Physica A 510 (2018) 415–425

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Influencer discovery algorithm in a multi-relational network

Ning Ma [a], Yijun Liu [a,*], Yuxue Chi [a,b]

[a] Institute of Science and Development, Chinese Academy of Sciences, No.15 ZhongGuanCun BeiYiTiao Alley, Haidian District, Beijing 100190, PR China

[b] University of Chinese Academy of Sciences, No.15 ZhongGuanCun BeiYiTiao Alley, Haidian District, Beijing 100190, PR China

HIGHLIGHTS

- The multi-relational network model is proposed to simulate the social platform.
- The node has an impact weigh and the linked edge has a link weight in this model.
- The discovery algorithm—InfluencerRank algorithm is first proposed.
- The identification results of InfluencerRank algorithm and PageRank algorithm are compared.

ARTICLE INFO

Article history:
Received 9 February 2018
Received in revised form 23 May 2018
Available online 30 June 2018

Keywords:
Social platform
Multi-relational network
InfluencerRank algorithm
Influential nodes

ABSTRACT

With the development of social networks, the interaction between users and the application of social platforms for communications has become increasingly diverse. The influence and authority of different users have also been distinguished in constant communications. To better research the dissemination mechanism of different users' views on social platforms, a multi-relational network model first had to be built that can retain the interactive relationship between social networks to the maximum extent. In this model, the node has an impact weight, while the linked edge has a link weight. Combining these features of a multi-relational network model, a discovery algorithm – the InfluencerRank algorithm – was proposed. This discovery algorithm accurately identifies the essential influential nodes in networks. By combining the data of cases with the InfluencerRank algorithm, we identified influencers and conducted a comparative analysis.

1. Introduction

With the development of Internet technology, a large-scale, complex and pluralistic networked society has been formed in which the dissemination of information has been transformed into the distributed mode from the traditional single centralized mode. Facebook, Twitter and LinkedIn, the most popular platforms in the world, as well as Sina Weibo, Wechat and Zhihu (which are the most influential platforms in China) have become indispensable social platforms in people's everyday lives. Social platforms have changed people's lifestyles and the ways in which information is disseminated. These platforms have also attracted more and more researchers to study the dissemination of information based on social network analyses, such as the network information dissemination model [1], the evolution of the network structure [2], the dynamic mechanism of network information dissemination [3] and the mining algorithms of influential users [4,5]. Among the above, identifying influential users has real-world significance, such as precision marketing [6,7], information propagation [8], opinion formation [9], rumor [10], searching [11], expertise recommendation [12] etc.

* Corresponding author.
E-mail addresses: maning@casipm.ac.cn (N. Ma), yijunliu@casipm.ac.cn (Y. Liu), chiyuxue15@mails.ucas.ac.cn (Y. Chi).

https://doi.org/10.1016/j.physa.2018.06.064
0378-4371/

Studies on the identification of influential users or opinion leaders on online social networks (OSNs) have attracted worldwide attention. There are three main research methods. (1) The first method identifies influential users based on users' behavior. By researching various user characteristics and categorizing netizens according to their posting frequency, the number of followers, time spent online, the user's level and so on, or categorizing them with several indexes, users who are at the top are influential users [13,14]. (2) The second method identifies influential users by constructing a complex network model, based on the network structure index [15,16]. In the social network analysis model (SNA), centrality [17,18], semi-local centrality [19], k-core [20], k-shell [21] and bridging [22] are the indexes most commonly applied. In the dynamic network analysis model (DNA) [23], knowledge diversity, the task exclusive right, and cognitive load are the most frequently applied indexes [24]. In the supernetwork model [25], the vertex hyperdegree, hyperedge density, and clustering coefficient are the most frequently applied indexes [26,27]. (3) The third method searches for influential users by establishing a complex network model and applying the network link relations algorithm. In the SNA model, there are HITS [28], PageRank [29,30], NodeRank [31,32], LeaderRank [33], PolarityRank [34], UserRank [35], ConformRank [36] and so on. In the supernetwork analysis model, there are HostRank [37], SuperedgeRank [38] and so on. In terms of the study objects, there are mainly empirical studies conducted on Twitter [39], Weibo [40] and the blogosphere [41].

Empirical studies of Sina Weibo have focused on the information spread mechanism [42,43], user behavior judgment [44], user interest prediction [45] and user emotional analysis [46]. Many studies constructed a public spread network model based on Sina Weibo. Some studies focused on the forward relationship in the network model [47], namely the edge between nodes as the forward relationship [48,49]. Some studies examined the follow relationship between users as the edge to construct the network mode [50,51]. Previous studies have different emphases, but few considered forward, comment and follow relationships in constructing the network model.

In the real networked society, information dissemination largely depends on the properties of nodes and the properties of linking relationships among nodes. The weight of such edges has advantages by providing more accurate methods of understanding the network structure and function, which means the information dissemination network has node differences and edge heterogeneity. Take Sina Weibo as an example. (1) In terms of the node properties, after providing information, Weibo users who have a larger number of followers, a higher level and a larger number of microblogs produce an obviously stronger diffusion effect. (2) In terms of the linked relationship, edges with different properties affect the network topology, and different properties have a mutual coupling function. For example, the connected forward and comment relationship is significantly stronger than the single forward relationship.

Therefore, the construction of a multi-relational network model (with node differences and edge heterogeneity) could solve the problem on which most studies currently focus—the construction of a single relationship, which ignores the absence of the node weight and the edge weight. In addition, after the complex network has been researched for more than 10 years, experts and scholars have accepted that the role defining microscopic individuals was one of the top 10 challenges in the current complex network research [52]. This challenge includes exploring and identifying which the nodes in a network play a crucial role and the specific function of such a role. Therefore, this paper proposes an algorithm for identifying the influential nodes based on this a model to identify the real influential users online The monthly active users of Sina Weibo total 0.39 billion and daily active users total about 0.17 billion. Users with high influence among these user groups have important practical significance. From the perspective of national development, cultivating influential users to propagandize national culture, establish topics with positive energy and participate in public event discussion has a role in enhancing national cohesion and promoting the legal system and institutions. From the government perspective, finding influential users who can communicate with the government plays an important role in shaping the government's image, improving government credibility, cultivating the use of Weibo in political activity and controlling the spread of false information. In terms of enterprises, influential users can be used for virus marketing, online product recommendations, online advertising investment, marketing effect evaluation or even enterprise crisis public relations

2. Multi-relational network model

2.1. Construction principle

The multi-relational network model is constructed by focusing on the heterogeneity of opinion agents and several linking relationships among such agents while they disseminate information to highlight the different authority and closeness of relationships among netizens as well as the their effect on information dissemination. The traditional information dissemination network model is constructed by extracting the replying relationship among netizens, which abstracts netizens as nodes and the replying relationship as edges (Fig. 1A). The nodes and edges in the traditional communication network are homogeneous, which means the nodes and edges have no weight However, the multi-relational network model is a social network containing weight with multiple nodes and multiple edge relationships, which could maximize the retention of all information related to the interaction network and the real world. Taking Sina Weibo into consideration, a multi-relational network model could be constructed for specific events (Fig. 1B).

In this network, the nodes (Weibo users) have their own influential weight. The more influential they are for the user, the larger their influential weight. In the network diagram, the nodes with influential weight are larger.

In this network, the edges (Weibo user relationship) have a linking weight. The closer the relationship for users, the larger the linking weight. In the network diagram, the linked edges representing the linking weight are thicker.

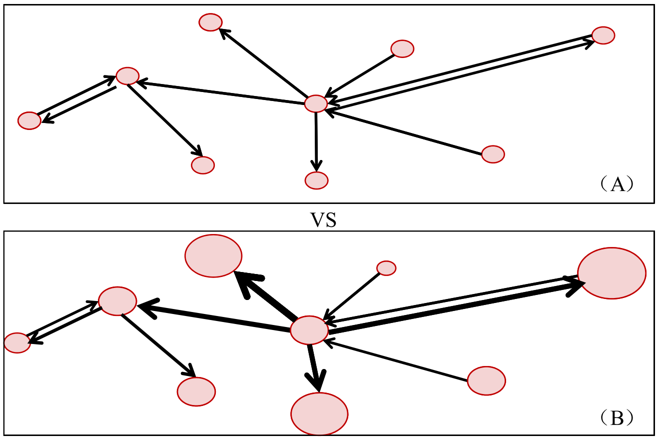


Fig. 1. Comparison of information dissemination networks. (A) Traditional dissemination network model. (B) Multi-relational network model.

To establish the multi-relational network model, the following indexes measure the node influential weight and the edge link weight. The following definitions are used for the Sina Weibo platform (shown in Fig. 2).

Definition 1. Node denotes Sina Weibo users.

Definition 2. Edge denotes the interactive relationship between different Sina Weibo users, including follow, forward and comment relationships

Definition 3. Microblog denotes the total number of microblogs issued by one Sina Weibo user.

Definition 4. Follower denotes the number of followers of one Sina Weibo user and represents the user's activity on the Sina Weibo platform for a long period

Definition 5. Level denotes the level of Sina Weibo users and it is the evaluation mechanism of the Sina Weibo platform for users. The higher the level, the larger the user's influence

Definition 6. Forward denotes a retweet caused by Sina Weibo users who post microblogs

Definition 7. Comment denotes a comment caused by Sina Weibo users who post microblogs.

Definition 8. Following denotes the follow relationship between different users.

2.2. Nodes' influential weight

Taking Sina Weibo as an example, during the construction of a multi-relational network model, the influential degree of nodes is determined jointly by the following four factors (Table 1).

$\boldsymbol{v}_{\mathbf{Follower}}$ means the nodes corresponding to the number of Weibo user followers, which is the direct reflection of the user's influence. The larger this number, the larger the number of others who forward and comment on the expression of specific events. The influence is greater as well. Moreover, a larger number of followers attracts more people to follow the user, which is called the "Matthew effect". Subject to the specific data on Sina Weibo followers' ranking (users who have more than 70 million followers are identified as users with the largest number of followers; the number of users whose followers exceed 2 million has reached 2000), the criteria for the weight classification of the number of Weibo user followers have been given.

$\boldsymbol{v}_{\mathbf{Microblog}}$ means the nodes corresponding to the number of microblogs posted by users. The larger the number, the more active the user. When a specific event occurs, the more actively the users participate in the event by posting related microblogs, the more the users affect the progress of this event. Similarly, combined with the analysis of the quantity of microblogs issued by users of Sina Weibo, the criteria for the weight classification of the number of microblogs have been given.

$\boldsymbol{v}_{\mathbf{Level}}$ means the nodes corresponding to the Weibo user level. On Sina Weibo, the user level is determined by active days. If a user logs in and uses Weibo every day to maintain his or her online duration, he or she obtains active days and improves his or her level. The higher the level, the more privileges the user has.

Fig. 2. Sina Weibo user page and microblog example.

Table 1
Determination of nodes' influential weight.

Nodes' influential weight		1	0.8	0.6	0.4	0.2
$v_{Follower}$	10000 units	>1000	500–1000	100–500	50–100	<50
$v_{Microblog}$	1000 units	>20	10–20	5–10	2–5	<2
v_{Level}	–	>10	10	9	8	<8
$v_{Activity}$	–	active days *100/(active days + inactive days)				

$\boldsymbol{v}_{\mathbf{Activity}}$ means the integrated nodes corresponding to the Weibo user's active days and inactive days, which in a certain context could better reflect the user's integrated activity than the user level identified by Weibo.

The calculation formulas for these four factors that affect the node weight are shown in Table 1.

These four factors affect the determination of the node weight and the criteria for the weight classification given in Table 1. The calculation formula for node influential weight v_{Node} is as follows:

$$v_{\text{Node}} = \frac{v_{\text{Follower}} + v_{\text{Microblog}} + v_{\text{Level}} + v_{\text{Activity}}}{4}$$

Using influential Weibo user XX Li as an example (Fig. 2), the calculation for the nodes' influential weight is $v_{\text{Follower}} = 1$, $v_{\text{Microblog}} = 0.8$, $v_{\text{Level}} = 1$, $v_{\text{Activity}} = 0.94$. As a result, the node weight of Weibo user XX Li in the multi-relational network model is $v_{\text{Node}} = (1 + 0.8 + 1 + 0.97)/4 = 0.94$

2.3. Edge link weight

In the multi-relational network model constructed for Sina Weibo, the linking weight of the edges is determined by two dimensions: quantity and quality.

Quantity dimension: Between two random Weibo users, there are three potential types of linked relationships: forward, comment and follow. Among these relationships, the quantity of forwards and comments varies, which means any Weibo user can forward or comment to other users several times

Quality dimension: The edge quality includes v_{Forward}, v_{Comment} and $v_{\text{Following}}$, in which v_{Forward} is determined by the number of followers of the users who forward the posting The more followers the users who forward the posting have, the greater the likelihood that the microblog posting will be forwarded and commented on by other users. The diffusion effect will be greater; v_{Comment} is determined by the number of comments posted by commentators. Generally, the longer the comment, the more opinions are reflected, and the more users follow the comment; $v_{\text{Following}}$ is determined by whether there is a following relationship among different users (Table 2).

Table 2
Determining the quality of the edges.

Edges "Quality"		1.0	0.8	0.6	0.4	0.2
v_{Forward}	The number of followers (10000 units)	>1000	500–1000	100–500	50–100	<50
v_{Comment}	The number of words	100–140	60–100	20–60	10–20	<10
$v_{\text{Following}}$	The following relationship	Existence: $v_{\text{Following}} = 1$			Nonexistence: $v_{\text{Following}} = 0$	

Based on the quantity and quality dimensions, the calculation formula for the linking weight of edges in the multi-relational network model w_{Edge} is as follows:

$$w_{\text{Edge}} = \alpha \times (n_{\text{Forward}} \times v_{\text{Forward}}) + \beta \times (n_{\text{Comment}} \times v_{\text{Comment}}) + \gamma \times v_{\text{Following}}, \quad \alpha + \beta + \gamma = 1$$

α, β and γ are tunable parameters. If the immediate relationship is the focus, α and β have a larger weight; if a stable relationship is the focus, γ has a larger weight. The forward and comment relationship (a transient immediate relationship) is generated from the topics or events in which the users are interested. When the specific event ends, these two relationships disappear as well, whereas the following relationship (a stable relationship) is a long-term relationship established between two users who have an intimate relationship. n_{Forward} and n_{Comment} are quantity dimensions; v_{Forward}, v_{Comment} and $v_{\text{Following}}$ are quality dimensions.

For example, Weibo user *A* and user *B* have three linked relations: User *A* follows user *B*. User *A* forwards user *B*'s microblog 4 times, and user *A* has 8 million followers. User *A* comments on user *B*'s microblog 2 times, and each comment has 30 words on average. Thus, the linking weight between user *A* and user *B* is $w_{A\to B} = \alpha \times (n_{\text{Forward}} \times v_{\text{Forward}}) + \beta \times (n_{\text{Comment}} \times v_{\text{Comment}}) + \gamma \times v_{\text{Following}} = (4 \times 0.8 + 2 \times 0.6 + 1) \times 0.33 = 1.8$ (assume that $\alpha = \beta = \gamma$, and $\alpha + \beta + \gamma = 1$).

3. Influencer discovery algorithms

3.1. Core concept

The PageRank algorithm is a technology calculated by a search engine in accordance with an interactive hyperlink among the webpages and determines the rank of a webpage by the network hyperlink relationships. The link from page A to page B is interpreted as votes made by page A for page B. Then, the new rank can be determined in accordance with the source of the votes (even the source of a source, that is the page linked to page A) and the rank of a source target. The PageRank score of one page is calculated by the PageRank score of other pages. Namely, the PageRank score that shows the importance of each page not only depends on the number of links to this page but is also influenced by quality and importance of this page, and the PageRank score of this page is distributed equally to the linked page. With constantly repetitive calculations, the PageRank scores of these pages tend to be stable.

Two important assumptions exist in the PageRank algorithm. One assumption is that each page has the same PageRank score by default in the initial phase. Namely, the influence of each page is similar by default. The other assumption is that each page distributes its existing PageRank score to link pages and can be determined by the number of links. These two assumptions conform to the facts of an interactive link relationship among search engine webpages. However, in the multi-relational network model, which is established based on an actual social network platform, each node (Weibo user) has its own influential weight, and each edge has its own linking weight. For example, a compound forward and comment relationship is closer than a simple forward relationship. Therefore, based on the established multi-relational network model and by referring to the PageRank algorithm concept, the InfluencerRank algorithm, which is more applicable to social network platforms, is proposed.

3.2. Specific algorithm

A specific algorithm of influential users integrated attraction when users are active on Sina Weibo for a long period and attention after a user posts about a specific event. The influence of specific events is closely related to the network structure of the spread of public opinion in events. The InfluencerRank algorithm measures the influence of a user by the followers, microblogs and a user's activity accumulated over a long period. By summarizing these aspects, the most influential users can be found.

3.2.1. Influence on specific events

The InfluencerRank algorithm, which is put forward in the multi-rational network model, mainly considers the diversity and tightness of edges among different nodes; namely, the introduction of $w_{j\to i}$ indicates the weight of node *j* linked to node *i*. In the improved algorithm, changes are made to the measurement of the link numbers in the initial PageRank algorithm. The link numbers are not defined by the outdegree of the quantity statistics. In the improved algorithm, it is defined by the strength of the weighted edges. Specifically, $w_{j\to i}$ is larger, which shows the larger ratio of the InfluencerRank score for node *j* distributed to node *i* and the InfluencerRank score of node *i* is higher. This is the core of the InfluencerRank algorithm, which is put forward with full consideration of the weight of the linked edges in the multi-rational network model. The InfluencerRank score is the score of all agents in a special event:

$$IR_i = \frac{1-d}{N} + d \sum_{j \in M(i)} \frac{IR_j * w_{j \to i}}{L_j}, \tag{1}$$

where IR_i refers to the InfluencerRank score of node i, d refers to the dampening factor set as 0.85 in normal cases, N refers to the numbers of all nodes, IR_j refers to the InfluencerRank score of node j, $w_{j \to i}$ refers to the linked edge weight of node j linked to node i, L_j refers to the total nodes connected to node j and $M(i)$ refers to the set linked to node i

3.2.2. Influence on Sina Weibo

Throughout the whole network platform, different users have their own influential weight; namely, the node's influences weight v_{Node}. This weight is not influenced by certain events.

Based on Table 1, the influence of the corresponding node i of Sina Weibo users is measured:

$$v_i = \frac{v_{\mathrm{Follower}_i} + v_{\mathrm{Microblog}_i} + v_{\mathrm{Level}_i} + v_{\mathrm{Activity}_i}}{4}, \tag{2}$$

where v_i is the average influence, of which node i is active on Sina Weibo for a long period. v_{Follower_i} refers to the weight value of followers. $v_{\mathrm{Microblog}_i}$ is the weight value of the number of microblogs, v_{Level_i} is the weight value of the Weibo level, v_{Activity_i} is the weight value of its activity v_i and the weight value range is 0 to 1.

3.2.3. Comprehensive influence

Observing the whole social platform network, the InfluencerRank score of certain events and the influence weight itself of the most influential node must be considered:

$$Influence_i = IR_i * v_i, \tag{3}$$

where $Influence_i$ refers to the total influence of node i, IR_i refers to the InfluencerRank score of node i and v_i refers to the influential weight of node i.

4. Experimental analysis and results

4.1. Data acquisition and processing

At present, when an emergency occurs, the dissemination of public opinion information on various self-media platforms has various characteristics, including a high speed of dissemination, a high degree of freedom and a large range of influence. In March 2016, an explosive report on tainted vaccines in a province of China triggered great concern among netizens. During this emergency, the largest Chinese social network platform Sina Weibo set up "vaccination" as a key word for relevant information. The data capacity was approximately 23 032, and the duration was about 14 days. The specific data acquisition and disposal process are shown as follows:

Data acquisition: based on the public opinion information data acquisition platform, we set up "vaccine" as the keyword in the event discussion time period to retrieve information.

Acquisition content: included the issuers (users), microblogs, personal information of issuers (number of microblogs, followers, and level), forwards, comments and likes by blogs.

Data pretreatment: Based on Tables 1 and 2, data acquisition is used to calculate users' influence weight for the corresponding nodes and the edge link weight.

Construct the multi-regional model: Based on the relationship between Weibo users, the multi-relationship network model was constructed. Nodes represent Weibo users, and edges represent the following, forward and comment relationships between different users.

Retain the core network parts: The initial network graph reflects the interactive relationships among Weibo users. Then, isolated nodes and leaf nodes with less influence are removed step by step.

Finally, taking into account the data visualization, the multi-relational network (Fig. 3) consisting of 109 core nodes and 385 edges can be obtained. In this model, the size of the nodes represents the influence weight ($v_{\mathrm{Node}} \in [0.2, 0.9]$), and the thickness of the linked edges represents the linking weight of the edges ($w_{\mathrm{Edge}} \in [1, 4]$).

4.2. Identification results for influencers

By applying the InfluencerRank algorithm, the top 10 influential nodes in this network are identified after the final influence of the nodes is calculated with MATLAB. At the same time, the top 10 influential nodes in this network can also be identified with the PageRank algorithm. All identification results are shown in Table 3. In comparison, the rank sequences of the other nodes' influence are different, unless node J ranks first in two algorithms, and node J is the most influential node in the network. The top 10 influential nodes identified by the InfluencerRank algorithm include node E and node I, but the

Table 3
Analysis of the identification results for influential nodes.

	InfluencerRank		PageRank		Ranking Difference	
Rank	Node	Score	Node	Score	Node	Difference
1	J	0.00477	J	0.15384	J	0
2	G	0.00396	D	0.12188	G	–4
3	K	0.00381	H	0.11866	K	–2
4	E	0.00361	L	0.11002	E	–
5	I	0.00348	K	0.08564	I	–
6	D	0.00347	G	0.04124	D	4
7	F	0.00315	C	0.03702	F	–1
8	C	0.00309	F	0.03360	C	1
9	B	0.00297	A	0.03218	B	–1
10	L	0.00293	B	0.03034	L	6

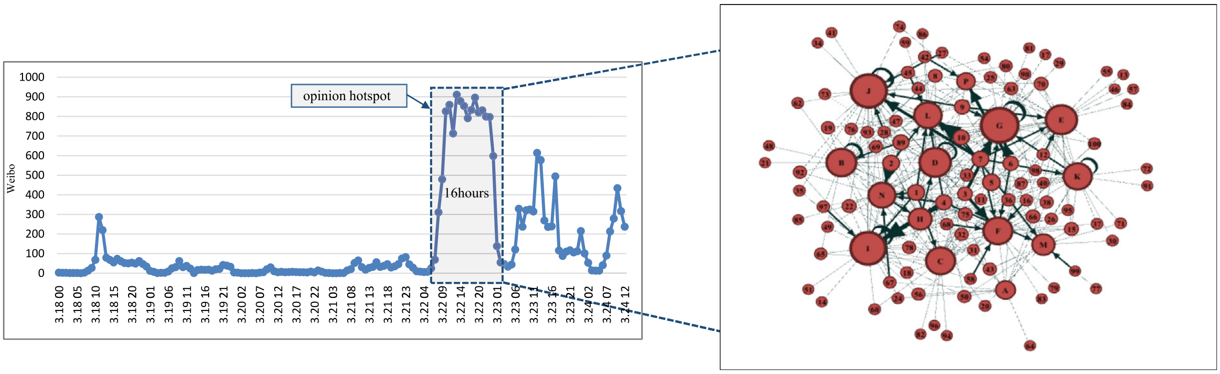
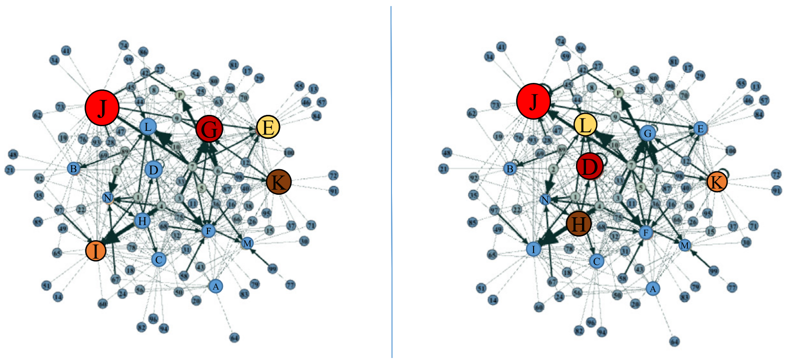


Fig. 3. Multi-relational network model of a specific case.



InfluencerRank algorithm PageRank algorithm

Fig. 4. Comparative analysis of the identification results of influencers.

top 10 influential nodes identified by the PageRank algorithm exclude these two nodes, and they are substituted by node H and node A (Fig. 4).

In the calculation of the InfluencerRank algorithm, the edge link weight is used to calculate $\alpha = \beta = \gamma$ $(\alpha + \beta + \gamma = 1)$ in the formula. The weight was set up in line with two situations: in the specific event, the influence of public opinion spread is generated. In other words, the immediate relationship of the specific events is observed. The weight of α and β is larger. Here, $\alpha = \beta = 0.4$, $\gamma = 0.2$, when users are active on Sina Weibo for a long period and construct stable long-term following relationships with other users. The γ weight is larger. Here, $\alpha = \beta = 0.2$, $\gamma = 0.6$. The results for the three weights are compared (Table 4). The partial nodes' InfluencerRank algorithm ranking results differ. By comparing with the weight $(\alpha = \beta = \gamma)$, as the immediate relationship $(\alpha = \beta = 0.4, \gamma = 0.2)$, the ranking increase for node E, node F and node L

Table 4
Identification results for different edge link weights.

Rank	$\alpha=\beta=\gamma,\ \alpha+\beta+\gamma=1$		$\alpha=\beta=0.4,\ \gamma=0.2$		$\alpha=\beta=0.2,\ \gamma=0.6$	
	Node	Score	Node	Score	Node	Score
1	J	0.00477	J	0.00976	G↑	0.00865
2	G	0.00396	E↑	0.00683	J	0.00604
3	K	0.00381	G	0.00321	D↑	0.00339
4	E	0.00361	F↑	0.00309	K	0.00349
5	I	0.00348	K	0.00308	I	0.00354
6	D	0.00347	D	0.00256	E	0.00240
7	F	0.00315	L↑	0.00226	B↑	0.00247
8	C	0.00309	C	0.00192	F	0.00185
9	B	0.00297	B	0.00188	L	0.00252
10	L	0.00293	H	0.00090	C	0.00095

Table 5
Comparison of influencers.

Node	Rank		Number of followers	Average attention of all microblogs	Attention of the case	Comprehensive evaluation of influence
	InfluencerRank	PageRank				
G	2	6	52646682	1250	69920	great
D	6	2	5708801	750	3322	good
L	10	4	304741	200	2765	normal

Table 6
Public opinion data for 10 typical cases.

Case	Case description	Nodes	Edges	Number of total followers	Number of total microblogs
1	A star died in a traffic accident	350	360	15137154	1835214
2	Trafficked girl find parents	607	612	2200907	820725
3	A girl lost search notice is a scam	4027	4140	104520139	19840211
4	The children in poverty mountain area need warm clothes	613	638	5955447	4015376
5	Horror game leads to emotional upheavals	192	196	8184911	1494470
6	The parents spare no money in finding their child	3547	3631	64002705	11660754
7	A comparative survey of Chinese and Western passport	740	783	2730653	7387144
8	shopping receipt contains highly carcinogenic substance	204	207	8609240	1796473
9	The Ebola outbreak in China	1314	1359	14224402	4461929
10	Dangerous cosmic rays will affect the earth	160	163	2445340	1028169

is obvious. This conforms to the high forward and comment numbers caused by users posting about this event. Similar to the following stable relationship ($\alpha = \beta = 0.2, \gamma = 0.6$), the ranking increase in node G, node D and node B is obvious. This conforms to the situation that the users in these nodes are active on Sina Weibo to attract numerous followers in stable relationships.

4.3. *Algorithm effectiveness evaluation*

According to Table 3, the nodes identified by the InfluencerRank and PageRank that have significant differences in ranking are node G (differs by 4), node D (differs by 4) and node L (differs by 6). When the three nodes are compared in terms of the number of followers, the average attention level of all microblogs and the attention level of the microblogs posted about this case (the sum of forwards, comments and likes) (Table 5), influencer G is the most influential among these influencers, and the second most influential is influencer D. Influencer L has the least influence among the three influencers. These comparison results are consistent with the ranking of the influential nodes identified by the InfluencerRank algorithm. Therefore, the InfluencerRank algorithm is more effective than the PageRank algorithm.

4.4. *Comparative analysis of more empirical analysis*

In order to further verify the effectiveness of the InfluencerRank algorithm, 10 typical public opinion cases that occurred recently in China were selected. Public opinion data were collected from Sina Weibo for the analysis of influential users. The basic information of the 10 typical cases is shown in Table 6. The network model for each case is shown in Fig. 5.

Based on the multi-relational network model, the InfluencerRank algorithm identified influential users in each case. In the calculation, the immediate relationship and the stable relationship should be considered. The edge link weight is set up as the equal weight. The top 10 influential users identified in each case are shown in Table 7. The same influence users have special marks (Δ).

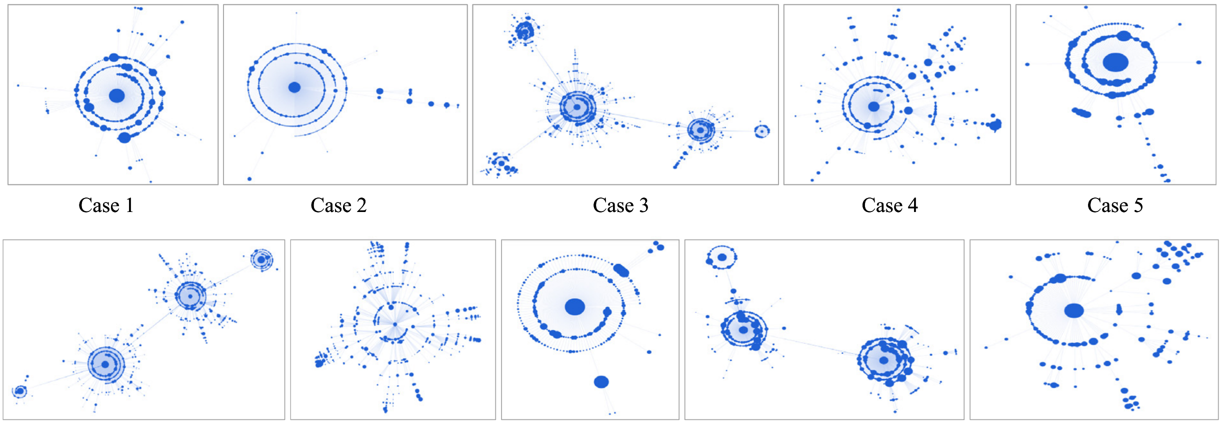


Fig. 5. The network model of 10 typical cases (Software: PKUVIS).

Table 7
The list of top 10 influential users in 10 typical cases.

Rank	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
1	ΔXin…	Weng…	Yan…	Kong…	ΔXin…	Jia…	Lei…	ΔZhong…	ΔXin…	Hua…
2	ΔXiao…	ΔZhong…	ΔChen…	Lu…	Su…	Let…	Hu…	ΔQing…	God…	Liu…
3	do…	Tao…	ΔZhong…	Si…	Duo…	ΔFu…	Liang…	Mi…	Mo…	Mou…
4	Tie…	Duan…	Anty…	Ni…	Shi…	Ri…	Li…	Shu…	Yue…	Mang…
5	Shui…	Gu…	Ao…	Mei…	Han…	Liao…	Xuan…	King…	Guang…	Colour…
6	Letter…	Fei…	C0…	ΔYong…	Zhi…	ΔYe…	Bei…	ΔRen…	Shen…	Bian…
7	Yang…	ju…	ΔRen…	Man…	Qu…	Zhen…	Dai…	ΔChen…	Yuan…	PE…
8	ΔLao…	Luo…	Xi…	ΔYe…	Pu…	ΔQing…	Yu…	lo…	Jun…	Jing…
9	Jiong…	ΔXin…	ΔFu…	Zi…	ΔXiao…	Yes…	ΔLao…	ΔMeng…	so…	ΔWu…
10	ΔYong…	ΔMeng…	Bi…	Hui…	ΔWu…	Yi…	Co…	An…	Be…	Moca…

Table 7 shows that in 10 different public opinion cases, the overlapping ratio of influential users is up to 27%, indicating that these users influence the spread of public opinion on Sina Weibo. These results also show the effectiveness of the InfluencerRank algorithm.

5. Conclusion and discussion

In this paper, we used Sina Weibo, the most influential social platform in China. We introduced the process of building a multi-relational network model in detail. In this model, the node has an influential weight, and the linked edge has a linking weight. Specifically, the different influence of Weibo users and the diversity of relationships among users are considered in the model. Then, combining the weights of the nodes and edges in the multi-relational network model, a discovery algorithm – the InfluencerRank algorithm – was proposed. By comparing the influential nodes identified by the InfluencerRank algorithm with those identified the PageRank algorithm, we found that influential nodes were identified more accurately by the InfluencerRank algorithm in the multi-relational network model as this algorithm focuses on the spreading effect of influential users. This research contributes to understanding the mechanism of key influential users online.

In this paper, an empirical study of Sina Weibo was conducted. Compared to the attention on the node influence weight in previous studies, the multi-relational network model constructed in this paper considers not only different influences of users (the node influence weight) but also the diverse relationships between users (the edge link weight) to embody users' interactive mechanism on Sina Weibo. Furthermore, in the process of identifying influential users, the model integrates the influence of users' long-term activity on Sina Weibo and considers the attention caused by users who comment on specific public opinion cases, namely influence on the spread of public opinion during a specific period. Previous related studies on influential user identification focused only on an aspect. This empirical study of Sina Weibo and the multi-relational network model also measured the user interactive mechanism on Sina Weibo. The study showed no matter whether Sina Weibo or Twitter or Facebook is the platform examined, the platform operation and user management mechanism are very similar. As a result, this method is easy to apply on other social media. The influential users identified by the algorithm could develop an important role in national development, government image construction, enterprise propaganda marketing and false information control. Further work is needed to propose new network indices or algorithms in the multi-relational network model. Then, the simulation analysis can be used to reflect the evolution of online public opinion.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) (71503246, 71573247).

References

[1] L. Sun, Y. Liu, M.R. Bartolacci, et al., A multiinformation dissemination model considering the interference of derivative information, Physica A 451 (2016) 541–548.
[2] F. Feng, X. Chen, L. Liu, et al., Social dilemmas in an online social network: the structure and evolution of cooperation, Phys. Lett. A 371 (2007) 58–64.
[3] H. Zhu, C. Huang, R. Lu, et al., Modelling information dissemination under privacy concerns in social media, Physica A 449 (2016) 53–63.
[4] A.B. Eliacik, N. Erdogan, Influential user weighted sentiment analysis on topic based microblogging community, Expert Syst. Appl. 92 (2018) 403–418.
[5] A. Sheikhahmadi, M.A. Nematbakhsh, A. Zareie, Identification of influential users by neighbors in online social networks, Physica A 486 (2017) 517–534.
[6] Z.Z. Alp, S.G. Öğüdücü, Identifying topical influencers on Twitter based on user behavior and network topology, Knowl.-Based Syst. 141 (2017) 211–221.
[7] J.Y.C. Ho, M. Dempsey, Viral marketing: motivations to forward online content, J. Bus. Res. 63 (2010) 1000–1006.
[8] F.D. Malliaros, M.E. Rossi, M. Vazirgiannis, Locating influential nodes in complex networks, Sci. Rep. 6 (2016) 19307.
[9] V. Kandiah, D.L. Shepelyansky, PageRank model of opinion formation on social networks, Physica A 391 (2012) 5779–5793.
[10] S. Tang, X. Teng, S. Pei, et al., Identification of highly susceptible individuals in complex networks, Physica A 432 (2015) 363–372.
[11] L. Adamic, E. Adar, How to search a social network, Soc. Netw. 27 (2012) 187–203.
[12] V. Carchiolo, A. Longheu, M. Malgeri, et al., Searching for experts in a context-aware recommendation network, Comput. Hum. Behav. 51 (2015) 1086–1091.
[13] O. Biran, S. Rosenthal, J. Andreas, et al., Detecting influencers in written online conversations, in: Proceedings of the Second Workshop on Language in Social Media, Association for Computational Linguistics, 2012, pp. 37–45.
[14] N. Agarwal, H. Liu, L. Tang, et al., Identifying the influential bloggers in a community, in: Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, 2008, pp. 207–218.
[15] S. Aral, D. Walker, Identifying influential and susceptible members of social networks, Science 337 (2012) 337–341.
[16] K. Klemm, M.Á. Serrano, V.M. Eguíluz, et al., A measure of individual role in collective dynamics, Sci. Rep. 2 (2012) 292.
[17] L.C. Freeman, Centrality in social networks: conceptual clarification, Soc. Netw. 1 (1979) 215–239.
[18] D. Gómezabc, Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems, European J. Oper. Res. 226 (2013) 354–365.
[19] D. Chen, L. Lü, M.S. Shang, et al., Identifying influential nodes in complex networks, Physica A 391 (2012) 1777–1787.
[20] M.A. Al-Garadi, K.D. Varathan, S.D. Ravana, Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method, Physica A 468 (2017) 278–288.
[21] M. Kitsak, L.K. Gallos, S. Havlin, et al., Identification of influential spreaders in complex networks, Nat. Phys. 6 (2010) 888–893.
[22] T.W. Valente, K. Fujimoto, Bridging: locating critical connectors in a network, Soc. Netw. 32 (2010) 212–220.
[23] K.M. Carley, Dynamic network analysis, 2003.
[24] K.M. Carley, Dynamic network analysis for counter-terrorism, Unpublished manuscript, 2005.
[25] A. Nagurney, J. Dong, Supernetworks: Decision-Making for the Information Age, Elgar, Edward, 2002.
[26] G. Ghoshal, V. Zlatić, G. Caldarelli, et al., Random hypergraphs and their applications, Phys. Rev. E 79 (2009) 066118.
[27] V. Zlatić, G. Ghoshal, G. Caldarelli, Hypergraph topological quantities for tagged social networks, Phys. Rev. E 80 (2009) 036118.
[28] J.M. Kleinberg, R. Kumar, P. Raghavan, et al., The web as a graph: measurements, models, and methods, in: Computing and Combinatorics, Springer, Berlin, 1999, pp. 1–17.
[29] L. Page, S. Brin, R. Motwani, et al., The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford InfoLab, 1999.
[30] Y.H. Eom, D.L. Shepelyansky, Opinion formation driven by PageRank node influence on directed networks, Physica A 436 (2015) 707–715.
[31] Z. Yan, M. Xing, D. Zhang, et al., EXPRS: An extended pagerank method for product feature extraction from online consumer reviews, Inf. Manag. 52 (2015) 850–858.
[32] X. Cheng, S. Su, Z. Zhang, et al., Virtual network embedding through topology-aware node ranking, ACM SIGCOMM Comp Commun Rev. 41 (2011) 38–47.
[33] L. Lü, Y.C. Zhang, C.H. Yeung, et al., Leaders in social networks the delicious case, PLoS One 6 (2011) e21202.
[34] F.L. Cruz, C.G. Vallejo, F. Enrı, et al., Polarityrank: finding an equilibrium between followers and contraries in a network, Inf. Process. Manage. 48 (2012) 271–282.
[35] T. Majer, M. Šimko, Leveraging microblogs for resource ranking, Lecture Notes Comput. Sci. 7147 (2012) 518–529.
[36] Q. Wang, Y. Jin, S. Cheng, et al., ConformRank: A conformity-based rank for finding top-k influential users, Physica A 474 (2017) 39–48.
[37] G.R. Xue, Q. Yang, H.J. Zeng, et al., Exploiting the hierarchical structure for link analysis, in: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2005, pp. 186–193.
[38] N. Ma, Y. Liu, SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork, Expert Syst. Appl. 41 (2014) 1357–1368.
[39] F. Riquelme, P. González-Cantergiani, Measuring user influence on twitter: a survey, Inf. Process. Manage. 52 (2016) 949–975.
[40] B. Liang, Y. Liu, M. Zhang, et al., Searching for people to follow in social networks, Exp. Syst. Appl. Inter. J. 41 (2014) 7455–7465.
[41] Y.M. Li, C.Y. Lai, C.W. Chen, Discovering influencers for marketing in the blogosphere, Inform. Sci. 181 (2011) 5143–5157.
[42] W.G. Yuan, Y. Liu, A mixing evolution model for bidirectional microblog user networks, Physica A 432 (2015) 167–179.
[43] Y. Zhang, Y. Su, W. Li, et al., Rumor and authoritative information propagation model considering super spreading in complex social networks, Physica A 506 (2018) 395–411.
[44] J. Hu, Y. Luo, J. Yu, An empirical study on selectivity of retweeting behaviors under multiple exposures in social networks, J. Comput. Sci. (2017) http://dx.doi.org/10.1016/j.jocs.2017.11.004.
[45] H. Bao, Q. Li, S.S. Liao, et al., A new temporal and social PMF-based method to predict users' interests in micro-blogging, Decis. Support Syst. 55 (2013) 698–709.

[46] X. Zou, J. Yang, J. Zhang, et al., Microblog sentiment analysis with weak dependency connections, Knowl.-Based Syst. 142 (2018) 170–180.
[47] J. Li, N. Wu, Z. Feng, Model-based non-gaussian interest topic distribution for user retweeting in social networks, Neurocomputing 278 (2018) 87–98.
[48] R. Wang, S. Rho, B.W. Chen, et al., Modeling of large-scale social network services based on mechanisms of information diffusion: Sina Weibo as a case study, Future Gener. Comput. Syst. 74 (2017) 291–301.
[49] Y. Liu, B. Wang, B. Wu, et al., Characterizing super-spreading in microblog: an epidemic-based information propagation model, Physica A 463 (2016) 202–218.
[50] Z. Wang, J. Liang, R. Li, Exploiting user-to-user topic inclusion degree for link prediction in social-information networks, Expert Syst. Appl. 108 (2018) 143–158.
[51] X. Zheng, Z. Zeng, Z. Chen, et al., Detecting spammers on social networks, Neurocomputing 159 (2015) 27–34.
[52] T. Zhou, Z.K. Zhang, G.R. Chen, The opportunities and challenges of complex networks research, J. Univ. Electron. Sci. Tech. China 1 (2014) 1–5.