WILEY | Hindawi

Retraction

Retracted: Group Relationship Mining of College Students Based on Predictive Social Network

Security and Communication Networks

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Liu, "Group Relationship Mining of College Students Based on Predictive Social Network," *Security and Communication Networks*, vol. 2022, Article ID 5443655, 9 pages, 2022.

WILEY | Hindawi

Research Article

Group Relationship Mining of College Students Based on Predictive Social Network

Huazhang Liu (ID)

Chongqing Industry Polytechnic College, Chonqing 401120, China

Correspondence should be addressed to Huazhang Liu; liuhz@cqipc.edu.cn

With the rapid development of the Internet, social networks have shown an unprecedented development trend among college students. Closer social activities among college students have led to the emergence of college students with new social characteristics. The traditional method of college students' group classification can no longer meet the current demand. Therefore, this paper proposes a social network link prediction method-combination algorithm, which combines neighbor information and a random block. By mining the social networks of college students' group relationships, the classification of college students' groups can be realized. Firstly, on the basis of complex network theory, the essential relationship of college student groups under a complex network is analyzed. Secondly, a new combination algorithm is proposed by using the simplest linear combination method to combine the proximity link prediction based on neighbor information and the likelihood analysis link prediction based on a random block. Finally, the proposed combination algorithm is verified by using the social data of college students' networks. Experimental results show that, compared with the traditional link prediction algorithm, the proposed combination algorithm can effectively dig out the group characteristics of social networks and improve the accuracy of college students' association classification.

1. Introduction

Nowadays, the rapid development of the Internet affects people's work, life, and study. People's communication has gradually changed from offline to online, resulting in a social network platform that meets the needs of all kinds of people. The emergence of various social networks has enriched people's daily life, provided effective ways to obtain information and made communication between people more convenient. Due to the popularity of the Internet, social networks begin to influence the life and study habits of college students [1–5]. At present, college students are already a major group active on social platforms. After entering the university, students communicate with each other through study and life, forming different student groups. How to classify college students scientifically and effectively is the basis of doing a good job in college student affairs.

As an important and complicated group, college students' traditional behavior analysis methods have several

disadvantages [6–9]: (1) there is a lack of theoretical analysis; (2) the focus is not obvious; (3) it is easy to cause management blind spots. Because the relationship between people, in reality, can be reflected in the interaction in social networks. Therefore, in a specific range, the actual connection relationship of members in the social network can be deduced from the interaction relationship in the social network. It is of great guiding significance to apply social network theory to the management of college students. By identifying and dividing student groups through social networks, we can dig deeper into students' internal connections and characteristics, thus improving the efficiency of university management.

As a popular subject, the research on the division of complex network communities has become one of the most challenging basic research topics in the computer field, which has very important theoretical significance and application value [10]. However, the theoretical research in this field is still in its infancy. In addition, the social network is

different from the traditional static network; it is a dynamic network [11–13]. With the change of time, new entities and connections are constantly added to the network, and old instances and connections are constantly disappearing. This dynamic nature means that there is a great difference between the research of social networks and the research of traditional static networks. Therefore, the dynamic characteristics of social networks become the difficulty of research.

Therefore, the intelligent system based on link prediction comes into being. Link prediction in social networks refers to predicting the probability of link generation between two user nodes that do not have a connection edge in the network through various information such as network user nodes and topology [14–16]. In fact, users and the relationships among users in social networks constitute a complex network, in which users are set as nodes and the relationships among users can be used as edges. Corbellini et al. [17] proposed a social network model based on link prediction, which used graph theory to extract similarity indexes. Then, the similarity is arranged from large to small so as to predict whether there is an edge between this pair of nodes. Ma et al. [18] proposed a link prediction method based on structural similarity information and community information for the Twitter network. Experiments show that this method can be effectively used in large-scale directed and asymmetric networks. Kagan et al. [19] put forward a new link prediction algorithm based on the maximum likelihood model, combining the interest characteristics of nodes and the network structure characteristics, and achieved a better bidirectional edge division effect.

In this paper, the relationship among college students on social networks is investigated from the perspective of link prediction, and the social network link prediction model is constructed by taking community division as the research object. This study combines proximity link prediction and likelihood analysis link prediction, trying to find the most suitable link prediction method for college students so as to improve the accuracy of college students' group classification. The main innovation of this study is that, unlike the traditional social network link prediction method, we try to combine proximity analysis with likelihood analysis so as to improve the mining system and improve the classification accuracy and provide a reference for the further development of the follow-up mining system.

2. Student Group Identification Based on Social Network

2.1. Community Structure of a Complex Network. For network representation, the most traditional method is graph. Graph theory originated from Euler's "Konigsberg Seven Bridges Problem" in 1736 [20–22]. A concrete network can be abstracted as a graph $G=(V, E)$ composed of a node set V and an edge set E. The vertex number of G is $N = |V|$, and the edge number is $M = |E|$. Each pair of nodes in the node set V corresponds to an edge in the edge set E. Networks can be divided into the following four categories: (1) unauthorized undirected network; (2) weighted undirected network; (3)

unauthorized directed network; and (4) weighted directed network, as shown in Figure 1.

In computer, the graph structure can be represented by the algebraic method of graph. Common methods are adjacency matrix and adjacency table [23], as shown in Formula (1) and Figure 2.

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{1}$$

Using the method of adjacency matrix to represent a graph can intuitively judge whether any two vertices are connected by edges. Another advantage of adjacency matrix representation is that it enables us to study many properties of graphs by matrix analysis. Actually, large-scale complex networks are often sparse, which means that most of the elements in the corresponding adjacency matrix are 0, that is, sparse matrix [24]. In order to save space, the adjacency table is often used to deal with the sparse matrix of an unweighted graph. In complex networks, there are mainly four kinds of abstract models for analyzing the actual network structure and behavior in real life, which are regular networks, random networks, small clever networks, and scale-free networks.

Complex networks, in reality, are not random networks. These complex networks can be described by network communities. The actual network is mainly composed of several communities, and the connections between each community have obvious characteristics and differences. A typical community structure is shown in Figure 3.

Figure 3 can be divided into three communities. The inside of each virtual coil can be regarded as a community. There is a close relationship between members in the community. A member can have a connection relationship with several members in the community, but only one member has a single-line connection between the community closure and the community closure. Different network associations have different practical significance. The main existing forms of network associations can be divided into the following three categories [25], as shown in Figures 4–6

(1) Nonoverlapping network community structure: this kind of community refers to that there are no common nodes between communities or subgraphs, as shown in Figure 4

(2) Overlapping network communities: it mainly refers to that some nodes in the network belong to two or more different communities at the same time, as shown in Figure 5

(3) Hierarchical network community: this kind of community refers to the existence of hierarchy or hierarchical structure in the community, as shown in Figure 6
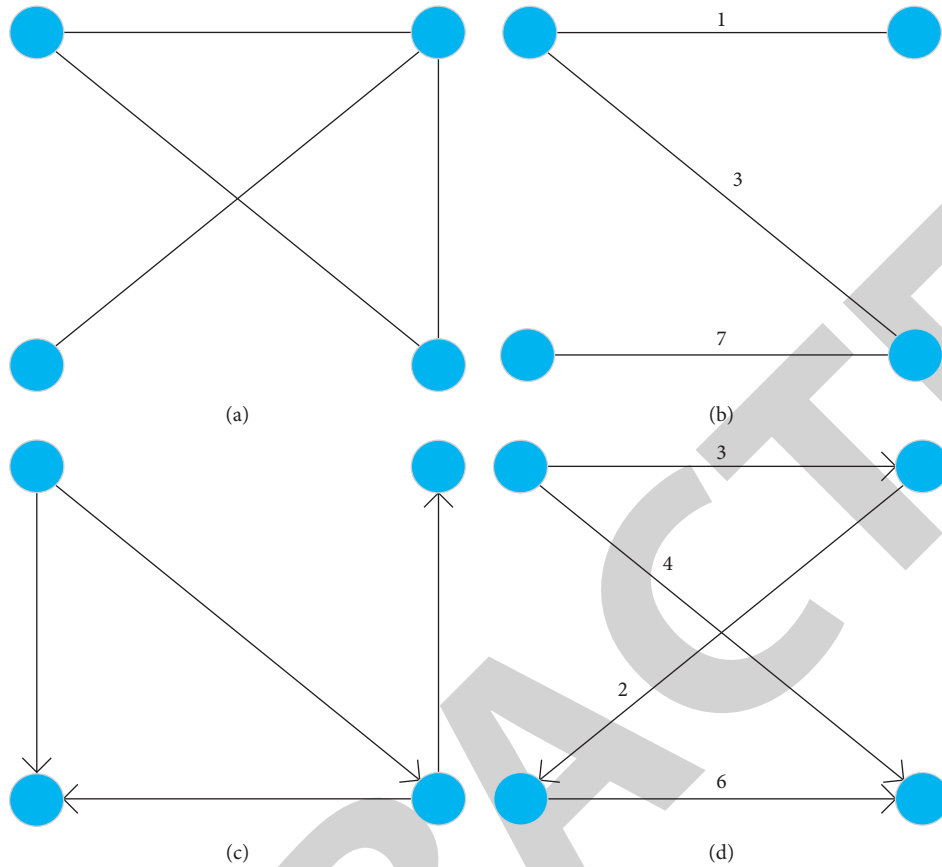
FIGURE 1: Four different networks. (a) Unweighted undirected networks. (b) Weighted undirected network. (c) Unweighted directed network. (d) Weighted directed network.
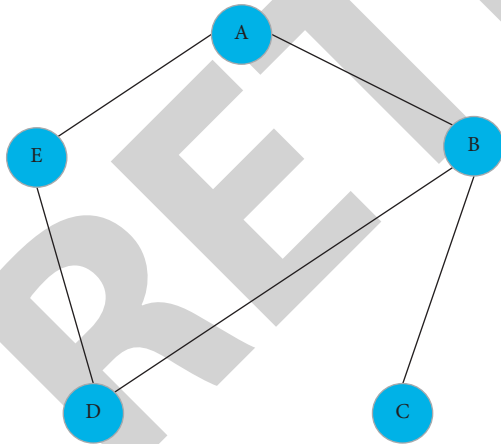

FIGURE 2: A simple network with 5 nodes.

2.2. Student Group Relations in the Complex Network. In this paper, students' connections are represented in the form of complex network diagrams. In the whole social relationship graph, because the nodes in the graph are students themselves, the community can fully represent a student subgroup. At the same time, the connection between the community can fully reflect the relationships among multiple student groups. An example of the student group relationship is shown in Figure 7.

In Figure 7, students ABC and students DEF form a community, respectively. However, in Figure 7(a), there is a connection between the community ABC and the community DEF. In Figure 7(b), there is no connection between the two associations. If two communities represent two classes, then the communities in Figure 7(a) are likely to represent two adjacent classes, and the communities in Figure 7(b) may be far apart. If two societies represent student interest organizations, then the organizations in Figure 7(a) may belong to the same category, such as football club and basketball club. However, the organization in Figure 7(b) does not matter. Therefore, community relations and node relations in complex networks are of great help for us to identify and analyze student groups.

3. Student Group Identification Based on Social Network Link Prediction

3.1. Problem Description. Description the of link prediction method [26]: a numerical value S_{xy} is assigned to all the node pairs v_x and v_y that have not been connected. This value can be regarded as a kind of proximity, and it is directly proportional to the probability of linking these two nodes. Because graph G is undirected, there is $S_{xy} = S_{yx}$. Then, the node pairs are sorted according to the numerical value, and the top node is considered to be more likely to have edge connection. The social network link prediction model is shown in Figure 8.
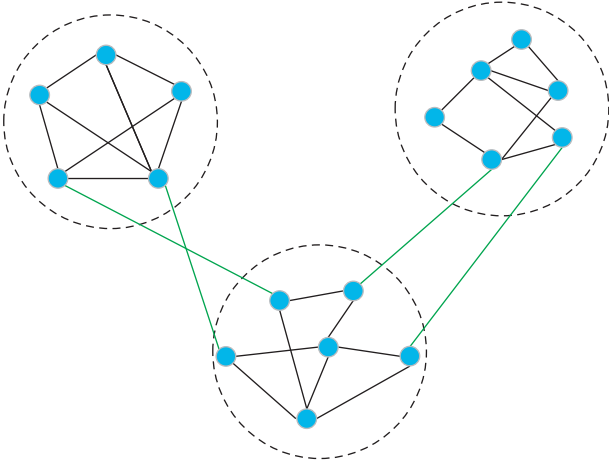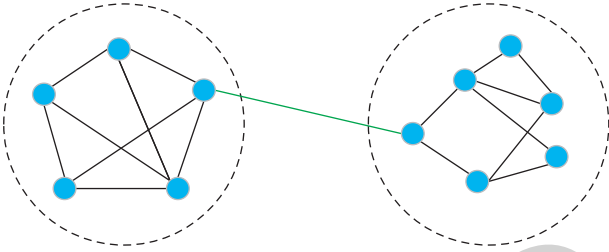
Figure 3: Community network structure diagram.

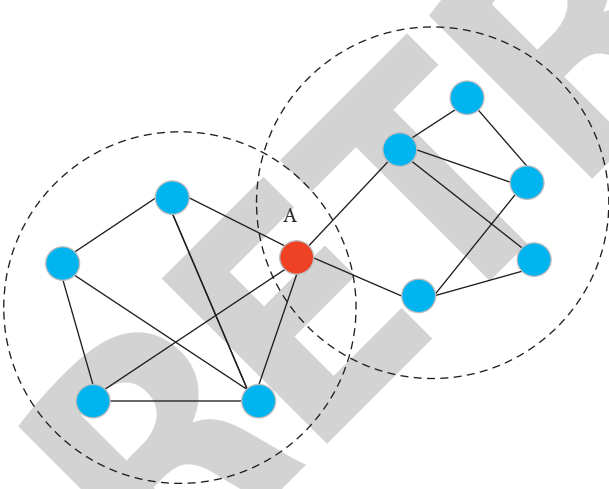Figure 4: Nonoverlapping network community structure.

Figure 5: Overlapping network community structure.

3.2. The Proposed Social Network Link Prediction Method-Combination Algorithm.

Link prediction based on neighbor information is a proximity method based on graph topology. Its basic premise is that if the intersection of the neighbor sets $\Gamma(x)$ and $\Gamma(y)$ of the student nodes v_x and v_y in the network is larger, the nodes v_x and v_y are more similar.

Common neighbor (CN) algorithm [27]: the simplest algorithm is considered from the perspective of common neighbors. CN is defined as follows: if the neighbor set of node v_x in the network is $\Gamma(x)$, then the proximity of the nodes v_x and v_y is defined as their common neighbor number, which is defined as follows:

$$s_{xv} = |\Gamma(x) \cap \Gamma(y)|. \tag{2}$$

The basic idea of link prediction based on a random block is to divide all nodes in the network into several groups [28]. The probability of whether two nodes are connected depends on the group in which they belong, that is, the status of nodes in the same group is the same. A block model $M = (P, Q)$ is given, where P represents the set of all clustering methods and Q represents the edge-connected probability matrix. The block model is defined as follows:

$$p\left(A^0 | P, Q\right) = \prod_{\alpha \leq \beta} Q_{\alpha\beta}\left(1 - Q_{\alpha\beta}\right)^{r_{\alpha\beta} - l_{\alpha\beta}^0}, \tag{3}$$

where A^0 represents the random matrix of the currently observed network, $Q_{\alpha\beta}$ represents the connection probability of the node v_x in group α and the node v_y in group β, $r_{\alpha\beta}$ represents the possible number of connected edges in two groups, and $l_{\alpha\beta}^0$ represents the number of connected edges connecting two groups in the observed network A^0.

According to the Bayesian theorem, the proximity of two nodes is defined as follows:

$$S_{xy} = p\left(A_{xy} = 1 | A^0\right)$$
$$= \sum_P \left(\frac{l_{\alpha\beta}^0 + 1}{r_{\alpha\beta} + 2}\right) \exp\left[-H(p)\right], \tag{4}$$

$$H(p) = \sum_{\alpha \leq \beta} \left[\ln\left(r_{\alpha\beta} + 1\right) + \ln\binom{r_{\alpha\beta}}{l_{\alpha\beta}^0}\right].$$

The accuracy of these two single social network link prediction methods is slightly better than that of the hierarchical structure model. However, it also has the disadvantages of inaccurate calculation and high calculation complexity, which is not suitable for the link prediction of large-scale networks.

From a scientific point of view, the existence of things is often the result of the interaction of many factors. We use the simplest linear method to combine the proximity link prediction based on neighbor information with the likelihood analysis link prediction based on a random block, which is called the combination algorithm.

$$S = \lambda \cdot S_1 + (1 - \lambda) \cdot S_2, \tag{5}$$

where S_1 represents the proximity link prediction based on neighbor information, S_2 represents the likelihood analysis link prediction based on a random block, and $\lambda \in [0, 1]$.

3.3. Student Group Division Process Based on Social Network Link Prediction.

According to the characteristics of college students' groups, according to the complex network theory, this paper uses the proposed combination algorithm to mine students' social network datasets so as to realize community division. This method can identify and analyze the student groups and their characteristics based on social network relationships. The specific division process is shown in Figure 9.
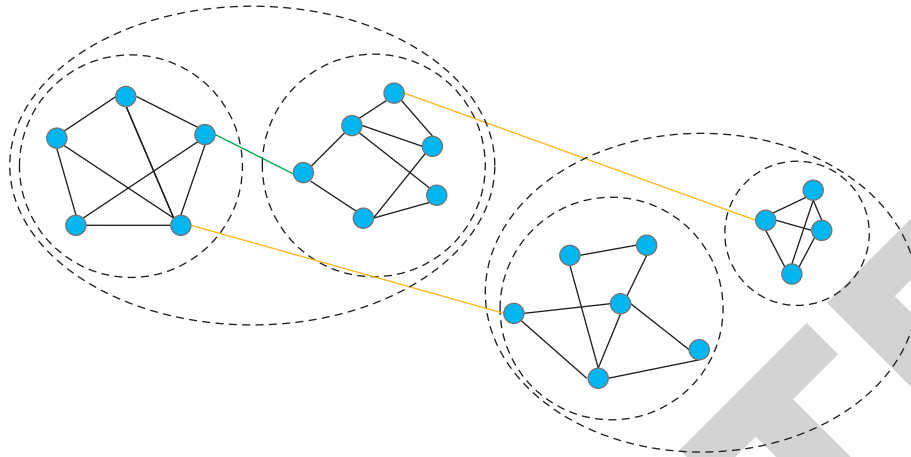
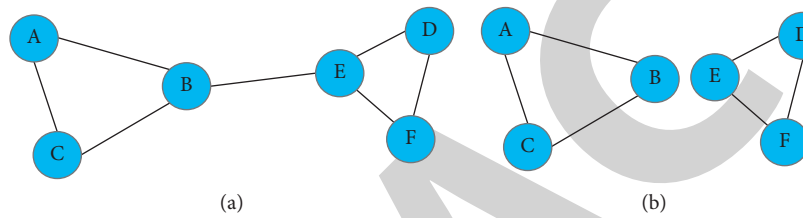Figure 6: Hierarchical network community structure.

(a)

(b)

Figure 7: Examples of student group relations. (a) Connected community. (b) Unconnected community.
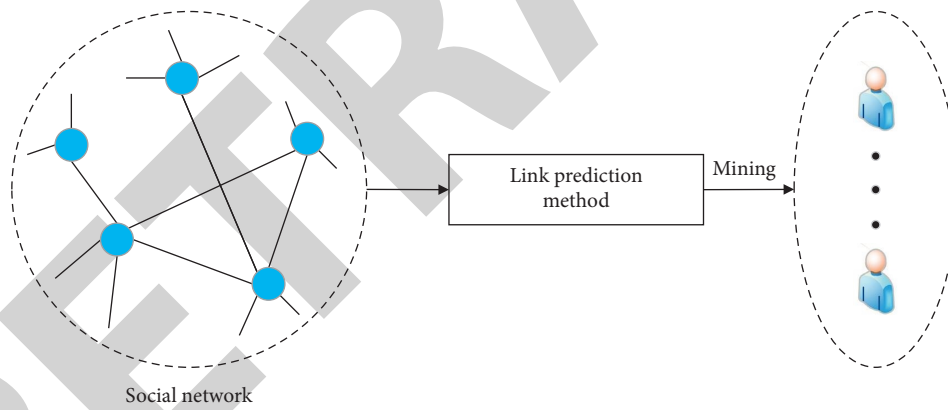
Social network

Figure 8: Social network link prediction model.

Basic data set of student social network relationship

Generation of Social Network Diagram

Class division based on social network diagram

Community division based on social network diagram

Mining special associations and subgraphs

Analysis of the characteristics of students' special groups
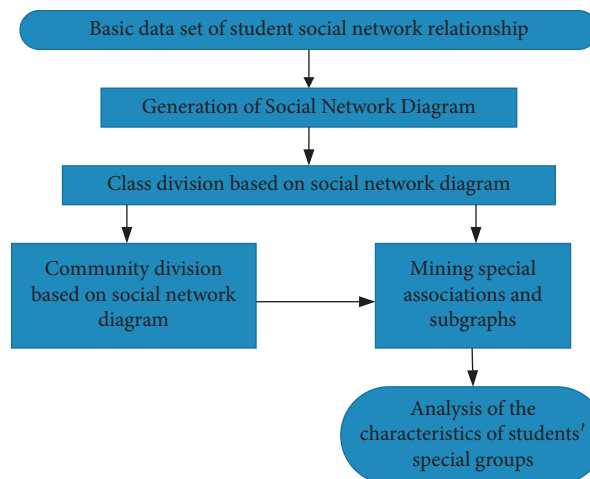
Figure 9: Social network-based student group division process.

By designing a questionnaire to collect students' QQ friends' information and WeChat friends' information, the basic dataset of social network relationships is constructed, and the social network relationship diagram of students is generated. Using the proposed combination algorithm, the social network diagram of students is divided into communities, and the results are analyzed and verified.

4. Experiment and Result Analysis

4.1. Evaluation Index. AUC is an evaluation index to measure the accuracy of the algorithm as a whole [29], and it is also the most common evaluation index. AUC is determined as follows:

$$\text{AUC} = \frac{n' + 0.5n''}{n}. \tag{6}$$

The AUC value reflects the accuracy of the proposed algorithm compared with the random algorithm.

4.2. Experimental Environment and Experimental Data. Hardware environment: CPU is Intel(R) Core(TM)2 E7200 @ 2.53 GHz, hard disk is SAMGUNG 320 GB, and memory is 4 GB. Environment: the operating system is Microsoft Windows 10, and the software used is MATLAB R2012a.

This paper collected data from 10 classes of students (300 students in total) and constructed the research dataset of the social network. The information of more than 300 people is sorted out, and a TXT document is generated; each line is a connecting edge, and the source node and the destination node are separated by spaces.

4.3. Generation of Student Social Network Diagram. The specific way to calculate the student social network diagram is as follows:

$$G = (V, E, W), \tag{7}$$

where V represents the node set of college students; E represents the set of edges representing the social relationship between students; that is, if there is a connection between two students QQ or WeChat, an edge will be generated between them; W represents the closeness set of students' network social contact, which is determined by the closeness of contact with others defined by students obtained through questionnaires.

4.4. Proximity of Combination Algorithm. The network structure composed of 300 student data has obvious hierarchy; that is, the network is composed of multiple communities interwoven together. The accuracy of the combination algorithm with different λ values is shown in Figure 10.

It is obvious from Figure 10 that the proposed combination algorithm has an optimal λ value. At this time, the value of AUC is the largest, that is, the highest accuracy. The optimal λ value of the proposed combination algorithm is 0.801, and the corresponding AUC is 0.9541. In the subsequent experiments, λ values were all 0.801.

4.5. Results of Student Class Division. In order to verify the mining effect of the proposed method on students' social network graph, the experimental datasets are simulated by proximity link prediction based on neighbor information, likelihood analysis link prediction based on a random block, and the combination algorithm. The visualization results of the class division are shown in Figure 11.

In Figure 11, the dots with different colors indicate the students in different classes, while the same colors indicate the same classes. It should be noted that different clusters express different community divisions. It can be seen from Figure 11 that after the student group relationship mining, for a group of 300 students, the number of results divided by the algorithm is equal to the actual value. However, the number of division results of a single social network link prediction is 6, far less than the actual category of 10, and the division error is large, which indicates that the division accuracy of the combination algorithm has been significantly improved.

4.6. Community Excavation Results. Besides the division of students' classes, the proposed method can also find some special associations in social networks. In order to better discover the community, we define the community disorder index H as follows:

$$H = \frac{N_{\text{class}}}{N_c} \times 100\%, \tag{8}$$

where N_{class} represents the number of classes in community C and N_c represents the total number of community members. We divided the social network diagram of 300 students into communities and explored the special groups among them. The discovered groups are all students with the same job or interest, such as student union and photography interest group. This shows that this method provides great help for us to discover the hidden relationship between students.

Using the proposed combination algorithm, we divided 300 students into communities, marked each student's class, and calculated the class disorder index of each community, as shown in Table 1.

From the results in Table 1, we can find that the class disorder is as high as 83.33% in Community 4. It is found that there are 6 students in observation Community 4, among whom 5 students are photographers and active groups. Therefore, it shows that, besides classes, the proposed method can also effectively dig out some
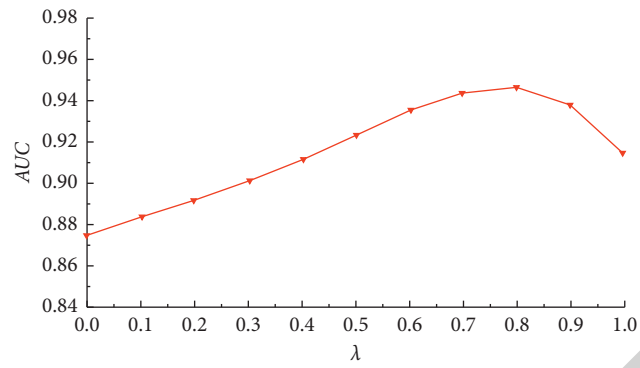
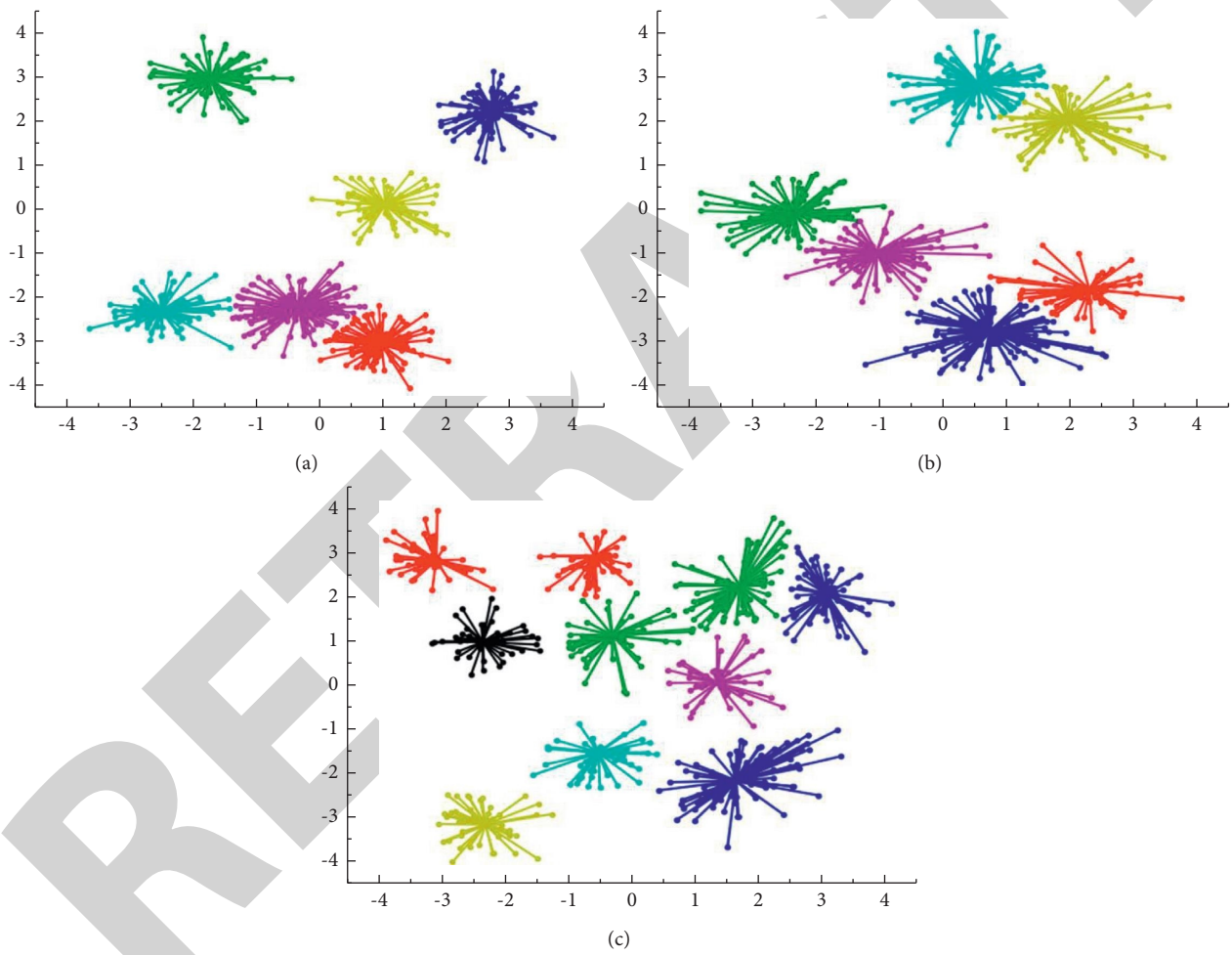FIGURE 10: The accuracy of the combination algorithm varying with the parameter λ.

FIGURE 11: Visualization results of class division. (a) Proximity link prediction based on neighbor information. (b) Likelihood link prediction based on random block. (c) Combination algorithm.

TABLE 1: Class disorder index H of each community.

Community number	Community disorder index H (%)
1	6.67
2	5.88
3	8.33
4	83.33
5	6.25
6	25.00
7	12.50
8	12.50
9	14.29

student groups with other common interests. Through the division of associations and the excavation of active groups of students, the methods of student management can be improved.

5. Conclusions

In this paper, we propose a social network link prediction method-combination algorithm, which combines neighbor information and a random block. By means of a questionnaire, students' social network information is obtained. A social network link prediction is used to identify students' groups, and the social network of students is divided into classes and communities accurately. This student group identification method based on social networks is verified and analyzed by experiments. Compared with the actual class, a variety of hidden relationships among students are obtained through graph analysis. Compared with the traditional link prediction algorithm, the proposed combination algorithm improves the accuracy of community classification. Later, we will try to study social networks based on link prediction from more angles and analyze the dynamics of social networks from the time dimension.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] Y.-L. Lai, M.-S. Hsu, F.-J. Lin, Y.-M. Chen, and Y.-H. Lin, "The effects of industry cluster knowledge management on innovation performance," *Journal of Business Research*, vol. 67, no. 5, pp. 734–739, 2014.

[2] M. Mcpherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[3] K. Xu, S. Zhang, H. Chen, and H. T. Li, "Measurement and analysis of online social networks," *Chinese Journal of Computers*, vol. 12, no. 4, pp. 29–42, 2014.

[4] M. A. Eitan and P. C. Renana, "The effect of social networks structure on innovation performance: a review and directions

for research," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 3–19, 2019.

[5] I. R. Gordon and P. Mccann, "Industrial clusters: complexes, agglomeration and/or social networks," *Urban Studies*, vol. 37, no. 3, pp. 513–532, 2013.

[6] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8788–8790, 2014.

[7] R. W. Helsley and Y. Zenou, "Social networks and interactions in cities," *Journal of Economic Theory*, vol. 150, no. 5, pp. 426–466, 2014.

[8] A. Susarla, J.-H. Oh, and Y. Tan, "Social networks and the diffusion of user-generated content: evidence from youtube," *Information Systems Research*, vol. 23, no. 1, pp. 23–41, 2012.

[9] S. Borg, "Social networks and health: models, methods, and applications," *Jama the Journal of the American Medical Association*, vol. 307, no. 11, p. 488, 2015.

[10] V. G. Eslick, "Book review: understanding social networks: theories, concepts, and findings," *Sociological Research Online*, vol. 17, no. 4, pp. 161-162, 2012.

[11] L. Atzori, A. Iera, G. Morabito, and M. Nitti, "The social internet of things (SIoT) - when social networks meet the internet of things: concept, architecture and network characterization," *Computer Networks*, vol. 56, no. 16, pp. 3594–3608, 2012.

[12] A. Forkosh-Baruch and A. Hershkovitz, "A case study of Israeli higher-education institutes sharing scholarly information with the community via social networks," *The Internet and Higher Education*, vol. 15, no. 1, pp. 58–68, 2012.

[13] L. Ye, Z. Li, and W. Breitung, "The social networks of new-generation migrants in China's urbanized villages: a case study of Guangzhou," *Habitat International*, vol. 36, no. 1, pp. 192–200, 2012.

[14] M. E. Zaglia, "Brand communities embedded in social networks☆," *Journal of Business Research*, vol. 66, no. 2-2, pp. 216–223, 2013.

[15] W. Chen, L. V. S. Lakshmanan, and C. Castillo, "Information and influence propagation in social networks," *Synthesis Lectures on Data Management*, vol. 5, no. 4, pp. 1–177, 2013.

[16] Z. Wang, J. Liao, and Q. Cao, "Friendbook: a semantic-based friend recommendation system for social networks," *Mobile Computing IEEE Transactions on*, vol. 14, no. 3, pp. 538–551, 2016.

[17] A. Corbellini, D. Godoy, and C. D. P. M. Mateos, "A novel distributed large-scale social graph processing framework for link prediction algorithms," *Future Generation Computer Systems*, vol. 78, no. 1, pp. 474–480, 2017.

[18] X. Ma, P. Sun, and G. Qin, "Nonnegative matrix factorization algorithms for link prediction in temporal networks using graph communicability," *Pattern Recognition*, vol. 71, pp. 361–374, 2017.

[19] D. Kagan, Y. Elovichi, and M. Fire, "Generic anomalous vertices detection utilizing a link prediction algorithm," *Social Network Analysis and Mining*, vol. 8, no. 1, pp. 27–35, 2018.

[20] S. Pulipati and M. Ramakrishnan, "Topological and attribute link prediction using firefly algorithm," *Open Computer Science*, vol. 10, no. 1, pp. 33–41, 2020.

[21] S. J. Devi, B. Singh, and H. Raza, "Link prediction evaluation using palette weisfeiler-lehman graph labelling algorithm," *International Journal of Knowledge and Systems Science*, vol. 10, no. 1, pp. 1–20, 2019.

[22] W. Wang, Y. Feng, P. Jiao, and W. Yu, "Kernel framework based on non-negative matrix factorization for networks

reconstruction and link prediction," *Knowledge-Based Systems*, vol. 137, no. 12, pp. 104–114, 2017.

[23] B. Moradabadi and M. R. Meybodi, "Link prediction in stochastic social networks: learning automata approach," *Journal of Computational Science*, vol. 24, no. 6, pp. 313–328, 2017.

[24] M. Naravani, N. D.G., S. Shinde, and M. M. Mulla, "A cross-layer routing metric with link prediction in wireless mesh networks," *Procedia Computer Science*, vol. 171, pp. 2215–2224, 2020.

[25] A. Rezaeipanah, G. Ahmadi, and S. S. Matoori, "A classification approach to link prediction in multiplex online ego-social networks," *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1–16, 2020.

[26] S. Pulipati, R. Somula, and B. R. Parvathala, "Nature inspired link prediction and community detection algorithms for social networks: a survey," *International Journal of Systems Assurance Engineering and Management*, vol. 8, no. 3, pp. 66–73, 2021.

[27] M. K. Manshad, M. R. Meybodi, and A. Salajegheh, "A new irregular cellular learning automata-based evolutionary computation for time series link prediction in social networks," *Applied Intelligence*, vol. 6, pp. 66–76, 2020.

[28] A. K. Singh and L. Kailasam, "Link prediction-based influence maximization in online social networks," *Neurocomputing*, vol. 453, pp. 11–18, 2021.

[29] F. Calderoni, S. Catanese, and P. D. Meo, "Robust link prediction in criminal networks: a case study of the Sicilian mafia," *Expert Systems with Applications*, vol. 161, no. 3, pp. 8–13, 2020.