Dynamic evaluation method on dissemination capability of microblog users based on topic segmentation

Kai Li[a], Hengmin Zhu[a,*], Yihan Zhang[a], Jing Wei[a]

[a] School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210003, People's Republic of China.

Highlights

- Different topics were considered when constructing the user interaction network.
- We studied the influence of neighbor nodes and time decay on information spreading.
- PageRank algorithm was improved to proportionally allocate *PR* value according to the relationship strength.
- SIR model was used to testify the ranking result of user dissemination capability.

Abstract

Dissemination capability refers to the ability of users facilitating information spreading in the network. It is a hot issue in the field of information spreading to evaluate users' dissemination capability. Previous studies hardly evaluated user dissemination capability from the perspective of information spreading, and did not consider the impact of neighbor nodes and time decay on user dissemination capability. By constructing the user interaction network of Sina Microblog, we achieved a dynamic evaluation of user dissemination capability. Firstly, based on the thought of topic segmentation, the relationship strength between users under different topics was calculated respectively according to their historical interaction records. Secondly, we analyzed the subsequent forwarding behaviors of neighbor nodes and the time decay effect on user dissemination capability, and proposed a dynamic evaluation model of user dissemination capability. Then, based on the improved PageRank algorithm, the ranking of user dissemination capability under different topics was calculated. Finally, experiments were conducted to verify the dynamic evaluation method on user dissemination capability. The results show that it can identify users with high dissemination capability under different topics, which might be helpful to intervene the information spreading.

Keywords

Online social networks; Dissemination capability evaluation; Topic segmentation; Time decay; Neighbor nodes

* Corresponding Author
E-mail address: hengminzhu@163.com (H. Zhu).

1. Introduction

Advances in Internet technology have led to the mushrooming of social networking platforms. Users can communicate easily through social networking platforms such as Sina Microblog, Twitter and Facebook, online social networks have gradually become the main forum for information spreading and public opinion fermentation. Identifying users with great influence in information dissemination from online social networks is of great significance for hot topic monitoring and network marketing.

At present, most of the related work measure the user influence from three aspects: network topology, user behavior and interaction content. Different from user influence in social platforms, which refers to the ability of a user's behavior causing changes other users' behavior, user dissemination capability refers to the ability of users boosting the spreading of information in the network. In recent years, there have been few studies on the evaluation of user dissemination capability, especially considering the influence of user interest topics, neighbor nodes and time decay on user dissemination capability.

In this paper, we proposed a dynamic evaluation model of user dissemination capability. Firstly, the user dissemination network was constructed based on the historical interaction records among users. Topics of the interaction records were identified by LDA (Latent Dirichlet Allocation) model, and the strength of user relationships under different topics was calculated, which can describe users' interests. Then, considering the subsequent forwarding behaviors of neighbor nodes and the time decay effect on users' interests, the PageRank algorithm was improved to realize the dynamic evaluation of users' information dissemination capability under different topics. Finally, the effectiveness of the method was experimentally verified.

This study makes several contributions in the field of information diffusion. First, we constructed a user interaction relationship network based on the thought of topic segmentation which can describe users' different interests. Second, considering the role of neighbor nodes and time decay, we proposed a dynamic evaluation model of user dissemination capability, which can be used to effectively identify users who can boost information spreading in social networks.

The remainder of the paper is organized as follows: In the next section we review related literature. In Section 3, we identify the topics of interaction records, and calculate the strength of user relationships to construct the user interaction relationship network. Section 4 discusses the influence of neighbor nodes and time decay on user dissemination capability, and constructs a dynamic evaluation model of user dissemination capability. Section 5 analyzes the experimental results and verifies the effectiveness of the evaluation model by means of the SIR model.

2. Related work

2.1 Node evaluation in network topology

There are many existing methods for evaluating the influence of nodes, and one of the most basic measurement indicators is degree centrality [1]. It believes that the more neighbors a node has, the greater its influence is. However, degree centrality does not take into account the location of nodes in the network, and this evaluation method of node influence based on local neighbor information is obviously inaccurate [2]. Considering more node domain information, Chen et al. proposed a semi-local centrality evaluation method, which uses multi-order neighbor nodes to evaluate the influence of nodes and is more accurate than degree centrality [3].

There are also some methods to evaluate the influence of nodes from a global perspective, such as betweenness centrality [4-6] and closeness centrality [7]. Although these two methods can measure the importance of nodes, they are not suitable for large-scale networks due to the complexity of calculation. In addition, Kitsak et al. proposed the K-shell decomposition method for node influence evaluation of topology, which clearly pointed out that the most influential nodes are not those with the maximum degree, and they think that the core node with large K-shell center value has great influence [8].

Heidemann et al. used PageRank algorithm to calculate the centrality of nodes and identify the key nodes of significant influence [9]. But PageRank algorithm also has some defects [10]: not excluding zombie fans, ignoring the relevance of the topic, the average distribution of the PR value of the node and so on. To address these problems, some scholars have improved the PageRank algorithm in terms of activity [11], specific topics [12] and information flow [13], optimizing the evaluation of node influence. It can be seen that most of the research on node evaluation was conducted on the basis of PageRank algorithm.

2.2 Evaluation of network nodes in information spreading

Some scholars study the influence of nodes from the perspective of information diffusion.

In terms of user behavior, by analyzing the data generated by user interaction (publishing information, commenting, forwarding, establishing friendships, etc.), some scholars established user relationship network [14], studied the path and scope of information diffusion [15], and evaluated user's information dissemination capability [16,17], so as to guide and control information diffusion scientifically. Yang et al. studied the forwarding behavior of Twitter users and found that it is affected by factors such as user influence, time and content [18]. Tan et al. used dynamically changing interaction data to measure user influence [19]. Most scholars have studied user influence from the perspective of user behaviors. However, such studies failed to consider the influence of the neighbor nodes' forwarding behaviors on dissemination capability. Besides that, Ji et al. adjusted the influence weight of neighbor nodes on the target node for measuring the node's dissemination capability [20], most scholars failed to consider the subsequent information spreading behaviors of neighbor nodes, which needs further research.

In terms of user interaction content, due to the large number of people in the social network, the interaction content between users is extremely abundant. By analyzing them, the topics and characteristic words of interactive content can be extracted and the user's interest preferences can be analyzed. Varshney et al. used the LDA (Latent Dirichlet Allocation) topic model to mine the text topics to obtain Twitter users' interests [21]. Ma et al. also used LDA to mine user interest preferences in the comments of products [22]. Zheng et al. judged the dynamic trend of user interests based on the LDA model combined with the timeliness and interactivity of posts [23]. It can be seen that LDA topic models are extensively used to mine users' interest preferences. However, previous studies on user influence have mostly focused on network structure and interaction behavior. Tang et al. proposed TAP (Topical Affinity Propagation) model based on topic similarity to measure user influence under different topic distributions [24]. Jiang et al. proposed an influence evaluation method for a specific topic [25]. Therefore, the thought of topic segmentation should be introduced into the field of dissemination capability evaluation, so as to more accurately evaluate users' dissemination capability under different topics.

2.3 Dynamic evaluation of nodes

The information released by users is timeliness, so the users' dissemination capability is dynamically changing [26,27]. However, few scholars consider the effect of time decay on user dissemination capability. Chun et al. and Liu et al. considered the dynamics of interaction behavior, and analyzed the change of user interests over time [28,29]. The above studies analyzed user interests taking time decay into account, while Behzad et al. provided a time-sensitive ranking to identify the most influential users in the Twitter network, dynamically investigating the influence of users [30]. Therefore, the effect of time decay needs to be considered when calculating users' dissemination capability to capture the dynamic changes in user interests.

Table 1 summarizes the studies related to the evaluation of user dissemination capability in social networks in recent years. Previous studies evaluated user influence from the perspective of network topology or information spreading, which mainly analyzed the historical posts of a user or the interaction records between users. However, few studies comprehensively consider the effect of topic segmentation, neighbor nodes, and time decay on dissemination capability. In fact, in the process of information spreading, users' interests and interaction intensity have been changing over time. The forwarding behaviors of neighbor nodes boost information spreading in the network. Therefore, we analyzed the dissemination capability of users under different topics, and studied the influence of neighbor nodes' subsequent forwarding and time decay. Finally, the users' information dissemination capability was evaluated based on the improved PageRank algorithm.

Table 1. Comparison of previous studies on user dissemination capability. The letter Y and N represent 'with and without the research point' respectively.

References	Research perspective	Methods	Topic segmentation	Neighbor nodes	Time decay
Heideman et al. [9]	Network topology	PageRank	N	N	N
Tan et al. [19]	Information spreading	Historical interaction analysis	N	N	Y
Ji et al. [20]	Network topology	Network structure evaluation	N	Y	N
Tang et al. [24]	Information spreading	Historical posts analysis	Y	N	N
Behzad et al. [30]	Information spreading	Historical posts analysis	N	N	Y
Current study	Information spreading	Improved PageRank Historical interaction analysis	Y	Y	Y

3. Construction of a multi-topic interaction relationship network

The topic preferences of users influence their information dissemination behaviors, and posts on interest topics are more prone to be forwarded by users. However, users' interest preferences in social networks are different, so it is extremely essential to study the information dissemination capability of users under different topics. In this section, we performed topic segmentation on the interaction records and calculated the strength of users' relationships under different topics, then a multi-topic interaction relationship network was constructed.

3.1 Processing of the interaction records between users

We crawled a sub-network of following relationships containing 645 users from Sina Microblog and obtained historical interaction records of these users in the three months from Seb. 1st, 2017 to Nov. 31st.

In this paper, we deleted the interaction records outside the following relationship network and got the user forwarding content based on the following relationship. 17509 interaction records were obtained, and each record can be represented as a one-dimensional vector S, S = (user, parent_user, content, time, Id, csd), where 'user' is the forwarder, 'parent_user' is the forwarded user, 'content' is the forwarded content, 'time' is the forwarding time, 'Id' is the number of the crawled posts, and 'csd' is the cascade of information diffusion.

3.2 Topic identification

First, we used the jieba toolkit in python to segment the interaction content and remove the stop words. Then the LDA of Python toolkit sklearn library was used for topic identification. In this paper, we introduced the perplexity to determine the optimal number of topics as 12. Table 2 lists the top 10 feature words in weight under the 12 topics. As can be seen from the table, tp1, tp2 and tp7 are related to emotion and marriage topics; tp3 is related to microblog female topics; tp4 is related to news; tp5 is related to entertainment and variety shows; tp6 is related to daily life and funny topics; tp8 is related to parenting; tp9 is related to studying abroad; tp10 is related to movies; tp11 is related to gaming; tp12 is related to stock investment.

Table 2. Topic and its weight words.

Topic	10 words with the highest weight	Topic	10 words with the highest weight
tp1	like; world; sentiment; gentle; boy; daily life with my husband; beauty; emotional research; maturation; honest	tp7	confession; sensation; the first time; boy; feelings; romantic relationship; single; break up; choice; lovers
tp2	love; girl; forever; boyfriend; expression; girlfriend; dating; youth; marriage； self-preservation	tp8	children; life; mother; baby; learn; emotion; school; suggestion; mood; parents
tp3	wish; girl; story; effort; garments; lifetime; relinquish; sadness; excitement; classmate	tp9	matter; success; Japan; students; school; mood; experience; multi-years; university; further study abroad
tp4	net friends; Beijing; photos; scene; kindergarten; recently; terribleness; RYB.US; news	tp10	film; recommendation; director; publish; future; movie trailer; growing-up; starring role; movie release; optimal
tp5	cute; Taiwan; actor; marriage; abroad; island country; pictures; show; 'Lu Han'; variety show	tp11	China; 2017; hero; tears; live streaming; competition; League of Legends; commentary; S7
tp6	share; daily; happy; star; story; feeling; fashion; music; funny; interesting	tp12	Hong Kong stocks; nationwide; today; information; yesterday; investment; opportunity; experience; correct; operation

By means of topic segmentation, all interaction records were expressed as a multidimensional vector: $v = (t_1, t_2, ..., t_n)$, where n is the total number of topics, $n = 12$; t_k represents the weight of records belonging to topic T_k, and the sum of the weights under all topics is 1.

3.3 Calculation of user relationship strength

In order to identify users who contributed a lot to the spread of information, we first calculated the relationship strength between users, because a great relationship strength usually mean there exist frequent interaction in the network. The relationship strength of users depends on the specific topic, so it can be expressed as a one-dimensional vector. Let user i be one of followers of user j, the relationship strength of the user pair $< u_i, u_j >$, ST_{ij} , can be denoted as $(S_{ij}^{T_1}, S_{ij}^{T_2}, \cdots, S_{ij}^{T_n})$, where T_k represents the kth segmented topic, and $S_{\mathrm{ij}}^{T_k}$ denotes strength component of a user pair $<u_i, u_j>$ under topic T_k, which is calculated as follows:

$$S_{ij}^{T_k} = \frac{\sum_{m=1}^{c_j} t_k^{m,j} \cdot f_{ij}^m}{\sum_{m=1}^{c_j} t_k^{m,j}} \tag{1}$$

Where, c_j represents the total number of posts published by user j, $t_k^{m,j}$ represents the component of the m-th post published by user j under topic k, and f_{ij}^m is used to determine whether user i forwarded the m-th post of user j. if forwarded, then $f_{ij}^m = 1$; otherwise, 0.

3.4 Visualization of interaction relationship network

We used Gephi to visualize interaction relationship network, as shown in Figure 1. There are 645 nodes and 2224 edges. Nodes in the network have following relationships but little or no interactions occur among them, so this network is sparse.

Fig. 1. Interaction relationship network of Microblog user. The network has 17509 interaction records.

4. The Model

In this section, we first analyzed the influence of neighboring nodes on user dissemination capability, and adjusted user relationship strength by counting the number of subsequent forwarded posts. Then, the correlation analysis of user interaction frequency explains that user relationship strength decays over time. Finally, considering the problem of *PR* value being averagely allocated according to the connectivity of nodes by PageRank algorithm, we proposed an improved PageRank which proportionally allocates *PR* value referring to the decayed relationship strength, and calculated the users' dissemination capability.

4.1 Influence of neighbor nodes on user dissemination capability

User dissemination capability refers to the ability of users facilitating information spreading in the social network. For example, both user A and B forward a post under a specific topic. The neighbor nodes of user A forward this post 10 times, while the neighbor nodes of user B forward this post 1000 times. Because of the large number of subsequent forwards resulting from user B, this post spreads more widely in the network. Therefore, the dissemination capability of user B is greater than that of user A.

In this paper, we counted the number of subsequent forwarding and the topic weight of each post, and adjusted the relationship strength $S_{\mathrm{ij}}^{T_k}$ calculated by Eq. (1). The more subsequent forwarding times resulting from neighbor nodes, the greater the relationship strength between users will be. Meanwhile, the weight of posts determines the proportion of increase of relationship strength between users. Therefore, the formula is as follows:

$$S'^{T_k}_{ij} = S^{T_k}_{ij} \times (1 + \sum_{l \in M} C_l \times t_k^l) \qquad (2)$$

Where, $S'^{T_k}_{ij}$ represents the adjusted strength component of a user pair $<u_i, u_j>$ under topic T_k. M is the set of all posts forwarded between the user pair $<u_i, u_j>$, and C_l is the sum of forwarded times of the *l*-th post resulting from the neighbor nodes of user j. t_k^l is the weight of the *l*-th post under topic *k*.

4.2 Influence of time decay on user dissemination capability

In a period of time, user A forwards the post of user B under a specific topic many times, and user A and B interacted frequently during this period, which means that there might be a greater relationship strength between these two users. However, it will gradually decay over time, which is called the time decay effect of user relationship strength.

Relationship strength between users can be described by interaction frequency. In order to testify that user relationship strength decays over time, we divided 17509 interaction records into 30 segments according to their interaction time, each spanning 3 days. First, we calculated the interaction frequency of all user pairs in different time intervals, which is expressed as 30 2224-dimensional interaction frequency vectors. Then we used the inner product to calculate the similarity between any two of these 30 vectors, which reflects their correlation. Finally, the three-dimensional graph of the interaction frequency similarity was plotted.

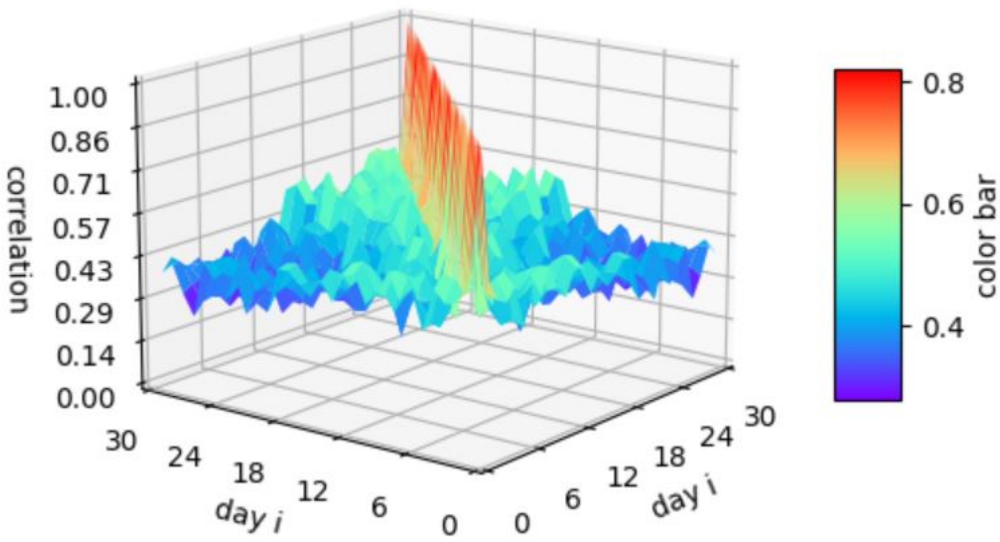

Fig. 2. Similarity of user interaction frequency vector. X, Y axis is time interval; Z axis is the correlation between any two of 30 vectors.

As can be seen from the figure, the interaction frequency on the diagonal occurs at the same time interval, so their interaction similarity is 1. The closer the two interaction frequency vectors are to the diagonal, the larger their similarity is, and the stronger the correlation of interaction frequency is. On the contrary, the farther away from the diagonal region which means the time span of two vectors is larger, the weaker the correlation between them. Therefore, the user's interaction strength will decay over time.

Considering the influence of time decay on user dissemination capability, we divided the user's three-month interaction records according to time, and calculated user relationship strength in each period of time. Then, we use the formula of Newton's cooling law to decay the relationship strength. The formula is as follows:

$$N(t) = N_0 e^{-\alpha t} \tag{3}$$

Where, t is the time interval, $N(t)$ is the value of N at time interval t, and N_0 is the initial value of N at time 0. α is the decay coefficient, and it decides the decay rate.

At last, the user relationship strength in three months was calculated according to the weight of each period of time. Suppose weight coefficient $\beta_1, \beta_2, ..., \beta_n$ are the weight of user relationship strength calculated in n period of time respectively, then $\beta_1+\beta_2+\cdots+\beta_n=1$. The user relationship strength in this whole period of time is the sum of the product of the user relationship strength and the corresponding weight in each period of time.

4.3 Dynamic evaluation process of user dissemination capability

In social networks, some users have an important role in promoting information diffusion, which is often evaluated by PageRank algorithm. In the PageRank algorithm, each user has an initial *PR* value that usually flows to the users with large dissemination capability. Considering that active users should be allocated a great PR value in information diffusion, therefore, we constructed a 645 × 645 matrix with reference to users' relationship strength which proportionally allocated *PR* value according to the relationship strength.

In a network with N users, the process of evaluating dissemination capability of each user is as follows:

(1) The user's interaction records are divided into n segments of t_1, t_2, ..., t_n according to the interaction time.

(2) The n segments are processed according to the following steps. First, topic segmentation is performed on interaction records among users, and relationship strength between users under k topics, $S_{ij}^{T_k}$, is calculated according to Eq. (1). Then, the number of subsequent forwarding of each post is counted and the relationship strength is adjusted according to Eq. (2) as $S'^{T_k}_{ij}$. Finally, the adjusted relationship strength is decayed to $S''^{T_k}_{\mathrm{ij}}$ according to Eq. (3).

(3) After the calculation of n-segment data, the relationship strength in each time period is obtained. The relationship strength between users under each topic in this whole period is finally calculated according to the weight of each time period, and the formula is as follows:

$$\tilde{S}''^{T_k}_{ij} = \beta_1 \times S''^{T_k}_{ij,t_1} + \beta_2 \times S''^{T_k}_{ij,t_2} + \cdots + \beta_n \times S''^{T_k}_{ij,t_n} \qquad (4)$$

(4) Constructing the matrix by referring to the final calculated relationship strength between users, and the user dissemination capability under k topics is calculated given by the following formula:

$$P_{n+1}^{k} = AP_n^k + \frac{1-\alpha}{N} e^T \qquad (5)$$

Where, P_n^k represents the *PR* value of the user after n iterations under topic k, and *PR* value represents the user dissemination capability. N is the number of all users in the network, and α is a constant, which is usually taken as 0.85. S is a relationship strength matrix for allocating users' *PR* values, and $A = \alpha \times S$.

5. Experiments and results

In our experiment, we used 17509 interaction records of 645 users from September to November 2017. First, we divided the records into three segments by month, and performed topic segmentation for the records of September, October and November respectively. After calculating the relationship strength among users, we added up the times of the subsequent forwarding of posts in each month, and then adjusted the relationship strength. Finally, we decayed it of each month, and calculated the relationship strength in the whole period of three months according to the weight of each month. Based on the improved PageRank algorithm, the user dissemination capability was calculated.

The time decay coefficient α was set to 0.15, and the relationship strength weights for September, October, and November were set to 0.25, 0.35, and 0.4.

5.1 Comparison of user dissemination capability under multi-topics

Based on the constructed multi-topic interaction relationship network, we obtained the ranking of user dissemination capability under twelve topics, and part of which is shown in table 3:

Table 3. Ranking of dissemination capability of some users under six topics.

User number	tp1	tp3	tp6	tp7	tp8	tp10
241	1	14	4	12	9	7
865	3	9	7	1	6	10
1396	7	5	10	20	9	5
1023	8	28	13	6	19	25
485	10	6	9	14	5	8
35	18	10	27	10	14	22
1479	22	68	25	70	38	28
1399	30	113	36	83	93	63
910	60	46	16	35	32	40
534	72	11	28	22	24	13

The table shows that there are significant differences in the ranking of most users' dissemination capability under different topics. For example, user 241 is ranked first under topic 1, but it is ranked lower under other topics. It is because that users in social platforms are willing to forward topics that they are interested in. Therefore, it is necessary to segment the topics of interaction record and calculate the dissemination capability of users under different topics.

5.2 Effect of neighbor nodes

The number of subsequent forwarding of neighbor nodes directly affects the information spreading in the social network. In order to study the influence of neighbor nodes on user dissemination capability, we adjusted user relationship strength based on the subsequent forwarding times of posts, and calculated the user dissemination capability ranking (Ranking A). 10 users were selected to compare the results of Ranking A and PageRank ranking (Ranking 1) under topic 1:

Table 4. Comparison of the ranking of user dissemination capability under topic 1.

User number	Ranking 1	Ranking A	Number of subsequent forwarding
268	3	4	0
1396	9	2	117
1390	12	15	22
534	16	33	19
594	20	10	65

48	29	38	0
108	35	39	3
877	37	12	153
872	57	14	82
931	101	60	114

Table 4 shows the differences in rankings for the same user under different methods. Compared to PageRank ranking, the ranking of most users under Ranking A rises. For example, users numbered 1396, 594, 877, and 872 all have risen significantly under Ranking A. That's due to the larger number of subsequent forwarding, which results in a wider spread of these posts in the network. Hence, the dissemination capability of those users is greater.

5.3 Effect of time decay

In this paper, we calculated the user relationship strength in the whole period of time and obtained the ranking of user dissemination capability considering time decay (Ranking B) according to improved PageRank algorithm. 10 users were selected to compare the results of Ranking B and Ranking 1 under topic 1:

Table 5. Comparison of the ranking of user dissemination capability under topic 1.

User number	Ranking 1	Ranking B
865	1	2
485	7	10
241	8	1
1396	9	7
1479	13	22
534	16	64
594	20	13
1277	30	176
1399	50	30
872	57	85

As shown in Table 5, the ranking of some users in Ranking B rises while that of others falls, compared to the result of Ranking 1. Take users 241 and 1277 for example, user 241 rises in the dissemination capability ranking because he is more active and generates more interaction with other users in recent time period. Whereas, user 1277 ranks lower as a result of that he interacting with other users in earlier time period, and the interaction strength is weakened by time decay, so the dissemination capability of users became smaller.

5.4 Comparative analysis of ranking

We compared the in-degree ranking of nodes (Ranking 0), the PageRank ranking (Ranking 1), the ranking of user dissemination capability considering neighbor nodes (Ranking 2), and the ranking of user dissemination capability comprehensively considering time decay and neighbor nodes (Ranking 3), as shown in Table 6:

Table 6. Comparison of different user ranking under topic 1.

User name number	Ranking 0	Ranking 1	Ranking 2	Ranking 3
865	5	1	1	3
1449	20	4	5	10
241	5	8	9	1
1396	1	9	2	7
1479	30	13	16	22
594	15	20	10	13
138	26	22	23	15
877	4	37	12	36
872	12	57	14	85
518	24	87	93	19

Table 6 shows that the ranking of the same user is different under topic 1. For example, the in-degree ranking of user 1396 is first, and his PageRank ranking is ninth. Taking the subsequent forwarding times of neighbor nodes into account, the ranking of dissemination capability rises to second, but drops to seventh when the time decay is further considered.

In order to compare the effectiveness of Ranking 1, 2 and 3, we used the SIR epidemic model to conduct the simulation experiment of information diffusion on different ranking results. As shown in Fig. 3, nodes in the SIR model can be divided into three states, susceptible, infectious and recovered state, which is denoted as S, I, R respectively. The susceptible state means a node has never received any information. The node who has received the information and is spreading it is in the infectious state, and the node who has received the information but is no longer interested in spreading it is in the recovered state.

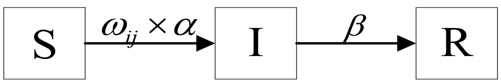

Fig. 3. State Change Diagram of SIR Model. α is the infection probability and β is the recovery probability. In SIR model, α and β are usually fixed. However, active users are more likely to forward posts in information diffusion. Therefore, the infection probability should take nodes' activity degree into account. When a node is spreading the information in the network, the nodes in state S who forward the information may turn into state I with the proportion of $\omega_{ij} \times \alpha$, in which ω_{ij} is the activity degree factor between users i and j. The nodes in state I who are no longer forward the information will change to state R with the proportion of β.

In this paper, we used users' historical interaction of a period to calculate the activity degree. The calculation formula is as follows:

$$\omega_{ij} = \frac{n \times r_{ij}^{T}}{\sum_{k=1}^{n} r_{kj}^{T}} \tag{6}$$

Where, ω_{ij} is the activity degree factor between a user pair <u_i, u_j>, which can be described by the frequency of user i forwarding the posts from user j. n is the total number of fan nodes of user j, and r_{ij}^{T} is the number of posts that user i forwards user j under topic T.

We chose the top 5 nodes of the three rankings under topic 1 respectively as the propagation sources. α is set at 0.5. There are common users in the top 5 of three rankings, but they are ranked differently. In order to make the users with higher ranking play a greater role in the process of information spreading, we set different immune probability β according to the ranking. Therefore, we set the β of the top 5 selected nodes in descending order of ranking as 0.1, 0.2, 0.3, 0.4, 0.5, and the β of other nodes is 0.6. The coverage is used to measure the effect of information spreading, which is the ratio of nodes with the state of I and R. The experimental results are shown in Fig. 4.

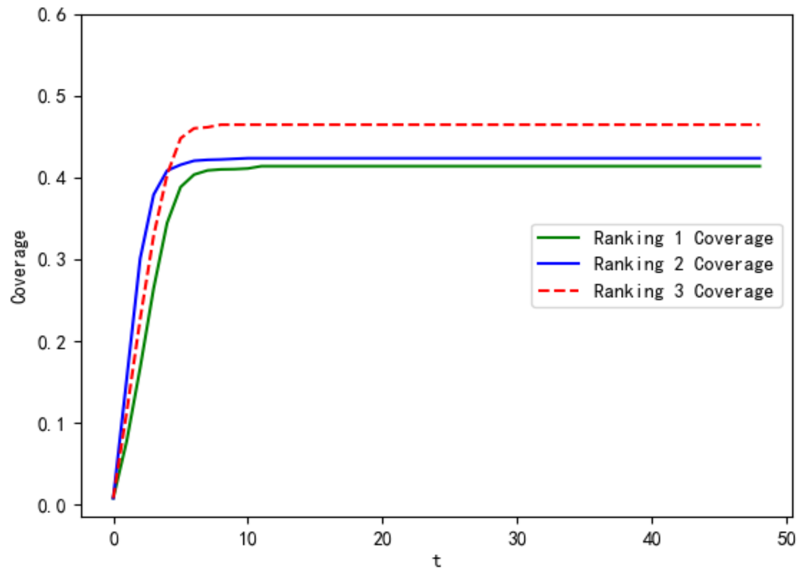

Fig. 4. Comparison of coverage of three rankings under topic 1. Horizontal axis is time step, vertical axis is coverage.

It can be seen from the figure that the final coverage of Ranking 1, 2 and 3 is 41.37 %, 42.34 %, and 46.44 % respectively. The user dissemination capability ranking considering neighbor nodes and time decay (Ranking 3) produces the biggest coverage of the information spreading, while PageRank ranking produces the smallest one. Therefore, it is reasonable to take the effect of neighbor nodes and time decay on information spreading into account when evaluating the influence of nodes in social network.

6. Conclusion

In the social platform, if posts can be spread in a wider range after forwarded by a user, then he can be seen as an influential user with great dissemination capability. Previous studies measured user influence by network topology, and rarely evaluated user influence from the perspective of information spreading. Considering the influence of neighbor nodes and time decay, we proposed a dynamic evaluation model of user dissemination capability based on topic segmentation, and realized the evaluation of user dissemination capability under different topics.

We constructed an interaction relationship network of Microblog users as the experimental network, and used the interaction records of users within three months as experimental data. Considering the influence of neighbor nodes, we counted the number of subsequent forwarding of posts by neighbor nodes under different topics, and adjusted the user relationship strength. By analyzing the correlation of user interaction frequency, it is verified that users' interaction strength decays over time. The experimental results show that the user dissemination capability is different under different topics; the effect of neighbor nodes and time decay should be taken into account when evaluating user dissemination capability.

Although the time span of the experimental data is small, it can reveal the differences in the influence of topic segmentation, neighbor nodes and time decay on user dissemination capability. The evaluation model of dissemination capability proposed from the perspective of information diffusion enriches the evaluation of node influence in social networks and provides scientific basis for intervening the spreading of information.

Acknowledgement

We are very thankful to the editor and people whose valuable suggestions and comments helped us to improve the article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This article is supported by the National Natural Science Foundation of China under Grant 71874088, 71704085, 71974102, Qing Lan Project.

References

[1] Bonacich P. Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology, 1972,2(1): 113-120.

[2] Christakis NA, Fowler JH. Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives. New York: Little, Brown And Company, 2009: 9-32.

[3] Chen D B, LüL Y, Ming S S, Zhang Y C, Zhou T. Identifying influential nodes in complex networks. Physica A, 2012, 391(4): 1777-1787.
[4] Freeman LC. A set of measures of centrality based on betweenness. Sociometry, 1977, 40(1): 35-41.
[5] Kahng B, Goh K I, Oh E, et al. Betweenness centrality correlation in social networks. Physical Review E, 2003, 67(1): 017101.
[6] Ortiz M G R, Hoyos J R C, López M G R. The social networks of academic performance in a student context of poverty in Mexico. Social networks, 2004, 26(2): 175-188.
[7] Sabidussi G, The Centrality Index of a Graph. Psychometrika, 1966, 31(4): 581-603.
[8] Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA. Identification of influential spreaders in complex networks. Nature Physics, 2010, 6(11): 888-893.
[9] Heidemann J, Klier M, Probst F. Identifying key users in online social networks: A pagerank based approach. 2010.
[10] Berkhin P. A survey on pagerank computing. Internet Mathematics, 2005, 2(1): 73-120.
[11] Shi L, Zhang C. Introducing an active index ranking mechanism for microblog users. Small Microcomputer System, 2012, 33(1): 110-114.
[12] Weng J, Lim E P, Jiang J, et al. Twitterrank: finding topic-sensitive influential twitterers. Proceedings of the third ACM international conference on Web search and data mining. 2010: 261-270.
[13] Yamaguchi Y, Takahashi T, Amagasa T, et al. Turank: Twitter user ranking based on user-tweet graph analysis. International Conference on Web Information Systems Engineering. Springer, Berlin, Heidelberg, 2010: 240-253.
[14] Zhu H, Yin X, Ma J, et al. Identifying the main paths of information diffusion in online social networks. Physica A: Statistical Mechanics and its Applications, 2016, 452:320-328.
[15] Hu Y, Song R J, Chen M. Modeling for information diffusion in online social networks via hydrodynamics. IEEE Access, 2017, 5: 128-135.
[16] Trusov M, Bodapati A V, Bucklin R E. Determining influential users in internet social networks. Journal of Marketing Research, 2010, 47 (4): 643-658.
[17] Leenders R T A. Modeling social influence through network autocorrelation: Constructing the weight matrix. Social Networks, 2002,24 (1): 21-47.
[18] Yang Z, Guo J, Cai K, Stern D. Understanding retweeting behaviors in social networks. Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010). New York: ACM Press, 2010:1633-1636.
[19] Tan C, Tang J, Sun J, Lin Q, Wang F. Social action tracking via noise tolerant time-varying factor graphs. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010). New York: ACM Press, 2010: 1049-1058.
[20] Ji Y, Zhang N. Node Propagation Ability Measurement Algorithm Based on Neighbor Degree Information. Information Technology, 2016(05): 19-21.

[21] Varshney D, Kumar S, Gupta V. Predicting information diffusion probabilities in social networks: A Bayesian networks based approach. Knowledge-Based Systems, 2017, 133: 66-76.

[22] Ma X, Lei X, Zhao G, et al. Rating prediction by exploring user's preference and sentiment. Multimedia Tools and Applications, 2018, 77(6): 6425-6444.

[23] Zheng W, Ge B, Wang C. Building a TIN-LDA model for mining microblog users' interest. IEEE Access, 2019, 7: 21795-21806.

[24] Tang J, Sun J, Wang C, Wang C, Yang Z. Social influence analysis in large-scale networks. Proc. of the 15th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD 2009). New York: ACM Press, 2009. 807-816.

[25] Jiang Q, Zhang X. Social network influence evaluation method based on specific topics. Journal of Applied Sciences, 2020, 38(03): 352-366.

[26] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in twitter: The million follower fallacy. Proceedings of the international AAAI conference on web and social media. 2010, 4(1).

[27] Li J, Peng W, Li T, et al. Social network user influence sense-making and dynamics prediction. Expert Systems with Applications, 2014, 41(11): 5115-5124.

[28] Ju C, Tao W. A novel relationship strength model for online social networks. Multimedia Tools & Applications, 2017, 76(16): 1-18.

[29] Liu W, He M, Wang L, Liu Y, et al. Research on microblog forwarding prediction based on user behavior characteristics. Journal of Computer, 2016, 39(10): 1992-2006.

[30] Rezaie B, Zahedi M, Mashayekhi H. Measuring time-sensitive user influence in Twitter. Knowledge and Information Systems, 2020, 62(9): 3481-350.