Sina microblogging", *Proceedings of the Global Communications Conference (GLOBECOM), IEEE, December*, pp. 2083-2088.

Wang, C.X., Guan, X.H., Qin, T. and Zhou, Y.D. (2015), "Modeling on opinion leader's influence in microblog message propagation and its application", *Journal of Software*, Vol. 26 No. 6, pp. 1473-1485 (in Chinese).

Weng, J., Lim, E.P., Jiang, J. and He, Q. (2010), "Twitterrank: finding topic-sensitive influential twitterers", *Proceedings of the Third ACM International Conference on Web Search and Data Mining, ACM, February*, pp. 261-270.

Xianhui, W., Hui, Z., Xujian, Z., Bo, L. and Chunming, Y. (2015), "Mining algorithm of microblogging opinion leaders based on user-behavior network", *Application Research of Computers*, Vol. 32 No. 4, pp. 2678-2683 (in Chinese).

Xindong, W., Yi, L. and Lei, L. (2014), "Influence analysis of online social networks", *Chinese Journal of Computers*, Vol. 37 No. 4, pp. 735-752 (in Chinese).

Yi, Y. (2011), "The analysis of structure, path and factor of microblog information communication", *Library and Information Service*, Vol. 55 No. 12, pp. 26-30 (in Chinese).

Yu, X., Wei, X. and Lin, X. (2010), "Algorithms of BBS opinion leader mining based on sentiment analysis", *Proceedings of the International Conference on Web Information Systems and Mining, Springer Berlin Heidelberg*, pp. 360-369.

Zhaoyun, D., Yan, J., Bin, Z., Jianfeng, Z., Yi, H. and Chunfeng, Y. (2013), "An influence strength measurement via time-aware probabilistic generative model for microblogs", *Proceedings of the 15th Asia-Pacific Web Conference, APWeb, Springer Berlin Heidelberg*, pp. 372-383.

Corresponding author
Dingguo Yu can be contacted at: yudg@zjicm.edu.cn