Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2022) 30: 592 – 608
© TÜBİTAK
doi:10.3906/elk-2105-163

Calculating influence based on the fusion of interest similarity and information dissemination ability

Shulin Cheng[1,2], **Ziming Wang**[1],**Meng Qian**[1,2,*],**Shan Yang**[1] **and Xin Zheng**[1,2]
[1]School of Computer and Information, Anqing Normal University, Anqing, China
[2]The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing, China

Received: 17.05.2021 • **Accepted/Published Online:** 07.10.2021 • **Final Version:** 21.03.2022

Abstract: With the popularization, in-depth development and application of the Internet, microblogs have become a mainstream social network platform. Several studies on social networks have conducted researches, and user influence evaluation is an important research hotspot. Most of the existing studies calculate user influence by improving PageRank and have achieved certain results. However, these studies ignored the fusion of users' interest theme similarity and information dissemination ability, and the analysis of interaction behaviors among users is not comprehensive. To address these issues, we propose a new microblog user influence algorithm called microblog user influence based on interest similarity and information dissemination ability (MUI-ISIDA), which fully integrates user's interest theme similarity and information dissemination ability. We construct the model of interest theme similarity and then allocate followers' contributions to the influence of bloggers reasonably. Considering the quality of microblogs, the numbers of forwarding, commenting, and effective interaction behaviors among users, the microblog quality coefficient and the assimilation effect coefficient are designed. On this basis, a user's information dissemination ability model was constructed. We verified the effectiveness of the proposed algorithm using a real dataset. According to these experimental results, our proposed algorithm achieved higher accuracy in ranking user influence than other state-of-the-art algorithms.

Key words: Microblog, user influence, PageRank, user interest, effective interaction

1. Introduction

Recently, with the rapid development and popularization of the Internet, social networks have played an increasingly important role in information dissemination and influence, and it has become an important medium for users to obtain and exchange information. Twitter is a famous social network representative of hundreds of millions of daily active users, and users can share and disseminate information through interaction behaviors every day [1]. Similarly, Facebook has created a new world of communication and cooperation for users. It is also a prevalent social network platform that covers all aspects of users' daily lives. In China, microblogs have become one of the most extensive social network platforms because of their openness and content simplicity. According to the 46th 'Statistical Report on China's Internet Development Status' reported by CNNIC [2], the usage rate of microblog has reached 40.4% as of June 2020. At present, microblogs have gradually become one of the most influential platforms in Chinese social networks.

As one of the most extensive social network platforms, microblog encourages users to publish short posts (usually no more than 140 words). Additionally, users can forward and comment on other users' microblogs, as

*Correspondence: tmpit@126.com; qianmeng@aqnu.edu.cn

well as give them a thumbs up. Different behaviors among users make information spread quickly, which forms the microblog social network. Certainly, the users here exclude abnormal users who engage 'zombie fans' to increase the spread of their microblogs.

Influence refers to the ability of an individual to influence others and change their thoughts or behaviors [3]. Ling (2016) believes that information spreads continuously through users' behaviors in social networks, while users receive and disseminate information; they are also in the process of influencing and being influenced [4]. Social networks can be used as an important information dissemination platform. Users of different influences have different effects on the speed and scope of information dissemination, and high-influence users in the network can be used as seed nodes to maximize the spread of influence [5]. Users with high influence play active roles in public opinion monitoring, advertising promotion, and other fields. For example, the real-time release of 2019-nCoV information through social networks can track epidemic trends. Therefore, it is of great significance to measure the influence of users on social networks. In this study, we propose a new algorithm, MUI-ISIDA, based on PageRank[6], which fuses the user's interest theme similarity and information dissemination ability to identify high-influence users in the microblog social network. The main contributions of this algorithm are as follows:

(1) Based on the connections among users, we combine users' microblogs with followers' opinions and comments and analyze the effectiveness of users' behaviors, which could reduce the interference of 'zombie fans' to a certain extent.

(2) The more similar the users' interest theme, the greater the mutual influences among users [7]. Therefore, inter-theme similarities among users are used to model the closeness of users' connections, overcoming the problem of even distribution of influence.

(3) Considering the quality of users' microblogs and the extent to which bloggers' behaviors affect users, a weighting vector model is constructed to assign different weights to different users.

The proposed algorithm considers the influencing factors of user influences more comprehensively and reasonably. The user influences were effectively measured, and the experimental results verified the effectiveness of this algorithm.

2. Related work

At present, with the continuous deepening of research on complex networks, the measurement of user influence has attracted the attention of many scholars and has achieved certain results and progress. Jianqiang et al. [8] proposed the UIRank algorithm through interactive information flow and interactive relationships among users in a microblog network. The results show that this algorithm is superior to other related algorithms in terms of precision and recall, however their study did not analyze the content of users' microblogs and explored the distribution of users' interests. Sun et al. [9] analyzed the actual behaviors of users comprehensively, introduced the user's own weight and proposed the MR-UIRank algorithm. However, they ignored the effectiveness of user behaviors. Wu et al. [10] proposed an MPPR algorithm based on relationships between microblog contents and user behaviors, and integrated the two relationships to evaluate user influences. However, the experiment was insufficient, only compared with PageRank. Qi et al. [11] analyzed users' behaviors comprehensively, and measured the contribution of each behavior to influence. They then proposed a method for assigning weights through different behaviors to improve the PageRank, but did not consider the users' implicit preferences. Lin et al. [12] proposed the IB-UIR algorithm based on PageRank, which considers users' behaviors and the time interval between behaviors comprehensively. Jun et al. [13] used a linear regression model to analyze changes in user influence over time, but they ignored the role of users' microblog content. Huang et al. [14] proposed a graph segmentation algorithm for identifying highly influential users by combining users' interest similarities with social interactions. Char et al. [15] proposed the BPR algorithm, which considers user behavioral characteristics and improves PageRank for the average distribution of influence. Sheikhahmadi et al. [16] proposed an improved clustering and sorting algorithm that considers the correlation between a user's connection structure and its neighbor nodes to estimate the influence value of the node. Zareie et al. [17] proposed a new algorithm based on domain diversity by analyzing the network structure and using the diversity of node neighbors to obtain the influence-ranking value. Zareie et al. [18] proposed a new standard for measuring users' interests in information, and conducted experiments on real and synthetic networks, while Xing et al. [19] proposed a weighted PageRank algorithm, which considers the importance of the internal and external links of the page and assigns ranking scores based on the popularity of the page. Ying [20] proposed a weighted web page ranking method based on the author's citation network and used principal component analysis (PCA) to detect the relationship between different measurement methods, and thirteen evaluation indicators were selected to analyze the experimental results. In [21] a topic related to ranking, which is based on a combination of topic model and weighted page ranking algorithm, using the ACT model to extract topics and associating topics with a single author, which constitutes the weighted vector of the PageRank algorithm, including different contextual information as a weighted vector into the web page-ranking algorithm, which brings finer granularity to expert ranking in various situations was proposed. Danil et al. [22] considered the limitations of page ranking and personalized page ranking, and freely weighted hyperlinks according to any possible preference behavior of users, especially considering the weights related to clustering, and proposed a weighted web page-ranking algorithm.

The above studies mainly focused on user attributes and interaction behaviors. However, they ignored the fusion of users' interest theme similarity and information dissemination ability, and the analysis of interaction behaviors among users is not comprehensive. To this end, we construct the model of interest theme similarity and then allocate followers' contributions to the influence of bloggers reasonably. Additionally, we consider the quality of microblogs, the numbers of forwarding, commenting, and effective interaction behaviors among users, thereby constructing a user information dissemination ability model. Finally, the MUI-ISIDA algorithm is proposed by fusing users' interest theme similarity and information dissemination ability.

3. Microblog social network model

3.1. Microblog social network diagram based on users' behaviors

The 'following' relationships among users form the information dissemination network, and the microblogs published by users can spread along the network. The interactive behaviors among users, such as forwarding and commenting, promoted the spread of information again [23]. To better study the algorithm, we first define the abstract model of the microblog social network.

In the microblog social network, there are multiple behaviors among users, and these behaviors are directional. We select the following, forwarding, and commenting in this study, and define the microblog network as $G < V, E, W >$, where V is the set of the user nodes. E is the edge set of the users' relationships, and W is the set of edge weights in the network. As shown in Figure 1, if user A follows user B, forwards or comments on his/her microblogs, it means that a directed edge is formed from user A to user B. The weight of the edge W_{AB} represents the closeness of the connection between A and B, and W_A represents the weight of the user.

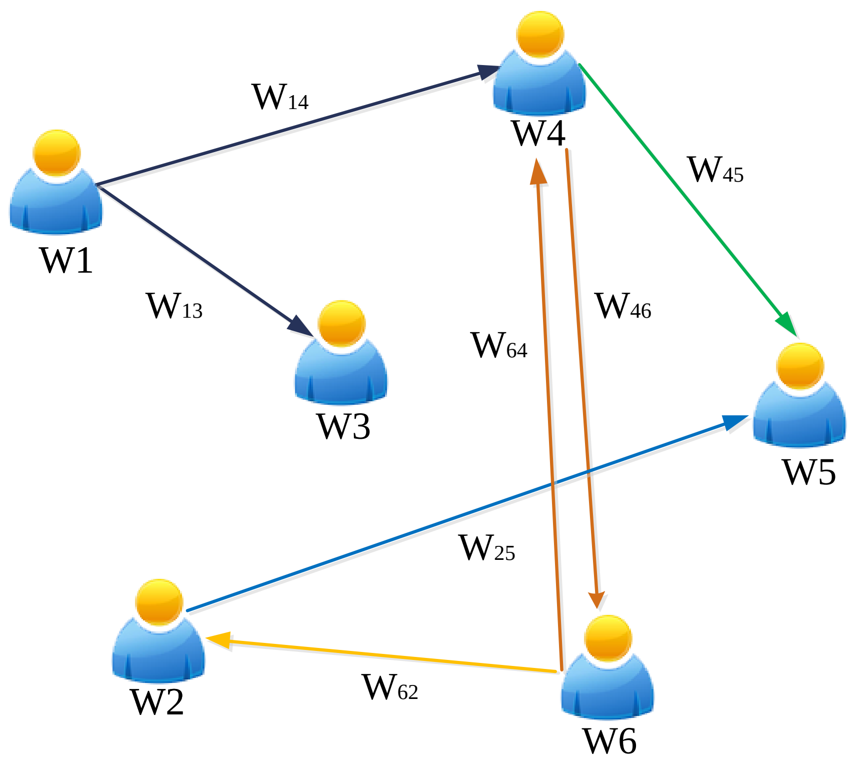


Figure 1. Microblog network.

3.2. The traditional calculation method for user influence

The PageRank algorithm proposed by Larry Page and Sergey Brin was the first to rank web page importance, and it was used to calculate microblog users' influences in the early days [24]. In PageRank, the link structure is used to calculate the importance of web pages, and a directed graph is leveraged to represent the relationships among these pages. Each node in the graph represents a webpage, and each edge indicates a link. As shown in Eq. (1)

$$\text{PageRank}(p_i) = \frac{1-\alpha}{N} + \alpha \sum_{p_j \in M_{p_i}} \frac{PageRank(p_j)}{L(p_j)} \tag{1}$$

where $p_1, p_2 \cdots \cdots p_N$ represents web pages, $M(p_i)$ is the collection of all pages linked to p_i, and $L(p_j)$ is the number of external links on the web page p_j $.\alpha$ is the damping factor $(0 \leq \alpha \leq 1)$, which indicates the probability of a page being accessed randomly.

Figure 2 shows the relationship between p_i and p_j when the PageRank algorithm is used. The PageRank algorithm ignores the contents and attributes of web pages, which only consider the link relationships among them. Therefore, the *PR* value contributes to the linked pages on average, where *PR* is an abbreviation of PageRank. The user behavior in the microblog social network is similar to the relationships among pages. Thus, the traditional way of evaluating user influence is to use PageRank directly, but this idea of a uniform distribution of *PR* values yields unsatisfactory results [25]. Therefore, we propose the MUI-ISIDA algorithm by improving PageRank, which overcomes its shortcomings.

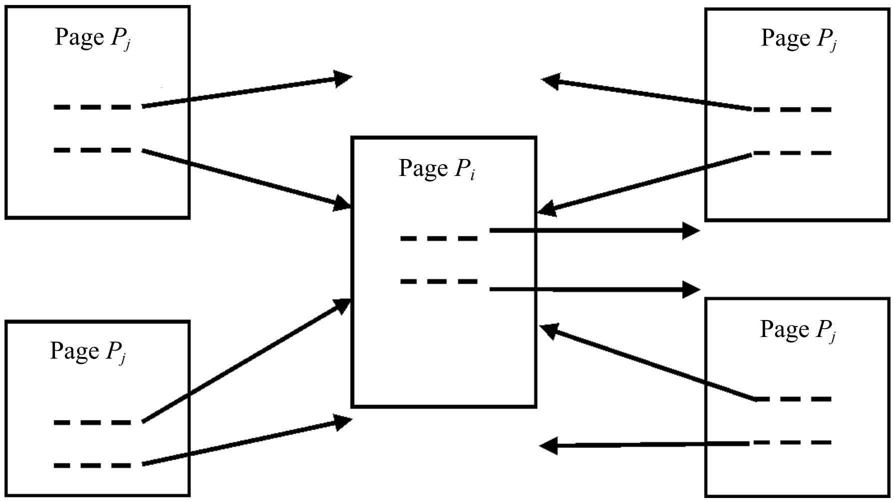


Figure 2. The relationship between p_i and p_j.

4. Main idea of MUI-ISIDA

4.1. User interest theme model

In a microblog network, it is an active task for users to publish microblogs. The 'following' relationship is an important indicator of interest similarity among users. Users follow bloggers in a certain field, mainly because they are interested in the contents of microblogs published by bloggers. Therefore, users' original microblogs can directly represent their own interests and preferences [26]. For each user, we integrate all original microblogs within a specified period into one document, and then build the user's interest theme model based on the *LDA* algorithm [27] .

4.1.1. Generation of interest themes

The purpose of interest theme generation is to identify topics of interest to each user based on the user' s original microblogs. We used the *LDA* model for topic mining. The *LDA* model is a probabilistic topic model, which is an unsupervised machine learning technology. The model contains a three-layer structure of word-topic-document, it treats each document as a 'bag of words', so each document emerges as a probability distribution over some topics. Figure 3 shows the structure of the *LDA* model.

where M represents a collection of documents. *Nd* denotes the set of words in the document, where T is the number of topics. φ and θ are the word multinomial distribution of each topic and the topic distribution of each document, and they have *Dirichlet* priors with hyper-parameters α and β, respectively. We use this model and normalize it to obtain the document-topic distribution. The result is represented by a matrix $D \times T$, which is denoted as DT.

$$DT = \begin{pmatrix} DT_{11} & DT_{12} & DT_{13} & \cdots & DT_{1j} \\ DT_{21} & DT_{22} & DT_{23} & \cdots & DT_{2j} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ DT_{k1} & DT_{k2} & DT_{k3} & \cdots & DT_{kj} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ DT_{i1} & DT_{i2} & DT_{i3} & \cdots & DT_{ij} \end{pmatrix}$$

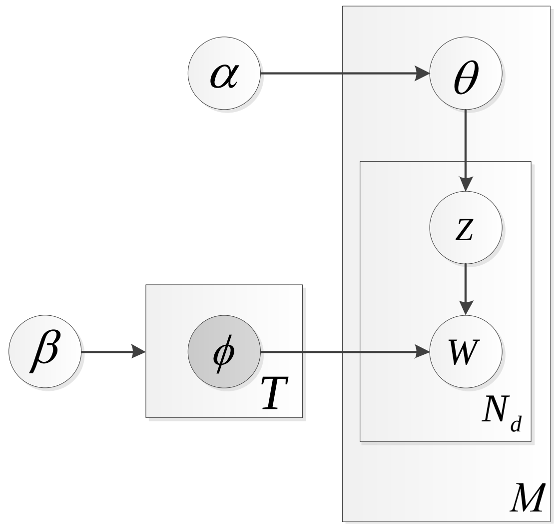


Figure 3. *LDA* model.

where D represents the number of users, and T denotes the number of topics. DT_{ij} stands for the degree to which user *i* is interested in topic *j*, that is, the topic probability distribution $\theta(i)$ of user *i*.

4.1.2. User interest similarity

Based on the document-topic distribution matrix *DT*, the interest similarities among users can be calculated by users' topic distributions through the *Pearson* correlation coefficient and is described by Eq. (2).

$$Sim < i,k >= \frac{\sum\limits_{j\in T}(DT_{ij}-\overline{DT_i})(DT_{kj}-\overline{DT_k})}{\sqrt{\sum\limits_{j\in T}\left(DT_{ij}-\overline{DT_i}\right)^2\sum\limits_{j\in T}\left(DT_{kj}-\overline{DT_k}\right)^2}} \tag{2}$$

where $Sim < i,k >$ denotes the interest similarity of user *i* and user *k*, and there is a 'following' relationship between them. $\overline{DT_i}$ represents the mean degree of interest of user *i* on these topics.

4.2. User information dissemination ability

In the microblog network, different users' microblogs spread at different rates. Some users' microblogs spread very quickly, while others spread very slowly because information dissemination is closely related to user forwarding and commenting. When a microblog is more valuable, it attracts more users' attention and it is more likely to be forwarded and commented, allowing information to spread smoothly. In addition to valuable microblog content, users' forwarding and commenting behaviors cause information to spread again among their followers, implying that the information spreads more conveniently among them. Therefore, we propose using the microblog quality coefficient to measure the quality of users' microblogs and the assimilation effect coefficient to calculate the extent to which bloggers' behaviors affect users. Finally, we integrate the two factors to evaluate the users' information dissemination ability.

4.2.1. Microblog quality coefficient

Intuitively, users often forward or comment on microblogs that they are interested in after reading them. If a microblog is forwarded or commented on more frequently, it spreads more widely in the network, implying that the user who published it has more influence. The types of forwarding are usually divided into commentary and noncommentary forwarding. Users can choose whether to share their own attitudes when they forward microblogs, and they tend to share their own opinions or participate in the discussion on the microblogs that they are keen on. The original microblog contents and their opinions are usually separated by "//@." Few users may engage 'zombie fans' to forward their microblogs to achieve certain purposes, so their microblogs are forwarded frequently. The users' interests are embodied by the behaviors of forwarding and commenting on some microblogs, which are probably related to the original microblogs. Therefore, we can integrate the user comments and opinions of forwarding and mine latent topics to find similarities to the original microblog contents. *A* higher degree of similarity implies that users participate in the interaction with the original content more effectively, and the original microblogs are of higher quality.

Therefore, we integrate all original microblogs by the same user into a document M_p and integrate these comments and opinions related to the original contents into a document M_c. The topic probability distributions of the two documents $\theta_c(k)$ and $\theta_p(k)$ are generated using the *LDA* model. The *KL* distance is used to measure the difference between two probability distributions, which can be expressed using Eq. (3).

$$D(\theta_c||\theta_p) = \sum \theta_c(k) \log \frac{\theta_c(k)}{\theta_p(k)} \tag{3}$$

KL divergence is an asymmetric measure [28]. Different results are obtained after swapping the positions of $\theta_c(k)$ and $\theta_p(k)$, namely, $D(\theta_c||\theta_p) \neq D(\theta_p||\theta_c)$. Because *KL* divergence is directional, the similarity between users is nondirectional, therefore, it is inappropriate to measure the theme similarity between users based on *KL* divergence directly. The *JS* divergence [29] is a smoother and symmetrical probability distribution measurement based on the *KL* divergence. Compared with *KL* divergence, *JS* divergence is more accurate in distinguishing similarity. Therefore, inspired by the *JS* divergence, we transform the *KL* divergence equation and use the Eq.(4) to symmetrize the *KL* divergence.

$$D_{KL}(\theta_c, \theta_p) = \frac{D(\theta_c||\theta_p) + D(\theta_p||\theta_c)}{2} \tag{4}$$

An increase in $D_{KL}(\theta_c, \theta_p)$ indicates that the similarity between M_p and M_c is decreasing. There may exist some possible invalidated forwarding activities in the user's microblogs. Consequently, the quality coefficient should be appropriately reduced. Therefore, the effective factor δ is defined using Eq. (5).

$$\delta = e^{-D_{KL}(\theta_c, \theta_p)} \tag{5}$$

The microblog quality coefficient is shown as in Eq. (6).

$$Q_k = \frac{R_k + C_k}{N_k} \times \delta \tag{6}$$

where Q_k represents the microblog quality coefficient of user k and N_k stands for the total number of microblogs published by user k, R_k and C_k denote the number of forwarded and commented of the same user, respectively. δ is the effective coefficient.

4.2.2. Assimilation effect coefficient

Users can browse the contents of bloggers' microblogs, and they tend to forward or comment on a microblog that interests them. Their interests and preferences may also be affected during the process. The greater proportion of forwarding and commenting a user has, the more easily the user is assimilated by others. Therefore, the calculation of the assimilation effect coefficient is given by Eq. (7).

$$S_k = \frac{r_k + c_k}{N_k} \tag{7}$$

where r_k, c_k and N_k indicate the number of forwarding, commenting and all microblogs for user k, respectively, and S_k represents the assimilation effect coefficient of the same user.

4.2.3. Information dissemination ability

Information is mainly disseminated through forwarding and commenting among users. User influence is the driving force for information dissemination, and information is the dissemination carrier of influence. With the spread of information, this influence continues to spread. Moreover, the larger the microblog quality coefficient and assimilation effect coefficient of users, the stronger the ability of users to control information transmission, which is conducive to information transmission in the microblog social network. Therefore, we define the information dissemination ability of user k as W_k. The expression is given by Eq. (8).

$$W_k = Q_k \times S_k \tag{8}$$

4.3. MUI-ISIDA algorithm model

The PageRank can be used to evaluate users' influences through the link relationships among users, that is the 'following' relationships in microblog network. However, it ignores user characteristics and behaviors. In the microblog social network, the relationships among users are also related to their themes of interest. The more similar the users' interest theme, the greater the mutual influence among users. Moreover, the spread of user influences is affected by their information dissemination ability. The stronger the user's ability to control information dissemination, the more conducive it is to influence dissemination. Therefore, we propose the MUI-ISIDA algorithm, which comprehensively considers the node attributes and propagation characteristics of the microblog social network. The main idea of the MUI-ISIDA algorithm is to take the user's information dissemination ability as a weight and, based on similarities among users, fairly allocate the contribution of followers to the influence of bloggers. The MUI-ISIDA value is computed using Eq. (9).

$$MUI-ISIDA(i) = \frac{1-\alpha}{N} + \alpha \sum_{j \in f(i)} MUI-ISIDA(j) \times \varphi(j,i) \times W_j \tag{9}$$

where α is the damping factor, which can be set between zero and one, usually set as 0.85 in the web page ranking algorithm [31], so we take the empirical value 0.85. W_j represents the information dissemination ability of user j, $f(i)$ is the follower set of user i. $\varphi(j,i)$ is the rate of MUI-ISIDA value that user j contributes to user i, and the value is determined by the sum of the similarities between user j and his following users. This value is expressed by Eq.(10)

$$\varphi(j,i) = \frac{sim\langle j,i\rangle}{\sum_{k=1}^{N} sim\langle j,k\rangle} (k = 1, 2, \cdots, N \quad k \in A(j)) \tag{10}$$

where $A(j)$ is the set of users followed by user j, k represents the k-th user in the set, and N denotes the total number of users followed by user j. We define the initial value of MUI-ISIDA as one, in the iterations of our algorithm, the followers contribute to the blogger, his/her MUI-ISIDA value will be updated. Therefore, the influence value of each user can be updated through the loop structure. When all users obtain the updated MUI-ISIDA values, one round of calculation is complete, and the new MUI-ISIDA values are taken as the initial values for the next round of calculation. Until each user's MUI-ISIDA value does not change significantly from the previous iteration, exit the loop. In this paper, we consider the value of the user's i-th iteration is P_0=MUI-ISIDA(i), and the value of the user's $(i+1)$th iteration is P=MUI-ISIDA$(i+1)$, iterate continuously to obtain the convergent MUI-ISIDA value when the inequality presented in Eq. (11) is satisfied:

$$|P - P_0| \leq 0.001 \tag{11}$$

The iteration ends, and the final influence values are obtained. Consequently, the MUI-ISIDA values of all users can be obtained. The proposed algorithm description of the MUI-ISIDA is given below.

MUI-ISIDA algorithm: Calculating influence based on the fusion of interest similarity and information dissemination ability

Input: Users' original microblogs M_p, the comment content and forwarding opinions M_c, the number of forwarded and commented R_k, C_k, the numbers of forwarding and commenting r_k, c_k, the number of users' microblogs N_k
Output: MUI-ISIDA value
Step 1. Initialize, get user-following matrix DT.
Step 2. Calculate the interest theme similarity between users i and k in DT using Eq.(2).
Step 3. According to Eq.(3) and Eq.(4), we calculate the differences in the topic distribution $D_{KL}(\theta_c, \theta_p)$ among users.
Step 4. Calculate the effective factor δ using the results of Step 3 and Eq.(5).
Step 5. Calculate the microblog quality coefficient Q_k according to the results obtained in Step 4 and Eq.(6).
Step 6. Calculate the assimilation effect coefficient S_k according to Eq.(7).
Step 7. Calculate the information dissemination ability W_k according to the results obtained in Step 5, Step 6 and Eq.(8).
Step 8. Initialize user' influence and define the initial value of MUI-ISIDA as one.
Step 9. According to Eqs.(9) and (10), calculate the new MUI-ISIDA values by using the results of Step 2 and Step 7 and use them as the initial MUI-ISIDA values for the next iteration.
Step 10. Repeat step 9 when the inequality in Eq.(11) is satisfied, the iteration is stopped.
Step 11. Finally, get the MUI-ISIDA influence value.

5. Experiment and analysis

In this section, we comprehensively evaluate the effectiveness of the proposed algorithm. The dataset, experimental environment, and evaluation criteria were introduced.

5.1. Dataset and experimental environment

5.1.1. Dataset

Sina microblog is a social network platform similar to Twitter. By October 2020, the number of monthly active users reached 523 million. Therefore, we selected the Sina microblog as the data source. We obtained a batch of real microblog data using a web crawler. The total number of 'following' records between users was 81,525. The statistical information is presented in Table 1

Table 1. Data set information table

Statistics	Data value
User counts	12,746
Time	2015.8
Original microblog counts	441,687
Forwarded microblog counts	66,289
Comment counts	55,784
Per capita microblog counts	39.85

Owing to the complexity and redundancy of the information, we preprocessed the dataset and selected 2000 users with more than 130 followers as dataset 1, and randomly selected 2000 users as dataset 2, of which 1428 users had less than 100 followers.

5.1.2. Experimental environment

We executed the experiments on a computer with a four-core 2.80-GHz processor, 8 GB GORE, and Windows 10 OS using Python 3.6 and Matlab2018a.

5.2. Experimental results and analysis

We compared and analyzed the experimental results with the classic PageRank algorithm and the MR-UIRank algorithm proposed in the literature [9], provide the top 15 users with the influence of the three algorithms and verify the accuracy of the algorithm. The MR-UIRank algorithm is an improvement of the traditional PageRank algorithm model. It considers the number of microblogs, forwarding, and commenting. This algorithm defines activity, microblog quality, and credibility. The weighet is defined as *W(i)*=activity(i) + quality(i) + credibility(i), and then added it to the original PageRank algorithm model to calculate the influence, truly, and effectively reflecting the actual influence of microblog users.

5.2.1. Analysis of influence ranking for the top-15 users

We first conducted experiments on dataset 1 and set the *LDA* model parameter values to T=10, α=0.5, and β =0.1, when quantifying user interest. To evaluate the validity of the ranking results, we sort the user influence, which is calculated using the traditional PageRank, MR-UIRank, and the MUI-ISIDA algorithm, and list the top 15 users in Tables 2, 3, and 4.

The three tables include user id, followers counts, the number of microblogs, microblog quality, and interaction counts. In the second and fourth columns, the values in parentheses indicate the interaction counts rankings and microblog quality rankings of these 15 users. The interaction counts include the total number of the user's microblogs forwarded and commented, and the microblog quality is defined as the ratio of the user's interaction counts to the number of microblogs. We define interaction counts and microblog counts of user k as I_k, N_k, respectively and the microblog quality is computed using Eq. (12).

$$Quality(k) = \frac{I_k}{N_k} \tag{12}$$

Table 2. Top 15 users of PageRank

User id	Followers counts	Interaction counts(ranking)	Microblog counts	Microblog quality(ranking)	PR value
2388	804	67(2)	35	1.91(4)	4.1309
2784	625	16(9)	6	2.67(3)	3.4485
6777	453	36(5)	10	3.60(1)	3.1501
3644	455	14(10)	28	0.50(12)	3.1263
5908	881	5(14)	27	0.19(15)	3.0019
8647	596	20(7)	25	0.80(9)	2.9966
2786	602	10(12)	25	0.40(13)	2.8629
1769	928	7(13)	2	3.50(2)	2.8158
4963	390	17(8)	31	0.55(11)	2.7824
258	488	95(1)	51	1.86(5)	2.7631
560	498	42(4)	28	1.50(7)	2.7631
7061	436	2(15)	10	0.20(14)	2.7577
3716	389	14(11)	11	1.27(8)	2.7547
1851	346	43(3)	28	1.53(6)	2.7546
253	366	34(6)	56	0.60(10)	2.7182

Table 3. Top 15 users of MUI-ISIDA

User id	Followers counts	Interaction counts(ranking)	Microblog counts	Microblog quality(ranking)	MUI-ISIDA value
258	488	95(3)	51	1.86(7)	6.1642
2415	267	144(1)	19	7.58(1)	5.3367
628	266	62(6)	42	1.48(9)	4.6187
2388	804	67(5)	35	1.91(5)	3.7657
8693	452	48(8)	20	2.40(4)	3.5735
560	498	42(10)	28	1.50(8)	3.3868
211	226	79(4)	24	3.29(2)	3.3493
3619	298	49(7)	17	2.88(3)	3.0384
3929	617	37(11)	28	1.32(10)	2.9471
7038	242	128(2)	68	1.88(6)	2.8498
8541	262	46(9)	40	1.15(12)	2.8362
4963	390	17(13)	31	0.55(13)	2.8312
5908	881	5(14)	27	0.19(14)	2.8245
8581	488	4(15)	22	0.18(15)	2.8191
7931	261	35(12)	28	1.25(11)	2.8052

As shown in Table 2, the calculation of influence depends a great deal on the number of followers of the users and the rankings are close to the number of followers. In microblog social network, there may exist some ‘zombie fans’ or ‘silent fans’ in users’ followers, therefore, the user’s followers counts cannot reflect his/her

Table 4. Top 15 users of MR-UIRank

User id	Followers counts	Interaction counts(ranking)	Microblog counts	Microblog quality(ranking)	MR-UIRank value
7038	242	128(1)	68	1.88(5)	7.9671
1250	302	30(10)	57	0.52(13)	7.2818
3614	520	7(13)	17	0.41(15)	6.6736
7929	727	6(15)	2	3.00(3)	5.9995
2873	276	35(8)	26	1.35(9)	5.7815
2388	804	67(3)	35	1.91(4)	5.1840
6777	453	36(7)	10	3.60(1)	4.9989
573	303	38(6)	35	1.08(10)	4.7777
253	366	34(9)	56	0.60(12)	4.6254
258	488	95(2)	51	1.86(6)	4.4751
560	498	42(5)	28	1.50(7)	4.3854
4449	325	13(12)	17	0.76(11)	4.1889
6460	317	27(11)	57	0.47(14)	3.9935
628	266	62(4)	42	1.48(8)	3.9558
1769	928	7(14)	2	3.50(2)	3.9245

real influence ranking. Take the users whose ids are '1769' and '5908' with many followers as examples, their followers did not interact with them frequently, so their PR values are not high. The contents of microblog are spread layer by layer through forwarding and commenting. In the results of the MUI-ISIDA, the top users seem relatively reasonable. More followers did not imply greater influence. Take the users whose ids are '2415' and '7038' as examples, although their followers counts are not high, but they have more interactions with other users, so they have higher influence in microblog network. That is, interaction counts are directly related to user influence, which eliminates the interference of 'zombie fans' and 'silent fans' to a certain extent.

Simultaneously, the user influence rankings calculated using the three algorithms are not positively correlated with the interaction counts and microblog quality. Although some users ranking in microblog quality are high, this does not mean that the influence values calculated using the algorithm are the same. Additionally, user interaction counts rankings and microblog quality rankings are not completely consistent. For example, the user whose id is '1250' in Table 4, although the ranking of the user's influence value is second, its interactive volume ranks tenth, and the microblog quality ranking is thirteenth. Therefore, it is necessary to further study the accuracy of algorithm sorting in terms of interaction counts and microblog quality.

5.2.2. The evaluation standard

In the microblog social network, users' forwarding and commenting are the main driving forces for information dissemination. With the spread of information, the influence of users will also be expanded. If a user's microblogs are forwarded or commented more frequently, this means that they have a higher influence. Therefore, the number of interactions can reflect the user's influence to some extent. The greater the number of interactions, the higher the user's influence ranking. Furthermore, users tend to browse microblogs with high quality and interact with them. The higher the quality of a user's microblogs, the more frequent the interactions will be attracted. Therefore, the quality of microblogs reflects the user's influence, and there is a positive correlation between them.

Therefore, we can use microblog quality, interaction counts as the standard to judge the level of users' influences. The accuracy reflects the degree of difference from the standard sorting; the higher the accuracy, the smaller the degree of difference, and we considered the hit rate as an indicator to judge the accuracy of sorting. Firstly, we find the top-*k* influence users of the corresponding algorithms in the experiment and set it as I_A, and then construct microblog quality and interaction counts top k standard sortings, and set them as I_M and I_N. When the user scale is top k, the hit rate indicates the ratio of the number of top k users correctly calculated using the target algorithm to the number of users k. Therefore, the hit rate of algorithm A in the microblog quality ranking is described as follows:

$$P_A = \frac{|I_A \cap I_M|}{|I_A|} \tag{13}$$

The hit rate of algorithm A in the ranking of the interaction counts is given by Eq.(13).

$$P_A{}' = \frac{|I_A \cap I_N|}{|I_A|} \tag{14}$$

5.2.3. Verification of the accuracy

Because the *LDA* model depends on three parameters, namely the number of topics T and *Dirichlet* hyper-parameter α and β , the different values of these parameters will have certain impacts on interest quantification; therefore, to evaluate the proposed algorithm more reasonably, in the research on the accuracy of the algorithm, we set two sets of parameter values:(1)T=10, α=0.5, and β =0.1, and (2), T=15, α=0.5, and β =0.1. We analyzed the results from the following three aspects:

(2.1)When MUI-ISIDA ($Q_k = 1$), we calculated the user's information dissemination ability, ignoring the interference of the microblog quality coefficient.

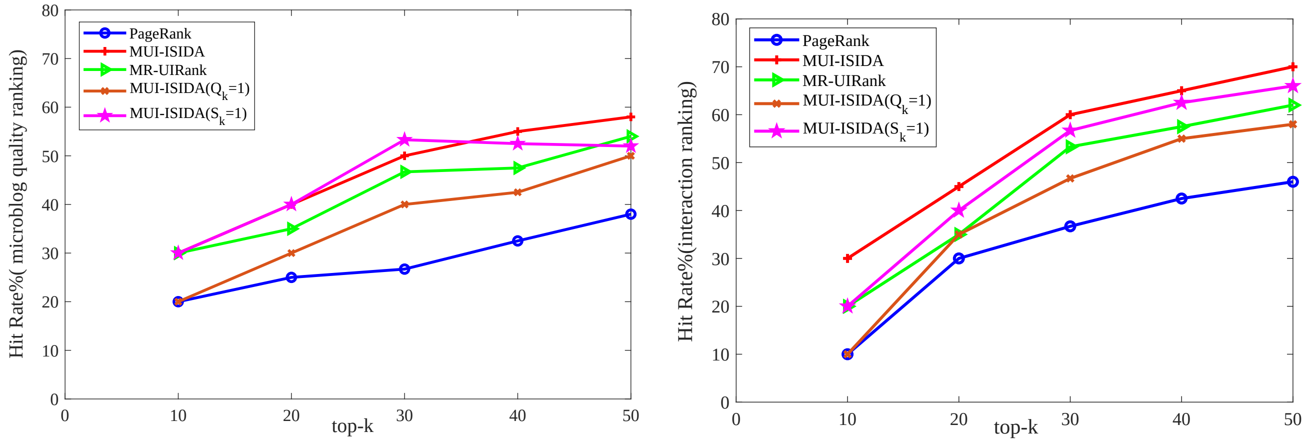
(2.2)When MUI-ISIDA ($S_k = 1$), we calculated the user's information dissemination ability and ignore the interference of the assimilation effect coefficient.

(2.3)When MUI-ISIDA (the values of Q_k and S_k are determined using our method), we calculated the user's information dissemination ability, fully consider the microblog quality and assimilation effect coefficients.

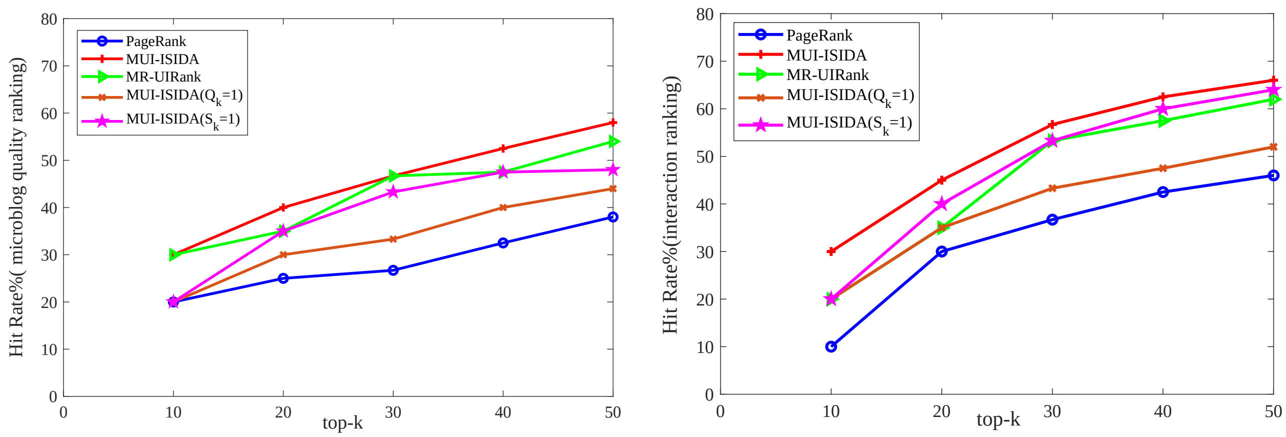
In the diagram below, the abscissa denotes the top-*k* users of the three methods, and the ordinate represents the hit rates of the corresponding results in the microblog interaction rankings and microblog quality rankings, respectively.

As shown in Figure 4, when the number of top-*k* was 30, if the number of topics T was 10, the hit rates both increased by 23.3% compared with the PageRank algorithm, and the hit rates increased by 3.4% and 6.7% compared with the MR-UIRank algorithm, respectively. If the number of topics T was 15, the hit rates both increased by 20% compared with the PageRank algorithm, and the hit rate increased by 3.4% compared with MR-UIRank in the ranking of interaction counts.

As shown in Figure 5, when the number of top-*k* was 30, if the number of topics T was 10, the hit rates increased by 16.7% and 20% compared with the PageRank algorithm, and the hit rates increased by 3.4% and 6.7% compared with the MR-UIRank algorithm, respectively. If the number of topics T was 15, the hit rates both increased by 20% compared with the PageRank algorithm, and the hit rates both increased by



$T = 10$

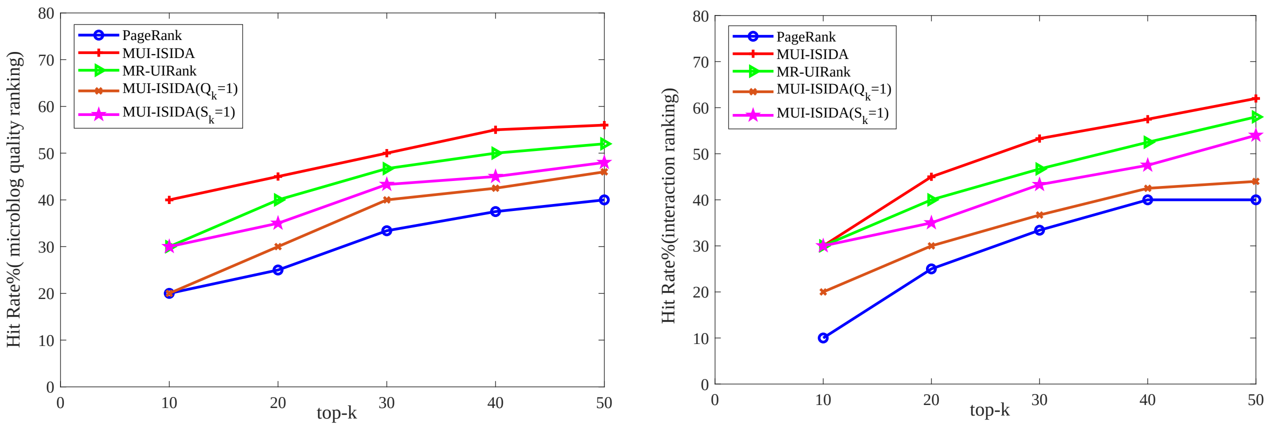


$T = 15$

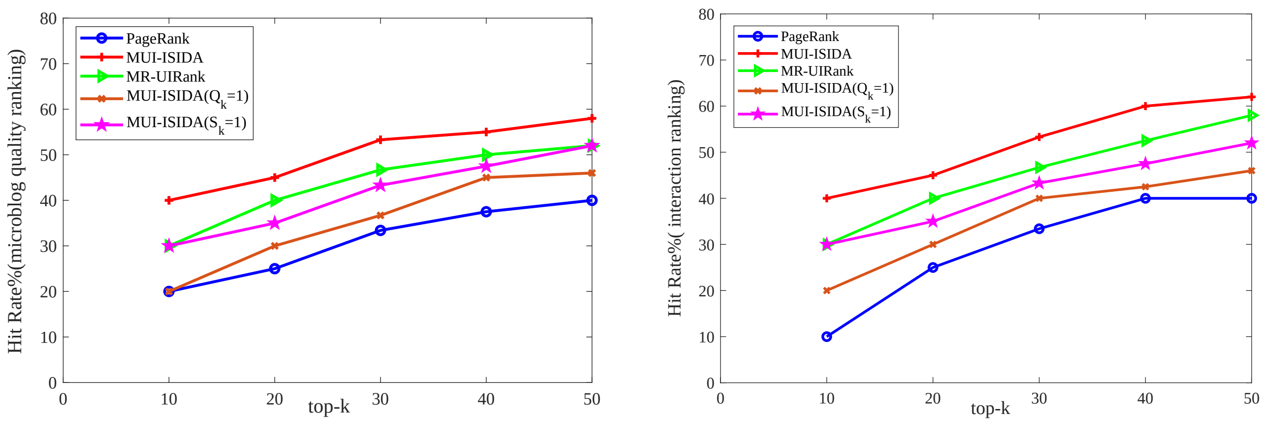
Figure 4. The hit rate of each algorithm in dataset 1

6.7% compared with MR-UIRank. By selecting different datasets for experiments, we concluded that different datasets for experiments would have different effects on the hit rates. When we selected dataset 1 with many followers, we were able to obtain higher accuracy in the rankings of interaction counts when the number of top-k was 50, if the number of topics T was 10, the hit rate reached 70%. However, if the number of topics T was 15, the advantage of the hit rate was not significant compared with MR-UIRank algorithm in microblog quality rankings. When we selected dataset 2, of which contains users with less followers the hit rates relatively reduced in the rankings of interaction counts, and did not achieve the desired result. However, compared with MR-UIRank algorithm, the hit rates got remarkable results.

The proposed algorithm considers multidimensional factors and reasonably integrates them into traditional PageRank, which overcomes the shortcomings of the uniform distribution of PR values and reduces the dependence on users' followers. We adopted the hit rate as an indicator to judge the accuracy of sortings. The higher the hit rate, the greater the accuracy. Experimental results show that, compared with the PageRank and MR-UIRank algorithms, our proposed algorithm achieves higher accuracy.



$T = 10$



$T = 15$

Figure 5. The hit rate of each algorithm in dataset 2

6. Conclusion

In social networks, it is of great practical significance to find users with high influence. In this study, we propose a new algorithm to calculate user influences, which fully integrate the user's interest theme similarity and information dissemination ability. On the one hand, we analyze users' original microblogs to capture their interests to allocate the influence fairly. On the other hand, we quantify user behaviors and verify their effectiveness. Finally, we evaluated our algorithm using the microblog dataset. According to the experimental results, our proposed algorithm achieves higher accuracy than other state of the art algorithms.

The proposed algorithm improves the effectiveness and objectivity of the user's influence calculation to a certain extent. Users' interests and preferences are considered in the influence contribution, and good experimental results have been achieved. If users' interests and preferences are considered in the calculation of information dissemination ability, better experimental results may be achieved. Furthermore, we will improve this work in future research and select more datasets for experiments.

Acknowledgement

The work was supported by grants from the Nature Science Foundation of Anhui Province in China, No. 2008085MF193 and No. 1908085MF194, the University Synergy Innovation Program of Anhui Province, No. GXXT-2019-008, the Outstanding Young Talents Program of Anhui Province, Grant Number gxyqZD2018060 the Major Science and Technology Project of Anhui Province, No. 201903a06020006. We would like to thank Editage (*www.editage.cn*) for English language editing, and we thank all the anonymous reviewers for their hard work and valuable comments.

References

[1] Riquelme F, González-Cantergiani P. Measuring user influence on Twitter: A survey. Information processing & management 2016; 52(5): 949-975. doi: 10.1016/j.ipm.2016.04.003

[2] CNNIC. The 46th Statistical Report on the Development of China's Internet. Beijing, China: Internet Network Information Center, 2020.

[3] Liu Q, Xiang B , Yuan NJ , Chen E , Xiong H et al. An influence propagation view of pagerank. ACM Transactions on Knowledge Discovery from Data 2017; 11(3): 1-30. doi: 10.1145/3046941

[4] Sun L, Liu Y, Bartolacci MR, Ting IH. A multi information dissemination model considering the interference of derivative information. Physica A: Statistical Mechanics and its Applications 2016; 451: 541-548. doi: 10.1016/j.physa.2016.01.094

[5] Wen Z, Kveton B, Valko M, Vaswani S. Online influence maximization under independent cascade model with semi-bandit feedback. In Advances in neural information processing systems; 2017. pp. 3022–3032.

[6] PAGE L, BRIN S, MOTWANI R, Winograd T. The pagerank citation ranking: Bringing order to the Web. Stanford InfoLab 1998; 9(1): 1-14. doi: 10.1007/978–3–319–08789–4_10

[7] Weng J, Lim E P, Jiang J, He Q. Twitterrank: Finding Topic-Sensitive Influential Twitterers. In: Proceedings of the Third International Conference on Web Search and Web Data Mining. WSDM, New York, USA; 2010, pp. 261–270.

[8] Jianqiang Z, Xiaolin G, Feng T. A New Method of Identifying Influential Users in the Micro-Blog Networks. IEEE Access 2017; 5: 3008–3015. doi: 10.1109/ACCESS.2017.2672680

[9] SUN H, ZUO T, Research and Realization of Influence Optimization in Cloud Computing Environment. Journal of Chinese Computer Systems 2018; 39(1): 42–47 (in Chinese).

[10] Wu L, Yang B, Jian M, Sang J, Zhang D et al. MPPR: Multi Perspective Page Rank for User Influence Estimation. In: IEEE International Conference on Big Data & Smart Computing. IEEE Computer Society; Shanghai, China; 2018. pp. 25–29. doi: 10.1109/BigComp.2018.00013

[11] Qi C, Chen H, Yu H. Method of evaluating micro-blog users' influence based on comprehensive analysis of user behavior. Application Research of Computers 2014; 31(7): 2004-2007 (in Chinese).

[12] Lin J, Wu Y, Wang L, Ai W, Zeng Y. A user influence rank algorithm based on interaction behaviors in cyber group events. In: International Conference on Cyber–Enabled Distributed Computing and Knowledge Discovery; Guilin, China; 2019. pp. 151-157. doi: 10.1109/CyberC.2019.00035

[13] Zhou J, Zhang Y, Wang B, Yan Y. Predicting user influence in microblogs. In: First IEEE International Conference on Computer Communication and the Internet; Wuhan, China; 2016. pp. 292–295. doi: 10.1109/CCI.2016.7778928

[14] Huang M, Zou G, Zhang B, Gan Y, Jiang S et al. Identifying influential individuals in microblogging networks using graph partitioning. Expert Systems with Applications 2018; 102: 70-82. doi: 10.1016/j.eswa.2018.02.021

[15] Cha M, Haddadi H, Benevenuto F, Gummadi KP. Measuring user influence in twitter: The million follower fallacy. In: 4th International AAAI Conference on Weblogs and Social Media; Washington D.C, USA; 2010. pp. 10-17.

[16] Zareie A, Sheikhahmadi A, Jalili M, Fasaei MSK. Finding influential nodes in social networks based on neighborhood correlation coefficient. Knowledge-Based Systems 2020; 194:105580. doi: 10.1016/j.knosys.2020.105580

[17] Zareie A, Sheikhahmadi A, Jalili M. Influential node ranking in social networks based on neighborhood diversity. Future generation computer systems 2019; 94: 120-129. doi: 10.1016/j.future.2018.11.023

[18] Zareie A, Sheikhahmadi A, Jalili M. Identification of influential users in social networks based on users' interest. Information Sciences 2019; 493: 217-231. doi: 10.1016/j.ins.2019.04.033

[19] Xing W, Ghorbani A. Weighted PageRank algorithm Proceedings. In: Second Annual Conference on Communication Networks and Services Research; Fredericton, NB, Canada; 2004.pp.305-314. doi: 10.1109/DNSR.2004.1344743

[20] Ding Y. Applying weighted PageRank to author citation networks. Journal of the American Society for Information Science and Technology 2011; 62(2): 236-245. doi: 10.1002/asi.21452

[21] Ding Y. Topic-based PageRank on author cocitation networks. Journal of the American Society for Information Science and Technology 2011; 62(3): 449-466. doi: 10.1002/asi.21467

[22] Nemirovsky D, Avrachenkov K. Weighted PageRank: Cluster-Related Weights. In: Proceedings of The Seventeenth Text REtrieval Conference; Maryland, USA; 2008.

[23] Jianqiang Z, Xiaolin G , Feng T. A New Method of Identifying Influential Users in the Micro-Blog Networks. IEEE Access 2017; 5: 3008-3015. doi: 10.1109/ACCESS.2017.2672680

[24] Huang L J, Xiong YM. Evaluation of Microblog Users' Influence Based on PageRank and Users Behavior Analysis. Advances in Internet of Things 2013; 3(2): 34-40. doi: 10.4236/ait.2013.32005

[25] Xiang B, Qi L, Chen E, Xiong H, Zheng Y et al. PageRank with priors: An influence propagation perspective. In: Proceedings of the Twenty-Third international joint conference on Artificial Intelligence; Menlo Park, USA; 2013. pp. 2740-2746.

[26] Bingyu L, Cuirong W, Cong W, Junwei W, Xinwei W et al. Microblog community discovery algorithm based on dynamic topic model with multidimensional data fusion. Journal of Software 2017; 28(2): 246-261 (in Chinese). doi: 10.13328/j.cnki.jos.005116

[27] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research 2003; 3: 993-1022. doi: 10.1162/ jmlr.2003.3.4-5.993

[28] Cheng S, Wang W. Rating prediction algorithm based on user time-sensitivity. Information 2020; 11(1): 4. doi: 10.3390/info11010004.

[29] HV Nguyen, J Vreeken. Non-parametric jensen-shannon divergence. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases; Springer,Cham; 2015. pp. 173-189.

[30] Huang M, Zhang B, Zou G, Gu C, Wu P et al. Socialized User Modeling in Microblogging Scenarios for Interest Prediction. In: IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress; Auckland, New Zealand; 2016. pp. 124-131. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.39.

[31] Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 1998; 30(1-7): 107-117. doi: 10.1016/S0169-7552(98)00110-X