Recognition of Self-organized Aggregation Behavior in Social Networks Based on Ant Colony Algorithm

Nan Hu(✉) and Hongjian Li

China University of Labor Relations, Beijing 100048, China
Yingsongvip@163.com

Abstract. In order to effectively detect the real network community structure and improve the accuracy of user stage partitioning to the corresponding self-organized community, a self-organized clustering behavior recognition method based on ant colony algorithm is proposed. According to user's individual attribute and collaborative attribute, the node with high aggregation coefficient under user's knowledge quality scale is chosen as the core to construct social network aggregation behavior community. The evolutionary types of group trajectory are divided into seven types. Ant colony algorithm is used to track the group trajectory. Abstract tagged basic events from user attributes, establish recognition model to identify abnormal behavior, and realize self-organized aggregation behavior recognition in social network. Experimental results show that the self-organized aggregation recognition method based on ant colony algorithm can get more reasonable group structure, better quality of community partition, and improve the accuracy of user stage partition to the corresponding self-organized community.

Keywords: Ant colony algorithm · Social network · Self-organization · Aggregation behavior · Behavior recognition · User characteristics

1 Introduction

With the vigorous development of information technology and computer technology, the world is developing e-commerce industry, electronic payment in China has also been widely popularized. The communication between people has become simpler and more convenient, and the spread of information has broken through the geographical restrictions to become faster and more convenient, so that people's daily lives are more and more colorful, but also to further promote people-to-people exchanges and information sharing. Various kinds of social media have sprung up, changing the way people live, and the circle of people's lives is becoming more and more complex. Microblogs, short videos, live broadcasts and other forms of information are spreading at an alarming rate. Because of the increasing number and complexity of these connections, the society of life is becoming more and more "networked", that is to say, these "invisible" connections make the individual, the collective, the city, the country and even the whole world, these

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 469, pp. 84–98, 2023.
https://doi.org/10.1007/978-3-031-28867-8_7

human societies, large and small, closely linked together, making the society a network covering every participant. With the development of specialization, longitudinalization and intellectualization of information media, the platform of information communication is gradually moving towards mobile Internet. Countless "networks" aimed at connecting everyone are being created and constructed so that everyone around the world can participate and benefit from them, and the development of networks is becoming more and more "social." A network community structure with the characteristics of high cohesion and low coupling is formed between the changes and interactions of network members, which is composed of user member nodes [1]. Mining the characteristics of these community structures is of great significance to network research.

There are some common characteristics among complex networks, such as small world, scale-free, high aggregation and strong community structure. The research on complex networks has been a hot topic in many fields. Nodes in the community usually have some common attributes, which reflect the local regularity and global order of the social network to some extent. Community structure refers to the cluster of similar nodes in a complex network, that is, there are frequent links between nodes in the community and scattered links between nodes outside the community. Analyzing the community structure of complex networks is not only helpful to discover the potential relationship and function of complex networks, but also has great theoretical research value and practical application. A social network is a complex network, which is based on human society. Man is the main participant, but it is not an individual, or a group of people, or even a combination of people and things. For example, the participants in a cooperative network can be researchers, research institutes, or enterprises. It is not only the basis of the evolution of social network, but also can promote the development of related applications, such as recommendation system, privacy protection, network marketing and so on. It is of great significance to study the fission and reconstruction of network community system for the development of its information ecology.

At present, scholars in related fields have carried out research on the identification of relevant behaviors in social networks. Reference [2] proposes a dynamic behavior analysis and online detection method for navy users in social networks. Construct the dynamic behavior characteristics of social network users, and analyze the differences between normal users and navy users. Based on the semi-supervised model, combined with dynamic and static behavior characteristics, an online detection model is constructed. Through static behavior feature clustering and dynamic behavior feature filtering, the semi-supervised model uses the most valuable unlabeled user data to perform incremental learning to detect navy users. The average training time of this method is shorter, but the recognition accuracy of this method is lower. Reference [3] proposed an accurate identification method of inappropriate behaviors of naval user teams in social networks. Describe the social network and extract the dynamic characteristics of user behavior. The problem of identifying inappropriate behaviors of naval user teams in social networks is regarded as a binary classification problem, and the corresponding samples of dynamic features are extracted as inputs to build a decision tree. The decision tree is used to identify the misconduct of the naval user team on the new social network data set. This method has certain feasibility, but its recognition effect is poor.

Therefore, this paper proposes an ant colony algorithm based recognition method of self-organized aggregation behavior in social networks, according to the user's individual attribute and collaborative attribute characteristics, build a social network aggregation behavior community, track the group trajectory using ant colony algorithm, establish a recognition model to identify abnormal behaviors, and realize self-organized aggregation behavior recognition in social networks. The group structure obtained by the proposed method is more reasonable, the quality of community division is higher, the community structure of the real network can be effectively detected, and the accuracy of user stage division into the corresponding self-organized community can be improved.

2 Ant Colony Algorithm Based Recognition Method for Self Organized Aggregation Behavior in Social Networks

2.1 Analysis of User Attribute Characteristics in Social Networks

When users choose to participate in self-organizing groups in social networks, they should consider not only the individual attributes of users, but also the cooperative attributes of members. The user individual attributes and collaboration attributes considered in this article are shown in Fig. 1.

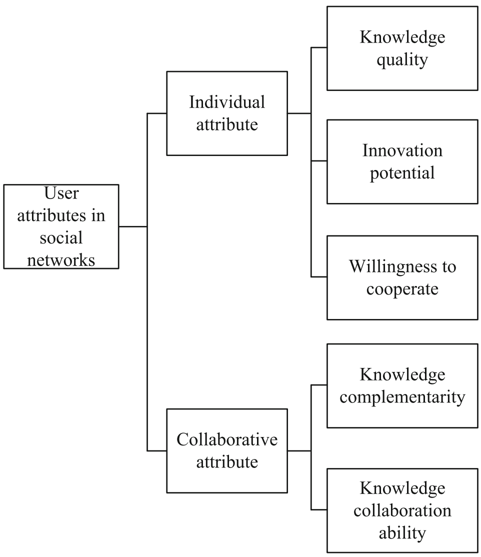

Fig. 1. Characteristics of user attributes in social networks

Individuals in social networks form a "relational structure" through various connections and produce a lot of information dissemination. This "relational structure" is the self-organized agglomeration behavior community. Self-organized agglomeration behavioural communities are an important structure of complex networks, which are composed of multiple communities, where nodes are very similar and compact, and where the connectivity between communities is fragmented [4]. Self-organizing aggregation in social networks requires a range of knowledge. User's knowledge quality includes situational knowledge quality, intrinsic knowledge quality and accessible knowledge quality. Since the target knowledge domain is not specified, only the intrinsic knowledge quality and accessible knowledge quality of user content are measured here. The knowledge quality of users can be quantified as follows:

$$A = \sum \alpha \ln \beta e^{-t/w} \tag{1}$$

In formula (1), A represents the user's knowledge quality; α represents the value coefficient; β represents the post over weight; t represents time; w represents the decay coefficient of knowledge; and e is a natural constant. Users have different knowledge potentials because of their knowledge structure. Some users have a deep understanding of a specific knowledge domain, while some users have only a little or no knowledge of that domain. Community structure can also be represented by user knowledge structure, and a matrix is constructed by the membership coefficient between each node and each community. In complex networks, communities can be divided by simply representing nodes of the same community with the same knowledge tag. Through the community structure, we can more intuitively reflect the relationship between the two sides of the network, and at the same time we can find the relationship between the community. The different structure of a user's knowledge determines its innovation potential. The innovation potential of users can be quantified from knowledge breadth and knowledge depth. When the user's interest in knowledge participation is high, the task of self-organization aggregation will advance more smoothly. The user engagement motivation indicator is used to quantify the user's willingness to collaborate, with the following expression:

$$B = \ln c \ln h \tag{2}$$

In formula (2), B represents the user's willingness to cooperate; c represents the number of replies of the user within the time t; and h represents the points of the user within the time t. In order to balance the distribution of data, both of them are processed logarithmically. The quality of users' knowledge in a social network has different impacts on the network depending on their location, and the closer the user is to the network center, the higher the importance of the user, the greater the impact on other nodes of the network, and the greater the value generated [5]. If the knowledge situation between users is too different, it will affect the knowledge communication between users, and even increase the cost of negotiation between users. Therefore, the complementarity of user knowledge will be considered. The knowledge complementarity coefficients of user s_1 and user s_2 can be expressed as follows:

$$\begin{cases} r = \sum g e^{-t/w} \\ g = |z(s_1) - z(s_2)| \\ z = \beta p \end{cases} \tag{3}$$

In formula (3), r represents the coefficient of knowledge complementarity between user s_1 and user; g represents the comparative advantage of knowledge; z represents the value of knowledge capability of user; and p represents the knowledge stock of post text, $z(s_1)$, $z(s_2)$ represents the knowledge ability value of user s_1 and user s_2, respectively. Using the replies of users to measure the degree of common interest between users. The knowledge interaction ability between users is quantified by the comment relation between users.

2.2 Building Social Network Aggregation Behavior Communities

Community can be regarded as the local epitome of the network, especially in the online social network with a large scale of nodes. Because of the large amount of data, it is difficult to study the whole network directly. Proximity is a measure of a node's ability to occupy the center of a network. It is defined as the length of the shortest path to any other node in the network. The higher the value near centrality, the higher the importance of the node in the whole network. However, it takes a lot of time to get the close-centricity by getting the structure information of the whole network. According to user attributes, the nodes with local maximal characteristics are selected as the core of each community to form the initial structure of each community. Then the external nodes are divided into more attractive communities by comparing the attractiveness of the internal nodes of each community to their connected external nodes, and the process is repeated until all nodes are divided into corresponding communities. The topology of the network is sensitive to proximity to the center. If there are many edge connections in a local area of the network, it can find the center node well, but it can't detect the result well in the sparse area. In this paper, the nodes with higher aggregation coefficient under certain user knowledge quality scale are selected as the core of community, and such nodes are called local maximal aggregation nodes. For a user node, if the product of its aggregation coefficient and its own knowledge quality is greater than or equal to all its neighbor nodes, the node is called local maximum aggregation node. In the random walk process of a node, the probability of jump will change with the change of the node. The importance of nodes in the set is sorted by descending order and the first node is selected as the initial core node. The whole self-organizing aggregation behavior community is regarded as a whole, and the ownership of nodes is determined by the attraction of the community to nodes [6]. The attraction between two user nodes in a social network can be expressed as:

$$u_n(s_1, s_2) = x(s_1, s_2)\, o\,(s_1)\, o\,(s_2) \tag{4}$$

In formula (4), u_n represents the attractiveness between the user s_1 and the user s_2; $x(s_1, s_2)$ represents the elements of the adjacency matrix of the social network in row s_1 and column; and $o(s_1)$, $o(s_2)$ represents the degrees of user s_1 and user s_2 nodes, respectively. When an unupdated node conflicts during the update phase, a label is chosen arbitrarily. The core node is partially expanded by random walk, so that the initial core community is constructed. Considering the closeness of the nodes in the community, the nodes which are three to four sides away from the core nodes are taken as the nodes of the random walk. This strategy of random label selection results in non- unique algorithm results. In the network, there is a two-way interaction between nodes. In the case of label update conflict, the label is selected according to the label weight. The label weight can be expressed as the quotient of attraction between nodes and length of neighbor nodes, and the final node is updated according to the label with the highest weight. At the end of the tagging process, the social network can be divided into two aggregation communities through an iteration, and the aggregation community structure is the same as the real community structure. In the process of traversing the non-community core nodes in the network, some nodes can be regarded as overlapping nodes because they are attracted by multiple communities simultaneously. The membership of overlapping nodes to each community is calculated to measure the degree of subordination of the node to each community. The formula for calculating community membership of overlapping nodes is as follows:

$$\chi(s_1) = \frac{u_n(s_1, s_2)}{\sum_n u_n(s_1, s_2)} \tag{5}$$

In formula (5), χ represents the membership of overlapping nodes to the community; n represents the number of self-organizing aggregation communities in a social network. The whole network is traversed many times, each traversal calculates the attraction between nodes and communities, divides nodes into corresponding communities until all nodes enter into the communities. Based on this, the self-organized aggregation behavior community structure is defined.

2.3 Tracking Self-organizing Group Trajectories Based on Ant Colony Algorithm

Individual tracking in self-organized agglomeration community in social networks involves multiple tracking objects, which is difficult to track. In this paper, the ant colony algorithm is used to match and track the trajectories of self-organized groups. If the core node gravity chain changes, the dissimilarity between nodes in two neighboring time sliced communities satisfies a certain threshold condition, then a certain evolutionary relationship is defined [7]. Generally speaking, the changes of social networks over time mainly include: exiting original nodes, joining new nodes, breaking old connections, and new links between nodes. The change of network structure will also make the internal community structure change correspondingly. In this paper, the evolution of population trajectory is divided into seven types: persistent, shrinking, growing, splitting, merging, disappearing and forming. The selection of data is also more general, which can well mine the types of evolution events under different time slices. Community formation is a process in which some unconnected nodes in the network form a community structure at the due to increasing contact with each other. Observing the changes of the core node chain of each community to judge the evolutionary behavior patterns of community groups. For example, if the core chain of a community breaks, the community splits; if the core chain of multiple communities is connected, the communities merge; and if a new core chain appears, a new community emerges. The way the community is structured is shown in Fig. 2.

The first stage is to partition the community structure of the network snapshot at the time of t using the static social network discovery algorithm based on the attraction

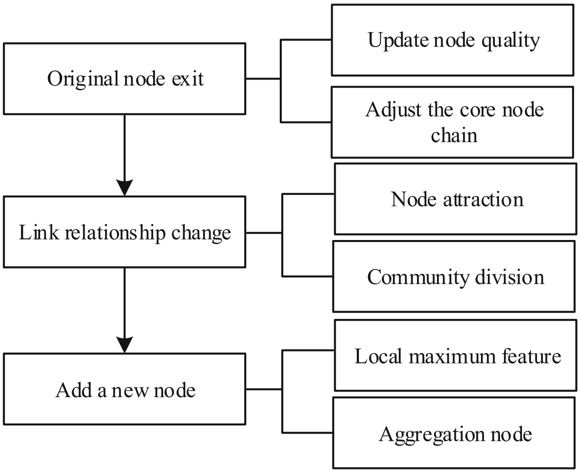

Fig. 2. Community restructuring approach

between nodes. In the second stage, the impact of network changes on social networks is fully considered, including the exit of original nodes, the change of link relationships and the addition of new nodes. In the original ant colony algorithm, the ant mainly relies on the probability function to select the next node. But in the initial stage of the algorithm, the difference of the pheromone concentration is very small, which can not guide the ant to choose the route effectively. Hence the pseudo-random state transition function shown in formula (6).

$$y = \arg\max\{v_1, v_2\} \tag{6}$$

In formula (6), y represents the selection function of candidate nodes; v_1 and v_2 represent the relative importance of pheromone and distance expectations, respectively. Candidate nodes will be selected based on the highest pheromone concentration in the selection function and the highest expected value of the path node. Because the pheromone is updated in the global direction in the ant cycle model, this paper uses the ant cycle model as the updating model, and all ants will update the pheromone content after the iteration. Then, the probability of any candidate node is calculated according to the probability formula, and the next node is determined by roulette. In different iteration periods, the requirements for pheromone increment are different. Therefore, in order to accelerate the convergence rate of ants and avoid local optimization, dynamic pheromone updating method is used here. The calculation formula is as follows:

$$P(t) = P_0 + \eta t \tag{7}$$

In formula (7), P represents the pheromone update value; P_0 is the initial value of the pheromone; and η is the pheromone update mechanism. The volatilization coefficient of pheromone also affects the global convergence and the optimal solution, and it also affects the pheromone content of the path to some extent. The size of pheromone volatilization factor indicates the degree of pheromone change with time for ants. In general, the setting of this factor can not be too large or too small, if set too large, although it can speed up the convergence of the algorithm but will also make the ant on the path that has not yet walked in the volatilization of pheromone faster, if set too small, the ant will fall into the current local optimal state when searching for the path so that the ant can not jump out to find more solutions. In addition, if the volatility factor is set to a constant amount during the search, the search will be reduced. By reserving the pheromone on each path, the global searching ability of the algorithm is ensured in the early stage, and the local optimization caused by excessive pheromone volatilization is avoided [8]. In each part, the quality (degree) of nodes and the attractiveness among nodes of the network should be updated, the core node chain of each community should be adjusted, and the evolutionary behavior of the community should be judged according to the changes of the core node chain of each community. Then, the structure of the self-organized aggregation network should be redivided or adjusted according to the attractiveness among the updated nodes. In the middle and later stages of the algorithm, a larger pheromone volatilization coefficient (i.e. a smaller pheromone residue coefficient) should be used, because there is a certain gap in pheromone on the path after the previous operation, and by using a larger pheromone volatilization coefficient, the difference in pheromone on each path should be increased to ensure the convergence speed of the algorithm. Finally, we consider the new nodes in the network, and add these nodes to the algorithm and related links. The addition of new nodes will make some nodes in the network more difficult. After adjusting the core node chain of each community, the rest nodes in the network should be redivided according to the attractiveness of the updated nodes.

2.4 Establishment of Self-organizing Aggregation Behavior Recognition Model

According to the tag of user attribute data, the self-organized aggregation behavior is transformed into a binary classification problem to judge whether the event is normal or abnormal. When there are only normal events in the detection model, the model belongs to the normal event detection model of clustering, and the abnormal events are far away from the center of clustering. The structure of the self-organizing aggregation behavior recognition model is established as shown in Fig. 3.

First, the tagged basic events are extracted from the user attributes, and then the events are learned to build a classifier. Finally, the events are classified by a classifier. Nodes will select the largest number of neighbor tags in the label update stage, and will select a tag randomly when unupdated nodes collide in the update stage. The occurrence of anomalous events does not exist independently at a certain point in time. Events may change slightly over time, and the whole event cannot be viewed from a single point in time. Therefore, when the events occurring at all points in time are combined together, a complete event can be seen [9]. In the network, there is a two-way interaction between nodes, so this paper considers the interaction between nodes on the impact of label selection, according to the interaction force index to improve the update strategy. Select the node j to be updated from the node update sequence. During propagation, the node j to be updated will select the label corresponding to the maximum force to update its own label. The formula is as follows:

$$f(j) = \arg\max \vartheta(\varphi) \tag{8}$$

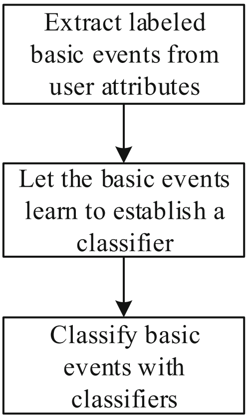

Fig. 3. Establishing the structure of the self-organizing aggregation behavior recognition model

In formula (8), f stands for node update function; ϑ stands for node force value; and φ stands for label. The node force value shall be comprehensively measured by the node's own properties and the relationship between nodes. The calculation formula is as follows:

$$\vartheta = \varepsilon\kappa + (1 - \varepsilon)\theta \tag{9}$$

In the formula (9), ε represents the damping coefficient, which is used to balance the influence of a node j on its neighbors and the influence of a neighbor on the node j; κ represents the similarity of a node j to its neighbors; θ represents the force of a neighbor on a node j, and the larger the value is, the greater the influence of a neighbor on a node is and the deeper the influence is. Based on the objective knowledge system, the key elements existing in the system are analyzed, including user element, text element and knowledge element. According to the mapping relationship between elements, the integrated model of user knowledge hyper-network model is established. The method based on feature reconstruction considers that there are correlations between normal events, and they are similar to a great extent. It is usually possible to represent a normal event linearly with other normal events, and to reconstruct its main features. Its process is to train the normal events, get the dictionary after learning, and calculate its reconstruction error for the sample to be tested. The normal events can be reconstructed with very small error, but the abnormal events are relatively large. Through the full expression of heterogeneity elements in the target knowledge system, it provides material and framework for the analysis of the elements such as users and knowledge. It is worth noting that most of the current research into user mining, often based on the user's contribution to the community of all knowledge modeling, ignoring the changes in user behavior online. It is worth noting that the earlier the historical behavior of the user occurs, the less its reference value. Therefore, the time dimension is added, that is, the panel data in the target knowledge system is used for dynamic modeling of user knowledge super-network. So we can judge whether the event is abnormal or not according to the result that the reconstruction error is larger than the set threshold. At present, the usual method is sparse representation, which makes a dictionary of the normal events, and then uses the sparse representation of the test events by the dictionary atoms. Thus, the self-organized aggregation behavior recognition method based on ant colony algorithm is designed.

3 Experimental Study

3.1 Experimental Settings

In this paper, three networks are used to verify the effectiveness of the proposed ant colony algorithm for self-organized clustering behavior recognition in social networks. The three representative datasets selected were the Karate Club Social Network (Karate), the Dolphin Social Network (Dolphin) and the Enron Social Network. This paper needs to analyze the structure and characteristics of complex network communities on these datasets. Each of their snapshots is treated as a static network and is identified as a self-organizing aggregation behavior. The Karate network consists of 34 nodes and 78 edges; the Dolphin network consists of 62 nodes and 159 edges. The nodes represent the members of the club, and the edges represent the social relationships among the members. The Enron Social Network Dataset describes the data that Enron employees exchange email. The Enron Social Network contains information on 56 employee nodes and 854 e-mails. Through preprocessing, it was divided into 12 time slice snapshots by month. This method not only ensures the high similarity of the original topology of the network community, but also effectively increases the number of evolutionary events to provide sample support for the subsequent experimental analysis. The self-organized aggregation behavior recognition method based on GN algorithm and Fast algorithm in social networks is used as a comparison method to carry out experimental tests. In this paper, modularity and Normalized mutual information are used to test the effectiveness of the proposed self-organized aggregation behavior recognition method in social networks.

3.2 Experimental Results and Analysis

Modularity, also known as modularity measure, is a commonly used method to measure the structural strength of a network community. The value of modularity represents the result of community partition of self-organized aggregation behavior, and the larger the value is, the better the result of community partition is. The modular results of the respective organization aggregation behavior recognition methods in the Karate, Dolphin, and Enron networks are shown in Tables 1, 2 and 3.

In Karate network dataset, the average modularity of self-organizing aggregation behavior recognition method based on ant colony algorithm is 0.383, which is 0.146 and 0.180 higher than that based on GN algorithm and Fast algorithm. It can be seen that the results of community division of self-organized aggregation behavior of the self-organized aggregation behavior identification method in social network based on ant colony algorithm are better.

In Dolphin network dataset, the average modularity of self-organizing aggregation behavior recognition method based on ant colony algorithm is 0.362, which is 0.153

Table 1. Comparison of modularity results for Karate networks

Number of tests	Recognition of self-organized aggregation behavior in social networks based on ant colony algorithm	Recognition of self-organized aggregation behavior in social networks based on GN algorithm	Fast algorithm based recognition method for self-organizing aggregation behavior in social networks
1	0.386	0.251	0.206
2	0.395	0.224	0.218
3	0.378	0.238	0.197
4	0.382	0.245	0.204
5	0.393	0.222	0.205
6	0.386	0.233	0.212
7	0.375	0.242	0.196
8	0.382	0.255	0.182
9	0.394	0.229	0.203
10	0.361	0.226	0.204

Table 2. Comparison of modularity results for Dolphin networks

Number of tests	Recognition of self-organized aggregation behavior in social networks based on ant colony algorithm	Recognition of self-organized aggregation behavior in social networks based on GN algorithm	Fast algorithm based recognition method for self-organizing aggregation behavior in social networks
1	0.374	0.206	0.208
2	0.342	0.188	0.217
3	0.358	0.185	0.234
4	0.366	0.193	0.226
5	0.373	0.186	0.202
6	0.385	0.222	0.213
7	0.352	0.235	0.225
8	0.361	0.219	0.202
9	0.344	0.232	0.191
10	0.369	0.224	0.194

and 0.151 higher than that based on GN algorithm and Fast algorithm. It can be seen that the results of community division of self-organized aggregation behavior of the

self-organized aggregation behavior identification method in social network based on ant colony algorithm are better.

Table 3. Comparison of modularity results for Enron networks

Number of tests	Recognition of self-organized aggregation behavior in social networks based on ant colony algorithm	Recognition of self-organized aggregation behavior in social networks based on GN Algorithm	Fast algorithm based recognition method for self-organizing aggregation behavior in social networks
1	0.401	0.218	0.219
2	0.389	0.209	0.232
3	0.416	0.225	0.226
4	0.407	0.236	0.203
5	0.404	0.218	0.245
6	0.391	0.247	0.267
7	0.388	0.252	0.234
8	0.392	0.234	0.181
9	0.386	0.242	0.195
10	0.373	0.223	0.228

In Enron network dataset, the average modularity of self-organizing aggregation behavior recognition method based on ant colony algorithm is 0.395, which is 0.165 and 0.172 higher than that based on GN algorithm and Fast algorithm. From the perspective of modularity, the self-organized aggregation behavior recognition method proposed in this paper is more sensitive to community structure and has achieved good results. The method of this paper can correctly divide the user stage into the corresponding self-organization community.

Normalized mutual information is used to measure the similarity of clustering results. Normalized mutual information is an index used to evaluate the similarity between the real network structure and the community structure generated by the algorithm partition. The higher the value is, the higher the similarity between real network and partitioned community structure is. The Normalized mutual information results of the aggregation behavior identification methods in the Karate, Dolphin, and Enron networks are shown in Tables 4, 5 and 6.

In Karate network dataset, the mean of Normalized mutual information is 0.745, which is 0.107 and 0.175 higher than the comparison method based on GN algorithm and Fast algorithm. It can be seen that the more similar the real network and the divided community structure of the self-organized aggregation behavior identification method in the social network based on the ant colony algorithm, the higher the similarity.

In Dolphin network dataset, the mean of Normalized mutual information is 0.701, which is 0.160 and 0.190 higher than the comparison between GN algorithm and Fast

Table 4. Comparison results of normalized mutual information for Karate networks

Number of tests	Recognition of self-organized aggregation behavior in social networks based on ant colony algorithm	Recognition of self-organized aggregation behavior in social networks based on GN algorithm	Fast algorithm based recognition method for self-organizing aggregation behavior in social networks
1	0.726	0.603	0.582
2	0.784	0.629	0.597
3	0.749	0.645	0.604
4	0.750	0.658	0.571
5	0.762	0.626	0.552
6	0.725	0.662	0.563
7	0.717	0.635	0.542
8	0.728	0.627	0.555
9	0.736	0.654	0.578
10	0.772	0.641	0.560

Table 5. Comparison results of normalized mutual information for Dolphin networks

Number of tests	Recognition of self-organized aggregation behavior in social networks based on ant colony algorithm	Recognition of self-organized aggregation behavior in social networks based on GN algorithm	Fast algorithm based recognition method for self-organizing aggregation behavior in social networks
1	0.687	0.559	0.518
2	0.688	0.525	0.497
3	0.695	0.536	0.483
4	0.716	0.542	0.526
5	0.705	0.553	0.519
6	0.722	0.525	0.505
7	0.694	0.531	0.512
8	0.681	0.540	0.523
9	0.703	0.552	0.510
10	0.722	0.545	0.521

algorithm. It can be seen that the more similar the real network and the divided community structure of the self-organized aggregation behavior identification method in the social network based on the ant colony algorithm, the higher the similarity.

Table 6. Comparison results of normalized mutual information for Enron networks

Number of tests	Recognition of self-organized aggregation behavior in social networks based on ant colony algorithm	Recognition of self-organized aggregation behavior in social networks based on GN algorithm	Fast algorithm based recognition method for self-organizing aggregation behavior in social networks
1	0.684	0.516	0.564
2	0.691	0.538	0.577
3	0.686	0.545	0.551
4	0.723	0.521	0.548
5	0.755	0.505	0.555
6	0.762	0.539	0.536
7	0.714	0.512	0.523
8	0.827	0.546	0.545
9	0.796	0.553	0.554
10	0.805	0.526	0.529

In Enron network data set, the mean of Normalized mutual information is 0.744, which is 0.214 and 0.196 higher than the comparison between GN algorithm and Fast algorithm. From the standardised mutual information point of view, the self-organized aggregation recognition method presented in this paper is stable and can effectively detect the real network community structure.

4 Conclusion

Based on ant colony algorithm, a recognition method of self-organized aggregation behavior in social networks is proposed. By analyzing the characteristics of user attributes in social networks, the nodes with high aggregation coefficient under the scale of user knowledge quality are selected as the core to build a social network aggregation behavior community. Ant colony algorithm is used to track the group trajectory, extract labeled basic events from user attributes, establish a recognition model to identify abnormal behaviors, and realize self-organized aggregation behavior recognition in social networks. This method can improve the accuracy of community partition and dynamically adjust the group structure in the network. With the coming of big data era, how to choose a better time window or automatically divide the time window according to the dataset is a problem worthy of study.

Fund Project. This paper is the phased achievement of Motivation Attribution and system solution (19ZYJS021) of collective action in the process of China University of Labor relations of the first-class program of characteristic disciplines of China Institute of labor relations (basic scientific research business expenses of Central Universities); General project of Beijing Social

Science Foundation: phased achievements of employment structure change and staff training in the development of digital economy (20JJB008).

References

1. Zhang, K., Sun, Y.J., Han, H.: Customer credit modeling and credit granting method based on hybrid algorithm of collaborative filtering and social network. Telecommun. Sci. **36**(2), 52–60 (2020)
2. Li, Y., Deng, S., Lin, J.: Dynamic behavior analysis and online detection of spammer user in social network. Comput. Eng. **45**(08), 287–295 (2019)
3. Qiu, G., Li, X., Cheng, X., et al.: Accurate identification of misconduct of water user team in social network. Sci. Technol. Eng. **19**(07), 177–182 (2019)
4. Yuan, L., Gu, Y., Zhao, D.: Research on abnormal user detection technology in social network based on XGBoost method. Appl. Res. Comput. **37**(3), 814–817 (2020)
5. Peng, Y., Jiang, R., Xu, L.: An algorithm for identifying multi-class abnormal behavior of population based on rough set model. Sci. Technol. Eng. **21**(11), 4524–4533 (2021)
6. Xie, Q., Li, Z.: Network malicious behavior identification and detection based on big data association rules. J. Hefei Univ. **38**(2), 85–91 (2021)
7. Rong, W., Jiang, Z., Xie, Z., et al.: Clustering relational network for group activity recognition. J. Comput. Appl. **40**(9), 2507–2513 (2020)
8. Jin, X.: Deep mining simulation of unstructured big data based on ant colony algorithm. Comput. Simul. **37**(11), 329–333 (2020)
9. Yang, X., Sun, Y.: A survey on user behavior of social network based on knowledge graph. J. Hebei Univ. (Nat. Sci. Ed.) **41**(1), 77–86 (2021)