

Article

Data Mining of Students' Consumption Behaviour Pattern Based on Self-Attention Graph Neural Network

Fangyao Xu [†] and Shaojie Qu ^{*} 

Beijing Institute of Technology, Beijing 100081, China; 1120180084@bit.edu.cn

^{*} Correspondence: qushaojie@bit.edu.cn

[†] Current address: No. 8, Liangxiang East Road, Fangshan District, Beijing 102488, China.

Abstract: Performance prediction is of significant importance. Previous mining of behaviour data was limited to machine learning models. Corresponding research has not made good use of the information of spatial location changes over time, in addition to discriminative students' behavioural patterns and tendentious behaviour. Thus, we establish students' behaviour networks, combine temporal and spatial information to mine behavioural patterns of academic performance discrimination, and predict student's performance. Firstly, we put forward some principles to build graphs with a topological structure based on consumption data; secondly, we propose an improved self-attention mechanism model; thirdly, we perform classification tasks related to academic performance, and determine discriminative learning and life behaviour sequence patterns. Results showed that the accuracy of the two-category classification reached 84.86% and that of the three-category classification reached 79.43%. In addition, students with good academic performance were observed to study in the classroom or library after dinner and lunch. Apart from returning to the dormitory in the evening, they tended to stay focused in the library and other learning venues during the day. Lastly, different nodes have different contributions to the prediction, thereby providing an approach for feature selection. Our research findings provide a method to grasp students' campus traces.

Keywords: self-attention mechanism; graph neural network; data mining; behaviour sequence pattern; behaviour network



Citation: Xu, F.; Qu, S. Data Mining of Students' Consumption Behaviour Pattern Based on Self-Attention Graph Neural Network. *Appl. Sci.* **2021**, *11*, 10784. <https://doi.org/10.3390/app112210784>

Academic Editor: Federico Divina

Received: 14 October 2021

Accepted: 12 November 2021

Published: 15 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Methods to improve education quality by mining off-line education and on-line learning platform data [1] has led to the development of educational data mining (EDM) [2]. Among the several problems in the field of EDM, predicting students' scholastic performance is a key issue [3,4], and various statistical methods [5,6] and tools [7] have been developed to perform this task. However, these methods could not reflect the learning conditions of students over a specific period, and do not facilitate the discovery of new knowledge patterns from the data set for the development of new and accurate models. Advancements in machine learning has led to the emergence of powerful data visualization methods and a variety of models, such as clustering, classification, and prediction, including algorithms that can dynamically process data streams [8], and other algorithms such as the support vector machine have been used to detect students who may fail in courses as an early warning [9]. Research on machine learning using behavioural data for performance prediction has attracted a significant amount of attention, such as grade prediction using online behaviour [10–12], gaming behaviour [13], consumption behaviour [14] and travel behaviour [15]. To conduct on-line behaviour mining more deeply, artificial neural networks have been used for log data mining and desirable results have been achieved [16]. However, with regard to knowledge tracking, such as the prediction of student's test questions, artificial neural networks have not yielded satisfactory result, and recurrent neural networks (RNNs) have had to be introduced [17]. Further, when dealing with long sequences, RNNs found that gradient explosion and gradient disappearance

were prone to happen [18]. Therefore, to overcome such shortcomings, long short term memory networks (LSTM) were introduced [19]. LSTM can well solve the prediction problem with temporal data, while it has defects in processing spatial data. Further, the same behaviour will lead to ambiguity in the behavioural purpose if the difference in spatial location is ignored. For example, the behavioural meanings of fetching boiled water in the teaching building, in the dormitory, and in the bathroom are obviously different. We can, respectively, speculate that the purposes of corresponding behaviours are to study, to play games in the dormitory, and to take a bath. The lack of spatial location information will affect the prediction accuracy of models and the analysis of discriminative behavioural patterns. If the information in the spatial position was considered to mine behavioural characteristics contained in consumption data, the graph topological structure formed by behaviour cannot be well mined, and the above methods lost the ability to have good performance for such data with certain special structures. Naturally mining the data with graph structure for performance prediction requires the introduction of new tools. At present, graph neural networks (GNNs) have undergone rapid developments. Further, such networks have been extensively used [20–22] and are suitable for dealing with graph structure data. We observed that the behaviour of students can be demonstrated in the form of a graph and particular behavioural pattern and tendency can be observed, which inspired us to use graph-related tools to mine hidden information in behavioural patterns using students' spatial location changes over time.

In this study, we aimed to utilize the information regarding spatial position changes over time reflected by consumption data collected from a school to extract behavioural pattern features and construct graph structures for mining discriminative behavioural characteristics and behavioural trends related to academic performance, as well as for performance prediction. Firstly, we proposed two guidelines for constructing the graph structures by extracting features from the consumption behaviour data. Secondly, we proposed an improved self-attention mechanism model based on previous graph self-attention mechanism; thirdly, we made use of graphs composed of behaviour characteristics to perform classification and determined discriminative learning and life behaviour sequence patterns.

The remainder of this article is as follows: In Section 2, we present some additional current research on consumption behaviour and discuss the possibility of using a GNN for graph mining. In Section 3, we explain the data we used in our study as well as the processing methods; we also describe the method of graph construction and the improved model. In Section 4, we report the results of the experiment and analyse the results obtained. In Section 5, we propose some open issues, summarize the paper and also present some shortcomings of our research.

2. Related Work

Continuously adapting to rapid development and educational innovation is highly crucial and has led to the emergence and application of EDM in various fields. Scholars investigated the three aspects of student performance, teaching equality, and policy making, and found student performance had the greatest significance [23,24]. Related research had also focused on performance prediction and the discussion of methods [25] and models [26], including data [27], and models and methods were considerably improved for different purposes.

From the perspective of students, performance prediction [28,29], early warning of failure in subjects [30], sentiment analysis [31] and course recommendation [32,33] are key issues; besides, the trajectory of school behaviour is rich in information of learning habits of students to mine, and analyses of behavioural patterns based on spatial location change have been widely carried out for performance prediction. The trajectory of students' behaviour at school helps us understand the various characteristics of learning status and lifestyle [34]. Dalvi [35] focused on students' green and low-carbon behaviour, and Islam [36] studied electronic product consumption; in both these studies, only the lifestyles of students were investigated, and the impact of such consumption behaviour on

academic performance was not studied. Mei [37] used campus behaviour data to predict academic performance. However, discriminative learning and life patterns that can help distinguish students with different learning levels was unclear. Further, time and location information has not yet been considered in most studies. Li [14] focused on behaviours that could reflect the regularity of students' lives; however, behavioural patterns are not discriminative enough by nature. In study [38], based on the behaviour records of undergraduates' smart cards, the authors studied the impact of students' diligence and the regularity of their daily life on grades. Due to a lack of spatial information, the authors had to regard different behaviours as the same behaviour and the prediction results were affected. The main reason for this is the lack of consideration about spatial information. Thus, consideration of behavioural patterns with spatial location information is necessary.

Advanced statistical methods were extensively applied in EDM during its early stage, such as the *t*-test [39], for the prediction of academic performance [40,41]. Statistical methods were suitable for small sample data; otherwise, it was necessary to put forward intelligent algorithms. Machine learning algorithms were utilized for predicting academic performance [42] to warn students who might fail in certain courses [43] and predict the graduation rates [44]. Regarding machine learning algorithms that did not perform well on time series data, the improved recurrent neural network (RNN) based on artificial neural networks showed a good mining effect [45], and appeared to be ineffective for dealing with non-European data, such as those with a graph structure. There was also a lack of in-depth mining on student behaviour tendency reflected by the spatial location information. Due to the particularity of graph structure, it was also difficult to use general models to process and analyse. We therefore studied behavioural characteristics from the perspectives of spatial topological structure and time dimension, and we introduced a new and powerful tool for dealing with non-European structure data.

As a powerful tool for processing non-European structure data, graph neural networks have undergone rapid development and been widely applied, such as the knowledge graph [46], natural language processing [47], graph-based text representation [48] and graph embedding techniques [49]. In particular, some scholars have proposed the graph attention mechanism to improve the performance of node classification [50]. Some scholars have recommended graphs for recommendation systems [51], such as the music recommendation system in mobile networks [52,53] because of graphs' powerful information representation abilities and wide applications. Specifically, Zhang [54] used bipartite graph to perform context-sensitive web service discovery. Notably, community detection is also a key task [55,56]. Consumption behaviour at school has structure and characteristics similar to those of social networks and graph. A topology structure must therefore be introduced to distinguish behavioural patterns and find students' behaviour tendency. Inspired by the node classification method, we improved the present self-attention GNN to mine consumption behaviour data from both time and spatial aspects.

3. Methods

3.1. Data Description

As the behaviour of students is often the same regardless of the semester, we only collected behaviour data over a single month. The activities of first-year students are relatively messy due to their curiosity and are difficult to analyse, while the third-year and fourth-year students need to engage in some social work outside the school, which leads to very short time at school and consumption behaviours are lacking and inconvenient to analyse. Therefore, we considered the consumption data of all second-year students at Beijing Institute of Technology, which is characterized by science and engineering majors and has a male to female ratio of 2.2 to 1. Students use campus cards for on-campus consumption, and the corresponding data are transmitted and stored in the school's campus card consumption system later, which is relatively convenient to access. Thus, from December 2020 to January 2021, we collected 752,725 pieces of consumption data generated during May 2020 and 3640 pieces of final exam scores of the course of

data structure from two data systems, including campus card consumption system and educational administration system. The collected consumption data are detailed in Table 1, and a more detailed explanation regarding the “action” is presented in Table 2.

Table 1. Collected raw data.

ID Number	Consumption Money (RMB Cent)	Time	Action
36984	200	2018/5/2 9:18	Breakfast
36984	400	2018/5/2 11:47	Lunch
36984	1500	2018/5/2 16:52	Dinner
17347	300	2018/5/31 11:24	Lunch
17347	250	2018/5/31 11:25	Lunch
17347	38	2018/5/31 11:26	Lunch
17347	150	2018/5/31 11:27	Lunch
10075	180	2018/5/20 11:16	Lunch
10075	200	2018/5/20 11:39	Lunch
10075	1500	2018/5/20 17:50	Dinner
10075	300	2018/5/21 07:56	Breakfast
10075	50	2018/5/21 07:56	Breakfast

Table 2. Specific explanation of different consumption behaviours.

Behaviour	Action Explanation
Dinner	Consumption after 4:00 p.m. in the cafeteria.
Lunch	Consumption between 10:00 a.m. and 4:00 p.m. in the cafeteria
Breakfast	Consumption before 10:00 a.m. in the cafeteria.
Supermarket	Consumption in the supermarket
Library	Consumption in the library
Dormitory bathroom	Consumption in the dormitory bathroom
Dormitory boiled water	Consumption on the dormitory water
Gym	Consumption in the school gym
School bus	Consumption caused by taking the school bus between campuses
Management office	Consumption in the school management office
Classroom boiled water	Consumption of water available in the classroom

3.2. Data Preprocessing

Based on the above data, we performed certain preprocessing measurements, using the following pseudo-code of Algorithm 1.

Algorithm 1 Data pre-processing**Input:** Raw data file**Output:** Processed data files for different students

```

1: Group data by ID
2: while Student ID equals some ID do
3:   if Two adjacent rows of data are exactly the same then
4:     Delete a row
5:   end if
6:   if The consumption behaviour is in the gym, management office or school bus then
7:     Delete the row
8:   end if
9:   if Two adjacent behaviours are the same then
10:    if Time difference is less than 5 minutes then
11:      Delete the second row
12:    end if
13:  end if
14: end while
15: if More than 15 consumption data for some student then
16:   Generate a new data file for the student
17: else
18:   Do not consider the student's behaviour data
19: end if

```

In collected data, “management office” refers to students’ short-term campus card recharging behaviour; “school bus” refers to the behaviour of commuting between campuses, and the uncertainty of destination campus makes it impossible to analyse behavioural purpose; as for “gym”, the overall time and purpose of staying in the gym based on this record cannot be inferred. Hence, we deleted related behaviours in the preprocessing algorithm. After the data preprocessing, we got the data of 3616 students and next considered extracting features from the following three aspects:

1. Indicators of regularity
2. Amount of consumption
3. Behaviour sequence pattern

After obtaining the features, we performed the chi-square test, f-test and other feature selection methods, and finally determined thirty relevant features related to the grades. Table 3 presents some of the information related to the features that were extracted.

We noted that the features related to consumption money were intermediate, which meant that excessive consumption and low consumption were both abnormal phenomena. Thus, this type of feature was transformed to the maximum feature using Formula (1):

$$x_i = 1 - \frac{|x_i - x_{\text{best}}|}{\max\{|x_i - x_{\text{best}}|\}} \quad (1)$$

The same method was applied to the other similar intermediate indicators, as presented in Table 3.

Table 3. Extracted features based on students' data and reality.

Type	Feature Name	Feature Explanation	Intermediate	Node ID
Indicators of regularity	Study-actual	Actual number of visits to library or classroom	No	0
	Paid-days	Number of days with consumption records	No	2
	Getting-up-num	Number of wake-ups	No	3
Amount of consumption	total-meals	Total number of meals	Yes	26
	Total-supermarket-num	Total number of times to go shopping	Yes	7
	Total-month-money	Total monthly cost	Yes	8
Behaviour sequence pattern	Dinner-dormitory-num	Number of times in the dormitory after dinner	No	13
	Lunch-study-num	Number of times to study after lunch	No	15
	Dinner-study-num	Number of times to study after dinner	No	16
	Dinner-study-dorm-bathroom-num	Number of times to study after dinner and go back to the dorm	No	21

3.3. Graph Construction

After the extraction of the behavioural sequence features, the data lost the natural graph structure. As such, construction of suitable graphs from these features became a challenge. To reflect the continuity and trend of students' behaviours, we constructed graphs based on the characteristics of students' behaviours at school and daily experience and regarded the features as nodes.

3.3.1. p-Clique

For a graph (V, E) , V refers to a set of elements called vertices and E is a multiset of unordered pairs (u, v) whose elements are called edges. Two vertices are said to be adjacent if there exists an edge between them. A clique in a graph refers to a set of pairwise adjacent vertices, and if the number of vertices involved is p , it is called a p -clique. The behavioural patterns of students in their daily school life are often consistent. While some patterns may not be related to one another obviously, the extracted behaviour sequence features will be interrelated. This interrelationship between the features is an important indicator that can be used to distinguish between students with different learning levels. Based on this, we arranged part of the extracted behaviour sequence features into a p -clique in the constructed graph.

3.3.2. Other Criterion

Two nodes are said to be connected and interrelated if there exists an edge between them. Hence, for two nodes to have a relationship, we considered that the following two criteria must be fulfilled:

1. Necessary connection: This condition means that the two nodes are interrelated, for example, the edge between the total-month-money and total-lunch-money. The cost of lunch must be a part of the total monthly cost, and there exists a connection between the two features.
2. Unnecessary connection: This condition means that if the values of the two potentially related features are non-zero at the same time, there exists an edge between the two features. For example, if getting-up and breakfast-study are both non-zero at the same time, then there is a connection between the two features, which means that the student may tend to get up early for breakfast.

Based on the above-mentioned criterion, we constructed graphs for each student, creating a total of 3616 graphs for graph-level classification, using 700 graphs as the test set. Below, we present two graphs that were constructed according to the above-mentioned method (Figure 1).

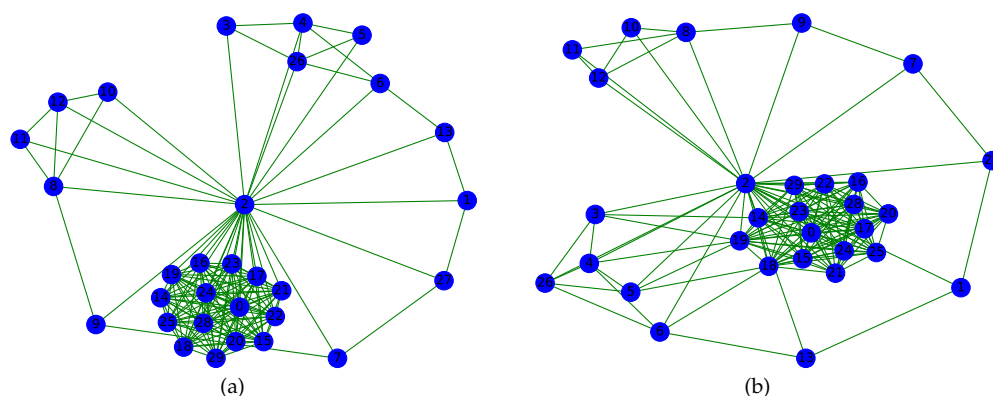


Figure 1. Different graph structures built for different students. The differences resulted from the students having different lifestyles. (a) Graph constructed for student with exam score 84.88 whose ID number equals 17347; (b) graph constructed for student with exam score 91.28 whose ID number equals 61465.

3.4. Model Description

We proposed an improved self-attention GNN based on previous research. The corresponding Algorithm 2 pseudo-code of improved self-attention model is as follows.

Algorithm 2 Graph classification based on improved self-attention GNN

Input: Adjacency matrix of constructed matrix W_{adj} , node feature vector V , graph indicators P and degree matrix D

Output: The prediction labels on the test set

- 1: Initialize: initialize weight matrix W and bias ε
 - 2: Normalize adjacency matrix by $L = D^{-\frac{1}{2}}(W_{adj} + I)D^{-\frac{1}{2}}$, in which D refers to the degree matrix
 - 3: Build a graph convolutional layer to obtain the attention scores by calculating $Z = GCN(L, V, W, \varepsilon)$
 - 4: Perform convolution twice, use activation function ReLu to process the scores, and concatenate the scores S
 - 5: Perform self-attention pooling by function pooling (W_{adj} , S , P) and update the graph structure and adjacency matrix W_{adj} according to the mask
 - 6: Perform maximum pooling and average pooling, and concatenate the results
 - 7: Predict the labels after the three fully connected layers
-

In the standard self-attention mechanism, given a group of nodes (x_1, x_2, \dots, x_k) and weight matrices W_Q, W_K, W_V that represent different linear transformations of features, the attention coefficients are computed to reflect the pair-wise importance of the nodes, as shown in Equation (2):

$$e_{ij} = (W_Q^T x_i)^T (W_K^T x_j), \forall 1 \leq i, j \leq k \quad (2)$$

Then e_{ij} is normalized by all possible values of j using the Softmax function as shown in Equation (3):

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{1 \leq l \leq k} \exp(e_{il})} \quad (3)$$

Finally, a weighted sum of transformed features is calculated as shown in Equation (4):

$$\vec{d}_i = \tanh \left(\sum_{1 \leq j \leq k} \alpha_{ij} W_V^T x_j \right) \quad (4)$$

Additionally, a new node embedding vector set can be obtained when using a multi-head graph attention layer. In graph attention neural networks, the node used for the attention mechanism generally only aggregates the information of the first-order neighbours to update the information. In the improved model, due to the characteristics of the extracted features, and to utilize the neighbouring information of node v_i better, we applied two graph convolution layers. In the convolution layers, we calculated the node embedding using a linear transformer W and used the activation function ReLu to calculate the raw attention score between pair-wise nodes using Equations (5) and (6):

$$Z = GCN(L, V, W, \varepsilon) \quad (5)$$

$$e_{ij} = \text{ReLU} \left(\vec{d}_i^T \left(z_i^{(l)} \| z_j^{(l)} \right) \right) \quad (6)$$

where \vec{a} is a weight vector for learning and \parallel refers to concatenation. Next, we applied the SoftMax function to normalize e_{ij} , and a weighted sum based on attention on the features of all the neighbour nodes, as shown in Equation (7):

$$h_i^{(l+1)} = \text{ReLu} \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} z_j^{(l)} \right) \quad (7)$$

$\mathcal{N}(i)$ refers to all the neighbours of node i . We also used the pooling method to handle redundant information and reduce the amount of calculations. The extracted features denote the occurrence frequency of consumption behaviour sequence patterns and the ReLu function shows the best performance on our dataset.

4. Result and Analysis

We compared our improved self-attention GNN to some other graph neural network variants on two classification methods, including some machine learning models. We will now discuss the difference in the loss reduction of the variant GCNs and the improved self-attention GNN mainly. Due to the uneven distribution of the labels, the results are weighted as demonstrated below.

4.1. Experiment Result

Since the proportion of students with a score of less than 60 was very small, we set a higher passing line. We used the score of 70 to divide the students' scores into two categories defined as $\{1, 2\}$ and conducted a training task. The student group whose scores are all greater than or equal to 70 is labeled 1, and the rest of the group whose scores are less than 70 is labeled 2. To construct a large graph and speed up calculations, we first batched all the training graphs, and then trained the self-attention GNN with 300 epochs, as shown in Figure 2. Compared with the other GNN variants trained using the same number of epochs, the loss of our improved model varied sharply during the training process. Table 4 lists the performance of different models.

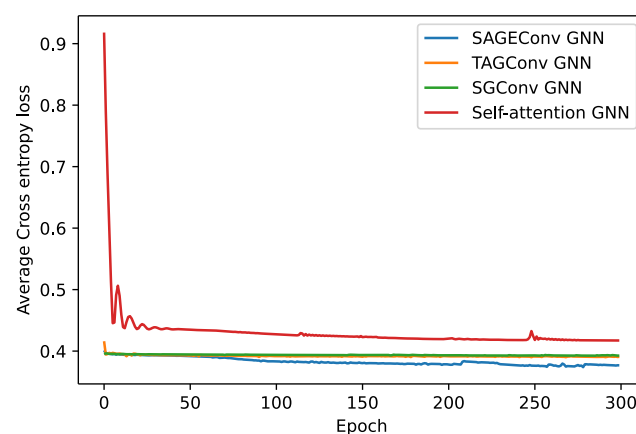


Figure 2. The processes aim at the two-category classification, in which “self-attention GNN” refers to our improved model.

Table 4. The two-category classification results on the test set.

Model	Accuracy	Precision	Recall	F1-Score
Improved self-attention graph neural network	84.86%	94.84%	84.86%	91.81%
Topology adaptive graph convolutional network (TAGConv GNN) [57]	68.57%	62.60%	68.57%	64.88%
GraphSAGE convolutional network (SAGEConv GNN) [58]	65.76%	65.29%	70.29%	67.04%
Simplified graph convolutional network (SGConv GNN) [59]	70.43%	66.19%	70.43%	67.71%
Logistic regression [60]	61.14%	81.39%	61.14%	66.64%
KNN [61]	84.00%	76.57%	84.00%	78.44%
Decision tree [62]	76.00%	77.00%	76.00%	77.00%

As can be seen from Figure 2, the cross-entropy loss reduction of the improved self-attention GNN was very fast, while it could quickly reach stability. While the descent process of the SAGEConv GNN exhibited fluctuations, the cross-entropy loss reduction was much smoother than our improved self-attention model. The remaining two graph neural network variants were both highly stable. This revealed the high sensitivity of our improved GNN model to variations in the data.

It was difficult to analyse the behavioural patterns of students who were outstanding in the two-category classification. Therefore, we further divided the students whose scores were greater than 70 into two categories and conducted a three-category classification task. Then, we used the scores of 70 and 85 to divide the students' scores into three categories defined as $\{1, 2, 3\}$ for multi-category classification. Specifically, students with a score of less than 70 are considered as failing, students with a score of 70 or above and not exceeding 85 are considered as good and students with a score of 85 or above are considered excellent, corresponding to label 1, label 2 and label 3, respectively. We then performed 300 epochs. We observed that the process of loss reduction gradually stabilized, as shown in Figure 3. The performances of different models are listed in Table 5.

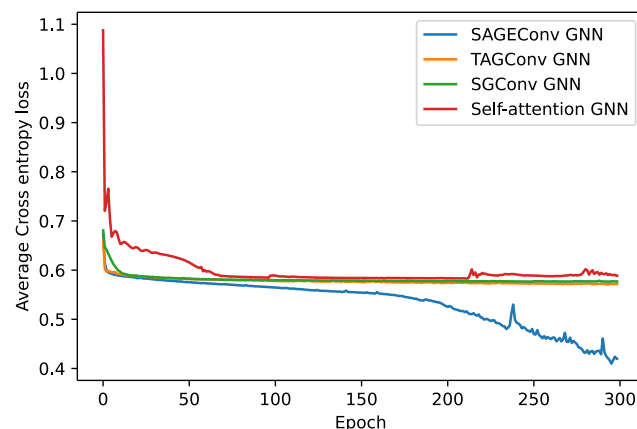
**Figure 3.** The processes aim at the three-category classification, in which “self-attention GNN” still means our improved model.

Table 5. The three-category classification results on the test set.

Model	Accuracy	Precision	Recall	F1-Score
Improved self-attention graph neural network	79.43%	97.20%	79.43%	87.28%
Topology adaptive graph convolutional network (TAGConv GNN) [57]	64.71%	62.44%	64.71%	62.99%
GraphSAGE convolutional network (SAGEConv GNN) [58]	66.06%	65.91%	65.86%	65.43%
Simplified graph convolutional network (SGConv GNN) [59]	67.00%	63.51%	67.00%	64.56%
Logistic regression [60]	48.29%	72.80%	48.29%	53.49%
KNN [61]	79.14%	68.93%	79.14%	68.00%
Decision tree [62]	68.05%	67.16%	68.09%	68.17%

The improved self-attention GNN showed a high degree of data sensitivity from Figure 3; after reaching stability, the cross-entropy loss of our model still showed some fluctuations. After training SAGEConv GNN for 300 epochs, the cross-entropy loss kept on declining and fluctuating from the declining process of the cross entropy loss. Therefore, our improved self-attention GNN was better than SAGEConv GNN in this regard, which implied that the two models may have high data requirements and data sensitivity.

Based on the idea of hypothesis testing, we constructed an indicator based on the discriminative behavioural patterns to reflect the differences in behavioural patterns using Formula (8):

$$\text{ratio} = \frac{\sum_{i=1}^n x_i / n}{\sum_{j=1}^m y_j / m} \quad (8)$$

In the two-category classification, x represents the number of occurrences of some behavioural patterns of outstanding students, while y represents the number of occurrences of corresponding behavioural pattern of lagging students; in the three-category classifications, three ratio indicators were used regarding three score categories, respectively. The numerator and denominator were divided by the corresponding number of people to make sure the number of people did not have an impact. Taking the two-category classification as an example, the related results are presented in Table 6.

4.2. Result Analysis

From the perspective of cross entropy loss decreasing process in Figures 2 and 3, compared with other models, the improved self-attention model had higher cross-entropy loss in the initial stage of training, while with certain fluctuations, it could decrease and converge rapidly in both classification tasks. The initial cross-entropy loss and training processes of the other three graph neural network model variants were similar in two-category classification, which may be related to the fact that these models treated the nodes indiscriminately during each iteration. However, the loss of SAGEConv GNN maintained the downward trend in three-category classification. Judging from the model's prediction accuracy and other performance, we could speculate that although the loss declined, the model might have been over-fitting, which led to poor performance on the test set. In addition, the training process of both our improved model and SAGEConv GNN demonstrated larger fluctuations, which revealed that they might be more sensitive to label distribution.

Table 6. Ratio of occurrences of different discriminative behavioural patterns in the two category classification tasks.

Behavioural Pattern Sequence	Average Times of Outstanding Students	Average Times of Lagging Students	Ratio	Node ID
Study-actual	5.3516	2.2573	2.3708	0
Paid-days	29.5463	27.5556	1.0722	2
Getting-up-num	10.7108	4.9357	2.1701	3
Dinner-dormitory-num	8.3119	6.0819	1.3667	13
Breakfast-study-num	1.2968	0.4269	3.0377	14
Lunch-study-num	1.0605	0.3275	3.2383	15
Dinner-study-num	0.7278	0.2982	2.4402	16
Classroom-lunch-or-dinner-num	1.8204	0.8655	2.1033	18
Breakfast-study-lunch-num	0.7675	0.2690	2.8530	19
Dinner-study-dorm-bathroom-num	0.2287	0.0526	4.4359	21
Total-meals	56.3516	42.7252	1.3189	26

From the perspective of prediction accuracy in Tables 4 and 5, the improved self-attention GNN performed better than any other model in both the classification tasks and achieved accuracy of 84.86% in two-category classification and 79.43% in three-category classification, respectively. Moreover, precision could reach 94.84% and 97.20%, and F1-score was able to achieve 91.81% and 87.28%, respectively. Indicators, such as precision, recall score and F1-score, revealed that improving the self-attention model could improve the prediction performance and converge to stability rapidly. Judging from several indicators, three variants performed similarly but worse than our improved model, in which the accuracy reached 68.57%, 65.76%, 70.43% and 64.71%, 66.06%, 67% in the two tasks, respectively. They focused on improving graph convolution methods. Compared with machine learning algorithms, the overall performance of the graph neural network variants was found to be superior, demonstrating the graph's strong and in-depth ability to mine mutual influences between local nodes. As comparison, the machine learning models equalized the input features, ignoring their potential mutual influence, which resulted in their relatively unsatisfactory performance in some extent. The KNN classification performance was also relatively good, in which accuracy and recall rate were the same, reaching 84% and 79.14% in the two tasks, respectively. However, the KNN could not predict the behavioural trend and identify discriminative behavioural patterns. Its performance was dependent on the training set and the way the distance between data points was defined, which limited its applicability. The decision tree can give judgments through internal nodes and determine classification results based on leaf nodes; however, the depth and the number of leaf nodes of the decision tree for good predictive effect had certain uncertainty and it could not predict students' behavioural tendency similar to the KNN. It was more suitable for ensemble learning to improve the accuracy, while in this research accuracy only achieved 76% and 68.05% in two tasks and bore a resemblance to the remaining three indicators. As for logistic regression, this method could be regarded as generalized linear regression, while a more complex non-linear relationship cannot be expressed, not to mention mining local mutual influence. The above shortcomings contributed to such a situation: Accuracy equaled to recall rate in both tasks by 61.14% and 48.29%, precisions were 81.39% and 72.8% and F1-scores were 66.64% and 53.49%. By paying more attention to the local structure information using the GNN, better local mutual information mining could be performed, except improving the prediction accuracy of our method. The improved self-attention GNN could further identify discriminative behavioural patterns and predict the behavioural trends of students from the node scores and the existence of edges, which provided an insight into the behavioural patterns of outstanding students.

Next we discussed the prediction of students' behavioural trends. Because of the use of the self-attention mechanism, different feature nodes contributed to the performance prediction differently. We utilize different colour shades to represent the scores of the different nodes as shown in Figure 4. The darker the colour, the higher the score, demonstrating the relationship between the different behavioural patterns and their contribution to predicting the academic performance of the student. The first figure (Figure 4a) represents one of the graph trainings resulting from the two-category classification task, followed by three-category classifications. In two different classification tasks, different nodes showed different contributions for grade prediction.

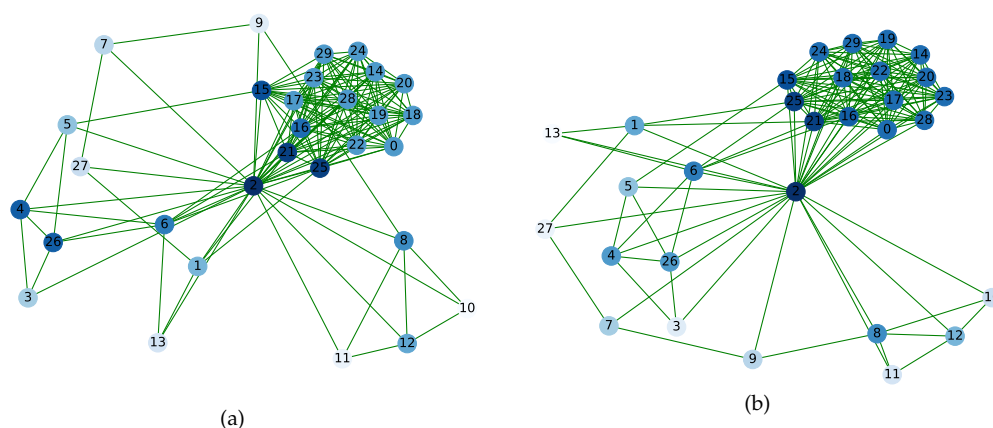


Figure 4. The score of the graph node after improved self-attention GNN training is indicated in each node by colour shade. (a) Two-category classification for students whose ID number is 61465 and exam score is 91.28; (b) Three-category classification for student whose ID number is 61465 and exam score is 91.28.

Among them, the paid-days (node 2) had the highest score, as can be seen from Figure 4. The results showed that daily life habits of the student were relatively consistent and regular. In addition, we noticed that the behavioural pattern of studying after dinner and then returning to the dormitory (node 25), and the number of times to study after lunch (node 15) influenced the prediction greatly. In addition, we noticed that whether the student at least ate lunch and dinner during the day impacted on their academic performance; this behavioural pattern was also a reflection of the regularity in their daily life. Additionally, taking the graph shown in Figure 4a as an example and considering the behavioural tendency, it was most likely that the student would study and then return to the dormitory after finishing dinner according to single node 21. Among the neighbours of node 4 (number of times eating both breakfast and lunch or both lunch and dinner), the influence of node 6 (number of dinners) was relatively large, which showed that if the student tended to enjoy dinner, the student was most likely to eat all three meals. It is therefore possible to create a similar analysis for other students as well.

Taking discriminative behavioural patterns into consideration, in Table 6, except for following behavioural patterns, such as the number of days with consumption records, returning to the dormitory after dinner, and the total number of meals over a month, the ratios of the remaining indicators exceeded 2. On average, the number of occurrences of these behavioural patterns of outstanding students exceeded twice that of lagging students. Further, we speculated that some corresponding behavioural patterns of some top students occurred more frequently. The visualization results were consistent with the above-mentioned results, thereby proving the correctness and rationality of our conclusions. Therefore, we believe that these characteristics could be used as a reference to distinguish students with regard to their learning levels.

5. Open Issues and Conclusions

5.1. Open Issues

In fact, judging from the results, GNN still needs to be improved in EDM. Below we list some open issues for further research:

1. How to use nodes and edges to reflect the relationship between features is still an open question.
2. The learning data is various, such as the data generated during MOOC online course learning, which inspires us to construct graphs to present more kinds of data for information mining.

3. As for node score based on self-attention mechanism, the features represented by different nodes can be ranked by importance. Thus, how to combine traditional feature selection methods and GNN to determine useful features is also worth exploring.
4. Through the comparison of scores, we can judge what the student's next behaviour pattern is most likely to be. However, based on the score values obtained from training, new quantification methods are needed.
5. How to build a behavioural network of all students, and integrate more information including curriculum arrangements, to conduct more diversified analysis, including skipping classes and social interactions, is encouraged.

5.2. Conclusions

In this study, we focused on mining on-campus consumption data to identify discriminative behavioural patterns based on spatial location change, analyse behaviour trends, and predict students' academic performance by constructing behaviour networks. Firstly, we preprocessed collected consumption data to extract features that could reflect students' living habits and their learning status. Secondly, we attempted to construct campus behaviour networks from reality and artificially, including introducing p-clique. Thirdly, combining with the pooling method, an improved self-attention GNN was utilized for training and prediction, and good prediction performance on the test set was achieved. For discriminative behavioural patterns, the habit of getting up early and behavioural patterns of continuous learning until meal time (lunch or dinner) in classroom and learning after three meals were discriminatory for distinguishing students with different learning levels, since the defined ratios of these behavioural pattern features are greater than 2 and we held certain beliefs that these characteristics are discriminatory in the sense of average. These new knowledge discoveries of behavioural patterns were consistent with the visualization results, conformed to the actual situation, and had certain reference meaning. For the behavioural trend analysis, judging from the behavioural pattern represented by single node 21, the student was more likely to continue studying after dinner than to return to the dormitory; from the perspective of node interaction and the existence of edges (e.g., node 4 and its neighbour nodes), three meals of students who often eat dinner were also relatively regular.

However, some limitations need noting regarding the present study. An arguable weakness is that all the graphs we consider are still not large samples. When we consider multi-category classification, the current method may fall into the shortcomings of few training samples for each category. In the meantime, few-shot learning may contribute to solving the above problem. Another weakness is that we discard some information, such as excluding behaviour related to gym and school busses. In addition, application limitation exists in our model and it is hard to apply in other universities.

As for future research directions, we were considering building a bigger behavioural network involving all students at school by their consumption record and some useful and possible video material to carry out analyses for different purposes, such as detecting absenteeism.

Author Contributions: Conceptualization, S.Q. and F.X.; methodology, F.X.; software, F.X.; validation, S.Q.; formal analysis, F.X. and S.Q.; investigation, S.Q.; resources, S.Q.; data curation, F.X.; writing—original draft preparation, F.X.; writing—review and editing, S.Q.; visualization, F.X.; supervision, S.Q.; project administration, S.Q.; funding acquisition, S.Q. 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, including code.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Jugo, I.; Kovai, B.; Slavuj, V. Increasing the adaptivity of an intelligent tutoring system with educational data mining: A system overview. *Int. J. Emerg. Technol. Learn.* **2016**, *11*, 67–70. [CrossRef]
2. Grigorova, K.; Malysheva, E.; Bobrovskiy, S. Application of Data Mining and Process Mining approaches for improving e-Learning Processes. In Proceedings of the 3rd International Conference on Information Technology and Nanotechnology, Samara, Russia, 24–27 April 2017; Volume 1903, pp. 115–121. [CrossRef]
3. Karthikeyan, V.G.; Thangaraj, P.; Karthik, S. Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation. *Soft Comput.* **2020**, *24*, 18477–18487. [CrossRef]
4. Anoopkumar, M.; Md Zubair Rahman, A. A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration. In Proceedings of the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), Ernakulam, India, 16–18 March 2016; pp. 122–133. [CrossRef]
5. Fernandes, E.; Carvalho, R.; Holanda, M.; Van Erven, G. Educational data mining: Discovery standards of academic performance by students in public high schools in the federal district of Brazil. In *World Conference on Information Systems and Technologies*; Springer: Cham, Switzerland, 2017; Volume 569, pp. 287–296. [CrossRef]
6. Nuankaew, W.; Nuankaew, P.; Teeraputon, D.; Phanniphong, K.; Bussaman, S. Perception and attitude toward self-regulated learning of Thailand’s students in educational data mining perspective. *Int. J. Emerg. Technol. Learn.* **2019**, *14*, 34–49. [CrossRef]
7. Sabourin, J.; McQuiggan, S.; de Waal, A. SAS Tools for educational data mining. In Proceedings of the EDM 2016, Raleigh, NC, USA, 29 June–2 July 2016; pp. 632–633.
8. Xu, S.; Wang, J. Dynamic extreme learning machine for data stream classification. *Neurocomputing* **2017**, *238*, 433–449. [CrossRef]
9. Costa, E.B.; Fonseca, B.; Santana, M.A.; de Araujo, F.F.; Rego, J. Evaluating the effectiveness of educational data mining techniques for early prediction of students academic failure in introductory programming courses. *Comput. Hum. Behav.* **2017**, *73*, 247–256. [CrossRef]
10. Ducange, P.; Pecori, R.; Sarti, L.; Vecchio, M. Educational big data mining: How to enhance virtual learning environments. In *International Conference on EUropean Transnational Education*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 527, pp. 681–690. [CrossRef]
11. Chen, J.; Zhao, J. An educational data mining model for supervision of network learning process. *Int. J. Emerg. Technol. Learn.* **2018**, *13*, 67–77. [CrossRef]
12. de J. Costa, J.; Bernardini, F.; Artigas, D.; Viterbo, J. Mining direct acyclic graphs to find frequent substructures—An experimental analysis on educational data. *Inf. Sci.* **2019**, *482*, 266–278. Available online: <https://www.sciencedirect.com/science/article/pii/S0020025519300398> (accessed on 11 January 2019). [CrossRef]
13. Malkiewicz, L.; Baker, R.S.; Shute, V.; Kai, S.; Paquette, L. Classifying behaviour to elucidate elegant problem solving in an educational game. In Proceedings of the Ninth International Conference on Educational Data Mining, Raleigh, NC, USA, 29 June–2 July 2016; pp. 448–453.
14. Li, Y.; Li, D. University students’ behaviour characteristics analysis and prediction method based on combined data mining model. In Proceedings of the 2020 3rd International Conference on Computers in Management and Business, Tokyo, Japan, 31 January–2 February 2020; pp. 9–13. [CrossRef]
15. Zheng, L.; Xia, D.; Zhao, X.; Tan, L.; Li, H.; Chen, L.; Liu, W. Spatial-temporal travel pattern mining using massive taxi trajectory data. *Phys. A Stat. Mech. Its Appl.* **2018**, *501*, 24–41. [CrossRef]
16. Altaf, S.; Soomro, W.; Rawi, M.I.M. Student Performance Prediction using Multi-Layers Artificial Neural Networks: A case study on educational data mining. In Proceedings of the 2019 3rd International Conference on Information System and Data Mining, Houston, TX, USA, 6–8 April 2019; pp. 59–64. [CrossRef]
17. Nakagawa, H.; Iwasawa, Y.; Matsuo, Y. End-to-end deep knowledge tracing by learning binary question-embedding. In Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 17–20 November 2018; pp. 334–342. [CrossRef]
18. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013; Number PART 3, pp. 2347–2355.
19. Tseng, C.W.; Chou, J.J.; Tsai, Y.C. Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members. *IEEE Access* **2018**, *6*, 72870–72879. [CrossRef]
20. Morsy, S.; Karypis, G. A study on curriculum planning and its relationship with graduation GPA and time to degree. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Tempe, AZ, USA, 4–8 March 2019; pp. 26–35. [CrossRef]
21. Hu, Q.; Polyzou, A.; Karypis, G.; Rangwala, H. Enriching course-Specific regression models with content features for grade prediction. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 19–21 October 2017; Volume 2018, pp. 504–513. [CrossRef]
22. Yang, Y.; Liu, H.; Carbonell, J.; Ma, W. Concept graph learning from educational data. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015; pp. 159–168. [CrossRef]

23. Aldowah, H.; Al-Samarraie, H.; Fauzy, W.M. Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telemat. Inform.* **2019**, *37*, 13–49. [\[CrossRef\]](#)
24. Jones, K.M.; Rubel, A.; LeClere, E. A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. *J. Assoc. Inf. Sci. Technol.* **2020**, *71*, 1227–1241. [\[CrossRef\]](#)
25. Amrieh, E.A.; Hamtini, T.; Aljarah, I. Mining educational data to predict student's academic performance using ensemble methods. *Int. J. Database Theory Appl.* **2016**, *9*, 119–136. [\[CrossRef\]](#)
26. Bhagavan, K.S.; Thangakumar, J.; Subramanian, D.V. Predictive analysis of student academic performance and employability chances using HLVQ algorithm. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 3789–3797. [\[CrossRef\]](#)
27. Gao, H.; Qi, G.; Ji, Q. Schema induction from incomplete semantic data. *Intell. Data Anal.* **2018**, *22*, 1337–1353. [\[CrossRef\]](#)
28. Wang, X.; Yu, X.; Guo, L.; Liu, F.; Xu, L. Student performance prediction with short-term sequential campus behaviours. *Information* **2020**, *11*, 201. [\[CrossRef\]](#)
29. Wu, Z.; He, T.; Mao, C.; Huang, C. Exam paper generation based on performance prediction of student group. *Inf. Sci.* **2020**, *532*, 72–90. Available online: <https://www.sciencedirect.com/science/article/pii/S0020025520303716> (accessed on 4 May 2020). [\[CrossRef\]](#)
30. Sun, Y.; Chai, R. An early-warning model for online learners based on user portrait. *Ing. Des Syst. D'Inf.* **2020**, *25*, 535–541. [\[CrossRef\]](#)
31. Onan, A. Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Comput. Appl. Eng. Educ.* **2021**, *29*, 572–589. [\[CrossRef\]](#)
32. Zhang, H.; Huang, T.; Lv, Z.; Liu, S.; Zhou, Z. MCRS: A course recommendation system for MOOCs. *Multimed. Tools Appl.* **2018**, *77*, 7051–7069. [\[CrossRef\]](#)
33. Kardan, A.A.; Ebrahimi, M. A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups. *Inf. Sci.* **2013**, *219*, 93–110. Available online: <https://www.sciencedirect.com/science/article/pii/S0020025512004756> (accessed on 24 July 2012). [\[CrossRef\]](#)
34. Xie, T.; Zheng, Q.; Zhang, W. Mining temporal characteristics of behaviours from interval events in e-learning. *Inf. Sci.* **2018**, *447*, 169–185. Available online: <https://www.sciencedirect.com/science/article/pii/S0020025518301993> (accessed on 1 June 2018). [\[CrossRef\]](#)
35. Dalvi-Esfahani, M.; Alaedini, Z.; Nilashi, M.; Samad, S.; Asadi, S.; Mohammadi, M. Students green information technology behaviour: Beliefs and personality traits. *J. Clean. Prod.* **2020**, *257*, 120406. [\[CrossRef\]](#)
36. Islam, M.T.; Dias, P.; Huda, N. Young consumers e-waste awareness, consumption, disposal, and recycling behaviour: A case study of university students in Sydney, Australia. *J. Clean. Prod.* **2021**, *282*, 124490. [\[CrossRef\]](#)
37. Mei, G.; Hou, Y.; Zhang, T.; Xu, W. Behaviour Represents Achievement: Academic Performance Analytics of Engineering Students via Campus Data. In *2020 Chinese Automation Congress (CAC)*; IEEE: Piscataway, NJ, USA, 2020; pp. 4348–4353. [\[CrossRef\]](#)
38. 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*, 20180210. [\[CrossRef\]](#)
39. Vijayalakshmi, M.; Salimath, S.; Shettar, A.S.; Bhadri, G. A study of team formation strategies and their impact on individual student learning using educational data mining (EDM). In *Proceedings of the 2018 IEEE Tenth International Conference on Technology for Education (T4E)*, Chennai, India, 10–13 December 2018; pp. 182–185. [\[CrossRef\]](#)
40. Hao, J.; Liu, L.; von Davier, A.A.; Kyllonen, P.; Kitchen, C. *Collaborative Problem Solving Skills versus Collaboration Outcomes: Findings from Statistical Analysis and Data Mining*; International Educational Data Mining Society: Raleigh, NC, USA, 2016; pp. 382–387.
41. Gowri, G.; Thulasiram, R.; Baburao, M.A. Educational Data Mining Application for Estimating Students Performance in Weka Environment. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2017; Volume 263. [\[CrossRef\]](#)
42. Jovanovic, M.; Vukicevic, M.; Milovanovic, M.; Minovic, M. Using data mining on student behaviour and cognitive style data for improving e-learning systems: A case study. *Int. J. Comput. Intell. Syst.* **2012**, *5*, 597–610. [\[CrossRef\]](#)
43. Vilorio, A.; Garcia Guliany, J.; Niebles Nuz, W.; Hernandez Palma, H.; Niebles Nuz, L. Data Mining Applied in School Dropout Prediction. *J. Phys. Conf. Ser.* **2020**, *1432*. [\[CrossRef\]](#)
44. Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Appl. Intell.* **2020**, *50*, 4506–4528. [\[CrossRef\]](#)
45. Matayoshi, J.; Cosyn, E.; Uzun, H. Are We There Yet? Evaluating the Effectiveness of a Recurrent Neural Network-Based Stopping Algorithm for an Adaptive Assessment. *Int. J. Artif. Intell. Educ.* **2021**, *31*, 304–336. [\[CrossRef\]](#)
46. Issa, S.; Adekunle, O.; Hamdi, F.; Cherfi, S.S.S.; Dumontier, M.; Zaveri, A. Knowledge Graph Completeness: A Systematic Literature Review. *IEEE Access* **2021**, *9*, 31322–31339. [\[CrossRef\]](#)
47. Vashishth, S.; Yadati, N.; Talukdar, P. Graph-based deep learning in natural language processing. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, Hyderabad, India, 5–7 January 2020; pp. 371–372. [\[CrossRef\]](#)
48. Osman, A.H.; Barukub, O.M. Graph-Based Text Representation and Matching: A Review of the State of the Art and Future Challenges. *IEEE Access* **2020**, *8*, 87562–87583. [\[CrossRef\]](#)
49. Chen, Y.; Wu, Y.; Ma, S.; King, I. A Literature Review of Recent Graph Embedding Techniques for Biomedical Data. In *International Conference on Neural Information Processing 2020*; Springer: Cham, Switzerland, 2020; Volume 1333, pp. 21–29. [\[CrossRef\]](#)
50. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.

51. Kherad, M.; Bidgoly, A.J. Recommendation system using a deep learning and graph analysis approach. *arXiv* **2020**, arXiv:2004.08100.
52. Wang, R.; Ma, X.; Jiang, C.; Ye, Y.; Zhang, Y. Heterogeneous information network-based music recommendation system in mobile networks. *Comput. Commun.* **2020**, *150*, 429–437. [[CrossRef](#)]
53. Durand, G.; Belacel, N.; LaPlante, F. Graph theory based model for learning path recommendation. *Inf. Sci.* **2013**, *251*, 10–21. Available online: <https://www.sciencedirect.com/science/article/pii/S0020025513003149> (accessed on 30 April 2013). [[CrossRef](#)]
54. Zhang, R.; Zettsu, K.; Kidawara, Y.; Kiyoki, Y.; Zhou, A. Context-sensitive Web service discovery over the bipartite graph model. *Front. Comput. Sci.* **2013**, *7*, 875–893. [[CrossRef](#)]
55. Zhao, X.; Liang, J.; Wang, J. A community detection algorithm based on graph compression for large-scale social networks. *Inf. Sci.* **2021**, *551*, 358–372. [[CrossRef](#)]
56. Chen, J.; Li, R.; Zhao, S.; Zhang, Y.P. A New Clustering Cover Algorithm Based on Graph Representation for Community Detection. *Tien Tzu Hsueh Pao/Acta Electron. Sin.* **2020**, *48*, 1680–1687. [[CrossRef](#)]
57. Du, J.; Zhang, S.; Wu, G.; Moura, J.M.; Kar, S. Topology adaptive graph convolutional networks. *arXiv* **2017**, arXiv:1710.10370.
58. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive representation learning on large graphs. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1025–1035.
59. Wu, F.; Souza, A.; Zhang, T.; Fifty, C.; Yu, T.; Weinberger, K. Simplifying graph convolutional networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 6861–6871.
60. Cuji Chacha, B.R.; Gavilanes Lopez, W.L.; Vicente Guerrero, V.X.; Villacis Villacis, W.G. Student Dropout Model Based on Logistic Regression. In *International Conference on Applied Technologies 2020*; Springer: Cham, Switzerland, 2020; Volume 1194, pp. 321–333. [[CrossRef](#)]
61. Dervisevic, O.; Zunic, E.; Donko, D.; Buza, E. Application of KNN and Decision Tree Classification Algorithms in the Prediction of Education Success from the Edu720 Platform. In Proceedings of the 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 18–21 June 2019. [[CrossRef](#)]
62. Mkwazu, H.R.; Yan, C. Grade Prediction Method for University Course Selection Based on Decision Tree. In Proceedings of the 2020 International Conference on Aviation Safety and Information Technology, Weihai, China, 14–16 October 2020; pp. 593–599. [[CrossRef](#)]