Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

Social Network Analysis Based on the Attention Preference of Weibo College Students

Xiao Yidin, Yan Jinyi*

School of Cyberspace Security, University of International Relations, Haidian District, Beijing, 100091, China

*Corresponding to: Yan Jinyi

Abstract: Social networks are characterized by speed, spread, equality and self-organization. With the help of social networks, users can obtain behavioral data based on massive data, which can help operators deeply understand the operation mode of social networks. Since the main active users of the Internet are young people, especially college students, it is of great significance to analyze the social network behavior of college students for the development of online social services in the future. This paper focuses on college students (aged 18-22) in the social platform "Weibo", uses web crawlers and information retrieval to obtain the data of college students' attention relationship and type preference, realizes the visual construction of the data with the help of social network analysis software such as Gephi, summarizes and infers the characteristics and trends of college students' attention preferences after comprehensive research and judgment, and provides help for the operation guidance of the social network platform.

Keywords: College Students' Attention Preferences, Social Networks, Data Visualization

1 INTRODUCTION

Online social networking is one of the main forms of entertainment for college students. Through the analysis of user behavior in social networks, we can gain an in-depth understanding of the characteristics of users and provide theoretical support for the operation of social network platforms.[1] For example, it is used for development, flow, application, and maintenance of upgrade services.

Weibo is one of the most influential social media platforms in China. According to the statistics of the "2023 Weibo Young User Development Report", there will be more than 130 million monthly active users of Weibo aged 16-22 in 2023. Among them, college students aged 18-22 are highly active and have rich interests and hobbies, covering celebrities, film and television variety, games, beauty and other fields. An in-depth understanding of college students' concerns and preferences plays an important guiding role in college education and teaching.

In the data acquisition part, this paper uses python to crawl the information of college students aged 18-22 on the Weibo open platform. In the visualization construction part, this paper uses the Social Network Analysis (SNA) method [2] Gephi was used to construct a social network on the acquired data, using the Louvain algorithm [3] The community is divided, and then the group structure and community characteristics of the network are analyzed. In the analysis of college students' attention preferences, this paper uses the case analysis method to select representative college students in the community to analyze the distribution of user attention.

There are three innovations in this paper: first, Weibo is used as the data collection source, focusing on the specific user group of college students, and obtaining real and rich social network data; the second is to take the common concern of college students as a way to generate social connections between users, rather than the traditional one; The third is to combine social network analysis with case analysis to conduct a comprehensive analysis of social networks from multiple dimensions, breaking through the limitations of traditional research from a single perspective.

2 RELATED THEORIES AND METHODS

2.1 BASIC CONCEPTS OF SOCIAL NETWORKS

In social network analysis, a complex network is usually represented by G=(V, E).[4] denoted as G, where represents the

Journal of Research in Multidisciplinary Methods and Applications
http://www.satursonpublishing.com/
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

set of nodes in the network and represents the set of edges in the network.$V = \{v_1, v_2, \dots, v_n\}$$E = \{e_1, e_2, \dots, e_m\}$

Common network structures such as Figure 1shown:

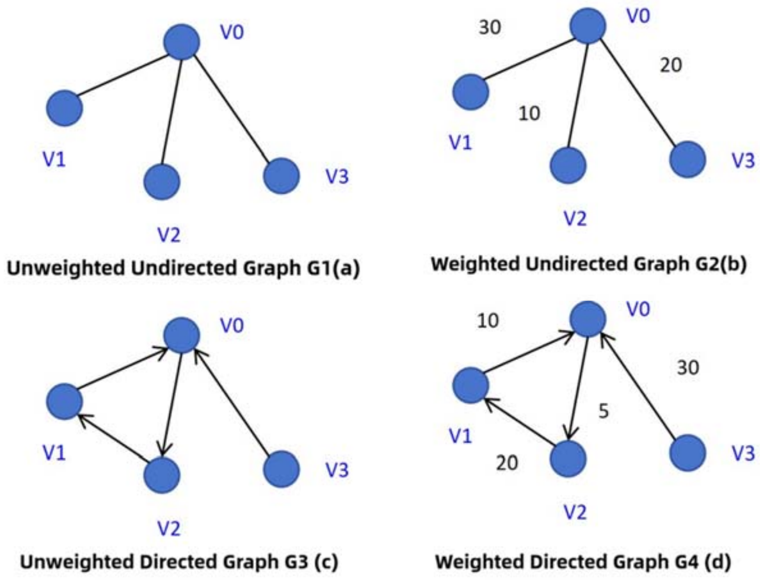

FIGURE 1 COMMON NETWORK STRUCTURE DIAGRAMS

2.2 *EVALUATION INDICATORS OF NODES*

Online social networks can generally be abstracted as G = (V, E), which indicates the number of direct connections between node i and other n-1 j nodes (j≠i, excluding the connection between i and itself), that is, the degree of the node$\sum_{j=1}^{n} x_{ij}$[5] . The evaluation index of nodes generally refers to degree centrality, mesocentrality and proximity centrality.

In order to eliminate the influence of network size on degree centrality, the method of wasserman and Faust was standardized to obtain the i-node degree Mental indicators [6] for

$$C_D(V_i) = \frac{1}{n-1}\sum_{j=1}^{n} x_{ij} \qquad (1)$$

$C_D(V_i) \in [0,1]$, the larger the value of the degree center of a node, the higher the degree of participation of the node in the network.

Intermediateness centrality [7] It focuses on how other nodes control or influence the relationship between two nodes that are not directly connected to it, which can be used as an important indicator to control network information exchange or resource flow. The mesonumber is measured by the number of times the shortest path between all nodes passes through a given node, and is normalized to the next section Point i's The intermediary indicator is:

$$C_B(V_i) = \frac{2}{(n-1)(n-2)}\sum_{j=1, j\neq k}^{n} \frac{n_{jk}(V_i)}{n_{jk}} \qquad (2)$$

$C_B(V_i) \in [0,1]$ to indicate the n_{jk} number of shortest paths between node j and node k, and the number of shortest paths between node j and node k passing through node$n_{jk}(V_i)$ i. The larger the intermediary value of a node, the more it can control or influence network relationships.

Similarly, from a global perspective, the ranking index of node importance in the network is close to centrality [8], which is mainly used to reflect the proximity of a node to other nodes in the network, measured by the shortest distance and size from one node to all other nodes. The normalized i-node is close to the centrality index for

$$C_C(V_i) = \left[\frac{1}{n-1}\sum_{j=1, j\neq i}^{n} d_{ij}\right]^{-1} \qquad (3)$$

$C_C(V_i) \in (0,1]$, the higher the value of the node near the center, the closer the relationship between the node and other nodes.

2.3 *COMMUNITY DISCOVERY ALGORITHMS*

Community Discovery (Community Detection) is a method of delving into communities of nodes found to have common characteristics in a multi-attribute network.[9] The application of community discovery algorithm in social network analysis is mainly used to discover user characteristics, predict user behavior, and help personalized social network recommendations.

Modularity is an important evaluation criterion for community discovery algorithms[10]It is used to evaluate the relative relationship between the tightness of the internal connections of the modules and the sparseness of the connections between the modules after the network is divided into different modules (communities).Louvain's algorithm[11]It is a cohesive clustering algorithm based on modularity theory proposed by

Journal of Research in Multidisciplinary Methods and Applications http://www.satursonpublishing.com/
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

Blondel et al., which is a typical method to divide communities in the network based on modularity maximization.

3 Data Acquisition And Preprocessing

In this paper, 300 were randomly selected Weibo user ID, crawled using python User information, strictly abide by relevant laws and regulations, and ensure the legal compliance of data. Data Collection The time span is from November 10 to 14, 2024, and the data obtained includes: UID, user nickname, gender, birthday, number of fans, number of followers, etc., such as Figure 2.

UID	Fans	Followers	Blog posts	Gender	Birthday	Registration time
7764330899	60	38	7	f	2004-10-29	2022/6/8 14:43
6636609042	21	478	239	f	2006-08-09	2018/8/15 15:57
6493638576	62	1025	142	f	2005-10-18	2018/2/22 20:22
7365731248	33	1082	1044	f	2005-11-25	2020/1/1 10:36
5493808261	87	224	400	f	2004-11-17	2015/1/24 12:51
7926224437	1	18	0	f	2005-11-19	2024/6/10 18:43
7739940037	11	175	70	f	2006-05-16	2022/2/7 15:27
7414566635	1	28	0	f	2006-03-15	2020/3/17 9:42
7396060522	4	308	1	f	2006-04-28	2020/2/20 20:44
7943801956	1	39	0	f	2006-08-09	2024/8/21 13:39
5073483906	21	61	168	f	2005-12-27	2014/3/16 18:29
6466494548	30	317	260	f	2003-01-13	2018/2/4 22:17
7672715650	10	34	6	f	2005-07-29	2021/8/17 10:02
7413097724	1	103	9	f	2005-01-29	2020/3/14 22:22
5642887152	1	87	56	f	2006-01-06	2015/6/29 17:33
7673975245	1	43	1	f	2006-04-19	2021/8/17 22:35
7947451129	1	49	3	f	2006-03-17	2024/9/1 17:32
7775505531	17	835	1330	f	2005-08-28	2022/7/22 22:35

Figure 2 Infographic Of Weibo Users

The above users were screened and 171 valid users were obtained. Set the number of followers/followers threshold to 3000 to remove abnormal data. Finally, 153 valid data were obtained.

Crawl the user's watchlist and follower list, including follow/fan ID, follow/fan nickname, etc., as shown in Figure 3.

UID	Follower's ID
7803104763	2166939692
7803104763	5259163942
7803104763	5576168164
7803104763	1112829033
7803104763	1597915381
7803104763	5636577946
7803104763	3893990834
7803104763	3193039841
7803104763	5578073848
7803104763	6436669966
7803104763	1673965152
7803104763	1669879400
7803104763	1878666200
7803104763	7595242776
7803104763	5472011311
7803104763	2794665971

Figure 3 User Watchlist

After the above steps, 307 valid tables (1 user personal information summary table, 153 user follow lists, and 153 user fan lists) were obtained, which was convenient for the subsequent construction of social networks.

4 Construction Of Social Networks For College Students On Weibo

4.1 *Construct An Adjacency Matrix Based On The Common Concern Relationship*

On the microblogging platform, the user's attention reflects the user's interests and preferences. If the adjacency matrix is constructed based on the relationship between users who follow and are followed, it is found that most of the elements are 0, which leads to poor network effect.

Therefore, this paper decides to use the common concern relationship as the basis for generating social connections between users. Let the total number of college students be N, and the common data set of college students is C, which is expressed as follows:

$$C = \{(i, j, \sigma(i \cup j)) | i, j \in \{1,2, \dots, N\} \land i \neq j\} \quad (4)$$

where i and j represent the UID of college students, and indicate the number of objects that user i and j are concerned about.$\sigma(i \cup j)$

If user i and user j have a matching follower, then +1 will be added to get the number of common concerns of user i and user j.$\sigma(i \cup j)$

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

The social network to be constructed in this paper takes college students as nodes, and each node has a unique user ID. If two users have more than one common interest object, there is an edge between the nodes. The number of objects of the same interest between two users is used as the weight of the edges. The greater the number of objects of the same concern, the greater the weight of the edges, indicating that the closer the relationship between the two users and the higher the similarity in terms of interests and hobbies.

adjacency matrix Adjacency Matrix is a matrix that represents the adjacency between vertices.[12] In the adjacency matrix A (G) of the proposed social network G = (V,E), the rows and columns represent the nodes, and the weights of the nodes and the edges between the nodes are, then A(G) can be expressed as:$V_i V_j \sigma(i \cup j)$

$$A_{ij} = \begin{cases} \sigma(i \cup j), & \text{If}, v-i.\ \text{with}, v-j.\ \sigma(i \cup j)\ \text{times adjacent in G} \\ 0, & \text{If}, v-i.\ \text{with}, v-j.\ \text{Not adjacent in G} \end{cases} \tag{5}$$

Obviously, the social relationships of college student users can be represented by a power undirected network diagram.

4.2 *Jointly Focus On The Visual Construction Of Social Networks*

Figure 4 shows that some of the adjacency matrices that college students are interested in are obtained.

	7803104763	5472534648	5736976152	7504005082	6720721144	5608966087
7803104763	0	0	0	0	0	1
5472534648	0	0	3	0	1	5
5736976152	0	3	0	0	0	5
7504005082	0	0	0	0	1	0
6720721144	0	1	0	1	0	1
5608966087	1	5	5	0	1	0

Figure 4 Adjacency Matrix Diagram Of College Students' Common Concerns

The co-adjacency matrix A (G) and the node table (the list of users' personal information) are imported into Gephi to draw a social network as shown in Figure 5, where one point represents a college student user, and if there is an edge between the two points, it means that the user has common concern.

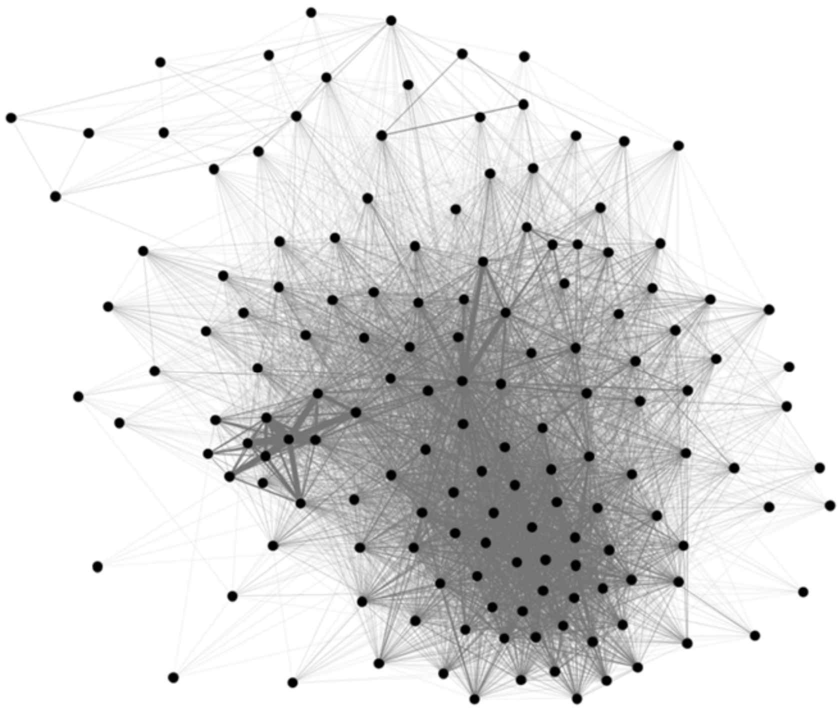

Figure 5 Importing The Social Network Graph Of Gephi College Students

It can be seen that there are many edges in the network, and the connection between nodes is too close. On this basis, the threshold is set to 15 according to the weight of the edges, and the above networks are filtered, and the weighting degree of the initial adjacency matrix A(G) nodes is used [13] The size is used as the size of the node to get the new social network as shown in Figure 6.

Journal of Research in Multidisciplinary Methods and Applications
http://www.satursonpublishing.com/
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

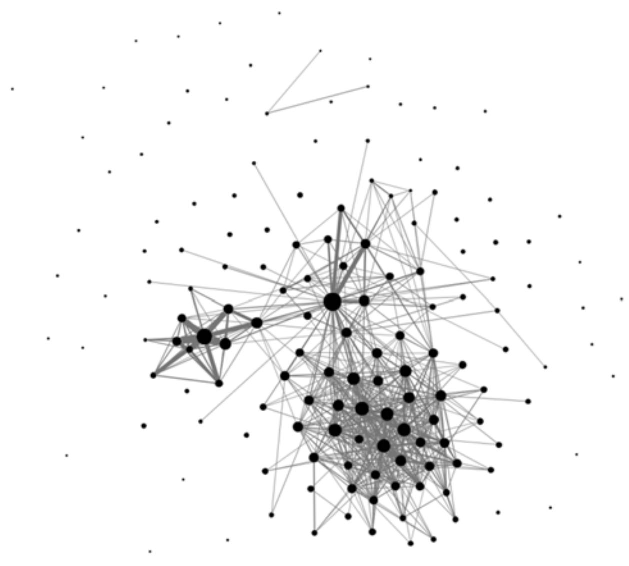

FIGURE 6 THE SOCIAL NETWORK OF COLLEGE STUDENTS AFTER FILTERING SOME VIRTUAL SOCIAL NETWORKS

The network in Figure 6 is partially filtered compared to Figure 5, where the node size represents the weighting degree of the node, and the thickness of the edge represents the weight of the edge. The larger the node, the more users have the same interest as more users, and their interests may be broad or popular; The thicker the edge between two nodes, the closer the relationship between the two users can be considered, and the higher the similarity in terms of interests and information acquisition.

4.3 *ANALYSIS OF THE BASIC CHARACTERISTICS OF SOCIAL NETWORKS*

4.3.1 *ANALYSIS OF COLLEGE STUDENT USER NODES*

In the undirected authority graph, the higher the degree centrality, the more direct connections the node has in the network, and the greater its importance and influence.[14].

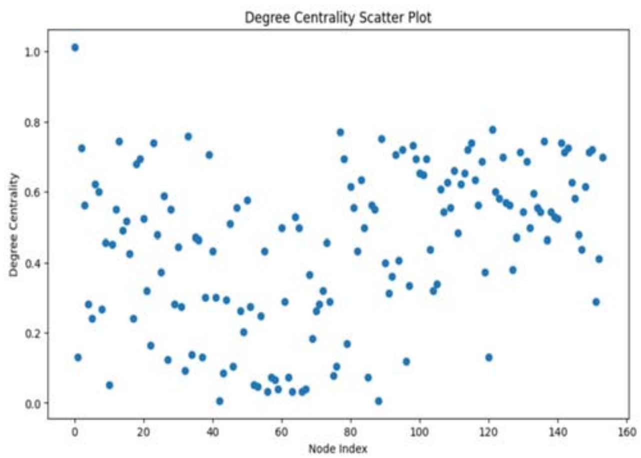

FIGURE 7 SCATTER PLOT OF NODE DEGREE CENTRALITY OF COLLEGE STUDENTS

The scatter plot of the centrality of the node degree of college students is shown in Figure 7. The 153 nodes were numbered from 1 to 153, and it was found that the nodes numbered 1 had the highest degree centrality, and the three nodes numbered 26, 78, and 121 had higher degree centrality, indicating that these college students had a wide range of social attention and influence.

Nodes with high intermediary centrality are often on the critical path of information propagation [15], which acts "Intermediary" The more times you go, the more you can guide the direction of the topic and the direction of the spread of the message.

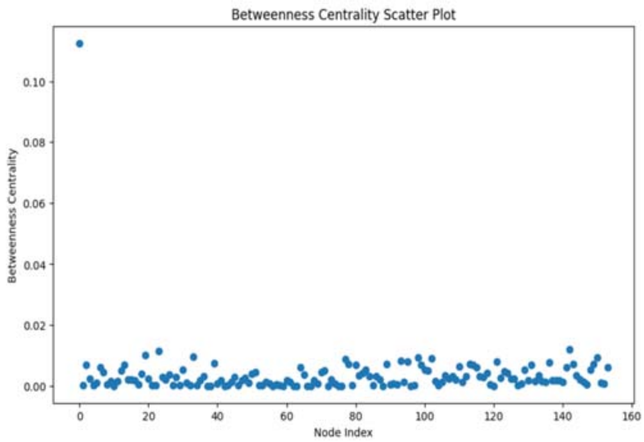

FIGURE 8 SCATTER PLOT OF THE INTERMEDIATENESS CENTRALITY OF THE NODE OF COLLEGE STUDENTS

Figure 8 is a scatter plot of the intermediateness centrality of the node of the college student user. Node 1 has the highest betweenness centrality and high degree centrality, indicating that the college student users not only have a wide range of social attention, but also play an important "intermediary" role in the network, which can influence the trend of other users' attention preferences.

The higher the proximity centrality of a node, the closer it is to other users.

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

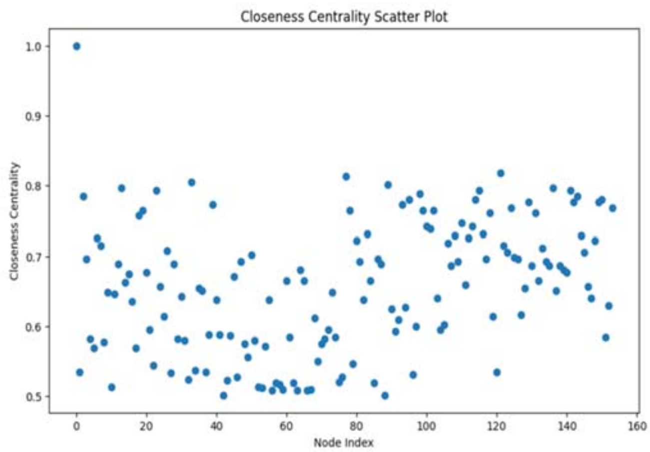

FIGURE 9 SCATTER PLOT OF COLLEGE STUDENT USER NODES APPROACHING CENTRALITY

The scatter plot of the near-centrality of college student user nodes is shown in Figure 9. Except for node 1, the proximity centrality of the other nodes is between 0.5 and 0.8. Nodes with higher near-centrality have higher similarity to the nodes they are connected to, and are likely to have similar attention preferences.

4.3.2 OVERALL SOCIAL NETWORK ANALYSIS

In a social network, network density density） [16] It is used to describe the density of interconnected edges between nodes, ranging from 0 to 1. For undirected the diagram defines the network density as:

$$Density = \frac{R}{N(N-1)/2} \quad (6)$$

where N represents the total number of nodes and R represents the number of edges.

The network density of the social network without the threshold is Density=0.439, and the social network Density' = 0.048 after the threshold is set. The results show that the network density is significantly reduced after removing some false connections in the social network of common interest. In fact, this may be due to the large number of Weibo college students, and the attention behavior has a certain randomness and individuality.

Social networks with low network density will lead to slower information dissemination and fewer opportunities for users in different circles to interact, which will reduce the activity of social networks. But it also gives social networks more flexibility.

The clustering coefficient is used to measure the degree of node aggregation in a social network, indicating the probability that two nodes adjacent to each other will be adjacent to each other, and its range is between 0 and 1.[17]

The clustering coefficient of node i in the network is defined as:$k_i C_i$

$$C_i = \frac{E_i}{(k_i(k_i-1)/2} = \frac{2E_i}{k_i(k_i-1)} \quad (7)$$

where is the number of edges that actually exist between the neighboring nodes of node i.$E_i k_i$

The clustering coefficient C of a social network is the average of the clustering coefficients of all nodes in the network.

The clustering coefficient of the social network without setting the threshold is C=0.740, and the scatter plot of the clustering function of the nodes is plotted as shown in Figure 10.

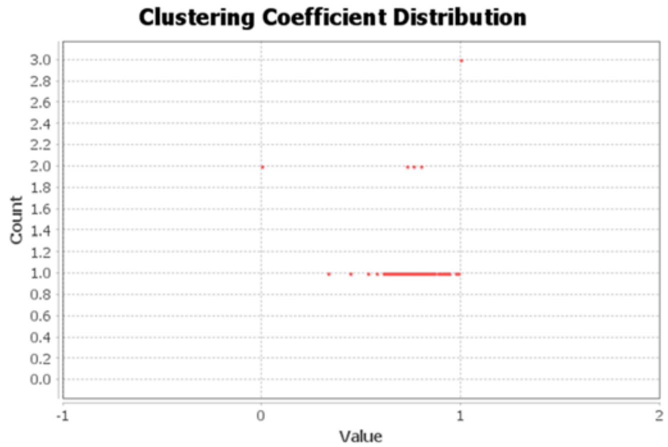

FIGURE 10 SCATTER PLOT OF NODE CLUSTERING FUNCTION WITHOUT THRESHOLD

The clustering coefficient C'=0.635 of the social network after the threshold is calculated, and the scatter plot of the clustering function of the nodes is plotted as shown in Figure 11.

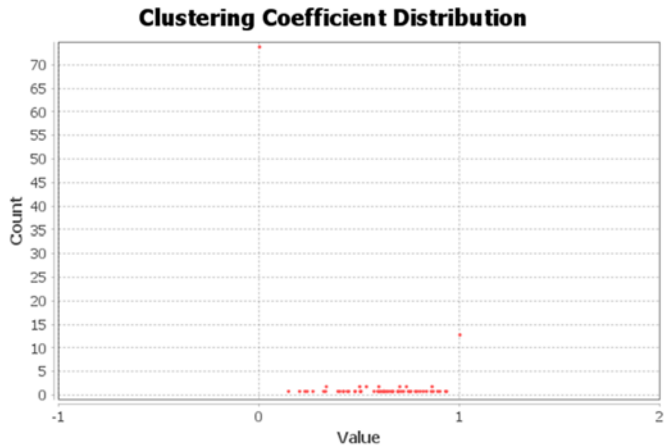

FIGURE 11 SCATTER PLOT OF NODE CLUSTERING FUNCTION AFTER SETTING THRESHOLDS

The results show that the social network has certain clustering characteristics regardless of whether the threshold is set or not. This reflects that the Weibo college student user group tends to form a close-knit small group, and the interaction and

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

communication between members of the group are more frequent.

The main reason for the formation of this clustering phenomenon is that the attention behavior of college students on Weibo is often based on factors such as hobbies and regions, and members of the group have common topics and hobbies. But this can also lead to limitations in the dissemination of information, so that some valuable information cannot be disseminated throughout the social network.

5 Segmentation And Analysis Of Social Network Communities

5.1 *Results Of Community Division*

In this paper, the Louvain algorithm is used to divide the community of the social network after the set threshold, and the distribution of the number of people in each community is obtained, as shown in Figure 12.

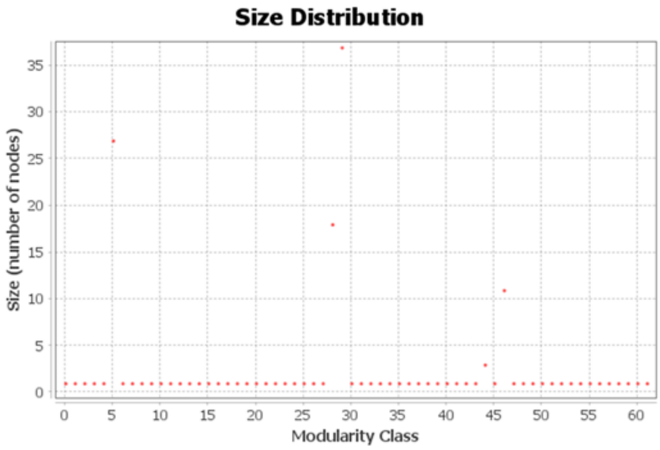

Figure 12 Distribution Of The Number Of People In Each Community After Setting The Threshold

After the community division, 62 communities were obtained, which were represented by 0-61. Since there are 57 individual nodes (no edges connected to them, and the number of nodes in the community is 1) after setting the threshold to remove some possible false social relationships, the study can be ignored and only focus on the analysis of the 5th, 28th, 29th, 44th, and 46th communities. Now different colors are used to represent different communities, and the optimized social network is obtained as shown in Figure 13.

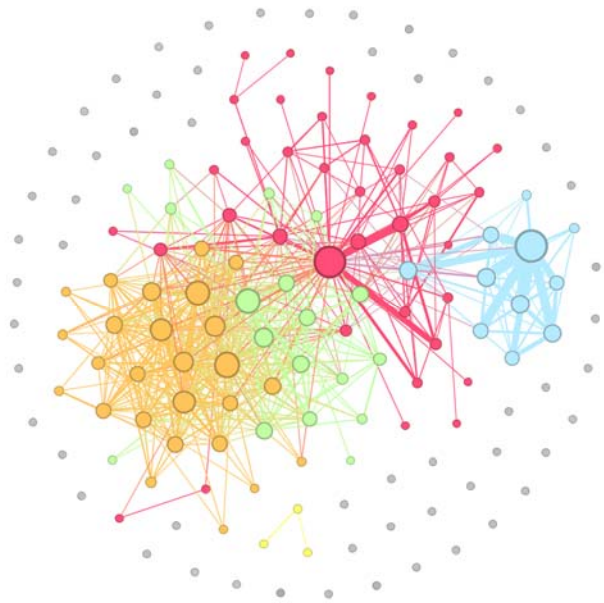

Figure 13 Social Network Diagram Of College Students After Community Division

Most of the nodes in the five communities are clustered, not only connected to nodes in their own community, but also connected to nodes in other communities.

In the community discovery algorithm, the larger the modularity Q, the better the community partitioning effect, and in the actual social network, the Q value is usually between 0.3 and 0.7[18] .

Figure 5-1-2Q'=0.452, and the community is divided into social networks without a threshold The distribution of the population in each community is shown in Figure 14.

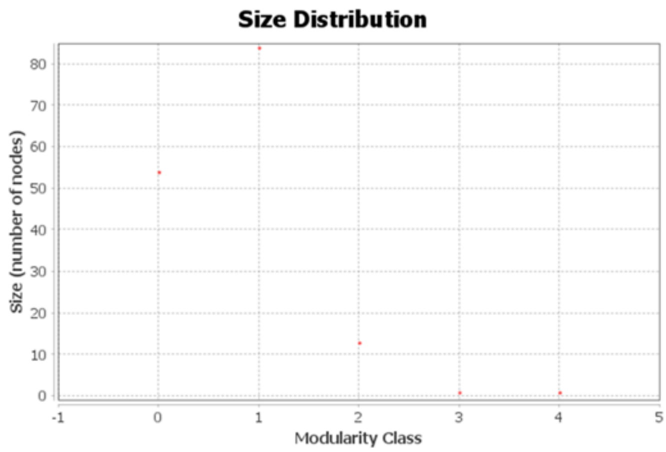

Figure 14 Distribution Of The Number Of People In Each Community After The Community Division Without Setting Thresholds

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

The social networks without the threshold were divided into 5 communities, and the module value Q=0.254. Among them, there are only three active communities, and two of them have more than 50 nodes. Therefore, setting thresholds can effectively remove virtual social relationships and make the community structure clearer. Since the module value Q'=0.452 is between 0.3 and 0.7, it indicates that the results meet the normal criteria.

5.2 ANALYSIS OF COMMUNITY STRUCTURE CHARACTERISTICS

After dividing the community, the number of nodes in each community was different: the proportion of nodes in the 5th, 28th, 29th, 44th, and 46th communities were 16.99%, 11.76%, 24.84%, 1.96%, and 7.19%, respectively, and the node colors were orange, green, red, yellow, and blue, respectively. The rest of the communities are single-node communities, accounting for 0.65%, and the node color is gray. Communities 5 and 29 are more populous, and such large communities typically cover a wide range of interests and may involve multiple hot areas. Smaller communities such as the 44th and 46th communities have more niche interests and hobbies, and the exchanges within the community are more in-depth.

In this paper, typical community members in five communities are selected for feature analysis, focusing on the attention preferences of college students. According to the weighting degree of nodes in each community, the top 20% of nodes with the largest weighting degree in each community were selected as typical individuals to analyze the neighborhoods of interest that they were concerned about.

The 5th, 28th, 29th, 44th, and 46th communities are selected as the top 5, 4, 8, 1, and 2 nodes with the largest weighting, respectively, and the corresponding college student user IDs are shown in Table 1 below.

TABLE 1 IDS OF TYPICAL COLLEGE STUDENTS IN FIVE COMMUNITIES

Community number	User ID	Weighted degree
5	5527854572	1020
5	5662331857	938
5	5865574382	834
5	6993372684	830
5	7550366779	744
28	6325844372	992
28	7015013638	546
28	5856254783	520
28	5863952669	676
29	6451253026	1394
29	7829623990	484
29	6466494548	432
29	5404010041	412
29	5472534648	350
29	7868222500	232
29	5225362387	230
29	1820562005	224
44	6079581443	40
46	7449316984	1426
46	5093116706	628

After analyzing the user areas of interest in the above table, the corresponding top 5 sectors of interest are drawn, as shown in Figure 15.

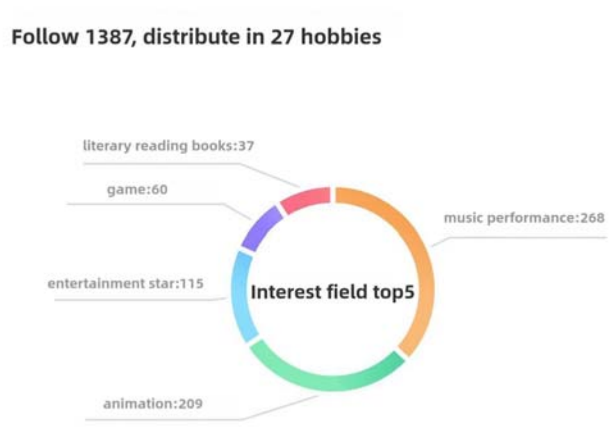

FIGURE 15 TOP 5 SECTORS OF INTEREST THAT USERS FOCUS ON

Figure 16 shows the distribution of areas of interest of typical users.

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

Community number	User ID	top1	top2	top3	top4	top5
5	5527854572	Music performance	Animation	Entertainment stars	Game	Cultural reading
5	5662331857	Animation	Entertainment stars	Cultural reading	Game	Life
5	5865574382	Animation	Life	Music performance	Game	Cultural reading
5	6993372684	Animation	Entertainment stars	Cultural reading	Life	Game
5	7550366779	Animation	Game	Entertainment stars	Cultural reading	Life
28	6325844372	Music performance	Entertainment stars	Animation	Life	Funny humor
28	7015013638	Animation	Entertainment stars	Music performance	Funny humor	Game
28	5856254783	Music performance	Entertainment stars	Game	Funny humor	Animation
28	5863952669	Animation	Music performance	Entertainment stars	Game	Life
29	6451253026	Entertainment stars	Music performance	Variety show	Cultural reading	TV play
29	7829623990	Entertainment stars	Music performance	Sports	Movie	Internet
29	6466494548	Entertainment stars	Cultural reading	Life	Animation	Face score
29	5404010041	Music performance	Entertainment stars	Funny humor	Animation	Face score
29	5472534648	Entertainment stars	Game	Music performance	Sports	Funny humor
29	7868222500	Life	Entertainment stars	Funny humor	Music performance	Movie
29	5225362387	Entertainment stars	Music performance	Internet	Game	Animation
29	1820562005	Entertainment stars	Game	Funny humor	Animation	Life
44	6079581443	Game	Animation	Sports	Entertainment stars	Life
46	7449316984	Entertainment stars	Internet	Variety show	Animation	Music performance
46	5093116706	Entertainment stars	Face score	Internet	Animation	Beauty makeup

FIGURE 16 TYPICAL USERS' AREAS OF INTEREST ARE DISTRIBUTED

The number of users in the fifth community is 26, the network density of the community is 1.000, the average weighting degree of nodes is 433.926, and the average clustering coefficient is 1.000. The members of this community communicate with each

Journal of Research in Multidisciplinary Methods and Applications http://www.satursonpublishing.com/
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 3, March 2025

other frequently, and the members gather together because they are also two-dimensional enthusiasts, and the social circle is active and stable.

The number of users in the 28th community is 18, the network density of the community is 0.987, the average weighting degree of nodes is 253.111, the average clustering coefficient is 0.987, and the user's attention preference is entertainment such as music performances and stars. Although college students in the community prefer entertainment topics, due to the difference in music taste and celebrity preferences, the degree of closeness and activity within the community is average.

The number of users in the 29th community is 38, the network density of the community is 0.730, the average weighting degree of nodes is 233.189, and the average clustering coefficient is 0.814. Although the community has the largest number of people, indicators such as network density are small, and it is found that members have a wide range of areas of interest, involving various genres such as celebrities, games, film and television, sports, etc., and there is relatively little contact between members.

The number of users in the 44th community is 3, the network density of the community is 1.000, the average weighting degree of nodes is 34.677, the average clustering coefficient is 1.000, and the users' attention preferences are mainly games, sports and other aspects of life. This community is the least numerous, probably due to the smaller niche of members' concern preferences. Members share common concerns, but the strength of the ties is modest.

The number of users in the 46th community is 11, the network density of the community is 1.000, the average weighting degree of nodes is 505.189, and the average clustering coefficient is 1.00. Although the number of people in the community is small, the connection between users is very close, and the analysis finds that most of the users in the community are fans of a specific star, gathered together because of their love for the same star, they frequently post star-related topics, and interact more with each other, the connection within the community is very close, and the information spreads rapidly.

The analysis of the connections between communities helps to understand the structure of the entire social network and the information dissemination path. In almost every community, there are users who follow popular areas such as celebrities and games. The 5th community is connected to the 28th, 29th, and 46th communities, and the 28th community is more closely connected, reflecting the frequent information transmission between the two communities. The 28th neighborhood has connected edges with the 5th and 29th neighborhoods. The 29th community has connected edges with the 5th, 29th, and 46th communities, which indicates that the community will transfer information to other communities. Community 44 has no connected edges with other communities, and this community has the smallest number of people, indicating that the members of this community are closely connected, the internal structure is stable, and it is not affected by the spread of information from other communities. Community 46 has connected edges with communities 5 and 29, but the number of connected edges is less, and there are fewer connections with other communities.

Through the in-depth analysis of the community size, the characteristics of typical members' attention preferences, internal connections and the connections between communities, we can have a more comprehensive and in-depth understanding of the structure and characteristics of Weibo college students' common attention to social networks, further explore the attention preferences and social behaviors of college students, provide educational guidance for schools and other organizations, and provide opinions for the operation of social platforms.

6 Summary And Outlook

Combined with the current situation of the deep integration of social networks and community divisions of college students' social concern, this paper collects and analyzes a large number of real microblog data, constructs an adjacency matrix of college students' common concern relationship, and further visualizes and clarifies it after setting thresholds, filtering and other steps, and obtains a social network composed of nodes and edges. Then, this paper analyzes the graph from scattered to whole, from micro to macro by computational centrality, median centrality, close centrality to network density, clustering coefficient and other indicators, and uses Louvain's algorithm to divide the communities, and deeply investigates the structural characteristics of the communities, as well as the modularity and the connection between the communities of each community. The results show that the social network of college students is relatively close, and it shows the characteristics of high community and centripetal. On the whole, the attention preferences of college students are diverse and dynamic. At the same time, it also shows that the community algorithm has certain advantages in the overall observation and classification of college students' social networks. However, there are still some shortcomings in this study, such as a small sample of users and a relatively single observation standard for users' social behaviors, so the sampling range and sample size can be expanded to observe different social behaviors of users, and the attention preference can be further investigated by combining text mining and other methods.

The results show that educational organizations, social software and push backend should pay special attention to key nodes, for example, by understanding the attention preferences of college students, appropriately adjusting the teaching methods, appropriately improving the field breadth and information density of users with high node indicators, or improving the efficiency of information dissemination, and to a certain extent, soften the hard junction caused by personal preferences and information cocoons in various communities, and promote in-depth communication between various groups.

Funding

Student Academic Research Training Project of University of International Relations (No.3262024SWA10)

REFERENCES

[1]Zhao Weiya, Tian Junjing, Zhu Jianxun, et al. Research and Application of Social Network Forensics Method Based on Microblogging Platform[J]. Network Security Technology and Application,2024,(1):129-132.

[2]ZHAO Qian. Research on Data Mining Method Based on Social Network Analysis[J]. Journal of Anhui Police Officer Vocational College,2024,23(5):124-128.

[3]HU Zhangrong. Research on Community Discovery in Social Networks Based on Louvain Algorithm[J]. Computer Knowledge and Technology,2020,16(23):197-198.

[4]ZHAO Lifei. Network characteristics and interpersonal relationship prediction of people with high social exclusion: based on social network analysis[D].Sichuan:Department of Psychology,Sichuan Normal University,2023.

[5]ZHANG Dayong,ZHANG Yifan. Statistical Analysis of Node Centrality and Correlation Degree of Online Social Network[J]. New Media Research,2022,8(24):8-15.

[6]WANG Tongtong,LI Shengen,WANG Gang. Mining its Community Framework Based on the Centrality of Social Network Nodes[J]. Journal of Computer Applications and Software,2016,33(7):83-87.

[7]WANG Tongtong,LI Shengen,WANG Gang. Mining its Community Framework Based on the Centrality of Social Network Nodes[J]. Journal of Computer Applications and Software,2016,33(7):83-87.

[8]Wu Ruizhong. Research on the calculation method of between-centeredness of road network[D]. Guangdong:Guangzhou University,2023(in Chinese).

[9]ZHANG Dayong,ZHANG Yifan. Statistical Analysis of Node Centrality and Correlation Degree of Online Social Network[J]. New Media Research,2022,8(24):8-15.

[10]XU Hefei. Research on local community discovery method in location information social network[D].Anhui:Department of Computer Technology,Anhui University,2023.

[11]Xu Wei, Lin Baigang, Lin Sijuan, et al. Research on Social Network Community Discovery Method Based on User Interaction Behavior and Similarity[J]. Netinfo Security,2015,(7):77-83.]

[12]ZHAO Meng, LI Zichao, GAO Mei, et al. Consensus Model of Large Group Decision-making Interaction in Social Network Based on Louvain Method[J]. Journal of Engineering and Management Engineering,2021,35(4):152-161.

[13]YAN Weimin,LI Dongmei,WU Weimin. Data structure[M].2nd Edition. Beijing: People's Posts and Telecommunications Press, 2015.

[14]Centrality Weighting Algorithm of Social Network Based on Spark Platform in Different Cultural Environments[J]. Journal of Guangdong University of Technology,2017,34(3):15-20,48.

[15]Ju Chunhua,Zhao Kaidi,Bao Fuguang. Computational Model of User Influence Intensity in Social Network Integrating Compactness, Centrality and Credit[J]. Journal of Information Technology,2019,38(2):170-177.

[16]ZHANG Sai,XU Ke,LI Haitao. Measurement and Analysis of Information Dissemination in Microblog Social Networks[J]. Journal of Xi'an Jiaotong University,2013,47(2):124-130.

[17]P. W. Holland and S. Leinhardt. Transitivity in structural models of small groups[J]. Comparative Group Studies, 2005,2（2）:107-124.

[18]SHAO Siqi. Research on the Nature of Clustering Snowball Sampling Algorithm in Large Social Networks[D]. Shandong:Qufu Normal University,2024(in Chinese).