Social-path embedding-based transformer for graduation development prediction

Guangze Yang¹ · Yong Ouyang¹ · Zhiwei Ye¹ · Rong Gao^{1,2} · Yawen Zeng³

Accepted: 18 January 2022 / Published online: 3 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

As the education of students attracts more and more attention, the task of graduation development prediction has gradually become a hot topic in academia and industry. The task of graduation development prediction aims to predict the employment category of students in advance via academic achievement data, which can help administrators understand students' learning status and set up a reasonable learning plan. However, existing research ignores the potential impact of social relationships on students' graduation development choices. To fully explore social relationships among students, we propose a Social-path Embedding-based Transformer Neural Network (SPE-TNN) for the task of graduation development prediction in this paper. Specifically, SPE-TNN is divided into the Social-path selection layer, the Social-path embedding layer, the Transformer layer, and the Multi-layer projection layer. Firstly, the Social-path selection layer is designed to find social relationships that impact graduation development and embed them into the student's performance features through the Social-path embedding layer. Secondly, the Transformer layer is adopted to balance the weights of the students' features. Finally, the Multi-layer projection layer is used to achieve the student graduation development prediction. Experimental results on the real-world datasets show that SPE-TNN outperforms the existing popular approaches.

Keywords Social relationship · Graduation development · GCN · Transformer

1 Introduction

Since the beginning of the 21st century, with the continuous enrollment expansion of colleges and universities, the employment rate of graduates has gradually attracted attention. Recently, due to the spread of the COVID-19, college graduates have to face more severe challenges than ever before [1]. Choosing the right direction for university students after graduation, whether to be directly employed or further their studies, etc., is related to their future career orientation and is critical to achieving sustainable

development in higher education [2, 3]. Only when students have a stable and clear goal, their career orientation can be formed gradually during their university career and future career exploration. Therefore, career guidance is an indispensable part of the university, which can help students to plan their life goals after graduation better [4].

Therefore, the task of providing advance guidance for graduates' career choice and development after graduation appears to be crucial [5]. Although many colleges and universities have started to provide career guidance for graduates, these career guidance efforts by colleges and universities for students are often formal and do not benefit students [6]. The job matching is based on the psychological assessment questionnaire or subjective willingness to graduate throughout the existing universities. However, this matching idea is still the traditional personality-job fit theory, which only collects students' subjective judgment data, which is immediate information but ignores the years of learning and social data accumulated by students in the university stage. For this reason, by predicting students' graduation development in advance through these data of students, colleges and universities can provide more targeted employment guidance to students, thus better

✉ Yong Ouyang
oyywuhan@163.com

Guangze Yang
101910729@hbut.edu.cn

¹ School of Computer Science, Hubei University of Technology, Wuhan, China

² State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

³ School of Computer Science, Hunan University, Hunan, China

guiding them to choose the development direction that suits them during their school years, realizing personalization and differentiation of employment and solving the problem of difficult employment for graduates [7, 8].

Many efforts have been made in predicting students' graduation development, mainly using traditional machine algorithms and deep learning methods. Traditional machine learning algorithms mainly include the Naive Bayes model [9], C4.5 Algorithm [10], etc. However, with the rise of educational data mining [11] and increased student data processing, traditional algorithms show disadvantages such as weak generalization and insufficient computational abilities. Benefiting from the merit of deep learning [12–14], neural networks such as Back Propagation Neural Network (BP) [15], Generative Adversarial Networks (GAN) [16], Long Short-Term Memory (LSTM) [16], Graph Convolutional Network (GCN) [8] and Attention mechanism [8] have also been applied to the prediction of graduation development. However, the above-mentioned research work only predicts students' graduation development based on their grades [8–10, 16], credits [8, 9] or their regular performances [10, 15], ignoring the potential impact of multi-dimensional social relations of college students on graduation development prediction research.

Some scholars have pointed out in the in-depth study of human behaviour data, including quantitative analysis of college students' campus behaviour is helpful to explain some complex social phenomena [17]. Humans are social creatures, and all products and activities are completed through interactions between people. In particular, this kind of interaction with different purposes tends to form various social circles. People in different social circles exchange different information, emotions, or opinions through different expressions such as language and behaviour and have different influences. To better understand the impact of social relations on the person, many studies [18–20] have analyzed social relations. Lin et al. [20] mentioned that everyone has different features in virtual social networks. The relations owned by each person are rich and multi-dimensional, and each social relationship has different influences on individuals. Generally, social relations have the following facts:

1. The social relationships that everyone has are so rich and complex that they are difficult to define and understand accurately.
2. Different social relationships are not equal, and there is a focus that has different impacts on people.

This provides new ideas for predicting the graduation development of students. As the carriers of the small-scale society, college students also have various social relationships to form different social circles among them. Different social circles have a particular impact on students'

studies, interpersonal relationships, and emotions [21]. There has been much work on the analysis of students based on social relations. For example, Kim et al. [22] analyzed the potential relationship between different types of social relations and students' burnout. They found that there is a negative correlation between social relations and students' burnout. Jr [23] studied the influence of social relations on students' grades by analyzing students' social circles and found that some social relations promoted students' grades, while others caused negative effects. Similarly, for seniors facing graduation, their graduation development decisions are usually influenced by other students in the multi-dimensional social circle [21, 24]. Therefore, their graduation decisions are highly dependent on their social relations, so mining students' social relations is crucial for the prediction of students' graduation development.

To explore the conditions for the generation of student social relationships, many scholars have explored the social relations of college students based on real-world datasets. For example, Eagle et al. [25] tracked students' mobile phones to mine the social network of students and first proposed in 2009 that clustering is an essential condition for students to have social relationships in real life. The so-called gathering refers to the exchange of information, emotions or opinions between people in a certain place for some reason in a certain place through expressions such as language and behaviour [26]. Crandall et al. [27] further showed that when different students have more clusterings in a period, the possibility of building social relationships between them also increases exponentially. Furthermore, Xu et al. [28] found that when two persons gather in the same place for a long time due to some coercive factors, they will have a greater chance of having a social relationship.

Therefore, we use the following two strategies to explore the social relationships between students fully.

The first one is to find out students' social relations influenced by the compulsory factors of school. Because the university campus is closed to a certain extent, students live in the dormitory for a long time due to the school's requirements, and students in the same class often participate in activities or attend classes together. In this way, students are forced to gather in the same class or the same dormitory for a long time, which leads to an increase in their contact opportunities and makes it easier to form a social circle among students.

The second is to find the social relationship formed by students' independent choices because of the same individuality. In this regard, we have mainly observed three points: First, the elective courses offered by the school, as prescribed courses arranged by students themselves, can cultivate students' individual interests and personal abilities. When different students choose multiple identical elective courses together, the same interests and individuality lead

them to form social relationships. Second, self-learning is an indispensable part of university life. Suppose some students may have the same learning purposes, such as applying for studying abroad, preparing for the postgraduate entrance examination or a civil service examination. In that case, they often go in and out of the library to study together, leading to a stable social relationship between them. Third, the college canteen provides the students' daily food and beverage service. Students would often dine in droves, and the probability of having social relationships between students often eat together also relatively high.

Based on the categorization described above, this paper regards relationships formed through students' compulsory clustering in the same space for a long time by the designated classes and dormitories as commonality clustering social relationship. On the other hand, the social relationships created by students who choose multiple same elective courses, often study together in the library, and often eat together in the cafeteria due to similar individualities is regarded as the individuality clustering social relationship. In addition, to better find the factors that affect students' graduation development, this paper also takes the students' scores in compulsory courses in the first two years of university.

However, the current research still has the following two challenges:

1. Each student has multi-dimensional social relationships. However, each social relationship impacts graduation development differently, and not every social relationship impacts it. Therefore, how to find social relationships that influence each student's graduation development relationship is important.
2. After searching for the social relationships that affect the graduation development of each student, it is critical to retain the unique information in the various social relationships and embed it in the student individuals since there are still multiple social relationships with different attributes.

In order to address the challenges mentioned above, we propose the Social-path Embedding-based Transformer Neural Network (SPE-TNN) predict students' graduation development, as shown in Fig. 1. Specifically, SPE-TNN is divided into the Social-path selection layer, the Social-path embedding layer, the Transformer layer, and the Multi-layer projection layer. Firstly, the Social-path selection layer is designed to find social relationships that impact graduation development and embed them into the student's performance features through the Social-path embedding layer. Secondly, the Transformer layer is adopted to balance the weights of the students' features. Finally, the Multi-layer projection layer is used to achieve the student graduation development prediction. SPE-TNN enables administrators

and educators to provide students with more accurate career guidance work, help students choose a more appropriate graduation development direction during their school years and prevent students from becoming unemployed after graduation.

The main contributions of this work are summarized as follow:

- To the best of our knowledge, this is the first work that introduces commonality clustering and individuality clustering to explore the social relationships among students.
- We contribute an SPE-TNN framework, which is designed to capture social relationships between students. This model finds and embeds the social path that impacts the graduation development of students and uses the Transformer to balance feature weights.
- Extensive experiments on the public dataset are performed to validate the effectiveness of our proposed method.

2 Related work

2.1 Graduation development prediction

The task of predicting the development of students' graduation development predicts the different development directions of graduates after graduation based on their previous academic achievement or daily performance, helping education administrators to provide more targeted guidance to graduates, helping students to understand their learning status in advance and formulate reasonable learning goals.

In recent years, many studies have been based on traditional machine learning methods to predict students' graduation development. For example, Hartatik et al. [9] utilized the Naive Bayes algorithm to analyze students' academic performance, thereby predicting students' graduation development tasks. In addition, Putri et al. [10] adopted the C4.5 decision tree method to improve the accuracy of graduation development predictions concerning student performance and usual performance. However, due to the complexity and diversity of student features, some machine learning algorithms have weak generalization capabilities and insufficient feature extraction capabilities, which leads to unsatisfactory performance in the prediction of graduation development when the amount of student data increases.

A'rifian et al. [29] used decision trees, logistic regression, and artificial neural networks to predict graduation development based on graduate employment datasets, among which artificial neural networks have the best effect. Compared with traditional machine learning methods, deep learning

Fig. 1 Graduation development prediction. The model predicts the development of graduates based on students' different social relationships and academic performance. A student's different possible graduation development destinations are influenced by commonality

has more robust feature extraction capabilities and higher learning efficiency in many fields [12]. Therefore, in recent years, researchers have focused more on applying deep learning to the prediction of student graduation development: Chen et al. [15] adopted Backpropagation Neural Network (BP) to analyze students' academic achievement to predict student employment. Guo et al. [16] used Generative Adversarial Networks (GAN) and Long Short-Term Memory Networks (LSTM) to process student academic achievement data to complete the task of predicting student employment. Ouyang et al. [8] utilized students' compulsory and elective course data and corresponding course credits to predict the graduation development of students based on Graph Convolutional Networks (GCN) and Attention and achieved better results.

2.2 Graph convolutional network

Graph Convolutional Network (GCN) [30] is a neural network based on the graph structure, which can be used to process non-Euclidean space data [31] such as social networks, knowledge graphs, and chemical molecular structures. Khayi and Rus [32] used GCN to process students' test answers to improve the accuracy of the

clustering social relationship, individuality clustering social relationship, and compulsory course scores. The red arrow pointed out by SPE-TNN is the most likely destination of the student predicted by the model

evaluation of test answers. In the online learning evaluation model of Karimi et al. [33], Graph Convolutional Network (GCN) was applied to extract the features of courses and students and embed them to achieve better results. However, the disadvantage of GCN is that it assumes that the network structure of the dataset graph has been given in advance and does not distinguish between different types of nodes and edges. In this way, the absence of connections or false connections has a significant impact on the results. Because university students are in different colleges and living environments, the social circles generated by their different social relationships also have different effects on their graduation development. Not every social relationship impacts their graduation development, so we need to find the Social-Path that has a real impact on students' graduation development. Yun et al. [34] proposed that Graph Transformer Networks uses Graph Transformer Layer (GTLayer) to find Meta-Path in heterogeneous graphs, which uses an end-to-end approach to learn network representations, and automatically generates the Meta-Path of graphs through soft selection.

In addition, because GCN is not suitable for dealing with multi-relational networks, it can only embed graph nodes and ignore the expression of relations. For this reason,

Vashishth et al. [35] proposed a learning model called CompGCN based on a Graph Convolutional Network. CompGCN is proposed to solve two problems. One is that GCN modelling training for multiple relationship graphs causes parameter overload due to too many relationship types; the other is that GCN only embeds the graph's nodes and ignores the expression of relationship vectors in the multiple relationship graphs. Therefore, to retain the unique information of the captured Social-Paths for graph embedding, we apply CompGCN as the Social-path embedding layer of the model. CompGCN can learn the representation of nodes and edges simultaneously, introduce edges in three directions: out, in, and self-loop, and use three learning matrices to update the vector of nodes.

2.3 Transformer

In recent years, Attention mechanism [36] has been successful in many fields of deep learning, which is inspired by the fact that human beings use limited visual systems to screen high-value information when observing objects. The essence of Attention is to assign weights to different elements, highlight essential features, and improve network performance. For example, Zeng et al. [37] utilized High-Level Attention to measure the importance of different courses and optimize academic abnormal prediction tasks. In addition, Ouyang et al. [8] adopted a multi-layer Attention network to balance the weights of different student academic achievement to improve the accuracy of graduation development prediction.

The Transformer [38] is a model that uses Attention to accelerate training. The traditional CNN and RNN are abandoned in the Transformer. The entire network structure is entirely composed of the Attention mechanism. More accurately, the Transformer only consists of self-Attention and Feed Forward Neural Network. The Transformer adopts the method of Attention and adds extra Position Embedding to model the dependency relationship between sequences. Such a network structure can be trained in parallel to improve training speed, and feature extraction capabilities are more robust, and thus it has better performance.

2.4 Our work

Inspired by the above research, this paper proposed a student graduation prediction model based on students' social relationships and grades in required courses. To better address students' employment problems, we have done the following work:

- Tapping into social relationships among students. This article brings social relations into the vision of educational researchers. As one of the essential

behaviours in a student's career, social relations influence the student's study, life, and graduation direction. Therefore, in addition to student graduation projections, we can also connect social relationships when conducting other analytical work with students. Such as the potential impact of a student's social relationships on the student's academic performance can be considered when conducting academic alerts tasks, and whether the inclusion of social relationships connections will improve the accuracy of student performance prediction tasks, etc. Integrating students' social relationships into the study of student education can better help administrators and educators understand students' school status, formulate and plan better growth goals and long-term plans for students, and effectively use school information resources to improve student management in universities.

- A more accurate graduation prediction model is proposed. When students reach their junior year, administrators and educators input students' required course data and social relationships in the first two years into the learned SPE-TNN model to obtain students' most likely graduation development destinations. Specifically, the SPE-TNN model incorporates students' social relationships into the grade features through CompGCN and uses the transformer to highlight the features that significantly impact students' graduation development. The network output has higher accuracy in predicting students' graduation direction than previous work. Based on the obtained students' development direction, administrators and educators can provide students with more targeted and personalized employment guidance, effectively alleviating the problem of difficult employment for college students. The graduation prediction model provides a vital decision basis for school administrators to take precise services or precise interventions and serves as an essential path to realize the precise employment services advocated by the Ministry of Education.

3 Preliminaries

3.1 Social-path selection layer and social-path embedding layer

Social-path selection layer: To select social relationships that impact students' graduation development, Social-path selection layer is inspired by GTLayer in the paper [34] and uses the 1*1 convolution kernel to learn Social-path in social networks based on given data, as shown in Fig. 2.

Fig. 2 Social-path selection layer

The GTLayer first utilizes the following formula to select two Q_1 and Q_2 from the adjacency matrix $\mathbb{A}$ of different edge relationship types, where the calculation formula of Q is:

$$Q = F(\mathbb{A}; W_\phi) = \phi(\mathbb{A}; \text{softmax}(W_\phi)) \quad (1)$$

ϕ is the convolutional layer and W is the convolutional layer parameter. Q_l can be expressed as $\sum_{t_l \in T^e} \alpha_{t_l}^{(l)} A_{t_l}$, T^e is the edge type set, $\alpha_{t_l}^{(l)}$ is the weight of the l -th type t_l of the edge in the l -th layer.

Therefore, the adjacency matrix A_P composed of l adjacency matrices for the Social-Path of length l can be expressed as:

$$A_P = (\sum_{t_1 \in T^e} \alpha_{t_1}^{(1)} A_{t_1}) (\sum_{t_2 \in T^e} \alpha_{t_2}^{(2)} A_{t_2}) \dots (\sum_{t_l \in T^e} \alpha_{t_l}^{(l)} A_{t_l}) \quad (2)$$

Social-path embedding layer: We follow the CompGCN strategy in [35] to introduce the Social-path embedding layer, which embeds Social-path into the student's academic achievement.

CompGCN is a graph neural network framework that can consider various relationship information to learn the representation of nodes and relationships simultaneously. Definition $G = \{V, R, E, X, Z\}$ represents the multi-relation graph, where V represents a set of nodes, R represents a set of relationships, E represents a set of edges, where edge (u, v, r) represents a relationship $r \in R$ between vertices u and v , $(u, v, r) \in E$, $X \in R^{|V| \times d_0}$ represents the d_0 dimensional input feature of each node, and $Z \in R^{|R| \times d_0}$ represents the initial relationship feature. CompGCN defines h_v as the updated representation of the

node v , then the update formula of h_v is:

$$h_v = f\left(\sum_{(u,r \in Num(v))} W_{\lambda(r)} \eta(x_u, z_r)\right) \quad (3)$$

x_u and z_r represent the initial features of the node u and the relationship r respectively, $u, r \in Num(v)$ is the set of neighbours of node u under relation r , and $W_{\lambda(r)} \in R^{d_1 \times d_0}$ is a specific parameter of the relationship type, and $R^d \times R^d \rightarrow R^d$ is a fusion operator.

3.2 Transformer

To balance the weight of student features embedded with Social-path, we apply the Transformer [38] model to process student features. The Transformer is composed of Num identical Encoder layers. Each layer comprises two sub-layers, namely Multi-head Attention Mechanism and Fully Connected Feed-forward Network, each of which adds Residual connection and Normalization: $LayerNorm(X + sublayer(X))$. X is the input feature. The Transformer structure is shown in Fig. 3.

Multi-head attention mechanism: The multi-head attention mechanism comprises multiple parallel Self-attentions, which can better capture relevant information through multiple calculation iterations. The main advantage of self-Attention is that it ignores the distance between different features and directly calculates its Attention weight to learn the internal structure of features. There are three different vectors in Self-attention, namely Query, Key, and Value. Multi-head attention is calculated as follows:

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h) W^0 \quad (4)$$

Q , K and V are obtained by the linear transformation of the input vector. $head_i$ is the output generated by multiple Attention Head and multiplied by a network weight W_0 to obtain the final output of Multi-Head Attention.

Fully connected feed-forward network: The linear transformation output by the Multi-head Attention Mechanism is obtained through the fully connected network, as follows:

$$Z' = \max(0, ZW_1 + b_1)W_2 + b_2 \quad (5)$$

In (5), W_1 , W_2 , b_1 and b_2 are the network weight and the bias value of Feed-forward network, Z is the output of Multi-head attention and Z' is the final output of Feed-forward network.

The Transformer continuously passes the input features through the Multi-head self-attention mechanism and Fully connected feed-forward network to obtain the final Attention-added feature expression.

4 The proposed model

To realize the prediction of students' graduation development based on social relationships and capture the influence of students' social relationships on graduation development, we construct the relationship of the graph from two aspects. Specifically, the social relationship formed by the students

who compulsorily gather in the same space for a long time according to the designated classes and dormitories is regarded as the commonality clustering social relationship. On the other hand, the social relationship formed by students who choose multiple same elective courses due to similar interests, often study together in the library and often eat together in the cafeteria is regarded as the individuality clustering social relationship.

As shown in Fig. 4, the SPE-TNN model aims to use social relationships and student's achievement to predict graduation development. It mainly contains 1) **Input layer**; 2) **Social-path selection layer**. Automatically find the appropriate Social-path; 3) **Social-path embedding layer**. Embed the Social-path into the student's academic achievement; 4) **Transformer layer**. Calculate the weight of each student's different features, balance the feature factors that affect graduation development; 5) **Multi-layer projection layers**. Used for feature fusion and prediction.

4.1 Problem formulation

We select the social relationship and student's course academic achievement as the student features of the student's graduation development prediction task.

Define the course grades of a college student S_1 as $S_1 = [T_1, T_2, \dots, T_j]$, where T represents the course grades of different courses, totalling j courses. The score matrix of all students can be expressed as $S = [S_1, S_2, \dots, S_N]^T$, and N is the total number of students.

Fig. 3 The architecture of the transformer encoder

Fig. 4 SPE-TNN Model

Five types of data are constructed with the rules of the commonality clustering social relationship and the individuality clustering social relationship: the dormitory dataset C_0 is composed of student dormitory information; the class dataset C_1 is composed of student class information; the elective course dataset I_0 is composed of elective course data; the library dataset I_1 is composed of the library visit record data in the campus card; the catering dataset I_2 is composed of the catering consumption record data in the campus card. Construct the social relationship adjacency matrix $A = [A_0, A_1, A_2, A_3, A_4]$ according to the six datasets, as shown in Fig. 5.

In addition, according to the division plan of Ouyang et al. [8], the graduation development direction of each student is defined as six categories: Students who choose to continue their studies after graduation; students who choose to study abroad after graduation; students who choose to contract with the company for normal employment; students who choose to be admitted to civil servants; Students who choose to start a business after graduation; and students who still have not found a job after graduation. Therefore, we simplify these graduation development directions as postgraduate, abroad, employed, civil servant, pioneer and unemployed, which are recorded as $target = \{a, b, c, d, e, f\}$.

Input : The student achievement matrix S and the social relationship adjacency matrix A .

Output : Future development after graduation predicts $Score = \arg\max(s|S, A)$. Where s is the probability score of each type of employment, and the highest probability is the post-graduation direction that the student may choose.

4.2 Social-path selection layer

Since the social relationship of each student is rich and complex, to better predict the development direction of

students' graduation through social relationships, we are inspired by [34] and apply the Social-path selection layer to find the Social-path in the social relationship network graph. The core idea of the Social-path selection layer is to apply the multi-channel 1×1 convolutional layer to assign weights to the social relationship edges. The social relationship edge with the high weight value is used as the Social-Path. The specific steps are shown in Fig. 6.

The social relationship adjacency matrix set A transferred from the input layer is divided into $A^C = [A_1, A_2]$ and $A^I = [A_3, A_4, A_5]$ according to the commonality and individuality, and send them into the Social-path selection layer to find the Social-path corresponding to the commonality clustering social relationship and the individuality clustering social relationship by using (6) and (7), and get the corresponding adjacency matrix. We set the number of channels of the 1×1 convolutional layer to 3, as shown below:

$$A_P^C = (\sum_{t_1^C \in T^C} \alpha_{t_1^C}^{(1)} A_{t_1^C}^C) (\sum_{t_2^C \in T^C} \alpha_{t_2^C}^{(2)} A_{t_2^C}^C) (\sum_{t_3^C \in T^C} \alpha_{t_3^C}^{(3)} A_{t_3^C}^C) \quad (6)$$

$$A_P^I = (\sum_{t_1^I \in T^I} \alpha_{t_1^I}^{(1)} A_{t_1^I}^I) (\sum_{t_2^I \in T^I} \alpha_{t_2^I}^{(2)} A_{t_2^I}^I) (\sum_{t_3^I \in T^I} \alpha_{t_3^I}^{(3)} A_{t_3^I}^I) \quad (7)$$

T^C is a set of edges of the commonality clustering social relationship type, T^I is a set of edges of the individuality clustering social relationship type, $\alpha_{t_i^C}^{(i)}$ and $\alpha_{t_i^I}^{(i)}$ represent the weights of the t_i^C and t_i^I edge types in the i -layer Social-path selection layer respectively. A_C^P and A_I^P are the adjacency matrix corresponding to the Social-path of the commonality clustering social relationship and the individuality clustering social relationship, respectively.

Then concat A_C^P and A_I^P according to (8) to obtain the final output $A_{Re}^P = \{A_{Re1}^P, A_{Re2}^P, \dots, A_{Re2ch}^P\}$ of Social-path

Fig. 5 Construct social relation adjacency matrix set

selection layer, where ch is the number of channels of 1×1 convolutional layer.

$$A_{Re}^P = \text{Concat}(A_P^C, A_P^I) \quad (8)$$

4.3 Social-path embedding layer

After obtaining the Social-path of the student's social relationship in the Social-path selection layer, this study uses CompGCN [35] to embed the Social-path into the student's academic achievement. CompGCN can simultaneously perform representation learning on Social-path and student nodes, and better learn the expression of students' social relationship vectors.

Suppose the adjacency matrix $A_{Re}^P = \{A_{Re1}^P, A_{Re2}^P, \dots, A_{Re2ch}^P\}$ corresponds to Social-path and the student's academic achievement feature $S = [S_1, S_2, \dots, S_n]^T$ constitutes the graph $G_S = \{V_S, R_S, E_S, X_S, Z_S\}$, V_S represents a group of nodes of a student, R_S represents a collection of student relations, and E_S represents a group of edges, where (u_s, v_s, r_s) represents the relationship $r_s \in R_S$ between student vertex u_s and vertex v_s . $X_S \in R^{|V_S| \times j}$

represents the j -dimensional input feature of each student node and $Z_S \in R^{|R_S| \times j}$ represents the student's initial relationship feature. At the same time, the following (9) is used to expand the student's E_S and R_S reversely:

$$E'_S = E_S \cup \{(v_s, u_s, r_s) | (u_s, v_s, r_s) \in E_S\} \cup \{(u_s, u_s, T_S) | u_s \in V_S\} \quad (9)$$

And $R'_S = R_S \cup R_{S_{inv}} \cup \{T_S\}$, Where $R_{S_{inv}} = \{r_s^{-1} | r_s \in R_S\}$ means reverse relationship, T_S means self-circulation.

Define the update expression of the student node v_s of the first of Social-path embedding layer as h_{v_s} :

$$h_{v_s} = f\left(\sum_{(u_s, r_s) \in \text{Num}(v_s)} W_{\lambda(r_s)} \eta(x_{u_s}, z_{r_s})\right) \quad (10)$$

x_{u_s} and z_{r_s} represent the initial features of the student node u_s and the social relationship r_s respectively, $u_s, r_s \in \text{Num}(v_s)$ is the set of neighbours of node v_s under relation r_s , and $W_{\lambda(r_s)} \in R^{d_1 \times j}$ is a specific parameter of the relationship type, and $\eta : R^d \times R^d \rightarrow R^d$ is a fusion operator. CompGCN does not define embeddings for each relationship separately, but expresses it as a linear combination of a set of basis vectors. Expressing

Fig. 6 Social-path selection layer

$\{v_1, v_2, \dots, v_{B_s}\}$ as a set of learnable basis vectors, the initial relationship is expressed as:

$$z_{r_s} = \sum_{b=1}^{B_s} \mu_{br_s} v_{sb} \quad (11)$$

$\mu_{br_s} \in R_S$ represents the learnable weight coefficient of the relation on the basis vector v_{sb} . In addition, different weights are assigned to edges in different directions, that is, $\lambda(r_s) = \text{dir}(r_s)$, which is expressed as:

$$W_{\text{dir}(r_s)} = \begin{cases} W_O, & r_s \in R_S \\ W_I, & r_s \in R_{S_{inv}} \\ W_S, & r_s = T_s(\text{self-loop}) \end{cases} \quad (12)$$

After the node embedding is updated, the relationship embedding is also transformed as follows:

$$h_{r_s} = W_{rel} z_{r_s} \quad (13)$$

The $W_{rel} \in R^{d_1 \times j}$ in the formula is a learnable transformation matrix, which projects each relationship into the same space as the node and makes it used in the next Social-path embedding layer.

Social-path embedding layer sets a total of k -layer CompGCN, then $H_{v_s}^{k+1}$ is the representation of the student node v_s after the k -th layer, $H_{r_s}^{k+1}$ represents the representation of the relationship in which r_s follows the k -layer, and then use the following (14) and (15) to update $H_{v_s}^{k+1}$ and $H_{r_s}^{k+1}$ respectively:

$$H_{v_s}^{k+1} = f\left(\sum_{(u_s, r_s) \in \text{Num}(v_s)} W_{\lambda(r_s)}^k \phi(H_{u_s}^k, H_{r_s}^k)\right) \quad (14)$$

$$H_{r_s}^{k+1} = W_{rel}^k h_{r_s}^k \quad (15)$$

Let the final student node set of CompGCN be denoted as $S' = [S'_1, S'_2, \dots, S'_N]$, the output result of the Social-path embedding layer.

4.4 Transformer layer

This paper uses the Transformer [38] to calculate the weight of each student's different features and balances the feature factors that affect the student's graduation development. The steps are shown in Fig. 7.

Fig. 7 Transformer layer

The input of the Transformer layer is $S' = [S'_1, S'_2, \dots, S'_N]$ and N represents the number of students. By multiplying three different weight matrices W_S^Q , W_S^K and W_S^V to get Q_S , K_S and V_S as the Q , K and V of the Multi-head Attention mechanism, and performs the dot product of (16) and scales it by $\sqrt{d_k}$ and normalized by *softmax* function. Then execute scaled dot-product Attention multiple times to get multiple $head_S$, connect them in series, and multiply them by a network weight W_S^0 , to get the final output Z_{S_0} of Multi-Head Attention, as the following (16)–(18):

$$Attention(Q_S, K_S, V_S) = softmax(\frac{Q_S K_S^T}{\sqrt{d_k}}) V_S \quad (16)$$

$$head_{S_i} = Attention(Q_S W_{S_i}^Q, K_S W_{S_i}^K, V_S W_{S_i}^V) \quad (17)$$

$$\begin{aligned} Z_{S_0} &= MultiHead(Q_S, K_S, V_S) \\ &= Concat(head_1, head_2, \dots, head_h) W_S^0 \end{aligned} \quad (18)$$

$W_{S_i}^Q$, $W_{S_i}^K$, $W_{S_i}^V$ are the corresponding network weights in scaled dot-product Attention and $i \in h$ is the number of executions. Then pass the result into the Fully connected feed-forward network and use the following (19) for linear transformation:

$$Z'_{S_0} = max(0, Z_{S_0} W_{S_1} + b_{S_1}) W_{S_2} + b_{S_2} \quad (19)$$

W_{S_1} , W_{S_2} are the network weight of the feed-forward network, and b_{S_1} , b_{S_2} are network biases. Z'_{S_0} is the output result of the first encoder, and it is passed to the encoder of the following transformer network.

Through the transformer network structure composed of multiple encoders, the final output Z'_S is obtained. Let $S'' = Z'_S$. Finally, the student feature expression S'' with added Attention is obtained.

4.5 The multi-layer projection layer

The function of The multi-layer projection layer is to perform feature fusion and prediction. For each student's $S''_i \in S''$ feature, such as (20):

$$\alpha(S'') = tanh(\sum_{k=1}^K w_k S''_k + \theta) \quad (20)$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (21)$$

$$s_i = \alpha_3(\alpha_2(\alpha_1(S''_i))) \quad (22)$$

$tanh(x)$ is the hyperbolic tangent function as the activation function. θ is the network bias value, and w is the network weight.

Then the s_i obtained through The multi-layer projection layer is normalized using the *softmax* function of (23), and the obtained *Score* is the probability value of the six categories corresponding to *target*. The category corresponding to the maximum probability is the prediction result.

$$Score_i = softmax(s_i) = \frac{exp(s_i)}{\sum_{i \in N} exp(s_i)} \quad (23)$$

Use the cross-entropy loss function *Loss* to adjust the network parameters, and use *L2* regularization, as shown in (24), and use the Adam optimization algorithm to optimize the network model.

$$\begin{aligned} Loss &= -\frac{1}{N} \sum (target_i \ln Score_i \\ &+ (1 - target_i) \ln(1 - Score_i)) + L2 \end{aligned} \quad (24)$$

5 Experiment

5.1 Dataset

The data set used in this article includes data from a total of 84314 students in 22 departments from 2014 to 2017. Each student's data contains information about the student while in school and whereabouts after graduation. The data are all statistics and input into the school system according to the real situation of students. This paper uses the augmented version of student dataset released in the paper [8] and [37] to test the model. The dataset contains all the information about the students, including dormitory classes, required courses and elective courses, campus card usage records, etc. Filter the valuable information in the dataset for this paper and save it to the student dataset U . U

includes student achievement dataset P , dormitory dataset B , campus card dataset Y , graduation development dataset Q , and $U = \{P, B, Y, Q\}$.

The student academic achievement dataset P includes CUR-NAME, TASK-NO, CURTYPE, CURDEP, CUR-CREDITH, GRADE, STUDEP, STUCLASS, STUID, STU-NAME, STUSEX represent course name, course number, course type, course affiliated college, credits, grade, student's college, class, student ID, student name, gender. The student attributes mainly used in this article includes course number, course type, student ID, class, college, and grade. The number distribution is shown in Table 1.

The student dormitory dataset B includes building number, dormitory number, Charges, student name, student ID, class, grade, Major, student's college, Student category, School system. The student dormitory dataset attributes mainly used in this article includes building number, dormitory number, student ID, and college. The number distribution is shown in Table 2.

The student campus card dataset Y includes the student account ID, student account, swipe time, total cost, card balance, amount of consumption, consumption type, number of consumption, trading terminal, access to records, student name, student ID, class. The student campus card attributes dataset mainly used in this article includes student ID, consumption time, consumption location, consumption type, and library visit time. The number distribution is shown in Table 3.

The student's graduation development dataset Q includes college, student number, name, graduation employment destination, company nature, company industry, employment status, unemployed category, specific company, location, gender, education, professional, school system, nationality, political appearance, birthday, remarks. The student's graduation development dataset attributes mainly used in this article includes college number, student ID, and graduation development direction. The number distribution is shown in Table 4.

In the graduation development dataset Q , the six types of students' graduation development direction $target = \{a, b, c, d, e, f\}$: postgraduate, abroad, employed, civil servant, pioneer, and unemployed, distributed as shown in Table 5.

5.2 Data processing

In order to obtain an input dataset suitable for the SPE-TNN model, it is requisite to perform basic cleaning of the dataset firstly. The process is mainly divided into five steps:

- The first is to clean the missing data: determine the range of missing values, calculate the missing value ratio for each field, and delete student data with a high missing ratio and missing important fields, such as missing student numbers, classes, grades, etc.
- The second is to clean data with wrong format and content: for example, data with inconsistent display formats such as time, date, numeric value, and half-width, there is data with characters that should not exist in the student field, and the content of an attribute in the student field does not match the expected content of the attribute.
- The third is to clean up logically incorrect data: remove duplicate student data caused by repeated entries; remove unreasonable data, such as data with a score less than 0 or greater than 100, and data whose access time is not during library opening hours.
- The fourth is to clean the attributes that are not needed: Since the attributes of the data set used in this article are less than the actual amount of attributes of the data set, the attributes that are not needed are cleaned. For details, refer Tables 1 to 4.
- Finally, there is relevance verification: after completing the data cleaning, because this article uses multiple data sets, it is necessary to verify the relevance of the different data of each student in the student achievement dataset, dormitory dataset, campus card dataset, graduation development dataset, to ensure that each student has all the attribute fields that need to be used.

After the data cleaning is completed, in the student dormitory dataset B , the dormitory dataset C_0 is constructed from the student number, building number, and dormitory number. In the student achievement dataset P , the class dataset C_1 is constructed by the student number and class. Select the course type as an elective course and combine it with the student number to construct the

Table 1 Academic achievement dataset

	Couse Number	Course-Type	Student-ID	Class	College	Grade
Type	String	String	String	String	String	Int
2014	18505	88	20781	172	25	102
2015	19834	91	20818	198	27	102
2016	23221	119	20870	188	14	102
2017	17809	93	21845	176	19	103

Table 2 Dormitory dataset

	Building Number	Dormitory Number	Student-ID	College
Type	String	String	String	String
2014	45	5280	20638	25
2015	45	5361	20722	27
2016	45	5276	20820	14
2017	45	5563	21751	17

elective course dataset I_0 . Select courses whose course types are compulsory courses and combine them with student numbers to construct S . In the student's campus card dataset Y , the library dataset I_1 is constructed by the student ID and the time of visiting the library; the catering data I_2 is constructed by the student ID, consumption time, consumption location, and consumption type.

In the student graduation development dataset Q , the student graduation development label $target = \{a, b, c, d, e, f\}$ is constructed from the student number and graduation development.

Then construct the social relationship adjacency matrix set $A = [A_0, A_1, A_2, A_3, A_4]$ according to the six datasets C_0, C_1, I_0, I_1, I_2 , as shown in Fig. 5 in Section 4.

5.3 Experiment setting

We used 70% of the dataset as the training set, 10% of the dataset as the validation set, and the remaining 20% of the data S_{test} as the test set to test results. The measurement indicators of this experiment use the general accuracy rate and recall rate. Acc represents the probability of correct prediction, as in (25), R represents the prediction accuracy rate of a specific target, as in (26), T_u represents correct prediction, and belongs to category $target = \{a, b, c, d, e, f\}$.

$$Acc = \frac{T_u(S_{test})}{S_{test}} \quad (25)$$

$$R(l) = \frac{T_u(S_{test} \in target_l)}{S_{test} \in target_l} \quad (26)$$

To verify the validity of the SGME-TNN model, the model is compared with the following mainstream advanced model models:

- **KNN** [39]. This model discovers the degree of difference in student's academic achievement in high-dimensional space mainly through clustering. We set the values in sequence in the test and select the best k value $k = 4$.
- **kNC** [40]. Compared with KNN, KNC provides a weighted solution method for unbalanced samples, which are simpler and more efficient. During the test, the best value is the same as in the KNN model, and the best value is 4.
- **GBDT** [41]. This model is a representative ensemble learning method, which finds the best classification point by constructing multiple attribute trees. When constructing, first use the training data to get the best GBDT model and test it on the test set..
- **EBP** [42]. This model is essentially an enhanced BP neural network used to extract students' academic achievement features and continuously enhance the attributes based on the labelling results. In the experiment, to obtain the network with the optimal number of layers, the number of network layers is set to [2,4,8,16,32]. Similarly, the hidden layer space is set to 64, and the embedded size is the same as SPE-TNN.
- **APAMT** [43]. This model is a variant of Long-Short Term Memory (LSTM) and a soft-attention mechanism, which can learn the student profile-aware representation from heterogeneous behavior sequences. In the experiment, the training data is used to train the model, and then test it on the test set.

Table 3 Campus card dataset

	College	Student-ID	Consumption time	Consumption place	Consumption type	Library visit time
Type	String	String	String	String	String	String
2014	25	20158	18536849	18536849	2	933589
2015	27	19975	15329412	15329412	2	892156
2016	14	20193	13579248	13579248	2	1154913
2017	17	20987	12851793	12851793	2	1327159

Table 4 Graduation development dataset

	College	Student-ID	Graduation development
Type	String	String	String
2014	25	20428	20428
2015	27	20355	20355
2016	14	20629	20629
2017	19	21423	21423

- **NCF** [44]. This model expands the matrix factorization by deep learning by mining potential factors. Use this framework directly for training and testing.
- **EESM** [45]. The model uses a continuous neural network method to predict students' graduation development. In the experiment, directly enter the softmax function to normalize and output the prediction result.
- **GAT** [46]. This model combines a graph convolutional neural network with a self-Attention mechanism. In the experiment, students are used as nodes, and social relationships are used as edges to be passed into the GAT model and output the prediction results.
- **HLVQ** [47]. This model is a Hybrid LVQ (learning vector quantization) neural network algorithm. It combines the AdaBoost optimization technique with the LVQ algorithm to improve the accuracy of predicting student employment. In the experiment, use this framework directly for training and testing.

5.4 Experiment analysis

5.4.1 Overall performance comparison

To verify the effectiveness and the accuracy of the SPE-TNN model, the following groups of standard other mainstream advanced models are designed for performance comparison. The results of the model comparison analysis experiment are shown in Table 6:

We use *Acc* to complete the overall evaluation of the model and use *R* to determine the prediction accuracy of a specific target. The overall performance of the SPE-TNN

model is superior to other models, with an accuracy rate of 87.07%.

Among other models, the prediction accuracy of the KNN and KNC clustering models is low, only 57.32% and 59.38%; The accuracy of the GBDT model is also low due to unobvious data features, only 65.90%; Although EBP improves the abstraction ability of the model, the 71.74% accuracy rate is not ideal because the ordinary neural network cannot handle the complex score feature data well. APAMT is a variant of Long-Short Term Memory and a soft-attention mechanism, but the 75.06% accuracy rate is not ideal; The NCF model and the EESM model take into account the hidden factors of student's academic achievement, and their effects are better, 77.12% and 78.83% respectively; The GAT model combines the GCN and the Attention mechanism and takes into account the two factors of the student's social relationship and the student's academic achievement, with an accuracy rate of 80.43%. As an improved clustering algorithm, HLVQ assists clustering by labeling student samples. Its effect is better than other models except for the model in this paper, with an accuracy rate of 82.72%.

In addition, the prediction accuracy rate of the specific target of the SGE-SANN model is also higher than that of other models. Among them, the accuracy rates of predicting the two types of targets a (for further education) and d (politics) are 83.64% and 85.51%, which are slightly lower than the overall accuracy; the accuracy of predicting the b (going abroad) target is the same as the overall accuracy, 87.10%; the prediction results of category c (employment), category e (entrepreneurship) and category f (unemployment) are relatively good, with 88.65%, 90.54%, and 89.74% respectively.

The SGME-TNN model uses the Social-path selection layer to find the Social-path that impacts the student's graduation development and uses CompGCN to characterize the student node and social relationship edge simultaneously. The social relationship is embedded in the student's academic achievement, and the transformer model is used to balance the weights of students' features. For the reasons mentioned above, the prediction results obtained by the SGE-SANN model are significantly higher than other models, which verifies the superiority of the model in this paper.

Table 5 Distribution of development target per year

	a	b	c	d	e	f
Type	String	String	String	String	String	String
2014	667	177	2888	782	350	318
2015	628	224	2715	731	323	293
2016	722	160	3163	635	245	298
2017	921	170	2977	641	123	217

Table 6 Overall performance comparison among various methods

Method	ACC	R(a)	R(b)	R(c)	R(d)	R(e)	R(f)
KNN	57.32	51.83	54.76	58.44	57.48	53.25	72.73
KNC	59.38	54.27	59.52	59.90	59.84	51.95	76.36
GBDT	65.90	66.46	64.29	63.33	64.57	68.83	83.71
EBP	71.74	68.90	71.43	72.13	71.65	70.13	78.18
APAMT	75.06	76.54	74.52	74.57	76.77	72.73	76.72
NCF	77.12	75.61	76.19	76.28	76.38	79.22	87.27
EESM	78.83	77.41	80.95	79.22	75.59	80.52	83.64
GAT	80.43	79.44	80.65	81.27	77.54	82.43	84.21
HLVQ	82.72	81.48	78.57	84.35	83.46	77.92	85.45
SPE-TNN	87.07	83.64	87.10	88.65	85.51	90.54	89.74

5.4.2 SPE-TNN parameter tuning

The 1*1 convolution kernel is the core part of the Social-path selection layer. The number of channels will affect the model's capture of Social-path to a certain extent. To obtain the best training parameters, we tested the effect of the convolution kernel on the model's accuracy under different channel numbers. In this experiment, the number of channels will be set $O \in \{1, 2, 3, 4, 5\}$ in sequence, and the model will be trained to convergence. The experimental results are shown in Fig. 8.

Different colours in the figure represent the number O of different 1*1 convolution kernel channels. The multi-dimensional adjacency matrix $A_S \in R^{N \times N \times m}$ composed of students' social relationships is transformed into a multi-dimensional matrix $Q_S \in R^{N \times N \times O}$ through the convolution kernel, where N represents the total number of students and m is the dimension of the social relationship adjacency matrix. It can be seen from Fig. 8 that as the number of channels increases, the convergence speed is

Fig. 8 Parameter tuning

slowly decreasing. When the number of channels reaches 3, the accuracy is the highest, 87.11%, and when it continues increasing, the accuracy decreases slightly. Experiments show that when the number of channels is 3, the model can best learn the Social-path in social relationships.

5.4.3 Ablation study

In order to verify that social relationships have a positive impact on the prediction of students' graduation development and that the two parts of Social-path and Transformer have the effect of improving the prediction effect of the model, this paper includes an ablation study:

- **No Social-path:** This model does not use Social-path selection layer to find Social-path, but uses Social-path embedding layer to embed social relationships.
- **No Embedding:** This model does not use Social-path embedding layer to embed social relationships based on No Social-path model, which is a further simplified model.
- **No Transformer** This model does not use the Transformer to train student's feature weight parameters.
- **No&&No** This model does not use the above methods and directly uses the most simple network model to predict the student's graduation development.

Under the same training set and test set, all models are trained. The accuracy rates are shown in Fig. 9.

The experimental results show that the SPE-TNN complete model has the best performance, which proves the effectiveness of the Social-path selection layer to find the Social-path and the Transformer to train the student's feature weight parameters. Firstly, by comparing the original model and the No Transformer model, when the lack of a transformer balances the weight parameters of the students' features, the effect of the model is slightly reduced, indicating that the model cannot learn the features of the students well without the Transformer, which reduces the accuracy

Fig. 9 Performance comparison of SPE-TNN and its variants

of the model's prediction. Secondly, by comparing the original model and the No Social-path model, when the model lacks the capture of Social-path, the accuracy of the model compared with the original model is significantly reduced, indicating that Social-path has a significant influence on the model, and the accuracy of the model can be significantly improved by capturing the Social-path that has an impact on the student's graduation development in the many social relationships of students. Thirdly, by comparing the original model with the No Embedding model, it is evident that after embedding no social relationship, the model only uses students' academic achievement to predict graduation development, and the accuracy rate is lower. Finally, when the social relationship is not embedded, and no Transformer assigns weights to the student's features, the No&&No model has the worst effect on the prediction of graduation development. In the end, the experiment proved that adding social relationships improved the accuracy of graduation development prediction. Meanwhile, searching for Social-path through Social-path selection layer and using the Transformer to train student feature weight parameters have significantly improved the model's task of predicting graduation development.

6 Conclusions

In this paper, we explore the task of graduation development prediction via gathering social relationships by introducing commonality and individual groups among students. Specifically, we proposed the SPE-TNN framework. The model uses Social-path selection layer to find social relationships that impact graduation development and embeds them into students' academic achievement through Social-path embedding layer. Then it introduces the Transformer to

balance the feature factors that affect the graduation development of students. Finally, the feature fusion and prediction are carried out by The multi-layer projection layer. Experiments on public datasets proved that the SPE-TNN model proposed in this paper has certain advantages over other models in accuracy and can provide more powerful support for graduates' employment guidance.

In the future, due to the richness and the complexity of social relationships, the social relationships mined in this paper may have many interfering factors, which will affect the prediction of graduation development. In our future work, we will consult more research literature, to find more representative social relationships among students, and to increase the accuracy of predicting student graduation development based on social relationships. These will help school administrators to provide better career guidance to students and enhance their employment rates.

Acknowledgements We gratefully acknowledge the support of National Natural Science Foundation of China(Grant No.61772180). This work is also supported by the Fujian Provincial Key Laboratory of Data Intensive Computing and Key Laboratory of Intelligent Computing and Information Processing (Grant No.BD201801), the Opening Project of State Key Laboratory for Novel Software Technology at Nanjing University (KFKT2021B12), the Key Research and Development plan of Hubei Province(2020BAB012).

References

1. Guan J, Xu H, Wang Y (2020) Impact of COVID-19 pandemic on the employment of chinese college graduates and countermeasures. *Basic Clin Pharmacol Toxicol* 127:233–234
2. Gati I, Levin N, Landman-Tal S (2019) Decision-Making Models and career guidance. *International Handbook of Career Guidance*, pp 115–145
3. Parola A (2020) Novel coronavirus outbreak and career development: A narrative approach into the meaning for Italian University Graduates. *Frontiers in psychology*
4. El Mrabet H, Moussa AA (2021) Iot-school guidance: A holistic approach to vocational self-awareness & career path. *Educ Inf Technol*, pp 1–18
5. Buchanan R (2017) ISocial media and social justice in the context of career guidance: is education enough?. *Career Guidance for Social Justice*, pp 109–124
6. El Mrabet H, Moussa AA (2019) Smart school guidance and vocational guidance system through the internet of things. *Proceedings of the 2nd International Conference on Netw, Inf Syst Secur - NISS19* 70:1–5
7. Belskaya E, Moldovanova E, Rozhkova S, Tsvetkova O, Chervach M (2016) University smart guidance counselling. *Smart Education and E-learning* 2016:39–49
8. Ouyang Y, Zeng Y, Gao R, Yu Y, Wang C (2020) Elective future: the influence factor mining of students' graduation development based on hierarchical attention neural network model with graph. *Appl Intell* 50:3023–3039
9. Hartatik, Kusriani K, Prasetyo AB (2020) Prediction of student graduation with naive bayes algorithm. *ICIC*, pp 1–5
10. Putri DY, Andreswari R, Hasibuan MA (2018) Analysis of students graduation target based on academic data record using

- c4.5 algorithm case study: Information systems students of telkom university. CITSM, pp 1–6
11. Baker R, RSJ (2010) Data mining for education. *International encyclopedia of education*. pp 122–118
 12. Sze V, Chen Y, Yang T, Emer JS (2017) Efficient processing of deep neural networks: a tutorial and survey. *Proc IEEE* 105(12):2295–2329
 13. Guo T, Yu K, Aloqaily M, Wan S (2022) Constructing a prior-dependent graph for data clustering and dimension reduction in the edge of AIoT. *Futur Gener Comput Syst* 128:381–394
 14. Wang L, Li M, Fang X, Nappi M, Wan S (2022) Improving random walker segmentation using a nonlocal bipartite graph. *Biomed Sign Process Control* 71:103154
 15. Chen P, Cheng W, Fan T (2016) Study on the BP neural network evaluation model of employability. Atlantis Press, pp 273–281
 16. Guo T, Xia F, Zhen S, Bai X, Zhang D, Liu Z, Tang J (2020) Graduate employment prediction with bias. *Comput Educ* 34(1):670–677
 17. Zhou T, Han X, Yan X, Yang Z, Zhao Z, Wang B (2013) Statistical mechanics on temporal and spatial activities of human. *J Univ Electron Sci Techno China* 42(04):481–540. (in Chinese)
 18. Kim Y, Choi TY, Yan T, Dooley K (2011) Structural investigation of supply networks: a social network analysis approach. *Journal of Operations Management* 29(3):194–211
 19. Wrzus C, Hänel M, Wagner J, Neyer FJ (2013) Social network changes and life events across the life span: a meta-analysis. *Psychol Bull* 139(1):53–80
 20. Lin J, Zhang L, He M, Zhang H, Liu G, Chen X, Chen Z (2019) Multi-Path relationship preserved social network embedding. *IEEE Access* 7:26507–26518
 21. John R, Jennifer BR (2015) Are students really connected? Predicting college adjustment from social network usage. *Educational Psychology* 35(7):819–834
 22. B Kim S, Jee J, Lee S, An S, Lee M (2018) Relationships between social support and student burnout: a meta-analytic approach. *Stress Health* 34(1):127–134
 23. Jr LG (2019) In-class social networks and academic performance: how good connections can improve grades. *ANAIS DO XXXIV SIMPÓSIO BRASILEIRO DE BANCO DE DADOS* 34:25–36
 24. Friedman BA, Mandel RG (2011) Motivation predictors of college student academic performance and retention. *J Coll Stud Retent: Res Theory Prac* 13(1):1–15
 25. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci* 106(36):15274–15278
 26. Cascio CJ, Moore D, McGlone F (2019) Social touch and human development. *Dev Cogn Neurosci* 35:5–11
 27. Crandall DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D, Kleinberg J (2010) Inferring social ties from geographic coincidences. *Proc Natl Acad Sci* 107(52):22436–22441
 28. Xu J, Yu S, Sun K, Ren J, Lee I, Pan S, Xia F (2020) Multivariate relations aggregation learning in social networks association for computing machinery. *Association for Computing Machinery*, pp 77–86
 29. A'rifian NINB, Daud NSABM, Romzi AFBM, Shahri NHHBM (2019) A comparative Study on Graduates' Employment in Malaysia by using Data Mining. *Journal of Physics: Conference Series*, vol 1366(1)
 30. Kipf TN, Welling M (2017) Semi-Supervised Classification with graph convolutional networks. *ICLR 2017*
 31. Yang LJ, Yang BS (1995) Research on common point problem of camber tangent plane in N-Dimension euclidean spaces. *Journal of Engineering of Heilongjiang University* 2:66–68
 32. Khayi NA, Rus V (2020) Graph convolutional networks for student answers assessment. In: *International conference on text*, pp 532–540
 33. Karimi H, Derr T, Huang J, Tang J (2020) Online academic course performance prediction using relational graph convolutional neural network. *Proceedings of The 13th international conference on educational data mining*, pp 444–450
 34. Yun S, Jeong M, Kim R, Kang J, Kim HJ (2019) Graph transformer networks. *Neural Information Processing Systems*, pp 11960–11970
 35. Vashishth S, Sanyal S, Nitin V, Talukdar P (2019) Composition-based Multi-Relational Graph Convolutional Networks. *ICLR 2020*
 36. Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. In: *Proceedings of the conference on empirical methods in natural language processing*, pp 1412–1421
 37. Zeng Y, Ouyang Y, Gao R, Qiu Y, Yu Y, Wang C (2020) “HHA: An attentive prediction model for academic abnormality,” *IEEE Access*, vol 8, pp IEEE Access
 38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *NIPS 2017*, pp 5998–6008
 39. Abeywickrama T, Cheema MA, Tanian D (2016) K-nearest Neighbors on Road networks: a journey in experimentation and in-memory implementation. *Proceedings of the VLDB Endowment* 9(6):492–503
 40. Zhang S, Li XM (2017) Learning k for kNN Classification. *Acm Trans Intell Syst Technol* 8:143–195
 41. Liang J, Yang J (2016) Big Data Application in education: Dropout Prediction in Edx MOOCs. In: *IEEE Second international conference on multimedia big data*
 42. Ajiboye R, Abdullah-Arshah R (2015) Using an enhanced Feed-Forward BP network for predictive model building from student's data. *Intell Autom Soft Comput* 22:169–175
 43. Liu H, Zhu Y, Zang T, Xu Y, Yu J, Tang F (2021) Jointly modeling heterogeneous student behaviors and interactions among multiple prediction tasks. In: *ACM TKDD2021*
 44. He X, Liao L, Zhang H, Nie L, Hu X, Chua T (2017) Neural collaborative filtering. *WWW*, pp 173–182
 45. Yu Su S, Liu Q (2018) Exercise-Enhanced Sequential modeling for student performance prediction. *AAAI*
 46. Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y (2017) Graph attention networks. *ICLR*
 47. Bhagavan KS, Thangakumar J, Subramanian DV (2021) Predictive analysis of student academic performance and employability chances using HLVQ algorithm. *J Ambient Intell Hum Comput* 12(3):3789–3797

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Guangze Yang received the M.S. degree from Hubei University of technology in 2022, Wuhan, China. His main research interests include Deep Learning and data mining.

Yong Ouyang received the M.S. degree from Hubei University of technology in 2007, Wuhan, China. He is currently a Associate Professor in School of Computer Science, Hubei University of Technology, Wuhan, China. He is also the director of the department of Computer Science and Technology in School of Computer Science, Hubei University of Technology, Wuhan, China. His research interests include data mining and intelligent education.

Rong Gao received the Ph.D. degree from Wuhan University in 2018, Wuhan, China. He is currently an Assistant Professor in School of Computer Science, Hubei University of Technology, Wuhan, China. His research interests include data mining and intelligent recommendation .

Zhiwei Ye received his Ph.D. degree in the school of remote sensing and information engineering at Wuhan University, Wuhan, China. He is a professor in the school of computer science, Hubei University of Technology, Wuhan, China. His major research interests include data mining, machine learning and evolutionary computation.

Yawen Zeng received the M.S. degree in Computer Science from Hunan University in 2022, Changsha, China. Her research interests include data mining and information retrieval.