Article

Visual Analytic Method for Students' Association via Modularity Optimization

XiaoYong Li ^{1,2} , QinYang Yu ³, Yong Zhang ^{1,*}, JinWei Dai ¹ and BaoCai Yin ¹

¹ Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Artificial Intelligence Institute, Faculty of Information Technology, Beijing 100124, China; lixiaoyong@bjut.edu.cn

² Information Technology Support Center, Beijing University of Technology, Beijing 100124, China

³ Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

* Correspondence: zhangyong2010@bjut.edu.cn

Received: 22 March 2020; Accepted: 15 April 2020; Published: 18 April 2020

Abstract: Students spend most of their time living and studying on campus, especially in Asia, and they form various types of associations in addition to those with classmates and roommates. It is necessary for university authorities to master these types of associations, so as to provide appropriate services, such as psychological guidance and academic advice. With the rapid development of the “smart campus,” many kinds of student behavior data are recorded, which provides an unprecedented opportunity to deeply analyze students’ associations. In this paper, we propose a visual analytic method to construct students’ association networks by computing the similarity of their behavior data. We discover student communities using the popular Louvain (or BGLL) algorithm, which can extract community structures based on modularity optimization. Using various visualization charts, we visualized associations among students so as to intuitively express them. We evaluated our method using the real behavior data of undergraduates in a university in Beijing. The experimental results indicate that this method is effective and intuitive for student association analysis.

Keywords: behavior data; Louvain algorithm; modularity optimization; students association; visualization

1. Introduction

Social relationships have important impacts on people’s lives, and as such, studying them can help to understand a person’s life. To understand a student’s social relationships, especially on campus, can help administrators to better serve the student. For example, we should pay more attention to lonely students who have few associations with others, and provide timely and useful psychological guidance if necessary. In addition, it is useful in team building to fully play the positive leading role of students who have greater associations with others. Hence, it is worth studying how to analyze the social relationships among students on campus.

Students’ social relationships, also called associations, can be reflected in their behaviors objectively. When two students are in a close association, their behavior should overlap significantly on campus, so we can study this relationship based on their behavior data. With the improvement of hardware and software on the “smart campus,” various daily activities of students are recorded, such as having meals, showering, borrowing books, and visiting the library. These data describe students’ daily behavior on campus from many aspects, which makes the multi-source behavior data available for in-depth analysis of their associations. Previous works based on behavior data has addressed topics such as constructing student social networks [1,2], predicting academic performance [3–7],

and forecasting career choices [8]. These works showed that it is possible to analyze students' lives through their behavior data.

To accurately and intuitively analyze associations, we propose a visual analytic method in which an association network is constructed based on the similarity of students' behavior. However, with a growing number of students, the association network becomes too complex to clearly observe. To overcome this issue, we use the Louvain algorithm [9], a kind of mostly adopted community detection algorithm based on modularity optimization, hierarchically partition the association network into communities. The students belonging to one community have higher probability of being friends with each other than with students of other communities. Identifying student communities offers insight on how the association network is organized. It allows us to focus on regions of interest and helps to classify the students based on their role with respect to the communities they belong to. For example, we can distinguish students who are embedded within the community from students at the boundary of the community, the two types of students play different roles in the community and may need different attention. Finally, we use visualization charts to express these associations intuitively. The main contributions and advantages of our method are as follows.

- The multi-source behavior data of students are collected from different information management systems, and then they are fused to construct an activity sequence for each student, which is the basis for analysis of association among students.
- Three kinds of similarity operators are proposed to compute the degree of similarity of behavior data among students. We take the similarity value as a weight of edges to generate association network representing social ties among students.
- The student communities were discovered by adopting the Louvain algorithm in the association network. We can further understand the organizational structure of association network through the hierarchical community structure.
- A visual analytic system was developed that interactively exhibits the association among students and allows us to intuitively explore the social relationships of specific students or communities of interest.

The remainder of this paper is organized as follows. Section 2 reviews the related work from the three aspects of how to infer social ties from spatiotemporal behavior data, detect communities from a complex network, and apply visualization techniques to detect communities. Section 3 introduces the framework of our method. Section 4 highlights privacy issues. Section 5 describes the experimental dataset and seven extracted behavior features. Section 6 explains how to construct the association network with three similarity operators and discover student communities via Louvain algorithm. Section 7 describes the visual expression of association. We conducted four experiments on the dataset and explain the results in Section 8. Finally, we conclude this paper in Section 9.

2. Related Work

We introduce related work from three perspectives.

2.1. Constructing Social Networks via Spatiotemporal Data

Much current research infers social ties between people from spatiotemporal behavior data, based on the assumption that two people should know each other if they have been in the same location at the same time on multiple occasions. However, is this assumption true? To answer this question, Crandall et al. [10] proposed a probabilistic model to quantify the correlation between the number of co-occurrences and the strength of social ties, and their experimental results proved that the number of co-occurrences has significant implications in forming inferences about social ties, which lays a foundation for related works.

To address the shortcomings of straightforward methods (richness and frequency), Huy et al. [11] proposed an entropy-based model that infers the social connections and quantifies their strength by

analyzing people's spatiotemporal co-occurrences. Their method uses Renyi entropy to measure the diversity of co-occurrences. A parameter controls how much coincidences contribute to diversity, and a weighted frequency can increase the impact of co-occurrences in uncrowded locations on the strength of social connections. It utilizes location semantics to improve the model. The experimental results show that it outperforms its counterparts and can be applied in networks, such as computing social strength in the social internet of things [12] and recommending friends in location-based social networks [13].

Based on the co-occurrences of behavior, He et al. [14] proposed that the impact of a given location on social strength should vary by time period. They exploited four spatiotemporal features, including fine-grained temporal features, weekday and weekend features, fine-grained location weight, and co-occurrence features, to infer friendship. Zhou et al. [15] proposed a theme-aware social strength inference method to extract themes from spatiotemporal behavior data and analyzes the contribution of each theme of behavior to social strength.

To analyze the social ties among large number of students over prolonged periods of time, Tao Liu [2] proposed an unsupervised statistical validation method based on the spatiotemporal behavior data on campus, taking spatiotemporal co-occurrence between two students as a measure to judge whether they have a tie, and validating ties against the null hypothesis that all co-occurrences are due to coincidence. They analyzed the structure of a network based on several metrics including degree distribution, degree assortativity, and attribute assortativity. Several differences exist between their work and our method, although both infer social ties from spatiotemporal behavior data. First, they only collected the check-in records in canteens and stores, while our behavior dataset also includes records of library entrance, book borrowing, and showering. Second, the community detection algorithms of the two methods are different. Our method uses the Louvain modularity optimization algorithm, while they used OSLOM. Third, the visualizations differ. We developed a visualization system, while they used Gephi to draw a force-directed graph.

2.2. Detecting Communities via Modularity

Identifying communities in networks, especially complex networks with huge numbers of nodes and edges, can offer insight on how a network is organized, and it has been a hot topic in various fields. There are many algorithms, which can be categorized by different criteria. In [16], the categories include spectral methods and methods based on statistical inference, optimization, and dynamics. Among them, optimization techniques have attracted the most attention. They aim to find an extremum of a quality function, which indicates the quality of clustering. Modularity is the most popular quality function. Newman in [9] proposed an algorithm to partition a complex network into densely connected subnetworks using the gain value of modularity. They iteratively remove edges that are selected by "betweenness" measures, and define the strength of the community structure to provide an objective metric to determine the number of communities into which a network should be divided.

However, tuning the resolution parameter is usually computationally intractable and time-consuming. To circumvent this issue, Blondel et al. [17] proposed the Louvain method to extract the community structure, which hierarchically performs a greedy optimization of modularity. Their experimental results show that the Louvain algorithm can accurately identify communities in network with millions of nodes in a short time, and can uncover the hierarchical community structure to facilitate observation at the desired resolution.

Many papers [16,18–21] compare the performance of the widely adopted community detection algorithms based on real-world and artificial datasets. Most conclude that the Louvain algorithm outperforms others or has similar detection results in terms of general metrics such as modularity, computing time, precision, and recall.

Due to its advantages, the Louvain algorithm is used in many applications. The modified Louvain algorithm was used in an online social network to discover the opinion leader [22], in economics

to delineate the regional economic geography [23], and in location-based networks to improve recommendation accuracy [24].

2.3. Visualization in Community Detection

Visualization can effectively represent data, especially those with a complex structure and huge volume, where it can facilitate the intuitive discovery of hidden information. Some work has incorporated visualization techniques in community detection.

To choose the most appropriate community-detection algorithms for a scenario, Claudio et al. [25] proposed a statistical visual methodology using visual interactive charts to present the detecting results of the commonly used community detection algorithms, enabling users to observe the distribution of nodes in a region of interest by zooming and panning. CRAMPES et al. [26] proposed a unified community detection method to handle bipartite graphs, directed graphs, and overlapping communities using the Louvain algorithm. Visualization is used to observe community partitioning, overlapping, and possible assignment contradictions. These visualizations help to understand the processes and results of community detection.

3. Proposed Methodologies

We propose a visual analytic method to construct association network of students by computing the degree of similarity of daily living behavior for pairs of students, and discover communities by using the Louvain algorithm. Students' behavior in a community should be similar, indicating high correlation. Visualization charts ensure an intuitive understanding of associations among students. Figure 1 shows the method's framework, with the three stages of data collection and pre-processing, qualitative analysis of associations, and visualization analysis of associations.

Figure 1. Framework of the proposed method.

3.1. Data Pre-Processing

Different types of behavior data of students are usually stored in independent information systems maintained by different departments. For example, book-borrowing records are stored in the library's database, and meal records paid via smart card are stored in the smart card system supported by the information support center. These "information islands" bring great difficulty to the analysis of students' overall behavior. To completely describe each student's behavior trajectory, we collect records of book borrowing records, library entrance, showering, and meals using extraction-transformation-loading (ETL) tools; we call these behaviors activities. After denoising,

we sort activities by time to construct the activity sequence for each student, where $D = \{S_z, T_z, L_z\}$ represents activity Z , S_z identifies a student, and T_z and L_z are respectively the time and location of activity Z , respectively. The activity sequence lays a foundation for subsequent association analysis.

3.2. Qualitative Analysis of Association

We propose three operators to measure the similarity of students' activity sequences. The higher the similarity the stronger the association. We construct an association network of N students, using the similarity value of two students as the weight of the edge connecting them. From this network, we can understand the overall relationship among students. However, the network will become increasingly complex with a growing number of students. To solve this problem, we used the Louvain algorithm to discover a hierarchical community structure, which reduces the complexity of the network while preserving the main information.

3.3. Visual Analysis of Association

Compared with qualitative analysis of association, visualization analysis can make the association more intuitive. We introduce some visualization techniques, such as chord diagram and Fruchterman–Reingold (FR) layout, which are widely used to represent association patterns such as in social and biological networks. These help us to intuitively and interactively understand the association among students.

4. Privacy Protection

The student services department provided support regarding privacy. During the student enrollment period, they asked students whether they would like to share their campus activity data for analysis. With students' consent, we took additional measures to protect their privacy. We created a mapping table between real student IDs and encoded student IDs; every real ID was encoded as a unique, anonymous, alphanumeric identifier. During data collection, real student IDs in every data source were replaced by their corresponding encoded identifiers, assuring anonymity throughout the experiment. A day was divided into 144 10-min bins. Behavior time was therefore encoded as a number in the range [1,144], and we could not obtain the precise behavior time. Verification was performed by the student services department; they invited some students as volunteers to verify the experimental results. Finally, the researchers involved in these experiments signed a confidentiality agreement.

5. Data Collecting and Feature Extracting

Our dataset includes records of book borrowing, showering, library entrance, and meals of 8685 students enrolled in the Spring 2017 semester.

To compute the behavior similarity among students, we extracted the following seven features from the original behavior data.

- Number of smart card transactions. In most Chinese universities, smart cards serve as the unique payment medium in canteens, bathrooms, and libraries. Therefore, smart card consumption records can describe the activities of each student on campus. We count the number of transactions via smart card at given locations in a specified period. The higher this number, the more active a student at a specified location and time.
- Amount of transactions via smart card. The sum of transaction amounts can evaluate the consumption level of a student in a specified location and period.(i.e., a semester or a month).
- Consumption frequency during peak period. A peak period is a time interval during which most students perform the same activity. According to the teaching schedule, we define three peak consumption periods: 7 a.m. to 9 a.m. for breakfast, 11 a.m. to 1 p.m. for lunch, and 5 p.m. to 8 p.m. for dinner. We defined a triple to store the frequency of consumption during the three peak periods, through which we can measure the level of orderliness of each student.

- Number of days that students have a regular lifestyle on campus. A regular lifestyle connotes consumption frequency during peak periods in the normal range.
- Entropy of activity locations. The degree of dispersion of activity locations reflects a student's level of disorder in a period. We acquire all the activity locations and compute the entropy as Equation (1), where L_u denotes all activity locations visited by student u , $L'_u = L_u \cap \{l\}$ represents the activity locations in the specified area l visited by student u , $|L'_u|$ is the total number of visits to area l by student u , and $P_u(l) = |L'_u|/|L_u|$ is the probability that student u visits area l .

$$\text{Entropy_Loc}(u) = - \sum_{l \in L_u} P_u(l) \log P_u(l). \quad (1)$$

- Entropy of activity time. This measures the degree of dispersion of each student's activity time. Similar to the entropy of activity locations, we acquired all activity times and locations, and computed this entropy as Equation (2), where T_u is overall time distribution of visits to the specified area by student u . $T'_u \subset T_u$ is the specified part of T_u , $|T'_u|$ is the number of times student u visits the area in the specified time interval, and $|T_u|$ is the total number of visits by student u to the given area. $P_u(t) = |T'_u|/|T_u|$ is the probability that student u visits the given area in time interval t .

$$\text{Entropy_Time}(u) = - \sum_{t \in T_u} P_u(t) \log P_u(t). \quad (2)$$

- Number of books borrowed from library. This can be computed through the book-borrowing records, which are filtered by borrowing time and book type.

6. Qualitative Analysis of Association among Students

When two students' activity sequences are similar, we generally think that they are closely associated. To measure this similarity, we propose spatial similarity, spatiotemporal similarity, and behavior features-based similarity operators, through which we can compute the degree of association among students, and take this as the weight of the edge in the association network. However, the network will become increasingly complicated with the number of nodes, so to clearly understand the association structure, we introduce the Louvain algorithm [17] to discover the student communities.

6.1. Three Similarity Operators

6.1.1. Similarity of Spatial Patterns

The behavior spatial pattern refers to students' spatial movement mode on campus, and does not consider the activity time. To express this pattern, we built a vector for each student to represent the frequency of activity at the main locations on campus. When two students are often active in the same locations, we think their behavior spatial patterns are similar. We compute the similarity by Equation (3), where $\text{SpaSim}(u, v)$ is this similarity operator, u and v are two students, L_u is the set of activity locations of student u , $l_u^i \in L_u$ is the i th activity location, w_u^i is the frequency of visits by user u at the i th location, and $\text{len}(L_u)$ is the cumulative visiting frequency at all activity locations of student u . If the co-visited location set is not empty, then $L_u \cap L_v \neq \phi$, and $\text{SpaSim}(u, v) > 0$; when $L_u = L_v$, $\text{SpaSim}(u, v) = 1$:

$$\begin{cases} \text{len}(L_u) = \sqrt{\sum_{l_u^i \in L_u} w_u^i} \\ \text{SpaSim}(u, v) = \sum_{l^i \in (L_u \cap L_v)} \frac{w_u^i w_v^i}{\text{len}(L_u) \text{len}(L_v)}. \end{cases} \quad (3)$$

6.1.2. Similarity of Spatiotemporal Patterns

Compared to the spatial pattern, the spatiotemporal pattern considers both the activity location and time. To express this pattern, we constructed the activity sequence including all activities. We divided a day into 144 time bins, so an activity time can be mapped to a discrete value. The activity location was labeled according to the campus map. After these pre-processing steps, we sorted the activity set by time and construct the activity sequence. When the sequences of two students overlap significantly, the similarity value of their spatiotemporal patterns should be high. We formulated this as Equation (4), where $\text{Act}(u, v)$ indicates the frequency of activities in which students u and v participated simultaneously, S_u is the activity sequence of student u , $\text{ActSim}(u, v)$ is the spatiotemporal similarity between u and v , ActNum is the total number of activities participated in by all students, and $\text{len}(A_{k,u,v})$ is the number of individuals appearing in the common sequence when $\text{Act}(u, v)$ occurs:

$$\begin{cases} \text{Act}(u, v) = |S_u \cap S_v| \\ \text{ActSim}(u, v) = \sum_{k=1}^{\text{ActNum}} \frac{\text{Act}(u, v)}{\text{len}(A_{k,u,v})} \end{cases} \quad (4)$$

6.1.3. Similarity of Behavior Features

The above two operators use the spatiotemporal co-occurrences to measure the similarity. However, whether we can use the behavior features to infer social association between students? To validate this hypothesis, we propose the behavior features-based similarity operator based on the features described in Section 5, and formulate this operator as Equation (5), where u and v denote two different students, m_u is the feature vector of student u , and m_u^d is the d -dimensional feature of m_u . We compute the Euclidean distance $\text{FeaDis}(u, v)$ and similarity of behavior features $\text{FeaSim}(u, v)$, which is the negative exponential function of $\text{FeaDis}(u, v)$, between m_u and m_v . δ is a tradeoff factor to restraint $\text{FeaSim}(u, v)$ to the range $[0, 1]$, $\delta = 1/2N(u, v) \times \sum_{u,v=1}^{N(u,v)} \text{FeaDis}(u, v)$, where $N(u, v)$ is the maximum number of pairwise students:

$$\begin{cases} \text{FeaDis}(u, v) = \sqrt{\sum_{d=1}^7 (m_u^d - m_v^d)^2} \\ \text{FeaSim}(u, v) = e^{-\text{Nor}(\text{FeaDis}(u, v)) - \delta} \end{cases} \quad (5)$$

6.2. Discovering Student Communities Based on Modularity Optimization

Based on the three operators, we can compute the similarity value among students, and then construct the association network. However, it becomes more difficult to retrieve the association information from the network as the number of nodes increases. A promising approach is to decompose the network into communities, whose nodes are highly interconnected and among which there are fewer connections [27]. Researchers have proposed many algorithms to reasonably partition the network. Among them, Louvain algorithm is one of mostly adopted community algorithms, and it has superior performance compared with its competitors, as stated in [16,18–21]. We use the algorithm to discover student communities.

6.2.1. Definition of Modularity

Newman highlighted the property of community structure and proposed the algorithms to divide the network into communities based on modularity optimization [9,27,28]. Modularity in a weighted network is defined as Equation (6) [29], where $w(i, j)$ is the weight of the edge-connecting nodes i and j , and $k_i = \sum_j w(i, j)$ is the sum of the weights of edges connected to node i , and c_i is the community to which node i is assigned. The function $\varphi(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise, and $m = \frac{1}{2} \sum_{i,j} w(i, j)$. The modularity Q is between -1 and 1 , and it can be used as an objective function to look for the divisions with high modularity over all possible partitions:

$$Q = \frac{1}{2m} \sum_{ij} [w(i, j) - \frac{k_i k_j}{2m}] \varphi(c_i, c_j). \quad (6)$$

6.2.2. Discovering Student Communities Using the Louvain Algorithm

The Louvain algorithm has two phases that are reapplied iteratively. For a weighted network of N nodes, it first views each node as an independent community. Then it takes the neighbors j of i for each node i and evaluates the gain of modularity from placing the node i in the community of node j ; placing occurs when the gain is positive and maximum; otherwise, node i remains in its original community. This algorithm executes this operation repeatedly on all nodes in a specified order until a local maximum of the modularity is attained. The gain in modularity ΔQ obtained by moving node i to community C can be computed as Equation (7) [17]:

$$\Delta Q = [\frac{\sum_{in} + k_{i,in}}{2m} - (\frac{\sum_{tot} + k_i}{2m})^2] - [\frac{\sum_{in}}{2m} - (\frac{\sum_{tot}}{2m})^2 - (\frac{k_i}{2m})^2], \quad (7)$$

where $\sum_{in}$ is the sum of the weights of the edges in community C , $\sum_{tot}$ is the sum of the weights of the edges connected to nodes in community C , k_i the sum of the weights of the edges connected to node i , $k_{i,in}$ is the sum of the weights of the edges from node i to nodes in community C , and m is the sum of the weights of all the edges in the network.

The second phase of the Louvain algorithm constructs a new network, whose nodes represent the communities discovered during the first phase. The weight of the edges between the new nodes is the sum of the weights of the edges between nodes in the corresponding communities, and the new nodes may have self-loops, whose weights are the sum of the weights of all the edges in the corresponding communities. When the second phase is completed, we can reuse the first phase in the new network; this iterative operation does not converge until the maximum of modularity is attained or the gain of modularity does not exceed a specified threshold. Through the iterative operation, the Louvain algorithm constructs the hierarchical community structure.

To discover student communities using the Louvain algorithm, we constructed a weighted network $G = (V, E)$ of 8685 nodes, where V is the set of nodes representing the students, and E is the set of edges, whose weight is computed using the above similarity operators. The process is shown as Algorithm 1.

We implemented this algorithm using Java and ran it in multi-threading mode. The network $G = (V, E)$ can be divided into k non-intersecting subnets $G_i = (V_i, E_i) (i = 1, 2, \dots, k)$. Based on the results, we can understand the association subnet for each student.

Algorithm 1 Algorithm to discover student communities.

Require: Weighted Association Network $G = (V, E)$
Ensure: $G_i = (V_i, E_i), (i = 1, 2, \dots, k)$

```

1: viewing each node as an independent community
2: Flag= False
3: while !Flag do
4:   Flag= True
5:   for  $i \in V$  do
6:     maxgain=0
7:     index=-1
8:     for  $j \in \text{neighbors of node } i$  do
9:       calculating  $\Delta Q$  by using Equation (7)
10:      if  $\Delta Q > \text{maxgain}$  then
11:        maxgain= $\Delta Q$ 
12:        index= $j$ 
13:      end if
14:    end for
15:    if maxgain>0 then
16:      placing node  $i$  in the community of corresponding node index
17:      Flag= False
18:    end if
19:  end for
20:  if !Flag then
21:    constructing the new network  $G'$ 
22:     $G = G'$ .
23:  end if
24: end while

```

7. Visual Analysis of Association

Different than qualitative analysis, visual analysis provides a more intuitive way for educational administrators to understand a student's social association. We introduce visualization charts, such as chord diagrams and the FR algorithm, to express the association among students. We next explain the usage of chord diagram and FR algorithm; see Section 8.5 for their specific application.

7.1. Chord Diagram

A chord diagram is a popular visualization chart that can express the relationship among nodes in complex social networks. It consists of nodes and chords, where a chord connects two nodes that have a relationship. We use a tuple $(G = (N, E, \omega), \gamma, \phi, \chi)$ to represent the chord diagram [30], where G is a social network graph; N is the set of nodes; $E \subset N \times N$ is the set of chords; ω is a function that assigns a non-negative, non-zero real number, known as a weight, to each chord; γ is the radius of the graph; ϕ is a function that computes the size of the node in the chord diagram; χ is a function that computes the width of the chord.

we used the chord diagram to develop an association visualization view representing the relationship among students, and take the spatiotemporal similarity operator as the function ω to compute the weight. When the similarity value exceeds a threshold, there is a chord connecting the nodes. The function ϕ is proportional to the number and weight of the chords connected to the node, and χ is identical to ω . As shown in Figure 2, one node denotes a student identified by a student ID. For example, the large size of the node representing student 150302cf indicates that this student has many associations with others; while student 150301ai has fewer associations. For better visualization effect, we used prior knowledge, such as class or grade, to categorize students, and we marked the corresponding nodes and their chords in different colors. Meanwhile, users can interact with the chart, when a user mouses over a node, the connected chords and their weights are highlighted.

Figure 3. Force-directed layout for relationships among students.

8. Experimental Results

To evaluate the performance of the proposed method, we carried out a set of experiments on a real dataset, and provided the experimental results to student services department for verification.

8.1. Comparison of Similarity Operators

We compared the accuracy of three similarity operators and their weighted combination to select the best one. The six operators being compared are listed in Table 1. We separately use these operators to calculate the similarity value among students. We take the top 50 most similar individuals for each student. We randomly invite some undergraduates as volunteers to label whether they were familiar with the 50 students, and computed the ratio. Table 2 presents the partial experimental results. For example, 78% of the 50 most similar students computed through the spatiotemporal operator are labeled by student 151441af as being familiar; this ratio is the highest among the results of the six operators, which indicates that the spatiotemporal similarity operator is the best one. We took the results of these volunteers as samples, and calculated the average accuracy ratio for each operator. This was 87% for the spatiotemporal similarity operator, which is the highest, so we used it to compute the similarity in subsequent experiments.

Table 1. Description of all similarity operators.

Similarity Operators	Description
SpaSim	Similarity of spatial patterns
SpaTemSim	Similarity of spatiotemporal patterns
FeaSim	Similarity of behavior features
Spa_SpaTemSim	Weighted combination operator of similarity of spatial patterns and similarity of spatiotemporal patterns
Spa_FeaSim	Weighted combination operator of similarity of spatial patterns and similarity of behavior features
SpaTem_FeaSim	Weighted combination operator of similarity of spatiotemporal patterns and similarity of behavior features

Table 2. Accuracy computed via six similarity operators.

Student ID	FeaSim	SpaSim	SpaTemSim	Spa_SpaTemSim	Spa_FeaSim	SpaTem_FeaSim
151441af	0.45	0.69	0.78	0.66	0.67	0.64
151441ag	0.45	0.69	0.89	0.66	0.67	0.64
151441ah	0.12	0.86	0.95	0.79	0.79	0.71
151441ai	0.33	0.85	0.94	0.66	0.8	0.75
151441ba	0.33	0.85	0.93	0.66	0.8	0.75
151441bd	0.41	0.73	0.92	0.6	0.7	0.67
151441bc	0.41	0.73	0.9	0.6	0.7	0.67
151441bf	0.37	0.94	0.92	0.6	0.88	0.82
160415gj	0.37	0.94	0.79	0.6	0.88	0.82
160416aa	0.26	0.86	0.83	0.64	0.8	0.74
160416ab	0.26	0.86	0.86	0.64	0.8	0.74
160416ae	0.32	0.73	0.91	0.6	0.69	0.65
160416ad	0.32	0.73	0.91	0.6	0.69	0.65
160416ag	0.29	0.9	0.91	0.59	0.84	0.78
160416ah	0.2	0.72	0.91	0.62	0.67	0.62
160416ai	0.2	0.72	0.89	0.62	0.67	0.62
..	..	..	..	..	..	..
Average Ratio	0.32	0.78	0.87	0.67	0.73	0.71

8.2. Discovering Communities via Louvain Algorithm

To illustrate the process of discovering student communities via the Louvain algorithm, we carried out five iterative operations; the results are shown in Table 3. After the first iteration operation, there are 1083 communities, the number of people in the largest community is 630, and the number in the smallest is 1. With increasing iteration operations, the number of communities decreases and the number of people in the largest community increases. After the fifth iteration, the gain of modularity approximately approaches zero, indicating that the algorithm has reached a stable state. We further combine the basic information of students, such as grade and class information, to analyze the discovery results, and conclude that students with lower grades tend to form large communities and those with higher grades tend to form small communities. This is consistent with common sense.

In addition, to manually verify the experiment, we constructed an association network of 88 students from a designated major, and discover the community structure via Algorithm 1. After two iterations, the iterative operation converged and the community structure stabilized at nine communities, as shown in Table 4. Students belonging to the same community should have similar spatiotemporal patterns. The no. 1 community consists of two students 150301ci and 150301dg, who live in the same room and often have meals and attend classes together, and the similarity value between them is 0.82. The no. 2 community has 42 members belonging to the same class. By communicating with the counselor, we found that the performance of this class is excellent. The 11 students in community no. 3 often play games and have meals together. The no. 4 community also has two members, 150301ab and 150301bd, who are lovers, and the similarity value is 0.63; they often have meals and participate in campus activities together. The no. 5 community also contains two students, 150302cj and 150241de, with a similarity value of 0.78, and while they have different majors, they are very familiar with each other and often do things together. The 14 students in no. 6 community are from same major, with similarity value 0.83, they belong to a learning team and hold regular study activities in library twice one week. The no. 7 community is the same as the no. 8 community, they have two girl students who are in a close friendship. The no. 9 community has 11 students who are from different majors, with a similarity value of 0.69, these students belongs to one basketball team, they often shower and have meals together after training. Through these offline verifications, we showed that the Louvain algorithm can effectively discover the associations among students based on their activity sequences.

Table 3. Results of discovering communities for all students via the Louvain algorithm.

Iteration No.	Number of Communities	Number of People in Largest Community	Number of People in Smallest Community
1	1083	630	1
2	334	801	1
3	199	880	1
4	173	961	1
5	172	1474	1

Table 4. Result of discovering community structure for 88 samples via the Louvain algorithm.

Result of First Iteration Operator		Result of Second Iteration Operator	
Community No.	Number of Members	Community No.	Number of Members
1	2	1	2
2	8	2	42
3	24	3	11
4	9	4	2
5	5	5	2
6	5	6	14
7	2	7	2
8	2	8	2
9	2	9	11
10	14		
11	2		
12	2		
13	11		

8.3. Features of Communities

For the communities discovered whose members have specific behavior features that are different from other communities, we used a parallel coordinate graph to simultaneously describe the distribution of values of the features. Figure 4 is a parallel coordinate graph that shows the features of a group. Through this graph, we can understand that the members of this group have lower social activity index and lower regular consumption index, which represent that these students have an irregular living style, have fewer social activities, and often are not on the campus.

Figure 4. Parallel coordinate graph representing behavior features of a group.

8.4. Exploring Associations via Chord Diagram

We used chord diagrams to visualize the association among students, as shown in Figures 5 and 6. The thickness and direction of a chord are important factors that represent the similarity of activity sequences of students. Through these diagrams, we can explore each individual's social association. In Figure 5, for example, the node representing student 150304bi only connects to one node belonging to the same category with a thicker chord, which means that this student has a close relationship with student 150304bd. They should have highly similar spatiotemporal activity sequences, but student 150304bi seldom gets involved in activities with other classmates. We suggest that this type of student should expand his or her social association by participating in various activities. In Figure 6, node 150304bf has more associations with others, and these associated nodes belong to different classes,

which means that this student has much richer social relationships, we may recommend this type of students as team leaders.

Figure 5. Chord diagram representing social relationship of student 150304bi.

Figure 6. Chord diagram representing social relationship of student 150304bf.

8.5. Exploring Associations via FR Algorithm

We used the FR layout to visualize the association network. When the similarity of the activity sequences between two students exceeds a threshold, there is an edge connecting the corresponding nodes. As shown in Figure 7, the node representing student 150303ba is at the periphery and has no connection with others, which means that this student's activity sequence is quite different from others in the specified group. Figure 8 is an enlargement of the center area of Figure 7; the nodes located in this area have many more connections with others, which means that the corresponding students have rich social associations. Through the FR layout, we can clearly observe the structure of network. For students who has fewer association with others, we can guide them how to link their social ties with others; and for students who have much associations, it is clear that which students they should build connection.

Figure 7. Layout of FR algorithm representing social relationship of student 150303ba.

Figure 8. Center area of Figure 7; the students located here have rich social relationships.

9. Conclusions and Future Work

We have proposed a visual analytic method to exploit the association among students based on their spatiotemporal behavior data on campus. We collected multi-source behavior data from different departments to construct a relatively complete activity sequence, which is superior to single-behavior data. We proposed three similarity operators with which to construct an association network. In addition to the commonly used spatiotemporal co-occurrences operators, we introduce an operator based on behavior features, although its performance is far lower than other spatiotemporal related operators. Based on the best spatiotemporal similarity operator, with 87% accuracy, we constructed an association network among students, and further discovered the student communities using the Louvain algorithm. In an experiment with 88 students, most detected communities are consistent with ground truth. Finally, we visualized the association network through chord diagram and FR layout. The experimental results prove that the proposed method can be helpful to educational administrators. For example, they can obtain some clues about outlier students who have less association with others, especially with classmates or roommates, and find the leaders at the centers of the communities. We plan to continue our research from the following aspects: (1) Integrate more types and longer periods of behavior data to construct a dynamic association network; (2) detect communities in dynamic network to understand the changes in students' social relationship; (3) analyze the correlation between measures such as social association, academic performance, and mental health.

Author Contributions: Conceptualization, Y.Z. and X.L.; methodology, X.L., Q.Y. and Y.Z.; software, X.L. and Q.Y.; validation, Q.Y., J.D. and X.L.; formal analysis, X.L. and Q.Y.; investigation, Y.Z.; resources, B.Y. and Y.Z.; data curation, X.L. and Q.Y.; writing—original draft preparation, X.L., Q.Y. and J.D.; writing—review and editing, X.L.

and Y.Z.; visualization, Q.Y. and J.D.; supervision, Y.Z.; project administration, B.Y.; funding acquisition, B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China under Grant 61632006, U19B2039.

Acknowledgments: The authors thank the volunteers for validating the experimental results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, J.; Liu, T.; Yang, L.; Davison, M.L.; Liu, S. Finding College Student Social Networks by Mining the Records of Student ID Transactions. *Symmetry* **2019**, *11*, 307. [\[CrossRef\]](#)
2. Liu, T.; Yang, L.; Liu, S.; Ge, S. Inferring and analysis of social networks using RFID check-in data in China. *PLoS ONE* **2017**. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Ghosh, S.; Ghosh, S.K. Exploring the association between mobility behaviours and academic performances of students: A context-aware traj-graph (CTG) analysis. *Prog. Artif. Intell.* **2018**, *7*, 307–326. [\[CrossRef\]](#)
4. Cao, Y.; Gao, J.; Lian, D.; Rong, Z.; Shi, J.; Wang, Q.; Wu, Y.; Yao, H.; Zhou, T. Orderliness predicts academic performance: Behavioural analysis on campus lifestyle. *J. R. Soc. Interface* **2018**, *15*. [\[CrossRef\]](#)
5. Zhang, X.; Sun, G.; Pan, Y.; Sun, H.; He, Y.; Tan, J. Students performance modeling based on behavior pattern. *J. Ambient Intell. Hum. Comput.* **2018**, *9*, 1659–1670. [\[CrossRef\]](#)
6. Wu, Y.; Gong, R.; Cao, Y.; Wen, C.; Teng, Z.; Pu, J. eduCircle: Visualizing Spatial Temporal Features of Student Performance from Campus Activity and Consumption Data. In Proceedings of the 13th International Conference, CDVE 2016, Sydney, NSW, Australia, 24–27 October 2016; pp. 313–321.
7. JO, I.; Park, Y.; Kim, J.; Song, J. Analysis of Online Behavior and Prediction of Learning Performance in Blended Learning Environments. *Educ. Technol. Int.* **2014**, *15*, 71–88.
8. Nie, M.; Yang, L.; Sun, J.; Su, H.; Xia, H.; Lian, D.; Yan, K. Advanced forecasting of career choices for college students based on campus big data. *Front. Comput. Sci.* **2018**, *12*, 494–503. [\[CrossRef\]](#)
9. Newman, M.E.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113. [\[CrossRef\]](#)
10. Crandall, D.J.; Backstrom, L.; Cosley, D. Inferring social ties from geographic coincidences. *Proc. Natl. Acad. Sci. USA* **2010**, *52*, 22436–22441. [\[CrossRef\]](#)
11. Huy, P.; Shahabi, C.; Liu, Y. Inferring Social Strength from Spatiotemporal Data. *ACM Trans. Database Syst.* **2016**, *71*, 1–47.
12. Jung, J.; Chun, S.; Jin, X.; Lee, K. Quantitative Computation of Social Strength in Social Internet of Things. *IEEE Internet Things J.* **2018**, *5*, 4066–4075. [\[CrossRef\]](#)
13. Rafailidis, D.; Crestani, F. Friend Recommendation in Location-based Social Networks via Deep Pairwise Learning. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, 28–31 August 2018; pp. 421–428.
14. He, C.; Peng, C.; Li, N.; Chen, X.; Guo, L.Y. Exploiting Spatiotemporal Features to Infer Friendship in Location-Based Social Networks. In Proceedings of the 15th Pacific Rim International Conference on Artificial Intelligence, Nanjing, China, 28–31 August 2018; pp. 395–403.
15. Zhou, N.N.; Zhang, X.; Wang, S. Theme-Aware Social Strength Inference from Spatiotemporal Data. In Proceedings of the 15th International Conference on Web-Age Information Management (WAIM), Macau, China, 16–18 June 2014; pp. 498–509.
16. Fortunato, S.; Hric, D. Community detection in networks: A user guide. *Phys. Rep.-Rev. Sect. Phys. Lett.* **2016**, *659*, 1–44. [\[CrossRef\]](#)
17. Blondel, V.D.; Guillaume, J.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory E* **2008**, *2008*, 10008–10012. [\[CrossRef\]](#)
18. Yang, Z.; Algesheimer, R.; Tessone, C.J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Sci. Rep.* **2016**, *6*. [\[CrossRef\]](#)
19. Lancichinetti, A.; Fortunato, S. Community detection algorithms: A comparative analysis. *Soft Comput. A Fusion Found. Methodol. Appl.* **2009**, *2*. [\[CrossRef\]](#)

20. Mothe, J.; Mkhitarian, K.; Haroutunian, M. Community detection: Comparison of state of the art algorithms. In Proceedings of the 2017 Computer Science and Information Technologies (CSIT), Yerevan, Armenia, 25–29 September 2017; pp. 125–129.
21. Chejara, P.; Godfrey, W.W. Comparative analysis of community detection algorithms. In Proceedings of the 2017 Conference on Information and Communication Technology (CICT), Gwalior, India, 3–5 November 2017; pp. 1–5.
22. Jain, L.; Katarya, R. Discover opinion leader in online social network using firefly algorithm. *Expert Syst. Appl.* **2019**, *122*, 1–15. [[CrossRef](#)]
23. Wu, K.; Tang, J.; Long, Y. Delineating the Regional Economic Geography of China by the Approach of Community Detection. *Sustainability* **2019**, *11*, 6053. [[CrossRef](#)]
24. Cai, W.; Wang, Y.; Lv, R.; Jin, Q. An efficient location recommendation scheme based on clustering and data fusion. *Comput. Elect. Eng.* **2019**, *77*, 289–299. [[CrossRef](#)]
25. Linhares, C.D.G.; Ponciano, J.R.; Pereira, F.S.F.; Rocha, L.E.C.; Paiva, J.G.S.; Travencolo, B.A.N. Visual analysis for evaluation of community detection algorithms. *Multimedia Tools Appl.* **2020**. [[CrossRef](#)]
26. Crampes, M.; Plantie, M. A Unified Community Detection, Visualization and Analysis Method. *Adv. Complex Syst.* **2014**, *17*. [[CrossRef](#)]
27. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2001**, *99*, 8271–8276. [[CrossRef](#)] [[PubMed](#)]
28. Newman, M.E.J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [[CrossRef](#)] [[PubMed](#)]
29. Newman, M. Analysis of weighted networks. *Phys. Rev. E* **2004**, *70*, 056131. [[CrossRef](#)] [[PubMed](#)]
30. Jalali, A. Supporting Social Network Analysis Using Chord Diagram in Process Mining. *Int. Conf. Bus. Inform. Res.* **2016**, 16–32. [[CrossRef](#)]
31. Battista, G.D.; Eades, P.; Tamassia, R.; Tollis, I.G. Algorithms for drawing graphs: An annotated bibliography. *Comput. Geom.* **1994**, *4*, 235–282. [[CrossRef](#)]
32. Fruchterman, T.M.J.; Reingold, E.M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **1991**, *21*, 1129–1164. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).