

Improved Key Node Recognition Method of Social Network Based on PageRank Algorithm

Lei Hong¹, Yiji Qian^{1,*}, Chaofan Gong², Yurui Zhang¹ and Xin Zhou³

¹Jiangsu Police Institute, Nanjing, 210000, China

²Nanjing Police Station, Nanjing, 210000, China

³System Consulting Pty Ltd., Sydney, 201101, Australia

*Corresponding Author: Yiji Qian. Email: qianyiji12345@163.com

Received: 27 February 2022; Accepted: 01 July 2022

Abstract: The types and functions of social networking sites are becoming more abundant with the prevalence of self-media culture, and the number of daily active users of social networking sites represented by Weibo and Zhihu continues to expand. There are key node users in social networks. Compared with ordinary users, their influence is greater, their radiation range is wider, and their information transmission capabilities are better. The key node users play important roles in public opinion monitoring and hot event prediction when evaluating the criticality of nodes in social networking sites. In order to solve the problems of incomplete evaluation factors, poor recognition rate and low accuracy of key nodes of social networking sites, this paper establishes a social networking site key node recognition algorithm (SNSKNIS) based on PageRank (PR) algorithm, and evaluates the importance of social networking site nodes in combination with the influence of nodes and the structure of nodes in social networks. This article takes the Sina Weibo platform as an example, uses the key node identification algorithm system of social networking sites to discover the key nodes in the social network, analyzes its importance in the social network, and displays it visually.

Keywords: Social networking site; PageRank algorithm; key node

1 Introduction

In recent years, social networking sites have continued to develop at a high speed with the large-scale popularization of the mobile Internet. The value contained in social networking site data has gradually become the focus of attention. Research on social networking sites such as user relationship mining, cluster division research, public opinion monitoring, user recommendation algorithms, etc. has gradually become a hot spot in the industry and academia.

The research on user nodes in social networking sites is the main aspect of data mining in social networking sites. The potential value information is extracted by mining the interrelationships between nodes to obtain important commercial value and social benefits. The huge user base and increasing



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

frequency of use have caused the amount of data generated by social networking sites to continue to grow at a high speed, the ability to disseminate information continues to increase, and the structure of user network relationships has become increasingly complex. Therefore, discovering key user nodes in social networking sites is very important for user analysis and research.

2 Research Summary

Evaluating the criticality of users in social networks is similar to the division of user influence in social networks. Ding believes that the more important the node location is, the more likely it [1]. Dalibor Fiala combines the PageRank algorithm with the co-occurrence of the author to explore the influence ranking of scholars in the academic network [2]. Current research on user influence mainly focuses on three mainstream levels: user post analysis, user characteristic information analysis, and social network topology analysis.

2.1 User Post Analysis

Social media posts are mainly based on current events or users' personal feelings, and is an important way for users to attract attention from others. The criticality of users is generated by the form, type and quality of the content of the post, and accumulates and changes according to the content of different posts. Scholars at home and abroad have conducted relevant research on this. Barbieri's research found that the topic of online posting is closely related to the credibility and prestige of the author, and established a cascade model of independent influence based on the topic [3]. Through research on WeChat public accounts, Qin Fen found that the title, length, picture and emotion of the content of the post are directly related to the influence of the post [4]. The rapid development of social networks, especially the role of user-generated content information in judging the criticality of users is becoming more and more obvious. It is necessary to dig deeper into the content of postings as an important carrier of the criticality of users to find its positive effect on the influence of nodes.

2.2 User Characteristic Information Analysis

User characteristics in social networks are important criteria for judging their influence. The influence of users can be quantified by indicators such as the number of fans, the number of likes, and the number of reposts. Scholars at home and abroad conduct research on various user characteristics and dig out user characteristics. The connection with influence. The more common user characteristic algorithms include the number of followers ranking algorithm based on the number of users' fans [5], the expert survey and judgment method based on the expert index [6], and the average forwarding algorithm based on the number of users' reposts [7], as well as some Multi-indicator fusion influence judgment method. User characteristics are important bases for user behavior analysis and modeling in social networks, and can provide data support for user influence research. By mining user characteristics, the method determines the strength of influence among users, and deeply analyzes the criticality of nodes in social networks.

2.3 Analysis of Social Network Topology

Social users form social networks through user attention, content browsing and forwarding, and their influence is spread and spread through the network. Social network topology is an important direction for studying user influence. Scholars study the location of nodes and network structure in social networks. Reference [8] To measure the criticality of users. The more common method based on

the network structure is the K-coverage algorithm that focuses on the cascading relationship between users [9]. Point centrality and betweenness centrality are also effective methods to determine user influence. Reference [10] The most widely used of all methods is the PageRank algorithm, which establishes a random walk model of node influence based on quantitative and quality assumptions [11]. On the basis of the PageRank algorithm and the characteristics of social networks, scholars evolve the mathematical model of the algorithm and derive many extended algorithms. Reference [12] The influence measurement method based on network topology is clear and direct, and it can grasp the overall structure of social network in a macroscopic way. In particular, the PageRank algorithm has played an important role in theory and practice [13].

In summary, the influence measurement based on network topology can grasp the overall structure of social networks in a macroscopic way, especially the PageRank algorithm has proved its feasibility in theory and practice [14]. Based on previous studies, this paper measures the influence and structure of nodes in the network based on the characteristics of social networks, uses the network to judge the influence and structure of social users, and finds key nodes in the network. Reference [15] Finally, a key node identification system for social networking sites based on PageRank algorithm is designed.

3 Research Methods

The key degree of user nodes of social networking sites is mainly measured from two aspects: the influence of users and the structure of users. Reference [16] The influence of users can be reflected by the number of users' fans, the popularity of microblog per unit time and other data. The user's structure is obtained by analyzing the user's position in the whole social networking website topology. Therefore, we can define the user's criticality as Eq. (1).

$$SKSKNI(U_i) = \alpha Influ(U_i) + \beta Struct(U_i) \quad (1)$$

In formula (1), $SKSKNI(U_i)$ is the criticality of user i , $Influ(U_i)$ is the influence of user i , and $Struct(U_i)$ is the structural value of user i . α and β are their respective weight values.

3.1 Social Networking Site Node Influence Measurement

The user influence of social networking sites is the degree of influence the user has on the crowd through the information released. Reference [17] In reality, the factors that affect the influence of user nodes on social networking sites include not only objective indicators such as the number of users' fans and likes, but also subjective indicators such as the activity of users' fans.

The topological structure of the user node network of social networking sites is analogous to the web page topology in the World Wide Web (WWW). Users in social networking sites are equivalent to web pages in WWW, and the behavior of different users following each other is equivalent to web links in WWW. PageRank algorithm can be used for reference. Measure the influence of user nodes on social networking sites [18]. However, the PageRank algorithm has problems in the evaluation of user influence of social networking sites such as inaccurate determination of the initial PR value, unreasonable ratio of PR value transmission, and unsuitable topological structure. This article divides the influence of social networking site user nodes into direct influence and indirect influence. The influence of social networking site nodes can be defined as formula (2).

$$Influ(U_i) = direct(U_i) + indirect(U_i) \quad (2)$$

In formula (2), $direct(U_i)$ is the direct influence of user i , and $indirect(U_i)$ is the indirect influence of user i .

3.1.1 Direct Influence Factors of Social Networking Site User Nodes

Due to the differences of users in social networking sites, it is unreasonable for the PageRank algorithm to give all users the same initial PR value before the iteration starts. Four elements can be added to evaluate the direct influence of users: the number of fans, the number of works, and the quality of the works. And authentication, give users a different initial PR value. At the same time, silent fans and zombie fans without any dynamics can be preliminarily screened out, and they will be excluded when constructing the topology.

The number of users' fans is a basic indicator of users' influence on social networking sites. The more fans a user has, the more links the node will have in the topology, and the greater its importance; Publishing works is the most important way for users to maintain their influence. The more works a user publishes, the greater its influence will be; In the social networking platform, work quality is an important factor for users to maintain their own influence. The number of video views, comments and forwarding can be used as the standard to measure video quality; Larger social networking sites such as microblog and Zhihu have launched authentication mechanisms. In the view of ordinary users, these authenticated users have been screened by social platforms and have higher credibility and authority than other users.

So we can define the direct influence of users as formula (3).

$$direct(U_i) = \alpha_1 fans(U_i) + \beta_1 works(U_i) + \gamma_1 qual(U_i) + \delta_1 auth(U_i) \quad (3)$$

In formula (3), $fans(U_i)$ is the number of fans of user i , $works(U_i)$ is the total number of works of user i , $qual(U_i)$ is the quality of user i 's works, and $auth(U_i)$ is whether the user is authenticated. α_1 , β_1 , γ_1 , and δ_1 are their respective weights. The quality of user work $qual(U_i)$ can be defined as formula (4).

$$qual(U_i) = \frac{\sum_{j \in m_i} (\alpha_2 like(j) + \beta_2 comment(j) + \gamma_2 forwarding(j))}{N_{m_i}} \quad (4)$$

In formula (4), m_i is the information collection released by user i in a unit time, j is any piece of information in the information collection released by user i in a unit time, $like(j)$ is the number of likes obtained by information j , $comment(j)$ is the number of comments obtained by information j , and $forwarding(j)$ is the number of forwarding obtained by information j . N_{m_i} is the number of microblogs posted by user i in a unit time. α_2 , β_2 , and γ_2 are their respective weights.

3.1.2 Indirect Influence Factors of User Nodes on Social Networking Sites

It is impossible for users on social networking sites to pay equal attention to users on the watch list. Ordinary users are often interested in a small number of users and are willing to invest in the promotion of their works. This preference is manifested in the interactive behavior between users, which can increase indirect influence for users, which is measured by the following behaviors between users and the likes, comments, and forwarding behaviors between users.

In the network topology, the following relationship forms a pair of link-in and link-out relationships, and the influence of social network nodes can be spread through the network. The three interactive behaviors of like, comment, and repost show how much the user likes the user they are following. Reposting indicates that the higher the importance of the user, the comment is the

second, and the like is the least. These three interactive behaviors quantify the close relationship between fans and followed users. The higher the degree of closeness, the fans will allocate a larger proportion of the contribution value to the user. Solve the problem of the unreasonable proportion of PR value transmission by quantifying the closeness between users.

So the user's indirect influence can be defined as formula (5):

$$indirect(U_i) = (1 - d) + d \sum_{U_j \in F_{U_i}} affect(U_i, U_j) * direct(U_j) \tag{5}$$

In formula (5), d is the damping coefficient, F_{U_i} is the set of fans of user i , j is any fan of user i , $affect(U_i, U_j)$ is the degree of interaction between user i and fan j , $affect(U_i, U_j)$ can be expressed as formula (6):

$$affect(U_i, U_j) = \frac{interact(U_i, U_j)}{\sum_{U_p \in I_{U_j}} interact(U_p, U_j)} \tag{6}$$

In Eq. (6), $interact(U_i, U_j)$ is the communication ability of user i to fan j , I_{U_j} is the set of users followed by fan j , and p is any one of the users followed by fan j . The communication ability can be expressed as formula (7):

$$interact(U_i, U_j) = \alpha_3 \frac{T_{like(U_i, U_j)}}{T_{like(U_j)}} + \beta_3 \frac{T_{comment(U_i, U_j)}}{T_{comment(U_j)}} + \gamma_3 \frac{T_{forwarding(U_i, U_j)}}{T_{forwarding(U_j)}} \tag{7}$$

In formula (7), $T_{like(U_i, U_j)}$ is the number of likes given to user i by fan j in a unit time, and $T_{like(U_j)}$ is the number of likes given to followers by fan j in a unit time, $T_{comment(U_i, U_j)}$ is the number of comments made by fan j to user i 's work in a unit time, $T_{comment(U_j)}$ is the total number of comments fan j has given to followers in a unit time, $T_{forwarding(U_i, U_j)}$ is the number of forwarding works from fan j to user i in a unit time, and $T_{forwarding(U_j)}$ is the total number of forwarding works to followers within a unit time.

3.2 Structural Measurement of User Nodes on Social Networking Sites

In addition to analyzing the influence of a single node, the current domestic and foreign research on social networking sites also focuses on the analysis of the entire social network. Social network analysis explains social phenomena by studying factors such as network structure and node location [11]. Measuring the structure of nodes in a social network composed of users of social networking sites can more objectively and accurately reflect the criticality of a single node compared to the entire social network. This article mainly measures the structure of nodes from seven indicators in the three dimensions of network structure, centrality, and scope of influence. Therefore, the structural index of the node can be defined as formula (8).

$$Struct(U_i) = \alpha_4 network(U_i) + \beta_4 center(U_i) + \gamma_4 impact(U_i) \tag{8}$$

In formula (8), $network(U_i)$ is the importance measurement value of node i in the network structure, $center(U_i)$ is the network centrality measurement value of node i , and $impact(U_i)$ is the influence range measurement value of node i in the network value. α_4 , β_4 , and γ_4 are the respective weight values.

3.2.1 Node's Network Structure Measurement Factor

The network structure of the node is mainly measured by two indexes of constraint degree and rank degree. The degree of constraint of a node is characterized by the degree of dependence of the network node on other nodes in the network. The smaller the degree of constraint, the stronger the independence of the node. The rank degree of a node is characterized by the degree to which the

restraint of the network is concentrated on a network member. The higher the rank degree, the greater the degree of restraint received by the member.

Combining these two points, we can define the network structure metric value of the node as Eq. (9), where α_5 and β_5 are their respective weight values.

$$network(U_i) = \alpha_5 strain(U_i) + \beta_5 class(U_i) \quad (9)$$

3.2.2 Node Centrality Measurement Factor

The degree centrality of a node is measured by the degree centrality, intermediate centrality and proximity centrality of the node. The degree centrality of a node refers to the number of other network nodes connected to the node, which is characterized by the information receiving and dissemination capabilities of the node. If the degree centrality of a node is greater, it means that the node is in the core position.

The centrality of a node is the ratio of the shortest path between any two nodes through the node, which is characterized as a measure of the ability of the node as a communication medium between other nodes. The greater the centrality of the node, the greater the contribution of the node in the process of information dissemination, and the stronger control over information resources.

The proximity centrality of a node indicates the degree to which a node is not dominated by other nodes. The smaller the proximity centrality of a node, it means that it is less dominated by other nodes during information interaction, thinking that it is at the center of the network.

Therefore, the centrality metric value of the node can be defined as Eq. (10), and α_6 , β_6 , and γ_6 are their respective weights.

$$center(U_i) = \alpha_6 degree(U_i) + \beta_6 between(U_i) + \gamma_6 closeness(U_i) \quad (10)$$

3.2.3 Node's Influence Range Measurement Factor

The influence range of the node is measured by the effective scale and local clustering coefficient. The effective scale of a node refers to the network scale of the node minus the redundancy of the network. The larger the effective scale of the node, the stronger the network influence of the node.

The local clustering coefficient of a node refers to the probability that the two neighboring nodes of node i are neighboring nodes to each other. The smaller the local clustering coefficient, the more likely the node will play a role in the spread of information in the process of dissemination.

Therefore, the metric value of the node's influence range can be defined as formula (11).

$$impact(U_i) = \alpha_7 size(U_i) + \beta_7 cluster(U_i) \quad (11)$$

From Eqs. (2) and (8), we can get the key degree measurement formula of social networking site user nodes, where the formula (2) is convergent, and the formula (8) is a linear relationship, so the key degree measurement value can be finally calculated.

3.3 Summary

Algorithm: SNSKNI algorithm for identifying key nodes of social networking sites

1. Calculate the quality of social networking site users' works:

$$qual(U_i) = \frac{\sum_{j \in m_i} (\alpha_2 like(j) + \beta_2 comment(j) + \gamma_2 forwarding(j))}{N_{m_i}}$$

2. Calculate the direct influence of users:

$$direct(U_i) = \alpha_1 fans(U_i) + \beta_1 works(U_i) + \gamma_1 qual(U_i) + \delta_1 auth(U_i)$$

3. Calculate user communication power:

$$interact(U_i, U_j) = \alpha_3 \frac{T_{like(U_i, U_j)}}{T_{like(U_j)}} + \beta_3 \frac{T_{comment(U_i, U_j)}}{T_{comment(U_j)}} + \gamma_3 \frac{T_{forwarding(U_i, U_j)}}{T_{forwarding(U_j)}}$$

4. Calculate the degree of interaction between users:

$$affect(U_i, U_j) = \frac{interact(U_i, U_j)}{\sum_{U_p \in I_{U_j}} interact(U_p, U_j)}$$

5. Calculate the indirect influence of users:

$$indirect(U_i) = (1 - d) + d \sum_{U_j \in F_{U_i}} affect(U_i, U_j) * direct(U_j)$$

6. Calculate the degree of constraint between social network nodes:

$$straint(U_i) = \sum_j \left(p_{ij} + \sum_q p_{iq} p_{qj} \right)^2$$

7. Calculate the constraints of social networks:

$$class(U_i) = \frac{\sum_j \left(\frac{straint(U_j)}{c} \right) \ln \left(\frac{straint(U_j)}{c} \right)}{N \ln(N)}$$

8. Calculate network structure metrics:

$$network(U_i) = \alpha_5 straint(U_i) + \beta_5 class(U_i)$$

9. Calculate the degree center value of the social node:

$$degree(U_i) = \sum_j X_{ij}$$

10. Betweenness centrality of computing nodes:

$$between(U_i) = \frac{\sum h_{jq}(i)}{g_{jq}}$$

11. Calculate the proximity centrality of nodes:

$$closeness(U_i) = \sum d_{ij}$$

(Continued)

Algorithm: Continued

12. Calculate the node centrality measure:

$$center(U_i) = \alpha_6 degree(U_i) + \beta_6 between(U_i) + \gamma_6 closeness(U_i)$$

13. Effective scale of computing nodes:

$$size(U_i) = \sum_j \left(1 - \sum_q p_{iq} p_{jq}\right)$$

14. Calculate the local clustering coefficient of the node:

$$cluster(U_i) = \frac{2E(i)}{k(i)(k(i) - 1)}$$

15. Calculate the influence range measure of the node:

$$impact(U_i) = \alpha_7 size(U_i) + \beta_7 cluster(U_i)$$

16. Calculate indirect influence:

$$Struct(U_i) = \alpha_4 network(U_i) + \beta_4 center(U_i) + \gamma_4 impact(U_i)$$

17. The criticality of user nodes of social networking sites is mainly measured by two aspects, namely, the user's influence and the user's structure. The user's influence can be measured by the number of fans of the user, the popularity of Weibo within a unit time and other data. to manifest. The structure of the user is obtained by analyzing the position of the user in the topology of the entire social networking site.

$$SKSKNI(U_i) = \alpha Influ(U_i) + \beta Struct(U_i)$$

4 Experiment and Result Analysis

This experiment uses the Sina Weibo platform as the source of experimental data. Set the research object as party and government agency accounts, select a total of 3000 party and government agencies' microblog accounts, first crawl their basic information and relationship lists, obtain 151,220 pieces of relationship information between the two, and build their social networking site user node relationship network. The user nodes that are not in the largest connected subnet are filtered out, and finally 2988 user nodes are obtained. Crawl all its basic information on Weibo from 0:00 on April 1, 2020 to 0:00 on October 1, 2020, and at the same time crawl the forwarding, like, and comment information between any two nodes. Finally, 4,389,100 pieces of basic Weibo information, 324,103 pieces of Weibo forwarding information, 141,917 pieces of Weibo comment information, and 947,761 pieces of Weibo like information were obtained.

4.1 Experiment Procedure

The number of users' fans, microblogs, and whether they are authenticated has been obtained when crawling the user's basic information. The quality of the user's work is calculated by weighting the total number of works, reposts, comments and likes. The number of reposts, comments, and likes of the user accounted for the proportions of all reposts, comments, and likes, respectively, and these three proportions were weighted to calculate the user's communication ability.

This paper uses the entropy method to determine the weight value needed in the system calculation process. After establishing a complex network based on the user relationship, use the function corresponding to the Python Networkx library to calculate seven indicators that affect the user's structure. The specific weight values are shown in the [Tab.1](#):

Table 1: Specific weight data

Weight formula	Weighted name	Weights
(1)	α	0.599240
	β	0.400760
(3)	$\alpha 1$	0.139581
	$\beta 1$	0.035303
	$\gamma 1$	0.496368
	$\delta 1$	0.328749
(4)	$\alpha 2$	0.346029
	$\beta 2$	0.264566
	$\gamma 2$	0.389405
(7)	$\alpha 3$	0.388005
	$\beta 3$	0.349658
	$\gamma 3$	0.262337
(8)	$\alpha 4$	0.133289
	$\beta 4$	0.617160
	$\gamma 4$	0.249551
(9)	$\alpha 5$	0.311558
	$\beta 5$	0.688442
(10)	$\alpha 6$	0.180407
	$\beta 6$	0.702955
	$\gamma 6$	0.116638
(11)	$\alpha 7$	0.542408
	$\beta 7$	0.457592

4.2 Experimental Results and Analysis

4.2.1 Experimental Results

The key node identification system of social networking sites based on the PageRank algorithm is used to measure the criticality of the selected 3000 party and government agencies' microblog official accounts. In the final ranking of criticality, the top 10 users are the Central Committee of the Communist Youth League, China Police Online, China Fire, Ministry of Public Security Transportation Administration, Emergency Management Department, Ministry of Public Security Criminal Investigation Bureau, Police and Civilian Hand in Hand, Healthy China, China Meteorological Administration and Shanghai release.

4.2.2 Analysis of User Influence Measurement Results

The SNSKNI algorithm is used to measure the influence of users. Compared with the simple ranking based on the number of fans and the traditional PageRank algorithm, the characteristics of different users are considered, the influence of zombie fans is excluded, and the interaction between users is considered.

It can be seen from Fig. 1 that the quality of the work with the highest weight in the user’s direct influence measurement is the key indicator to evaluate the user’s influence. Except for the “Central European Courier” ranked 4th in terms of user work quality, which ranks in the top 100 in terms of user influence, all other users rank in the top 60 in terms of influence. In addition, user authentication information is also one of the important indicators of user influence evaluation. The top 50 users of user influence are all Weibo authenticated users.

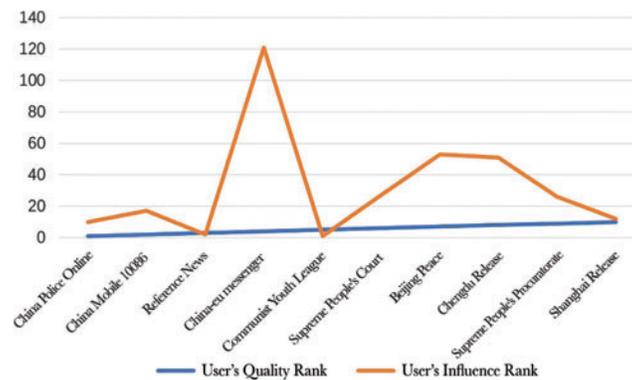


Figure 1: Comparison of user work quality ranking and user influence ranking

Fig. 2 shows that among the top 10 users with fans, the user influence of “China Mobile 10086”, “Central European Courier”, and “China National Geographic” ranks outside the 100. Analysis found that only a small part of the fans of these 3 users are among the 3000 users we selected, so in the user network we constructed, most of their fans are silent fans. In comparison, users with high influential rankings tend to have more fans among 3,000 users. For example, the “Communist Youth League Central Committee” has 1072 fans among these 3,000 users, and “China Fire” has 822 fans.

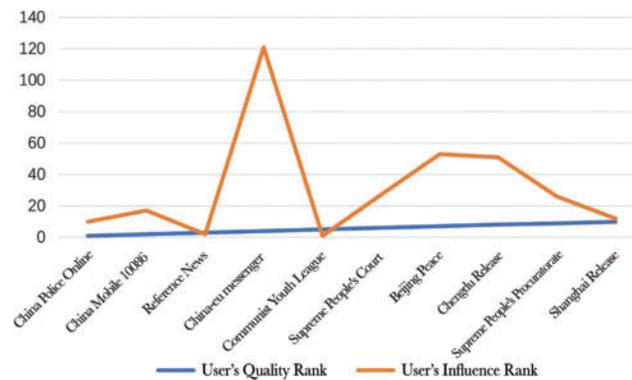


Figure 2: Comparison of user fan ranking and user influence ranking

In the user network of social networking sites, different levels of user interaction with each follower will lead to different distribution ratios. Among the top 10 users with direct influence, only the No. 1 “China Police Online” is not among the top 10 users. The “China Police Online” has 952 user networks, ranking second only to the “Central Committee of the Communist Youth League”. If calculated according to the traditional PageRank algorithm, its indirect influence should also be ranked high. However, through analysis, it is found that “China Police Online” frequently interacts with individual users. Among the 3000 government agencies’ official accounts, there are only 127 reposts, 23 comments, and 85 likes with “China Police Online”. Its indirect influence accounts for only 0.2% of the total influence. It can objectively reflect the transmission process of influence in social networks.

4.2.3 Analysis of User Structural Measurement Results

To analyze the results of user structural indicators, in order to better demonstrate the role of users in the entire social network, we use the louvain community discovery algorithm to divide 3000 users into 7 communities, of which community 1 has 1718 users, and community 2. There are 634 users, community 3 has 1154 users, community 4 has 494 users, community 5 has 224 users, community 6 has 164 users, and community 7 has 116 users. If a user connects with more users in different communities, it means that the structure in the network is stronger. Fig. 3 shows the connections between the top 5 structural users and 7 communities.

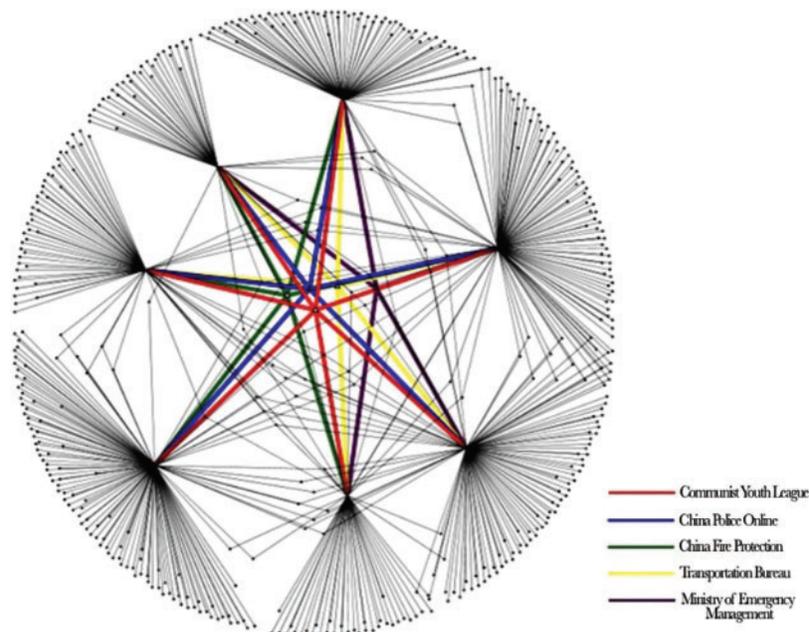


Figure 3: The top 5 structural indicators are the connections between users and communities

It can be found that the top 5 structural users are all connected to at least 5 communities, and these users assume the role of intermediary and transmitter in information dissemination. It plays an important role in the entire network structure of social networking site users.

4.3 Algorithm Result Comparison

The influence results obtained by the SNSKNIS algorithm and the FOLLOWERS RANK algorithm, the average forwarding algorithm, the v-index algorithm, and the linear threshold algorithm are compared, and the specific data of Weibo is compared and analyzed to compare the performance of various algorithms. Use the visualization method to display the corresponding rankings of the 10 most influential users in the SNSKNIS algorithm in other algorithms, as Fig. 4:

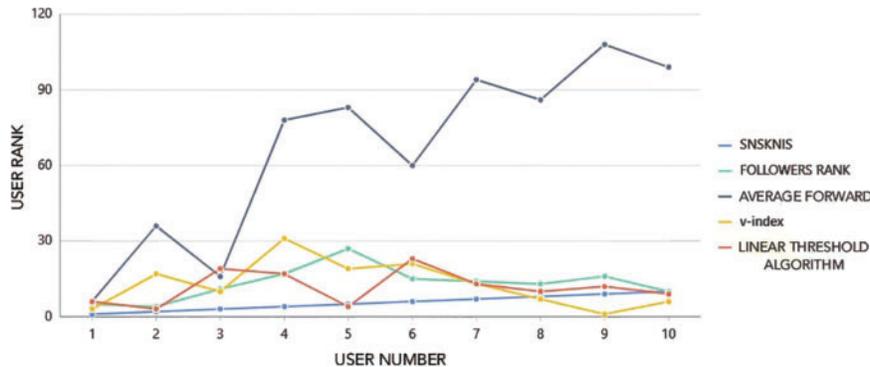


Figure 4: Ranking display diagram corresponding to the results of SNSKNIS algorithm and other algorithms

4.3.1 Comparison of SNSKNIS Algorithm and FOLLOWERS RANK Algorithm

The Tab. 2 below shows the comparison of the top 10 user rankings calculated by the SNSKNIS algorithm and the FOLLOWERS RANK algorithm.

Table 2: Top 10 comparison between SNSKNIS algorithm and FOLLOWERS RANK algorithm

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
1	Central Committee of the Communist Youth League	1	Weibo Constellation
2	Chinese police online	2	Central Securities Open Class
3	China Fire	3	Weibo Movement
4	Traffic Administration of the Ministry of Public Security	4	Chinese police online
5	Emergency Management Department	5	Douban movie
6	Criminal Investigation Bureau of the Ministry of Public Security	6	Central Committee of the Communist Youth League
7	Police and Citizens Together	7	Governor of Chang'an Avenue

(Continued)

Table 2: Continued

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
8	Healthy China	8	National Geographic
9	China Meteorological Administration	9	Tiger Punch Basketball
10	Released in Shanghai	10	Released in Shanghai

Through comparison, it is found that the rankings of most users in the two algorithms are not very different, but there are also cases where some users have a large difference in ranking between the two algorithms. The emergency management department, which ranks 5th in the PageRank algorithm, ranks 27th in the FOLLOWERS RANK algorithm, due to the relatively small number of fans of this user. After comparing the two algorithms, it is found that the FOLLOWERS RANK algorithm has certain limitations. Since the standard it measures is limited to the number of fans of the user, it cannot comprehensively measure the influence of the user. The SNSKNIS algorithm can comprehensively reflect the user's influence by comprehensively comparing the number of fans of the user, the quality and quantity of the user's works.

4.3.2 Comparison of SNSKNIS Algorithm and Average Forwarding Algorithm

The [Tab. 3](#) below shows the comparison of the top 10 user rankings calculated by the SNSKNIS algorithm and the average forwarding algorithm.

Table 3: Top 10 comparison between SNSKNIS algorithm and average forwarding algorithm

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
1	Central Committee of the Communist Youth League	1	Weibo task
2	Chinese police online	2	King of Wishes
3	China Fire	3	Mafengwo Official Weibo
4	Traffic Administration of the Ministry of Public Security	4	Estee Lauder
5	Emergency Management Department	5	National gang fight
6	Criminal Investigation Bureau of the Ministry of Public Security	6	Central Committee of the Communist Youth League

(Continued)

Table 3: Continued

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
7	Police and Citizens Together	7	Air Force release
8	Healthy China	8	Shanghai Disney Resort
9	China Meteorological Administration	9	National People's Congress
10	Released in Shanghai	10	torch of thought

Through comparison, it is found that the results of the two algorithms are quite different. Among the top 10 users of the SNSKNIS algorithm, only one user from the Central Committee of the Communist Youth League ranks in the top 10 of the average forwarding algorithm. In the average forwarding algorithm, the content posted by the top 10 users is the content that the masses are interested in, which is easy to arouse people's curiosity and desire to share, so they rank higher in the algorithm. However, this algorithm can only reflect the attractiveness of the user's Weibo, and cannot fully reflect the degree of influence of the user.

4.3.3 Comparison between SNSKNIS Algorithm and V-index Algorithm

The following [Tab. 4](#) shows the comparison of the top 10 user rankings calculated by the SNSKNIS algorithm and the v-index algorithm.

Table 4: Top 10 comparison between SNSKNIS algorithm and v-index algorithm

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
1	Central Committee of the Communist Youth League	1	First Financial
2	Chinese police online	2	Micro blog service
3	China Fire	3	Chinese police online
4	Traffic Administration of the Ministry of Public Security	4	Emergency Management Department
5	Emergency Management Department	5	China Telecom
6	Criminal Investigation Bureau of the Ministry of Public Security	6	Central Committee of the Communist Youth League
7	Police and Citizens Together	7	Didi Chuxing
8	Healthy China	8	Ping An

(Continued)

Table 4: Continued

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
9	China Meteorological Administration	9	Released in Shanghai
10	Released in Shanghai	10	Healthy China

Through comparison, it is found that the results obtained by the SNSKNIS algorithm and the v-index algorithm show that although the influence rankings of some users are relatively close, there are still certain differences. After analysis, it can be seen that the threshold of the v-index algorithm is the total number of microblogs multiplied by the set value, and the number of microblogs whose forwarding volume is greater than the threshold is calculated, but this algorithm does not consider the situation that the threshold is too large due to the excessive number of users' total microblogs, such as Public accounts such as China Police Online that publish multiple Weibo posts every day have too high a threshold, resulting in too few Weibo posts with a forwarding volume greater than the threshold. Therefore, compared with the SNSKNIS algorithm, the influence ranking of the v-index algorithm is less convincing.

4.3.4 Comparison of SNSKNIS Algorithm and Linear Threshold Algorithm

The [Tab. 5](#) below shows the comparison of the top 10 user rankings calculated by the SNSKNIS algorithm and the linear threshold algorithm.

Table 5: Top 10 comparison between SNSKNIS algorithm and linear threshold algorithm

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
1	Central Committee of the Communist Youth League	1	China Meteorological Administration
2	Chinese police online	2	Ping An
3	China Fire	3	Central Committee of the Communist Youth League
4	Traffic Administration of the Ministry of Public Security	4	Chinese environment
5	Emergency Management Department	5	China Telecom
6	Criminal Investigation Bureau of the Ministry of Public Security	6	Meituan takeaway

(Continued)

Table 5: Continued

SNSKNIS algorithm ranking	Username	Linear thresholding algorithm ranking	Username
7	Police and Citizens Together	7	Healthy China
8	Healthy China	8	China Transportation
9	China Meteorological Administration	9	China International Airlines
10	Released in Shanghai	10	Meteorological knowledge

Through comparison, it is found that the results obtained by the improved SNSKNIS algorithm and the linear threshold algorithm are partially similar, but the linear threshold algorithm also has significant defects. First, the threshold evaluation features are few. The linear threshold algorithm does not consider the hierarchical structure after the user's multi-dimensional features exceed the threshold, resulting in many influential official Weibo rankings with similar rankings and no good discrimination. Therefore, the influence ranking obtained by the SNSKNIS algorithm is more reliable.

5 Summary

This paper improves the traditional PageRank algorithm, introduces some structural indicators of nodes in complex networks, and designs a key node identification system for social networking sites based on PageRank. Through the experimental analysis of the Sina Weibo platform, the system can more accurately and objectively reflect the criticality of nodes in the entire network. The basic functions of the system meet the demand, but there are still some shortcomings. If the system interaction is low, user nodes cannot be added dynamically, and can only be imported through the back-end database; the supported social networking platform is single, currently the crawler script only supports the Sina Weibo platform as a data source, and it is necessary to write a data crawler or search for the corresponding social networking site. The corresponding api interface can continue to be studied for corresponding problems in the future.

Funding Statement: This study has been supported by Jiangsu Social Science Foundation Project (Grant No: 20TQC005). Philosophy Social Science Research Project Fund of Jiangsu University (Grant No: 2020SJA0500). The National Natural Science Foundation of China (Grant No. 61802155). The Innovation and Entrepreneurship Project Fund for College Students of Jiangsu Police Academy (Grant No. 202110329028Y). The “qinglan Project” of Jiangsu Universities.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] D. Fiala, F. Rousselot and K. Jezek, “PageRank for bibliographic networks,” *Scientometrics*, vol. 76, no. 1, pp. 135–158, 2008.

- [2] Y. D. Zhao, Y. Jia, B. Zhou and Y. Han, "Mining of topic influence individuals in Weibo based on multi-relational network," *China Communications*, vol. 10, no. 1, pp. 93–104, 2013.
- [3] N. Barbieri, F. Bonchi and G. Manco, "Topic-aware social influence propagation models," *Knowledge and Information Systems*, vol. 37, no. 3, pp. 555–584, 2013.
- [4] F. Qin, J. Y. Yan and K. Li, "Research on the influence of the content features of knowledge-based WeChat official accounts on personal use behavior," *Information Theory and Practice*, vol. 42, no. 7, pp. 106–112, 2019.
- [5] H. Chen, "Research on Weibo user influence ranking algorithm based on Hadoop," Ph.D. dissertation, East China University of Science and Technology, 2014.
- [6] K. Nallusamy and K. S. Easwarakumar, "Cgram: Enhanced algorithm for community detection in social networks," *Intelligent Automation & Soft Computing*, vol. 31, no. 2, pp. 749–765, 2022.
- [7] F. Amin, J. Choi and G. S. Choi, "Advanced community identification model for social networks," *Computers, Materials & Continua*, vol. 69, no. 2, pp. 1687–1707, 2021.
- [8] T. Maksim and K. Alexander, "Social network analysis methods and practice," in *Social Network and Connection Analysis*, 1st ed., vol. 1. China: Machinery Industry Press, pp. 8–10, 2013.
- [9] T. Li, "Research on the influence evaluation of university microblog and its application value," Ph.D. dissertation, Harbin Institute of Technology, 2015.
- [10] X. R. Zhang, X. Sun, W. Sun, T. Xu and P. P. Wang, "Deformation expression of soft T-tissue based on BP neural network," *Intelligent Automation & Soft Computing*, vol. 32, no. 2, pp. 1041–1053, 2022.
- [11] C. L. Gao, Z. Y. Sun, X. F. Xing and C. F. Li, "An energy efficient k-degree coverage algorithm in wireless sensor networks," *Computer Engineering and Applications*, vol. 52, no. 23, pp. 142–147+235, 2016.
- [12] L. Hong, J. Yin, L. Xia, C. Gong and Q. Huang, "Improved short-video user impact assessment method based on pagerank algorithm," *Intelligent Automation & Soft Computing*, vol. 29, no. 2, pp. 437–449, 2021.
- [13] Q. Wang, Y. Luo, H. Guo, P. Guo, J. Wei *et al.*, "A pagerank-based wechat user impact assessment algorithm," *Journal of New Media*, vol. 3, no. 3, pp. 89–98, 2021.
- [14] L. Page, *The Pagerank Citation Ranking: Bringing Order to the Web*. USA: Stanford Digital Libraries Working Paper, 1988. [Online]. Available: <http://google.stanford.edu/~backrub/pageranksub.ps>.
- [15] L. Guo, "Social network rumor recognition based on enhanced naive bayes," *Journal of New Media*, vol. 3, no. 3, pp. 99–107, 2021.
- [16] X. R. Zhang, J. Zhou, W. Sun and S. K. Jha, "A lightweight CNN based on transfer learning for COVID-19 diagnosis," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1123–1137, 2022.
- [17] M. M. Iqbal and K. Latha, "An effective community-based link prediction model for improving accuracy in social networks," *Journal of Intelligent and Fuzzy Systems*, vol. 42, no. 3, pp. 2695–2711, 2022.
- [18] P. Hui Li, J. Xu, Z. Yi Xu, S. Chen, B. Wei Niu *et al.*, "Automatic botnet attack identification based on machine learning," *Computers, Materials & Continua*, vol. 73, no. 2, pp. 3847–3860, 2022.