applied sciences

MDPI

Article

From Social to Academic: Associations and Predictions Between Different Types of Peer Relationships and Academic Performance Among College Students

Jiadong Tian [1,2], Jiali Lin [3] and Dagang Li [1,2,*]

[1] School of Computer Science and Engineering, Macau University of Science and Technology, Macau 999078, China; 2009853gia30009@student.must.edu.mo

[2] Zhuhai-M.U.S.T. Science and Technology Research Institute, Zhuhai 519031, China

[3] Key Laboratory of Neuroeconomics, Guangzhou Huashang College, Guangzhou 511300, China; jiali_lin@gdhsc.edu.cn

* Correspondence: dgli@must.edu.mo

Featured Application: The results are applicable to smart campus-related applications, including social behavior analysis, academic performance prediction, academic difficulty prediction, and other scenarios.

Abstract: This study aims to expose the correlation between different types of social behaviors and the academic performance of college students, and then to predict the academic performance of college students based on their social characteristics. We extracted and computed information on social relationships for roommates, classmates, members of the opposite sex, and others, based on real consumption data of 5597 freshmen students. The correlations between different types of peer relationships and academic performance were compared. Subsequently, we used Random Forests and Neural Networks as baseline methods, and introduced Graph Convolutional Network and Dynamic Graph Convolutional Network algorithms, on top of a graph network model based on social characteristics, to predict students' academic performances. The results show that the quantity and quality of all types of socialization are positively correlated with academic performance, and socialization among classmates and roommates demonstrates a stronger correlation. In addition, with the construction of the graph model and the integration of time-series information, the prediction accuracy of the dynamic graph convolution method improved compared to other methods. The findings demonstrate the advantages of using social characteristics for academic performance prediction, and reveal the significant potential of AI applications in supporting the field of school management.

Keywords: social behavior; academic performance; graph network modeling; educational data analysis

Academic Editors: Eloy López Meneses and César Bernal-Bravo

Received: 22 December 2024
Revised: 5 February 2025
Accepted: 17 February 2025
Published: 20 February 2025

Citation: Tian, J.; Lin, J.; Li, D. From Social to Academic: Associations and Predictions Between Different Types of Peer Relationships and Academic Performance Among College Students. *Appl. Sci.* **2025**, *15*, 2262. https://doi.org/10.3390/app15052262

1. Introduction

With the help of academic predictions, students can make well-informed decisions about their academic and career paths, and colleges and universities can proactively identify students who may not be able to graduate and provide them with the support they need to ensure their success [1,2]. Numerous studies have shown that academic support for college students is critical to their academic performance [3–6], and social support from close others is significantly associated with higher academic performance [7,8]. In related studies, peer relationships are typically categorized into three dimensions: classmates, roommates, and interactions with the opposite sex. However, few studies have simultaneously explored these relationships and compared their associations with academic performance. The present study uses the same methodology to extract these three types of peer relationships based on real-world data. We conducted two correlation analyses aimed at exploring the correlations between the quantity and quality of peer relationships and academic performance. In addition, we conducted a prediction experiment to assess the feasibility of predicting academic performance based on peer relationships.

1.1. Correlation Between Peer Relationships and Academic Performance

Numerous academic studies have confirmed that peer relationships are significantly associated with academic performance. Common peer relationship dynamics consist primarily of peer acceptance and peer rejection. Peer acceptance is usually associated with positive affect and peer rejection with negative affect [9]. Positive and negative peer relationships tend to predict high and low academic performance, respectively [10,11]. These studies have focused on the correlation between peer relationship quality and academic performance.

As an important component of peer relationships, classmate relationships are significant predictors of academic performance [12–17]. For example, in a Korean middle school where students were randomly assigned to different classes, researchers examined the presence of peer academic interactions and their structure in the classroom. The results showed that high-quality peer interactions enhance students' performance, with high-performing students being more inclined to interact with their peers [12]. In addition, both the quantity and quality of peers have been shown to improve student performance, and the breadth and cohesiveness of peer networks positively affect academic performance [17].

The correlation between roommate relationships and academic performance has also been confirmed [18–21]. High-quality roommate relationships are linked to the development of students' cognitive skills. For example, one study examined the distribution and changes in peer effects using regression methods based on longitudinal data from 5272 undergraduate students. The study found that social interactions between roommates were significantly more frequent than expected by chance and that further assimilation occurred over time, especially when roommates behaved similarly [18]. In a study of peer effects among graduate students at a Chinese university, master's students' GPAs were positively correlated with their roommates' university GPAs [19].

The correlation between opposite sex relationships and academic performance has also been supported by research [22–26]. Studies have shown that a higher proportion of female peers in a class can improve students' test scores and non-cognitive outcomes, including social adjustment and school satisfaction [23–26]. This may be due to the fact that a higher proportion of female classmates brings about healthier behavioral patterns. Alternatively, it has been shown that students in classes with a higher proportion of male classmates perform lower in math and English [24].

Nonetheless, the findings are not entirely consistent. One study found that peer problems were not associated with math achievement test results after considering ADHD, behavioral problems, and emotional problems [27]. This contrasts with the findings of Gerbino et al., who found that pro-social behaviors were significantly associated with overall achievement, even after accounting for multiple other factors such as intelligence and personality traits [10]. Additionally, one study noted that students' perceived academic engagement was positively related to student achievement, while there was no clear link between perceived peer support and academic performance [28]. Another study found that being in a romantic relationship during college was significantly associated with missing classes but not with grade point average [29].

1.2. Comparative Study of Different Types of Peer Relationships

Despite the large number of studies that have explored the association between different types of peer relationships and academic performance, there are fewer comparative studies of different types of peer relationships [30,31]. However, research has found that the impacts of different types of peer relationships are different [32]. In order to deepen theoretical understanding and support educational practice planning, comparative studies between peer relationships are essential.

Several studies have compared same sex and opposite sex relationships. A study of 246 emerging adults at a Midwestern university found that academic support from friends and academic support from partners are two different concepts. Academic support from friends was positively correlated with a student's GPA, while that from a partner was not [31]. Results from another study of 1436 high school students (670 boys, 756 girls) showed similar conclusions: adolescents' same sex and opposite sex peer relationships positively affected their academic performance in different ways [30]. A social network is a social structure comprising a set of participants and the mutual ties between these participants [33]. The first of the above studies limited its focus to the closest friends and romantic relationships and did not cover the overall comparison between same sex and opposite sex relationships. The second study, while comparing the effects of same sex and opposite sex relationships on students' academic performance, failed to further analyze the structure of social networks. In fact, different social network structures have significant effects on students' academic performance [6,16,34]. Additionally, all of the above studies relied on self-reporting, which may not objectively reflect the real situation of social network structures.

Another study compared the differences between roommate and classmate relationships. One study analyzed peer relationship networks based on the daily behavioral data of 4738 undergraduate students enrolled in 2018. The results showed that students in the same dormitory were the first to form peer relationships, but then relationships shifted towards being classmate-centered. The number of classmates in the social network was positively correlated with the probability of receiving a scholarship [35], revealing the different effects of roommate and classmate relationships on academic performance.

Based on existing research, this study contributed improvements in objectivity, the holistic nature of relationship networks, and category segmentation. We comprehensively compared four types of peer relationships (classmates, roommates, opposite sex, and others), both qualitatively and quantitatively, on an objective dataset. The 'other' category refers to relationships that appear in the overall social relationships but do not fall into the first three categories, such as peer relationships with students from other colleges. This comparison can help us recognize which type of relationship is most strongly associated with academic performance.

1.3. Prediction of Academic Performance Through Peer Relationships

There has been a proliferation of studies using social behavior to predict academic performance. For example, there have been studies on academic prediction using interaction data from online learning platforms [36,37] and academic prediction using social characteristics [38–40]. However, studies using social data, from campus big data, for academic prediction are still relatively few. The existing studies [7,41,42] have shown that academic support provided by peers is a predictor of students' academic performance.

In the above studies, in terms of prediction path, compared with questionnaires or self-reports, consumption data have the advantage of being large-scale and objective, and can reflect students' social activities in a more realistic and detailed way, which provides a guarantee for their accuracy. In terms of prediction methods, most existing studies rely on traditional machine learning methods, such as decision trees and support vector machines, which only consider students as isolated individuals, resulting in a lack of structural information. In contrast, predictions based on social relationships can be introduced into graphical models, which allow structural and temporal information to be more fully utilized. In addition, in terms of popularity, consumption record systems also have the advantage of being more popular in universities than other big data systems, such as web access record systems, which makes it easier to promote and apply consumer data-based research in practice.

1.4. Current Research

Existing research has mainly focused on the association between online social networks or a single type of social behavior and academic performance [17,43]. Achievement prediction studies are mostly based on behavioral patterns, such as life disorders and internet behavioral traits. These studies have paid less attention to comparisons between different categories of peer relationships and their different effects on academic performance, which limits our understanding of the overall social behavior of college students and its impact on academic performance.

We built a social network among students based on co-dining relationships. The university campus is not only a place of study, but also a core area of college students' lives. Daily activities on campus provide rich scenarios for peer-to-peer interactions. In particular, dining with peers is seen as an important social activity [10]. Although students additionally engage in a variety of peer activities, such as group discussions, library study, or stadium exercises, these activities are less prevalent than dining in the canteen. Therefore, dining together in the cafeteria can be regarded as the main activity of peer interaction among college students, which lays the foundation for the subsequent analysis in this paper [35].

This study categorized social relationships and explored the link between various types of social relationships and academic performance in both quantitative and qualitative dimensions. Finally, we also tried to predict their academic performance through social relationships.

The specific research questions are as follows:

Research Question 1: How does the frequency of various types of socialization vary among students in different academic performance groups?

Research Question 2: Does the quantity and quality of different types of peers correlate with academic performance?

Research Question 3: Is it possible to predict students' academic performance through social behavior?

The contributions of this study are mainly as follows: (1) this study analyzes and compares students' various peer relationships and their associations with academic performance, and explores the possibility of predicting students' academic performance through social networks. These findings can enrich our understanding of the social behavior of college students and extend methods of academic performance prediction. (2) At the theoretical level, this study provides a graph-model paradigm for the study of student behavior on campus. To explore the influence of other factors on students, this can be achieved by adding node attributes to the social graph network. These results help advance data analysis in education and provide empirical support for educational theory. (3) From a practical perspective, the results of this study can assist educational administrators in identifying abnormal behaviors, such as academic difficulties, at the individual level and in making more effective decisions at the overall level.

The rest of this paper is structured as follows: Section 2 shows the process of data collection and analysis, Section 3 presents the results of the experiments, Section 4 discusses the three topics presented in this paper, and Section 5 presents the conclusions and limitations of this study.

2. Materials and Methods

2.1. Datasets

Our data came from a university in southern China. To ensure the security of the data, the university replaced the identity ID in the data with random UIDs, hiding the identity information of the students. The study was approved by the Key Laboratory of Neuroeconomics, Guangzhou Huashang College (GZHSC-202312001).

This study focused on the weighted grade point average of the final grades of first-year undergraduate courses. The weighted GPA is very important for students and serves as a critical for various aspects such as scholarships, graduate school retention, graduation, and employment. It is calculated as follows:

$$G = \frac{\sum_{i=1}^{n}(w_i * g_i)}{\sum_{i=1}^{n} w_i}. \tag{1}$$

where n is the total number of subjects, g is the grade for a single subject, and w is the number of credits for that subject.

The academic performance data of the undergraduate students were obtained from the university's Academic Management System (AMS). Students were classified into different subgroups based on their weighted average grades: $G < 60$; $60 \leq G < 70$; $70 \leq G < 80$; $80 \leq G < 90$; $90 \leq G < 100$. These multiple subgroups help us understand and identify the behavioral characteristics of students in different achievement bands in a more nuanced way, while also allowing us to study academic difficulties.

The social data came from the university's campus card system. This system records students' spending behaviors related to their campus cards, such as recharging, dining hall purchases, supermarket transactions, taking the bus, visiting the doctor, and so on. The dataset includes consumption time, consumption location, service window, spending amount, and card balance. The consumption time is generated in timestamp format; the consumption location indicates where the behavior occurs and is recorded using a uniform location identification system for the entire campus; the service window indicates the specific window where the consumption occurs; and the consumption amount indicates the size of the transaction. Table 1 shows the raw data segments.

Table 1. The raw data segments.

UID	Location	Time	Service Window	Spending	Balance
a	Canteen I	2019/09/10 12:16:03	7	15	96
b	Canteen I	2019/09/10 12:16:16	7	12	156.14
c	Hospital	2019/09/11 06:10:14	1	3	165
d	...	...			
e	Supermarket I	2019/09/20 16:26:54	1	6	43
f	Supermarket II	2019/09/24 11:22:32	1	9.9	96.14

All data are anonymized using UIDs.

Inspired by the concept of co-occurrence in the Apriori algorithm [44], we computed the co-occurring relationships of students in consumption records. Apriori is one of the classical methods for computing frequent itemsets. Frequent itemset mining is the basis for important data mining tasks such as association rules, correlation analysis, etc., and its essence is in calculating the number of co-occurrences of two objects. Similarly, in the consumption record data, if A and B appear in the same location at the same time, they are considered to co-occur, indicating that a social relationship exists, as shown in the first two records in Table 1. To rule out chance occurrences, we needed to set a threshold for the number of co-occurrences, i.e., to determine how many times the co-occurrence relationship should appear within one month for it to be considered valid. In some recent studies, Zhou et al. [35] explored the changes in peer network characteristics under different thresholds, and the changes in the association between academic performance, but they did not give the best choice of thresholds or a method for determining them. Ren and Yang [45] used all co-occurring relationships directly to represent the campus friendship network, employed exponential decay to compute the degree between nodes, and did not set a threshold value. These studies did not help us determine the threshold value. Therefore, we conducted a practical survey in which we randomly followed 10 students and asked them to report the frequency of their shared meals for one month. The results showed that the average number of co-occurrences per student during that month was 22.1. In the data-based calculations, this value is the closest to the one calculated when the threshold is 3 (number of co-occurrences per capita = 25.9), as shown in Figure 1. Consequently, in subsequent experiments, we chose 3 as the threshold value, as it better reflects the actual number of shared meals among students.

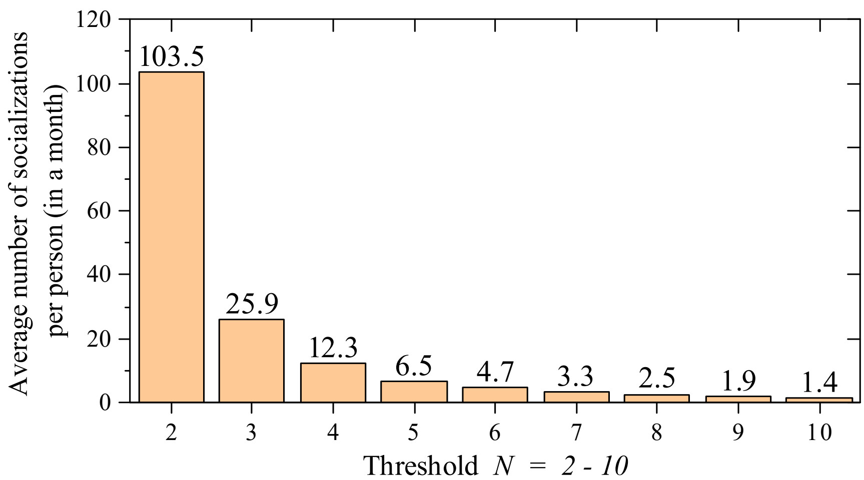


Figure 1. Data-based calculations. The average number of co-occurrences per person in a month when the threshold is taken from 2 to 10.

2.2. Population and Samplings

The university has a total of 5820 undergraduate students in the class of 2019, with a valid sample of 5597 students taking the freshman-level final exams.

In these samples, the total number of male students is slightly larger (n = 2871; 51.30%) and the rest are female students (n = 2726; 48.70%) as shown in Figure 2a. The population of the study consists of students enrolled in the year 2019. There is little difference in the subjects of exams taken and credits earned by students in the same grade, which makes their performance more comparable. The overall age distribution as of July 2020 was as follows: 22 and above (159; 2.84%), 21 (588; 10.51%), 20 (2257; 40.33%), 19 (2436; 43.52%), and 18 and below (157; 2.80%), as shown in Figure 2b.

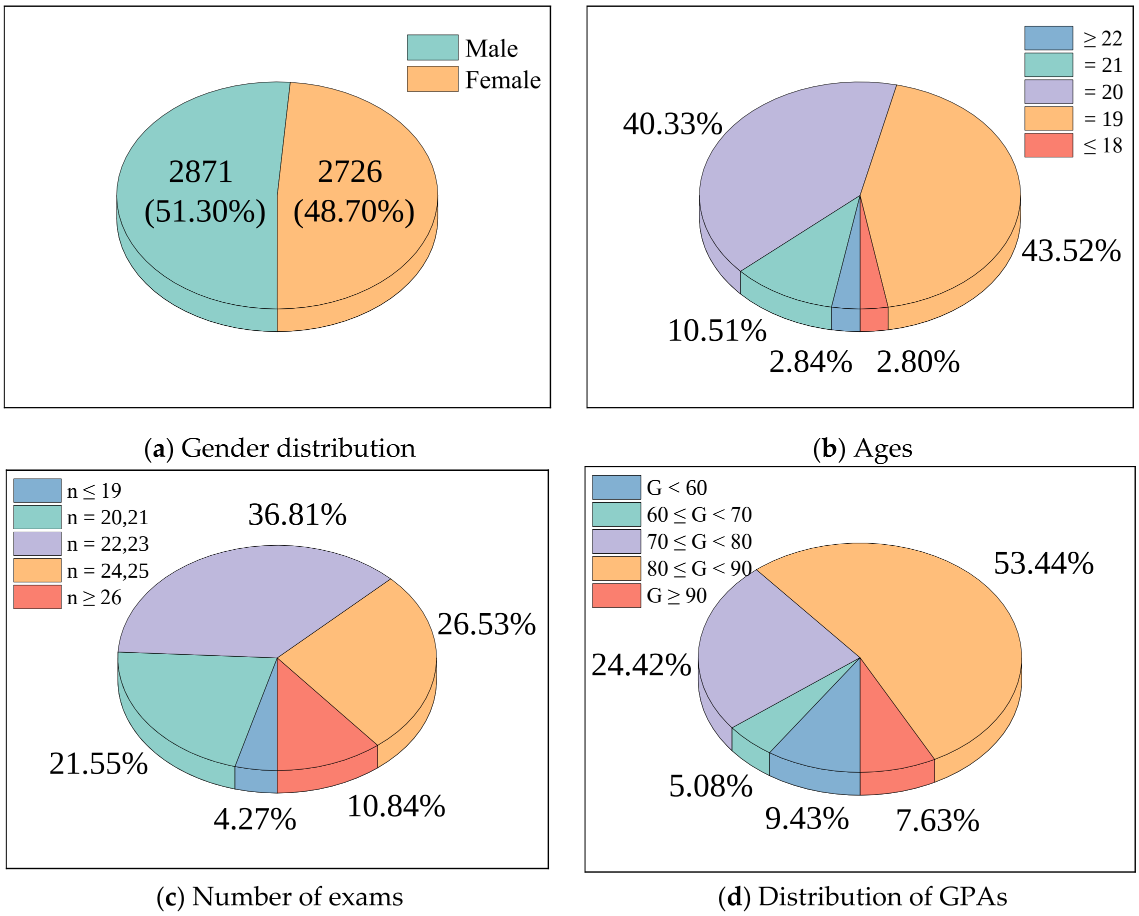
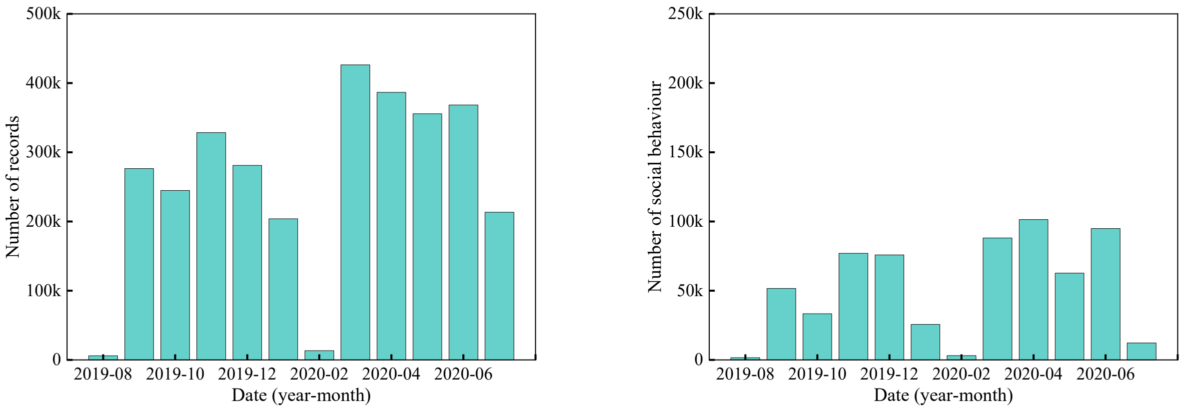


Figure 2. Population statistics.

We categorized students into five categories based on weighted GPA: $G < 60$; $60 \leq G < 70$; $70 \leq G < 80$; $80 \leq G < 90$; and $90 \leq G < 100$. Totals of 528, 284, 1367, 2991, and 427 students were enrolled in each of these categories, with the percentages shown in Figure 2d. In terms of the number of subjects, the vast majority of students took between 20 and 26 exams, which represents a sufficiently large and concentrated number of subjects, as shown in Figure 2c.

Next, we obtained the student campus card spending data from 1 August 2019, to 31 July 2020, totaling more than 3.09 million records. The variation in the number of records for each month is shown in Figure 3a. It can be seen that the number of spending records is low during vacation periods, including the summer holiday in July–August, the Spring Festival in January–February, and the impact of mini-vacations in May and October; the number is normal in all other months.

As described in Section 2.1, we calculated the number of co-occurrences between students based on time and place based on the campus card system and used it as a numerical representation of social behavior between students. The number of social behavior occurrences per month is shown in Figure 3b. We also categorized social relationships into four groups: total social relationships, roommates, classmates, and opposite sex peers based on the students' class information, dormitory information, and gender information. The total social relationship is the total co-occurring relationship presented by each student. If two individuals are both roommates and classmates, their relationship will be counted in both the roommate and classmate categories; the same applies to other relationships.



(**a**) Total number of consumption records per month (**b**) Total number of co-occurrences per month

Figure 3. Total number of consumption records and co-occurences per month.

2.3. Methods

Stage 1: Difference analysis. In order to reveal the relationship between social characteristics and achievement levels, we first compared the differences in social traits among groups with varying achievement levels at the numerical level.

Stage 2: Correlation analysis. In this stage, Pearson Correlation was applied to quantify the correlation between academic performance and social characteristics. Social characteristics include both the quantity of peers and the quality of peers.

For the quantity of peers indicator, we use the number of socializations directly. For the indicator of peer quality, we use a weighted average calculation. In this calculation, the academic performance of the peers is combined with the number of socializations as a weight and finally normalized.

Stage 3: Predictive validity validation. Finally, to explore the feasibility of predicting grades from social data, we used four common machine learning classification algorithms: Random Forest (RF) [46], a Neural Network (DNN) [47], a Graph Convolutional Network (GCN) [48], and a Dynamic Graph Convolutional Network (DGCN) [49], to predict the academic performance categories of students.

All four algorithms are applicable to multi-classification tasks. RF belongs to ensemble learning, where it obtains the classification of samples by voting after inputting sample feature data. The DNN learns sample features and performs classification through a multi-layer network structure. The GCN requires not only node features as input, but also the connections between nodes. The DGCN goes a step further by incorporating temporal information. These methods are based on different theoretical foundations and progressively use more information, which can help us to observe the impact of differences in information richness on the classification results.

For the Random Forest and Neural Network models, we used the number of socializations of students per month in five categories, including totals, roommates, classmates, and opposite sex, along with an other category, totaling 60 feature values, with academic categories as labels for classification experiments.

For graph networks, we considered students as nodes and the social relationships generated between them as edges, with the number of social relationships as degrees. If the number of co-occurrences of Student A and Student B was greater than the threshold, an edge was constructed to connect node A to node B. The number of times A and B co-occurred was the degree of this edge. The Dynamic Graph Network incorporated time information compared to the standard Graph Network; specifically, the network was split by month, resulting in a dynamic network that changed over time, as shown in Figure 4.

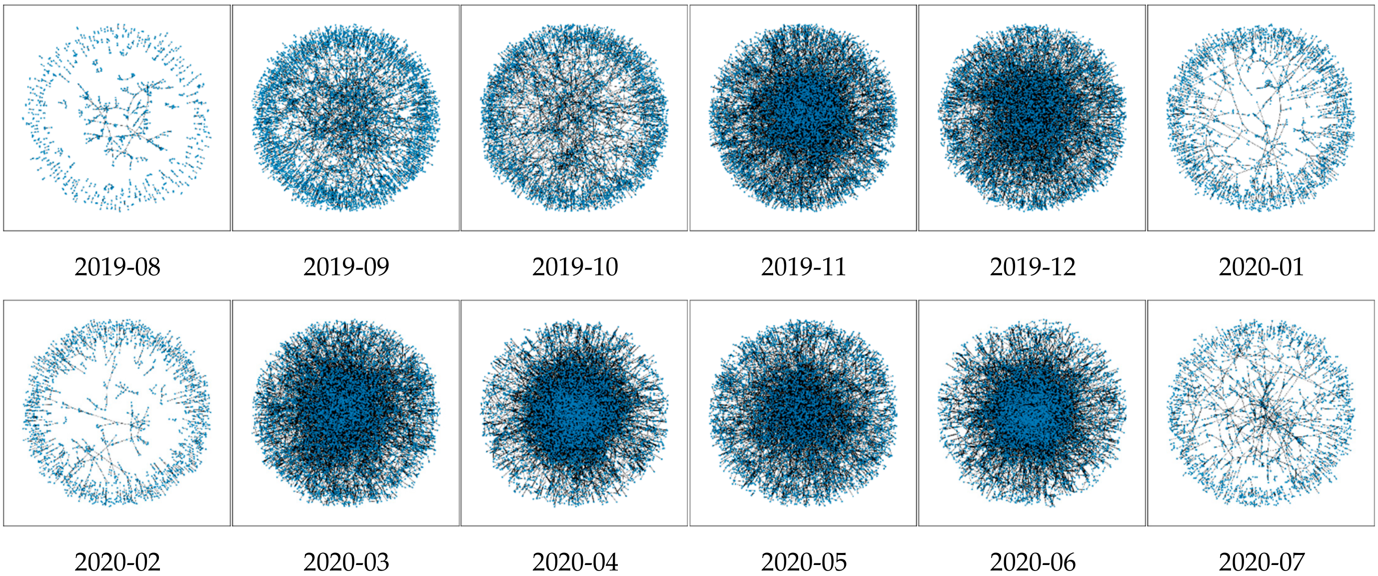


Figure 4. Social graph networks over the months.

We evaluated all methods using ten-fold cross-validation. To account for the random nature of the computation, we ran each experiment ten times and calculated the average as the final result for each method. The problem is multi-categorical and has a data imbalance, so we chose four metrics for model evaluation: accuracy, weighted average precision, weighted average recall, and weighted average F1 score. The weighted averages use the percentage of each category in the total number of samples as weights.

3. Results

3.1. Differences in the Frequency of Socialization Between Academic Performance Groups

To investigate whether there are differences in social behaviors among students in different academic performance groups, we first employed a numerical comparison method. We present the mean values of each socialization category for different subgroups of students each month, including total, classmate, roommate, opposite sex, and other, as shown in Figure 5, covering the time span from August 2019 to July 2020.

Overall, students with higher G values, i.e., better academic performance, generally socialize more, or, conversely, students who socialize more tend to perform better academically. This finding is more pronounced in classmate and roommate relationships. However, the difference in opposite sex socialization is not significant. In other forms of socialization, the pattern is reversed, with students who have higher G values engaging in less of the other types of socialization.

Over the course of the semester, the frequency of socialization as a whole is on the rise. Socialization frequency shows a significant decrease in January and February due to the Chinese New Year holiday. Additionally, in months such as October and May, there is a slight decrease, which may be related to the National Day holiday and Labor Day holiday.

Observations 5-a and 5-b indicate that classmate relationships account for the vast majority of all social relationships. Since dormitories at the university are assigned, most roommates are also classmates, resulting in a somewhat contained relationship; therefore, the value of classmates is slightly higher than that of roommates. Interestingly, while social behaviors occurring between members of the opposite sex are relatively rare, they have increased dramatically over time.

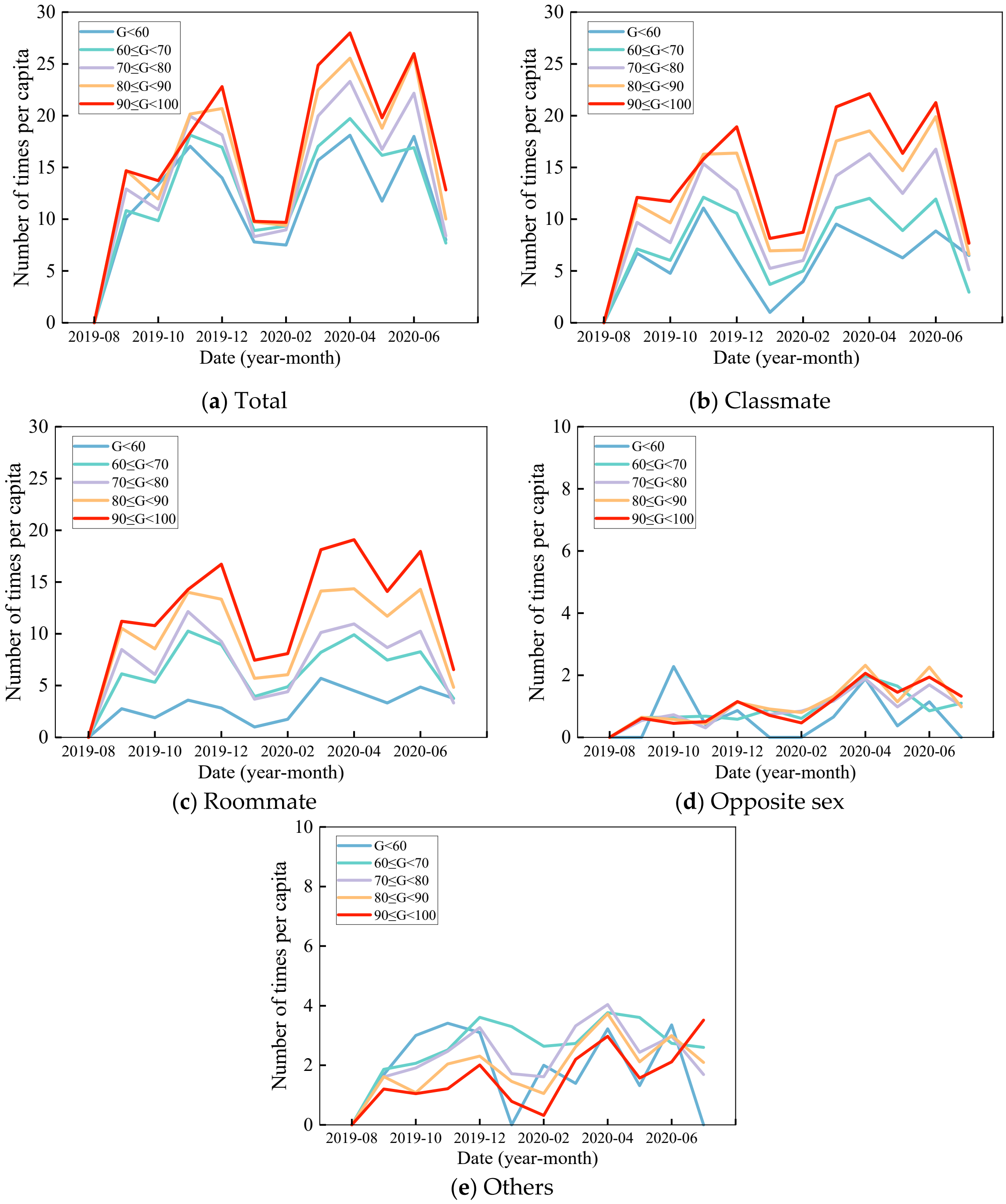


Figure 5. Comparison of socialization frequency between academic performance groups.

3.2. Correlation Between Academic Performance and the Quantity and Quality of Different Types of Peers

As shown in Figure 6, the overall frequency of social relationships is positively correlated with the academic category, and both correlations are significant ($p < 0.001$).

Regarding the timeline, there are no data for August 2019, as the freshmen of the class of 2019 have not yet enrolled. The months of January 2020, February 2020, and July 2020, which include long vacation periods, have less data, resulting in relatively small correlations, all around 0.18. In contrast, the other months exhibit more normal conditions, with correlations between total social frequency and academic category overwhelmingly greater than 0.25, and significant.

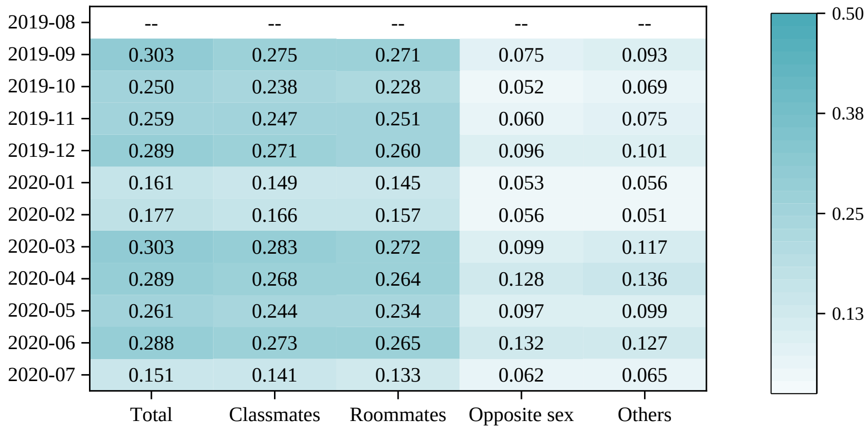


Figure 6. Correlation between socialization frequency and academic performance (p-values are all less than 0.001).

In terms of socialization categories, the correlations between total, roommate, and classmate socialization frequencies and academic categories are generally consistent, with all being around 0.25 in most months without vacations. The correlations between socialization frequency and academic category for the opposite sex and others socialization categories remain smaller compared to the first three, staying below 0.1 in most months.

Does the quality of peers correlate with individual achievement? In addition to exploring this in terms of frequency, we are also interested in the impact of peer quality. We calculated the correlation between this indicator and students' academic performance, as shown in Figure 7.

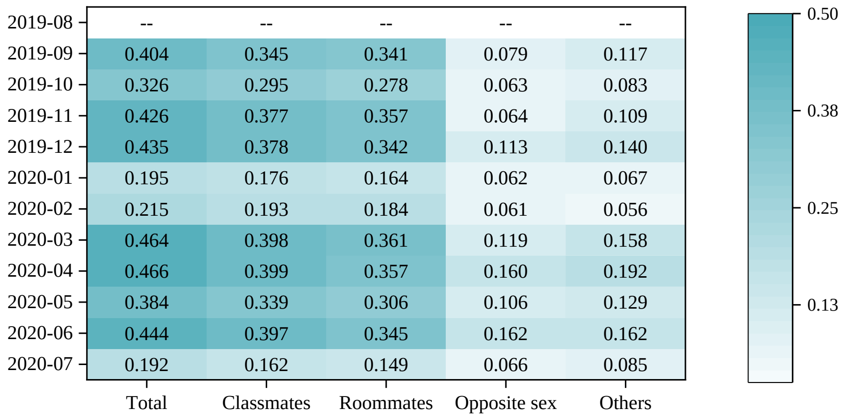


Figure 7. Correlations between peer quality and academic performance (p-values are all less than 0.001).

Overall, the quality of peers is positively correlated with the academic category. These correlations are greater than those for the frequency of socialization and are all significant ($p < 0.001$).

On the time axis, the correlations behave similarly to socialization frequency, and all correlations in the table are significant. The months of January 2020, February 2020, and July 2020 have long vacation periods, resulting in limited data, and the correlations are relatively small, all around 0.2. The months with shorter vacations (October 2019 and May 2020) exhibit slightly lower correlations, between 0.3 and 0.4. In contrast, the other months show more typical correlations, all greater than 0.4, and trending upward.

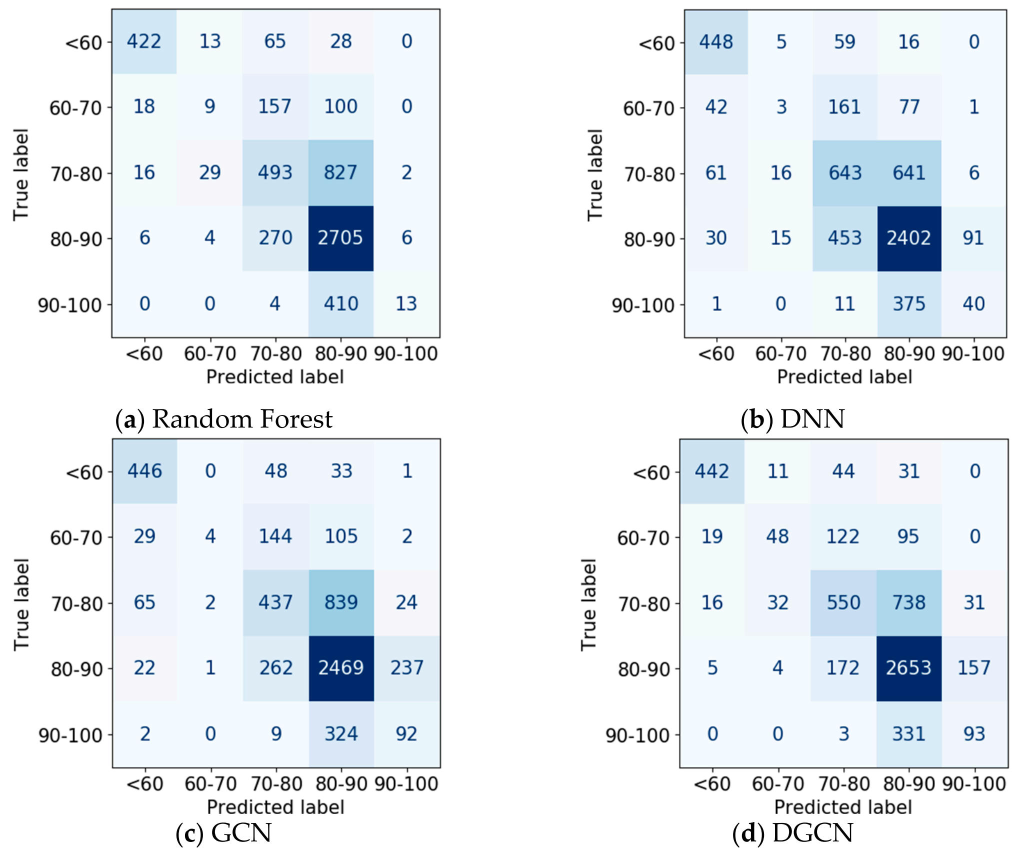
Horizontally, the quality of peers in the classmate category is slightly more relevant than that in the roommate category. The correlations are smaller for opposite sex and other types of peer quality.

3.3. Predicting Academic Performance from Social Characteristics

Although significant differences in social behaviors among students in different academic categories have been observed, these methods are not directly applicable to predicting academic performance. Therefore, we tried to use several machine learning models to predict students' academic performance categories based on their social behaviors.

We evaluated all methods using ten-fold cross-validation. Considering the randomness of the computation, we performed each experiment ten times and calculated the average as the final result for each method. The problem is multicategorical and exhibits a data imbalance, so we selected four metrics for model evaluation: accuracy, weighted average precision, weighted average recall, and weighted average F1-score. The weighted averages use the percentage of each category in the total number of samples as weights.

Figure 8 shows the confusion matrix for the four methods. The results indicate that the accuracy of all four methods exceeds 0.6.



(**a**) Random Forest (**b**) DNN (**c**) GCN (**d**) DGCN

Figure 8. Confusion matrix for four types of methods. The intensity of the color indicates the magnitude of the value.

Table 2 shows the predictive performance of each algorithm for academic performance. In comparison, the DGCN shows the highest accuracy of 0.68 for multi-categorization, and performed optimally in the other three metrics. This aligns with expectations, as its increased structural information compared to Random Forests and Neural Networks, as well as its improved ability to use temporal information compared to GCN, provides greater flexibility for fitting the data distribution.

Meanwhile, the prediction performance of each algorithm for academic difficulties (G < 60) can be calculated, as shown in Table 3. It can be seen that for the binary problem of predicting academic difficulties, the Random Forest and DGCN perform relatively well in terms of F1-score, reaching 0.85 and 0.87, respectively, but the Random Forest performs poorly in precision. The DNN and GCN achieve slightly better precision than the DGCN, but they are significantly behind in recall and F1-score indicators. Overall, the DGCN still performs optimally.

Table 2. Performance of each algorithm in predicting academic performance.

Methods	Accuracy	Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
RF	0.66	0.63	0.66	0.60
DNN	0.63	0.58	0.63	0.60
GCN	0.62	0.59	0.62	0.58
DGCN	0.68	0.66	0.68	0.65

Optimal results are highlighted using shadows.

Table 3. Performance of each algorithm in predicting academic difficulties.

Methods	Accuracy	Precision	Recall	F1-Score
RF	0.97	0.79	0.91	0.85
DNN	0.96	0.84	0.76	0.80
GCN	0.96	0.84	0.79	0.81
DGCN	0.97	0.83	0.91	0.87

Optimal results are highlighted using shadows.

4. Discussion

The purpose of this study is to investigate differences in social behaviors, including peer type, quantity, and quality, among college students with varying levels of academic achievement. Additionally, this study has explored whether social behavior predicts academic performance. It was found that individuals' social behavioral traits were effective in predicting academic performance.

4.1. How Does the Frequency of Various Types of Socialization Differ Among Students at Different Levels of Academic Performance?

The results reveal significant differences in the overall frequency of socialization among student groups with different academic performances. Students with higher academic performance socialized more often, which is consistent with findings from several existing studies. This connection may be due to the fact that more interaction leads to more information and support, and thus improved academic performance of students. Some studies have explained peer performance similarity through peer influence and peer choice. Peer influence means that students achieve better academic performance under the influence of peers who perform better academically, and peer choice is that students who perform better academically are more inclined to seek out peers who perform similarly [50]. However, all of these theories have yet to fully explain this relationship, and the association between students' academic performance and socialization is often the result of multiple mechanisms.

Roommate and classmate socialization constitute a major portion of the socialization composition. For example, in April 2020, the average total number of socializations for students in the $80 \leq G < 90$ group during that month is 25.5 times, with 14.3 instances for roommate socialization and 18.5 for classmate socialization. In contrast, opposite sex and other socialization categories show averages of only 2.3 and 3.7 times, respectively, which are significantly lower than the first two categories. This proves the importance of classmate relationships and roommate relationships on campus. Students who perform better academically engage more frequently in classmate and roommate socialization; however, this trend is not evident in opposite sex socialization. Instead, students with higher academic performance socialize less frequently in the other socialization categories, indicating that not all social relationships positively impact academic performance.

Furthermore, it can be observed that the overall number of socializations in the second semester is significantly higher than in the first semester. This aligns with findings from some existing studies suggesting that social networks tend to become more diverse and complex as students become more familiar with their environment and the people around them over time [35].

4.2. Does the Quantity and Quality of Different Types of Peers Correlate with Academic Performance?

Overall, the quantity and quality of peers is positively related to academic performance. This finding suggests that enhancing the quantity and quality of peer relationships can help improve the academic performance of college students, which is consistent with the findings of many studies [12,17,19]. If a student has a positive attitude toward academics, this will lead to better performance in classroom activities, discussions, and practical exercises. For this student's close friends, these types of academic behaviors tend to influence each other, thereby promoting higher academic performance among peers [50]. Additionally, the homogeneity effect makes students more inclined to socialize with friends who perform similarly [51], which explains, to some extent, the influence of peer quantity and quality on academic performance.

It is worth noting that although the quality of all four types of social relationships is positively associated with academic performance, the quality and quantity of classmate and roommate relationships are more highly correlated with academic performance, whereas opposite sex and other categories of relationships are less correlated. This is consistent with the findings of some studies that have noted an indirect positive association between opposite sex peer relationships and academic performance; however, this association is not as strong as that among same sex peer relationships [30]. Notably, this is not entirely consistent with some existing literature, as one study suggests that academic support from romantic partners does not predict GPA [31]. Furthermore, other relationships generally implies involvement in on-campus activities such as clubs, and compared to this type of socialization, interactions with roommates or classmates—who often share the same or similar majors—may facilitate knowledge transfer and academic collaboration.

4.3. Is It Possible to Predict Students' Academic Performance Through Social Behavior?

In the prediction experiments of this study, we used Random Forests and Neural Networks as baseline methods and selected two graph model-based methods: a Graph Neural Network (GCN) and a Dynamic Graph Convolutional Network (DGCN), which are capable of processing temporal information. The overall accuracies of all four methods exceed 62% in the multi-classification task. When focusing on predicting academic difficulties ($G < 60$), these methods achieve accuracies of 79.9%, 84.8%, 84.5%, and 83.7%, respectively. This performance outperforms prediction studies that routinely use Learning Management System (LMS) and Student Information System (SIS) data, which have accuracies ranging from 71% to 76% [52]. Similarly, the prediction accuracy of this study outperforms that of studies that utilize smartphone usage data [53] and internet usage behavior [54]. Although some studies achieve higher accuracy rates, they are more limited in scope. For example, predictions are made only for a single course [55], or only for courses offered on online platforms [36,56], whereas the prediction model in this study has wider applicability in real-world contexts. Overall, student socialization data are effective in predicting academic performance.

The DGCN achieves the best results in multi-classification prediction and demonstrated superior overall performance in predicting academic difficulties. This result aligns with expectations, because the DGCN not only handles the feature information of the nodes, but also incorporates structural information, unlike Random Forests and Neural Networks, as well as temporal information, unlike GCNs. More information leads to better fitting. This not only illustrates the significance of information richness in improving prediction accuracy but also confirms the effectiveness of the graphical modeling paradigm in campus student behavior research.

4.4. Significance of the Research

This study is important at both the theoretical and practical levels.

At the Theoretical Level:

Association Between Peer Relationships and Academic Performance: This study explored the link between peer relationships and academic performance and found different effects of various types of peer relationships on academic performance. This provides new evidence for understanding the impact of peer relationships on academic performance.

Complexity of Peer Relationship Networks: The careful differentiation of students' social types reveals the complexity of the composition of peer relationships, and comparisons and analyses across semesters reveal the complexity of the evolution of peer relationships. This analysis has important theoretical value for a deeper understanding of the mechanisms of formation and evolution of peer relationship networks.

Construction and Application of Graph Networks: This study extracts social relationships among students through spatio-temporal co-occurrence data and constructs a graph network model. This method effectively maps the students on campus and their various behavioral characteristics in the graph network model. It establishes a new research paradigm for behavioral big data in the field of higher education research and provides new research ideas for future smart campus-related studies.

At the Practical Level:

Peer Mentoring: Research has indicated that strong peer relationships tend to help students achieve better academic performance. This finding suggests that schools can implement interventions such as peer mentoring to promote better academic performers to assist students who are struggling academically.

Prediction of Academic Difficulties: It was found that academic difficulties can be identified using a prediction model based on social characteristics. This suggests that the occurrence of academic difficulties can be prevented and mitigated in advance.

Optimization of Educational Resources: Instructional administrators can utilize big data analytics to identify students in need of additional support, optimize the allocation of educational resources, and ensure effective resource usage.

5. Conclusions, Limitations and Future Research

This study proposed a new approach to using social behaviors, reflected in consumption data, as a means to assess the academic performance of college students. The findings indicate significant differences in social behaviors among student groups with varying academic performance, and that the richness and quality of social behaviors have a substantial impact on academic performance. Additionally, social data have proven effective in predicting academic performance. Campus card usage data are easy to collect and provide rich behavioral information. Combined with the generalizability of machine learning techniques, the predictive model developed in this study has significant practical value for educational management, especially in residential universities. We expect that this study will stimulate further research interest in the rational application of campus big data in educational management.

There are several limitations to this study. First, despite the large volume of consumption data, these data still only capture some of the students' social behaviors, while others, such as playing games together, going to the library, and exercising, cannot be captured. Second, although this study compared changes in the number of students' social relationships across semesters, it did not analyze them from a community perspective. Despite these limitations, this study demonstrates the potential of behavioral big data to analyze college student peer relationships and provide insights into instructional management practices.

Future research could incorporate more behavioral characteristics and explore additional directions to provide educators with more comprehensive references for student management. With the gradual improvement of campus informatization, a wider variety of data will become available, such as demographic data, internet usage, movement trajectories, book borrowing records, and reward and punishment histories. All of these can be used to analyze student characteristics and behaviors according to actual situations. In terms of application direction, the research should not be limited to predicting academic performance, but can also explore predictions regarding mental health status, emergencies, and other aspects.

Author Contributions: Conceptualization, J.T.; Data curation, J.T.; Investigation, J.L.; Methodology, J.T.; Project administration, D.L.; Resources, J.L.; Software, J.T.; Supervision, D.L.; Writing—original draft, J.T.; Writing—review and editing, J.T. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: We thank the reviewers for their insightful comments on the restructuring of the article, which greatly optimised the article structure.

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

References

1. Balaji, P.; Alelyani, S.; Qahmash, A.; Mohana, M. Contributions of Machine Learning Models towards Student Academic Performance Prediction: A Systematic Review. *Appl. Sci.* **2021**, *11*, 10007. [CrossRef]
2. Pelima, L.R.; Sukmana, Y.; Rosmansyah, Y. Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review. *IEEE Access* **2024**, *12*, 23451–23465. [CrossRef]
3. Chen, C.; Bian, F.; Zhu, Y. The Relationship between Social Support and Academic Engagement among University Students: The Chain Mediating Effects of Life Satisfaction and Academic Motivation. *BMC Public Health* **2023**, *23*, 2368. [CrossRef] [PubMed]
4. Tinajero, C.; Martínez-López, Z.; Rodríguez, M.S.; Páramo, M.F. Perceived Social Support as a Predictor of Academic Success in Spanish University Students. *An. Psicol.* **2019**, *36*, 134–142. [CrossRef]
5. Wong, M.D.; Strom, D.; Guerrero, L.R.; Chung, P.J.; Lopez, D.; Arellano, K.; Dudovitz, R.N. The Role of Social-Emotional and Social Network Factors in the Relationship between Academic Achievement and Risky Behaviors. *Acad. Pediatr.* **2017**, *17*, 633–641. [CrossRef] [PubMed]
6. Kiss, D. The Impact of Peer Achievement and Peer Heterogeneity on Own Achievement Growth: Evidence from School Transitions. *Econ. Educ. Rev.* **2013**, *37*, 58–65. [CrossRef]
7. Altermatt, E.R. Academic Support from Peers as a Predictor of Academic Self-Efficacy among College Students. *J. Coll. Stud. Retent. Res. Theory Pract.* **2016**, *21*, 21–37. [CrossRef]
8. Becker, B.E.; Luthar, S.S. Social-Emotional Factors Affecting Achievement Outcomes among Disadvantaged Students: Closing the Achievement Gap. *Educ. Psychol.* **2002**, *37*, 197–214. [CrossRef] [PubMed]
9. Gifford-Smith, M.E.; Brownell, C.A. Childhood Peer Relationships: Social Acceptance, Friendships, and Peer Networks. *J. Sch. Psychol.* **2003**, *41*, 235–284. [CrossRef]

10. Gerbino, M.; Zuffianò, A.; Eisenberg, N.; Castellani, V.; Luengo Kanacri, B.P.; Pastorelli, C.; Caprara, G.V. Adolescents' Prosocial Behavior Predicts Good Grades beyond Intelligence and Personality Traits. *J. Personal.* **2017**, *86*, 247–260. [CrossRef] [PubMed]
11. Brady, R.R.; Insler, M.A.; Rahman, A.S. Bad Company: Understanding Negative Peer Effects in College Achievement. *Eur. Econ. Rev.* **2017**, *98*, 144–168. [CrossRef]
12. Kang, C. Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea. *J. Urban Econ.* **2007**, *61*, 458–495. [CrossRef]
13. Golsteyn, B.H.H.; Non, A.; Zoelitz, U. The Impact of Peer Personality on Academic Achievement. *J. Political Econ.* **2020**, *129*, 1052–1099. [CrossRef]
14. Griffith, A.L.; Main, J.B. First Impressions in the Classroom: How Do Class Characteristics Affect Student Grades and Majors? *Econ. Educ. Rev.* **2019**, *69*, 125–137. [CrossRef]
15. Humlum, M.K.; Thorsager, M. The Importance of Peer Quality for Completion of Higher Education. *Econ. Educ. Rev.* **2021**, *83*, 102120. [CrossRef]
16. Ma, L.; Ha, W.; Cao, Y. Peer Effects on Learning Behaviors in Synchronous Online Education: Evidence from a Chinese Research University. *China Econ. Rev.* **2024**, *86*, 102206. [CrossRef]
17. Berthelon, M.; Bettinger, E.; Kruger, D.I.; Montecinos-Pearce, A. The Structure of Peers: The Impact of Peer Networks on Academic Achievement. *Res. High. Educ.* **2019**, *60*, 931–959. [CrossRef]
18. Cao, Y.; Zhou, T.; Gao, J. Heterogeneous Peer Effects of College Roommates on Academic Performance. *Nat. Commun.* **2024**, *15*, 4785. [CrossRef]
19. Gong, J.; Lu, Y.; Song, H. Gender Peer Effects on Students' Academic and Noncognitive Outcomes: Evidence and Mechanisms. *J. Hum. Resour.* **2019**, *56*, 686–710. [CrossRef]
20. Zhang, L.; Pu, S. It Takes Two Shining Lights to Brighten the Room: Peer Effects with Random Roommate Assignments. *Educ. Econ.* **2016**, *25*, 3–21. [CrossRef]
21. Pu, S.; Yan, Y.; Zhang, L. Peers, Study Effort, and Academic Performance in College Education: Evidence from Randomly Assigned Roommates in a Flipped Classroom. *Res. High. Educ.* **2019**, *61*, 248–269. [CrossRef]
22. Fang, G.; Wan, S. Peer Effects among Graduate Students: Evidence from China. *China Econ. Rev.* **2020**, *60*, 101406. [CrossRef]
23. Guo, Y.; Li, S.; Chen, S.; Tang, Y.; Liu, C. Health Benefits of Having More Female Classmates: Quasi-Experimental Evidence from China. *Econ. Educ. Rev.* **2022**, *91*, 102330. [CrossRef]
24. Lao, Y. The More Male Classmates, the Worse: How Male Peers Harm Academic Performance of a Student. *Int. J. Educ. Dev.* **2023**, *103*, 102880. [CrossRef]
25. Hoxby, C. *Peer Effects in the Classroom: Learning from Gender and Race Variation*; National Bureau of Economic Research: Cambridge, MA, USA, 2000. [CrossRef]
26. Hu, F. Do Girl Peers Improve Your Academic Performance? *Econ. Lett.* **2015**, *137*, 54–58. [CrossRef]
27. Adams, J.W.; Snowling, M.J.; Hennessy, S.M.; Kind, P. Problems of Behaviour, Reading and Arithmetic: Assessments of Comorbidity Using the Strengths and Difficulties Questionnaire. *Br. J. Educ. Psychol.* **1999**, *69*, 571–585. [CrossRef] [PubMed]
28. Chen, J.J.-L. Grade-Level Differences. *Sch. Psychol. Int.* **2008**, *29*, 183–198. [CrossRef]
29. Schmidt, J.; Lockwood, B. Love and Other Grades. *J. Coll. Stud. Retent. Res. Theory Pract.* **2015**, *19*, 81–97. [CrossRef]
30. Liem, G.A.D.; Martin, A.J. Peer Relationships and Adolescents' Academic and Non-Academic Outcomes: Same-Sex and Opposite-Sex Peer Effects and the Mediating Role of School Engagement. *Br. J. Educ. Psychol.* **2011**, *81*, 183–206. [CrossRef]
31. Fuhrman, R.W.; Flannagan, D.; Matamoros, M. Behavior Expectations in Cross-Sex Friendships, Same-Sex Friendships, and Romantic Relationships. *Pers. Relatsh.* **2009**, *16*, 575–596. [CrossRef]
32. Poldin, O.; Valeeva, D.; Yudkevich, M. Which Peers Matter: How Social Ties Affect Peer-Group Effects. *Res. High. Educ.* **2015**, *57*, 448–468. [CrossRef]
33. Nettleton, D.F. Data Mining of Social Networks Represented as Graphs. *Comput. Sci. Rev.* **2013**, *7*, 1–34. [CrossRef]
34. Li, Y.; Li, S.; Wei, C.; Liu, J. How Students' Friendship Network Affects Their GPA Ranking. *Inf. Technol. People* **2019**, *33*, 535–553. [CrossRef]
35. Zhou, Y.; Meng, X.; Wang, J.; Mo, X.; Jiang, S.; Dai, C.; Liu, M. Daily Peer Relationships and Academic Achievement among College Students: A Social Network Analysis Based on Behavioral Big Data. *Sustainability* **2023**, *15*, 15762. [CrossRef]
36. Zhu, X.; Ye, Y.; Zhao, L.; Shen, C. MOOC Behavior Analysis and Academic Performance Prediction Based on Entropy. *Sensors* **2021**, *21*, 6629. [CrossRef] [PubMed]
37. Kukkar, A.; Mohana, R.; Sharma, A.; Nayyar, A. Prediction of Student Academic Performance Based on Their Emotional Wellbeing and Interaction on Various E-Learning Platforms. *Educ. Inf. Technol.* **2023**, *28*, 9655–9684. [CrossRef]
38. Zorza, J.P.; Marino, J.; de Lemus, S.; Acosta Mesas, A. Academic Performance and Social Competence of Adolescents: Predictions Based on Effortful Control and Empathy. *Span. J. Psychol.* **2013**, *16*, E87. [CrossRef]
39. Tang, K.H.D. Personality Traits, Teamwork Competencies and Academic Performance among First-Year Engineering Students. *High. Educ. Ski. Work-Based Learn.* **2020**, *11*, 367–385. [CrossRef]

40. Ezziani, M.; Benhayoun, S.; Chtouki, Y. Data Analysis of Student Academic Performance Using Social and Academic Indicators. In Proceedings of the 2022 20th International Conference on Information Technology Based Higher Education and Training (ITHET), Antalya, Turkey, 7–9 November 2022; pp. 1–7. [CrossRef]
41. Guo, Q.; Zhou, J.; Feng, L. Pro-Social Behavior Is Predictive of Academic Success via Peer Acceptance: A Study of Chinese Primary School Children. *Learn. Individ. Differ.* **2018**, *65*, 187–194. [CrossRef]
42. Kishore, K.; Venkatramaphanikumar, S.; Alekhya, S. Prediction of Student Academic Progression: A Case Study on Vignan University. In Proceedings of the 2014 International Conference on Computer Communication and Informatics, Coimbatore, India, 3–5 January 2014. [CrossRef]
43. Vignery, K. Findings and Methodologies about Student Networks, Learning, Performance and Academic Achievement in Higher Education: A Literature Review on Quantitative Studies. *Rev. Educ.* **2021**, *9*, e3308. [CrossRef]
44. Agrawal, R.; Srikant, R. Fast Algorithms for Mining Association Rules in Large Databases. In Proceedings of the 20th International Conference on Very Large Data Bases 1994, VLDB '94, Santiago de Chile, Chile, 12–15 September 1994; pp. 487–499.
45. Chao, R.; Menghui, Y. Information Flow Simulation Community Detection of Weighted-Directed Campus Friendship Network in Continuous Time. *arXiv* **2023**, arXiv:2305.01958. [CrossRef]
46. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
47. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]
48. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:1609.02907. [CrossRef]
49. Phan, A.V.; Nguyen, M.L.; Nguyen, Y.L.H.; Bui, L.T. DGCNN: A Convolutional Neural Network over Large-Scale Labeled Graphs. *Neural Netw.* **2018**, *108*, 533–543. [CrossRef]
50. Leung, R.K.; Toumbourou, J.W.; Hemphill, S.A. The Effect of Peer Influence and Selection Processes on Adolescent Alcohol Use: A Systematic Review of Longitudinal Studies. *Health Psychol. Rev.* **2014**, *8*, 426–457. [CrossRef] [PubMed]
51. Zhang, Y.; Zhang, Q.; Zhang, L.; Ren, P.; Qi, X.; Chang, R. The Covariant Relationship between Adolescent Friendship Networks and Bullying: A Longitudinal Social Network Analysis. *Acta Psychol. Sin.* **2022**, *54*, 1048. [CrossRef]
52. Vignery, K.; Laurier, W. Achievement in Student Peer Networks: A Study of the Selection Process, Peer Effects and Student Centrality. *Int. J. Educ. Res.* **2020**, *99*, 101499. [CrossRef]
53. Kassarnig, V.; Mones, E.; Bjerre-Nielsen, A.; Sapiezynski, P.; Dreyer Lassen, D.; Lehmann, S. Academic Performance and Behavioral Patterns. *EPJ Data Sci.* **2018**, *7*, 10. [CrossRef]
54. Xu, X.; Wang, J.; Peng, H.; Wu, R. Prediction of Academic Performance Associated with Internet Usage Behaviors Using Machine Learning Algorithms. *Comput. Hum. Behav.* **2019**, *98*, 166–173. [CrossRef]
55. Romero, C.; López, M.-I.; Luna, J.-M.; Ventura, S. Predicting Students' Final Performance from Participation in On-Line Discussion Forums. *Comput. Educ.* **2013**, *68*, 458–472. [CrossRef]
56. Xing, W.; Chen, X.; Stein, J.; Marcinkowski, M. Temporal Predication of Dropouts in MOOCs: Reaching the Low Hanging Fruit through Stacking Generalization. *Comput. Hum. Behav.* **2016**, *58*, 119–129. [CrossRef]