Received 7 July 2023, accepted 27 July 2023, date of publication 4 August 2023, date of current version 28 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3301878

RESEARCH ARTICLE

Prediction of Graduation Development Based on Hypergraph Contrastive Learning With Imbalanced Sampling

YONG OUYANG[1], TUO FENG[1], RONG GAO[1,2], YUBIN XU[3], AND JINGHANG LIU[1]

[1]School of Computer Science, Hubei University of Technology, Wuhan 430068, China
[2]State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China
[3]Biology Department, South China Normal University, Guangzhou 510006, China

Corresponding author: Rong Gao (gaorong@hbut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62106070.

ABSTRACT With the increasingly competitive job market, the employment issue for college graduates has received more and more attention. Predicting graduation development can help students understand their suitable graduation development, thus easing the pressure of finding employment after graduation. However, existing research must look into the issue of imbalance and long-tail distribution in student graduation development. This paper proposes a novel hypergraph contrastive learning model based on imbalanced sampling (IS-HGCL) that enables us to address this problem. First, construct a hypergraph using students' school performance and social behavior. Then, our proposed imbalanced sampling strategy is applied to optimize the hypergraph structure and alleviate the imbalance issue. A self-updating hypergraph neural network is designed to optimize hyperedge representation and alleviate the long-tail distribution issue to enhance the hypergraph representation further. Finally, the structural consistency between the two optimized hypergraphs is maximized via node-level contrastive learning. Experiments on a real-world campus dataset demonstrate the superiority of the IS-HGCL model.

INDEX TERMS Graduation development prediction, hypergraph neural network, data imbalance, long tail distribution.

I. INTRODUCTION

As China attaches more importance to higher education, colleges and universities continue to expand yearly, and the number of graduates continues to increase. Data released by the China Bureau of Statistics shows that the number of general undergraduate graduates in China exceeded 9.67 million in 2022, an increase of 17.02% year-on-year from the previous year. In addition, with the outbreak of the COVID-19, the employment situation has become more severe, and the unemployment rate is rising yearly. In 2022, the national unemployment rate for college students increased by 1.2 percentage points compared to the previous year, currently standing at 6.7%. This surge in the unemployment rate has surpassed the levels recorded in recent years. Although the country has introduced employment policies and universities have launched employment counseling work for college students, the employment situation of college students still needs improvement.

Meanwhile, with the development of information technology, the campus information system in modern colleges and universities records a large amount of students' behavioral data during school. These data can reflect students' learning abilities, behavioral habits. Using data mining techniques to analyze the big data of campus, we can predict college students' employability in advance and guide them to employment in a targeted way. Therefore, the task of graduation development prediction by mining the information in students' school performance data to predict the employment intention and development direction of graduates after

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

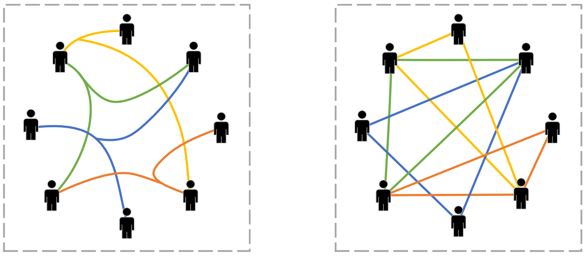

FIGURE 1. Hypergraph(left) and graph(right) representation of social relationships. (Different color sides represent different social relationships).

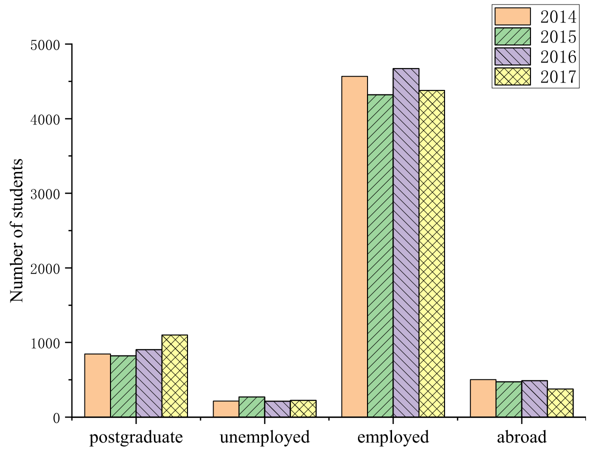

FIGURE 2. Graduation development distribution.

graduation is crucial to solving the difficult employment situation of college students.

In recent years, many studies have been conducted on student behavior. In the field of graduation development prediction, Nie et al. [1] proposed a central clustering method based on XGBOOST using behavioral data of more than 4000 students to predict students' career choices; Consedine et al. [2] investigated the importance of disgust sensitivity (DS) in medical career prediction by statistical methods; Khampirat [3] investigated the direct and indirect links between father's education, self-esteem, resilience, future goals, and college students' career expectations, and used structural equation modeling to develop and validate a model for predicting career expectations; Gao et al. [4] proposed an enhanced slime mold algorithm based on multiple cluster strategies for predicting graduate students' employment stability; Wang [5] examined the student-level and parental level to evaluate vocational education evaluation indexes and proposed a method for predicting vocational education employment rate based on the big data model.

Unfortunately, the above studies predict students' graduation development only by a single factor, such as students' grades, credits, or regular performance, ignoring social relationships that significantly impact students' graduation development choices [6]. Social relationships among students are vital for graduation development choice. Yang et al. [7] embedded social relationships based on students' elective and compulsory courses and balanced the weights of student characteristics by Transformer to predict students' graduation development; Cheng et al. [8] collected data on academic performance, English proficiency, and other activities to predict college students' employment outcomes using deep neural networks, providing guidelines for college students' career development.

However, the above studies represent complex social relationships through simple graphs, which may ignore some higher-order social information. As in Figure 1, In a simple graph, the edges only represent the connection between two nodes, which can only reflect a single binary relationship. The number of students' social relationships in a complex social relationship network is often substantial. Representing social relationships with simple graphs may lead to an excessive number of nodes and edges and overly complex graphs, which could be more conducive to graph visualization and analysis.

In contrast, hypergraphs are an extension of the simple graph model, as in Figure 1, and the hyperedges in hypergraphs can connect any number of nodes. They can describe the relationships between objects more comprehensively [9]. Therefore, using hypergraphs to represent social relationships is crucial, which can provide a more comprehensive basis and more accurate results for social network analysis and prediction.

Although modeling complex social relationships by adopting hypergraphs has great potential, there are still the following challenges:

1. Imbalanced distribution of graduation development data

Despite the annual increase in the enrollment of master's degree students in China, the graduation development of college students still presents an imbalance [10]. This paper utilizes the graduation development dataset [11] obtained from a university in China, which spans from 2014 to 2017. As depicted in Figure 2, the majority of students opted for direct employment, with a much smaller number of students pursuing non-employed paths such as studying abroad, postgraduate, and unemployment. In addition, the graph-based approach is susceptible to data sparsity [12], which may lead to the model being biased to learning the data characteristics of employed students and the prediction results being biased to employed, reducing the model accuracy. Therefore, solving data distribution imbalance is a crucial in graduation development prediction research.

Among the existing studies on data imbalance, Ajinkya [13] proposed an oversampling method to balance the number of different classes by sampling minority class samples; Zhao [14] used interpolation between minority class samples and neighboring samples to generate new minority class samples. The virtual minority class samples generated by the sampling methods of the above studies may be of low quality, leading to the model's bias toward the lesser number of non-employed students' characteristics being poorly learned, which leads to lower model accuracy. In contrast, existing studies on graduation development prediction [7] preprocess the dataset to alleviate the imbalance problem.

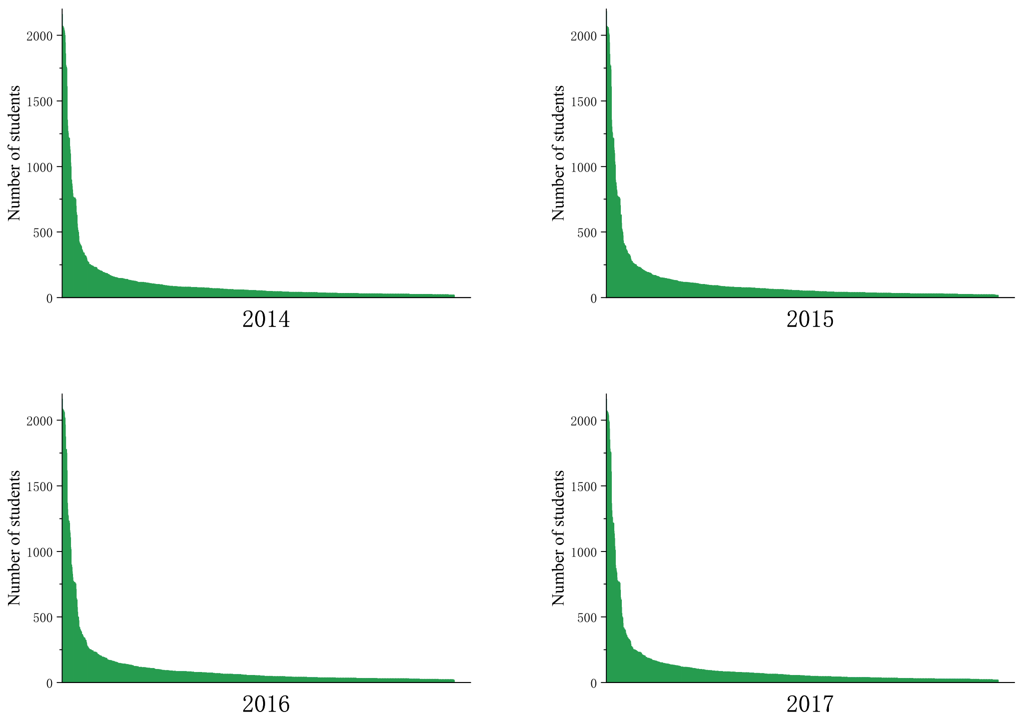

FIGURE 3. Student elective courses distribution(x-axis denotes the different electives and y-axis denotes the number of students who chose the electives).

2.The long-tail distribution of student social

With the development of higher education, there are more and more kinds and numbers of elective courses. Due to the lack of guidance, inspection, and incentives for teaching in schools to select courses, students often choose courses based on their interests or whether it is easy to get credits, etc. As a result, many students take a few courses while few take others. Zhao et al. [15] pointed out that social networks in the real world tend to be imbalanced, with certain social relationships being more intense than others.

As shown in Figure 3, more students take popular courses and fewer take other courses, which results in a long-tail distribution of students' social associations (Figure 4). The long-tail distribution causes the deep learning model to learn mainly the characteristics of students with high social associations, and the performance of students with fewer social associations will be significantly reduced.

In order to solve the above problems, this paper proposes a hypergraph contrastive learning model based on an imbalanced sampling graduation development prediction model (IS-HGCL). First, the fusion modeling of students' required courses, elective courses, and social behaviors is obtained by deeply mining the social relationships among students to obtain the joint hypergraph embedding. This paper proposes using an imbalanced sampling model based on Generative Adversarial Networks (GAN) to mitigate the issue of imbalanced data in the study of non-employed students.

Specifically, the proposed approach generates virtual nodes through GAN, effectively connecting them to the joint hypergraph to achieve a balanced network representation of each graduation development class. Furthermore, feature averaging is executed among the neighbors of virtual nodes to acquire virtual node characteristics. This proposed method attenuates the adverse implications of imbalanced data on model outcomes. Moreover, a hypergraph structure learning method is designed to learn the importance of hyperedges as well as node features through a self-updating hypergraph neural network, enhance the hypergraph by masking the unimportant information, and finally use the obtained view as a contrastive view and parameterize the balanced sampled hypergraph as another view. The node-level contrastive learning maximizes the structural consistency among different views, mitigates the effect of long-tail distribution, and improves the model's effectiveness for graduation development prediction.

Experiments on real campus datasets show that the method proposed in this paper outperforms other graduation development prediction models.

The main contributions in this paper are summarized as follows:

1. This paper proposes a GAN-based hypergraph node generation strategy aimed at alleviating the imbalanced problem in graduation development data. The generator is used to produce a set of virtual nodes representing graduation development of abroad, postgraduate, and unemployed, which

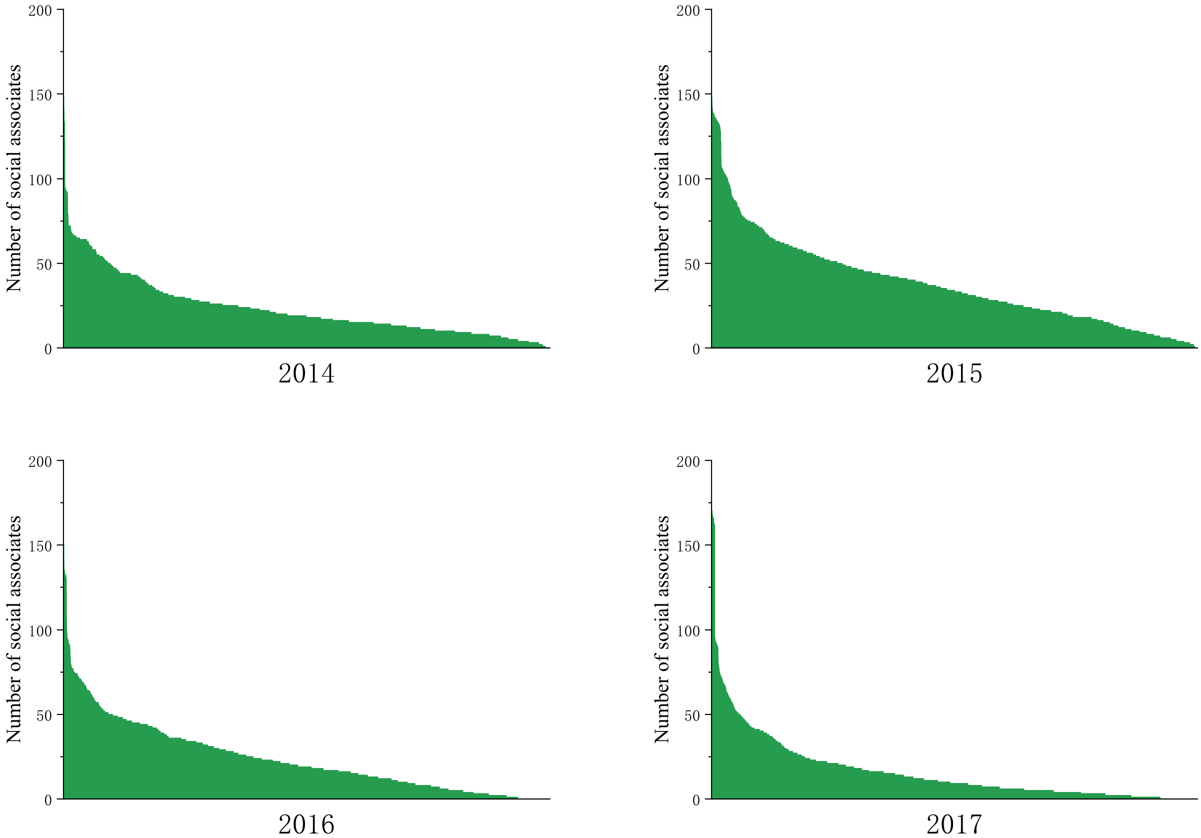

FIGURE 4. Student social association distribution(x-axis represents the student with a social relationship, y-axis represents the number of people with a social association with the student).

are added to the original network. As a result, a network is obtained in which the number of nodes representing each graduation development is balanced. Subsequently, the hypergraph neural network discriminator is used for adversarial training to enhance the reality of the generated virtual nodes, thereby improving the accuracy of the graduation development prediction task.

2. Using a node-based contrastive learning framework, the redundant hyperedges are removed through structured hypergraph learning, which enhances node representations and produces an optimized hypergraph. Meanwhile, node-level contrastive learning is performed between the hypergraph with balanced numbers of nodes and the optimized hypergraph, to maximize the structural consistency between different views adopting the contrastive loss. This strategy effectively mitigates the social long-tail distribution problem caused by an excess of unimportant social relationships among some students, and enhances the model's performance.

3. Empirical evidence demonstrates that graduation development prediction studies conducted after hypergraph sampling are more effective, and our method outperforms existing mainstream approaches on graduation data sets from four different years of actual campuses.

II. RELATED WORK

A. GRADUATION DEVELOPMENT PREDICTION

Research on predicting student graduation development is primarily based on traditional machine learning and deep learning methods. Nie et al. [1] proposed a data-driven framework for predicting students' career choices based on the behaviors of over 4,000 students on and around campus. Yu et al. [16] used the BP algorithm to construct a network layer and six types of feature label values to predict the employment and entrepreneurship development direction of 853 college students from 2010 to 2018. Kumar et al. [17] used a hybrid method to evaluate the impact of demographic characteristics on student employment arrangements and predicted students' employment status using support vector machines (SVM) and Random Forests (RF). Wang et al. [18] used an improved support vector machine based on communication and Gaussian simplification mechanisms with the enhanced butterfly optimization algorithm (CBBOA) to predict university students' career decisions. Peker et al. [19] developed an automated career guidance system based on fuzzy logic web services that incorporates students' previous academic achievements and teacher perspectives into employment predictions for vocational school students. However, traditional machine learning methods have limited learning capabilities when dealing with complex student data.

PetekŠter et al. [20] identified 31 predictive factors through survey analysis and used an artificial neural network to predict whether students choose a career in family medicine. Li et al. [21] adopted a statistical approach based on cluster analysis technology models, and established a graduate employment prediction model based on the Long-Short-Term Memory (LSTM) recurrent neural network.

Previous studies were conducted based on the assumption of an ideal distribution of student data and did not consider the possible variance in the actual data distribution.

B. CONTRASTIVE LEARNING

Contrastive learning [22] is a type of self-supervised learning that clusters positive samples together and pushes negative samples apart. MVGRL [23] learns node representations from two structural views and contrasts the encoding embeddings between the two graph views. Wang et al. [22] proposed a new HGNN collaborative contrastive learning mechanism, HeCo, inspired by self-supervised HGNN, which learns node embeddings from network motif views and meta-path views and performs contrastive learning between the two views.

Contrastive learning has achieved great success in computer vision (CV) [24] and natural language processing (NLP) [25]. Due to its powerful representation ability, it has been introduced into graph data and developed into graph contrastive learning [25]. In 2020, Khosla et al. [26] extended self-supervised contrastive learning to supervised learning and proposed supervised contrastive learning. Since then, supervised contrastive learning has been applied to graph classification and computer vision [27], [28]. However, the above methods could perform better in cases of imbalanced data distribution. Zhang et al. [27] proposed a new class-aware supervised contrastive learning framework for imbalanced fault diagnosis, which optimizes feature differences between any two classes using class information. Li et al. [28] proposed targeted supervised contrastive learning to improve the long-tail distribution problem in data. Xia et al. [29] proposed a graph contrastive learning framework, SimGRACE, which enhances graph data while preserving the original semantic information. They use the original graph as input, the GNN model and its perturbation version as two encoders to obtain two related views for contrast. Lee et al. [30] proposed a general framework, TriCL, for contrastive learning on hypergraphs, that performs three-way contrast between the same nodes, node groups, and groups and nodes in different views, and captures structural information of nodes and node groups in node embeddings.

This paper proposes to enhance the structure of hypergraphs to alleviate the long-tail distribution problem. The enhanced structure is compared with that of the original hypergraphs across nodes, and the model performance is improved by enhancing the structural consistency of different views.

C. DATA IMBALANCE

Data imbalance is common in real-world applications and has been a traditional research topic in the field of machine learning. Many tasks encounter this problem, such as traffic recognition [31] or image classification [32]. The class with more instances is usually called the majority class, while the class with fewer instances is usually called the minority class. Many methods now use oversampling or undersampling to adjust the size of each class directly. The standard oversampling method is to copy existing samples, which alleviates the impact of data imbalance but may lead to overfitting as no additional information is introduced.

SMOTE [33] is the most common oversampling method that solves this problem by generating new samples and performing interpolation between the minority class samples and their nearest neighbors. BOSME [34] generates artificial instances for the minority class based on the probability distribution of Bayesian networks, which learn from the original minority class through likelihood maximization. Zhao et al. [14] extended previous research on imbalanced learning to imbalanced node classification tasks on graphs, constructing an embedding space to encode the similarity between nodes to enhance the fidelity of generated nodes.

This paper utilizes GAN to sample hypergraphs and generate virtual nodes and improves the representational power of balanced hypergraphs by enhancing the reliability of virtual nodes through adversarial learning.

III. IS-HGCL

This paper proposes a hypergraph contrastive learning network based on imbalanced sampling to predict student graduation development by exploiting social information and academic performance data. The model consists of three main modules: a hypergraph construction module, a hypergraph imbalanced sampling module, and a hypergraph structure contrastive learning module.

The specific steps are as follows: 1) a hypergraph is constructed using preprocessed student performance and social data. 2) GAN is used to oversample the hypergraph to alleviate node imbalances and generate the joint hypergraph. Based on this, the hypergraph structure is learned to enhance hyperedge and node features and generate a contrastive view. 3) node-level contrastive learning is employed to maximize the structural.

A. PROBLEM FORMULATION

The task of graduation development prediction is based on various aspects, such as the students' historical academic performance and on-campus behavioral information. This study focuses on exploiting social relationships and course grades as the student features to output the graduation development of the students.

Let $G_1 = [R_1, R_2, \ldots, R_T]$ be the compulsory course grades of a student G_1 in a university, and R as the grades of different compulsory courses, and T as the number of

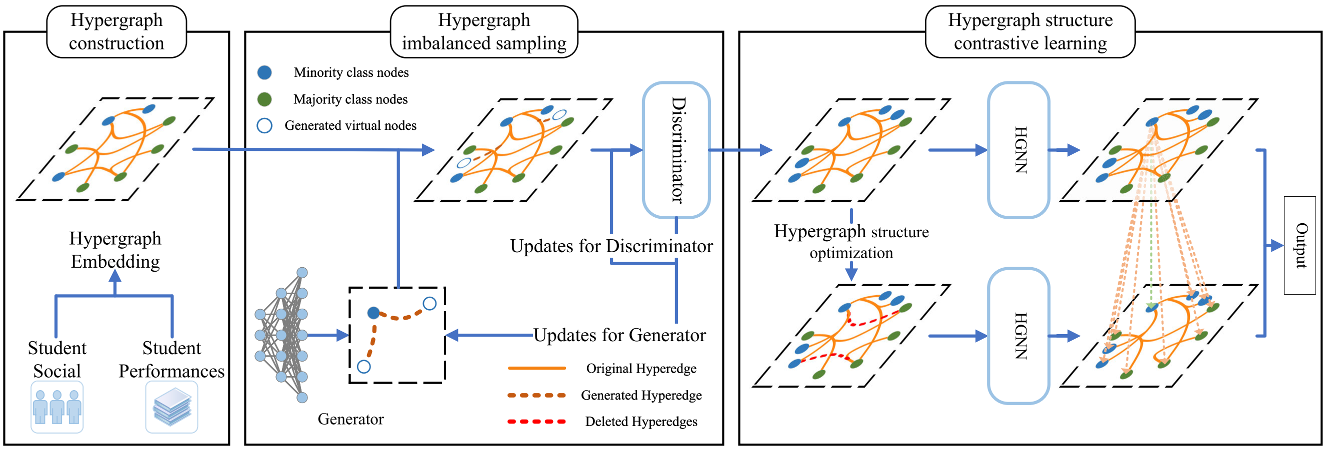

FIGURE 5. Overall model framework.

required courses. The compulsory course grade matrix for all students is denoted as $G = [G_1, G_2, \ldots, G_n]$, and n is the total number of students. In addition, following the classification plan of Wang et al. [35], the graduation development of students is classified into four categories: postgraduate, abroad, employed, and unemployed, denoted as $destination = \{a, b, c, d\}$.

Elective courses and dormitory data are used as social relationships to construct a hypergraph. The elective course grade matrix can be represented as $S = [S_1, S_2, \ldots, S_n]$, where S_i represents the elective course grade information of student i, and dormitory data is denoted as $Dor = [Dor_1, Dor_2, \ldots, Dor_n]$, where Dor_i represents the dormitory information of student i.

B. HYPERGRAPH CONSTRUCTION

Social relationships between students manifest as diverse higher-order relationships, which simple graph modeling cannot accurately represent. However, hyperedges in hypergraphs can associate with multiple nodes and thus better represent these multidimensional higher-order relationships. First, a hypergraph X is constructed based on the behavioral characteristics. Let $X = (V, E, A, F, C, W)$ be the hypergraph, where V represents the set of student nodes, node $v \in V$, E represents the set of hyperedges, and each hyperedge e is a subset of V. n_{maj} and n_{min} respectively represent the number of majority and minority nodes in hypergraph X, and the total number of nodes is $n = n_{maj} + n_{min}$, while the number of hyperedges is n_e. The adjacency matrix $A \in \mathbb{R}^{n \times n_e}$ is the matrix representation of the hypergraph:

$$A(v, e) = \begin{cases} 1, & if\ v \in e \\ 0, & otherwise \end{cases} \tag{1}$$

The feature matrix F represents the embedding of G's compulsory course grades, while $C = \{a, b, c, d\}$ is the node class matrix, with employed being the majority class $c_{min} = \{a, b, d\}$ and postgraduate, abroad, and unemployed being minority classes $c_{maj} = \{c\}$. The hyperedge weight matrix is denoted by W as a unit matrix. Each student is treated as a node, and adjacency matrix A is constructed using dormitory data Dor and elective course grade data S to determine whether a hyperedge e exists between two nodes.

C. GAN-BASED HYPERGRAPH IMBALANCED SAMPLING

In the task of graduation development prediction, the number of students who choose direct employment is always much more significant than those opting for postgraduate, abroad, or unemployed. This situation is an inevitable economic and societal development phenomenon, but it causes severe data imbalance issues that significantly affect graduation development prediction. The current imbalance issue is resolved through simple oversampling, which generates node information that needs more authenticity and provides limited assistance in predicting graduation development.

This paper proposes a hypergraph data generator, HypergraphGenerator, combined with HGNN to form a virtual node strategy based on GAN. This strategy can generate a minority class of student nodes to balance the number of student nodes choosing different graduation development. HypergraphGenerator can learn both the attribute distribution and the topological structure of the network. Afterward, the HGNN discriminator is trained to distinguish between actual nodes and generated virtual nodes and update HypergraphGenerator to increase the authenticity of the generated nodes. The GAN-based hypergraph imbalanced sampling module consists of two parts: a generator and a discriminator.

1) GENERATOR(G)

HypergraphGenerator: $Z \rightarrow F' \times A'$ is a fully connected neural network, which takes in a d_z dimensional noise space to obtain the feature space F', and the structure space A', of the generator network. The number of generated virtual nodes is $n_g = n_{maj} - n_{min}$. The number of input layer units $d_z = n_{\min}$, and the number of output layer units $d_o = n_g \times n_{\min}$ represent the topological relationship between the generated nodes and the original minority-class nodes. The

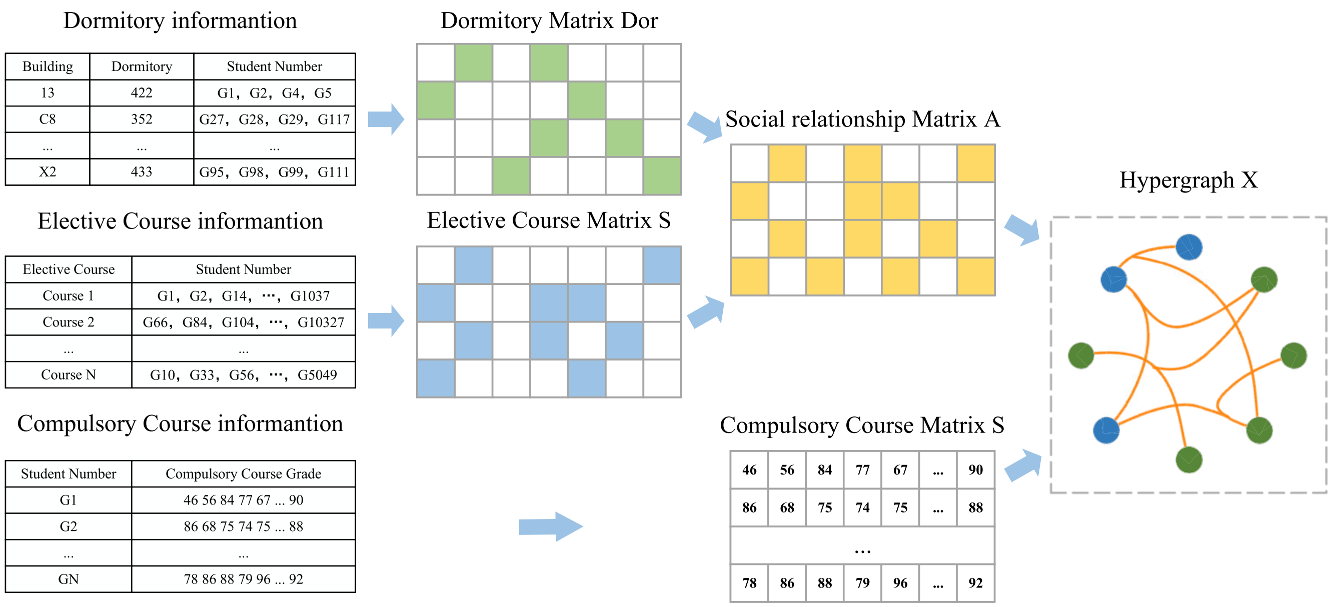

FIGURE 6. Hypergraph construction module.

hypergraph topological relationship is calculated as shown in equation 2:

$$o_i = ReLU(\hat{W}_i Z + \hat{b}_i), \quad i = 1, 2, 3, 4 \tag{2}$$

where $\hat{W}_i$ and $\hat{b}_i$ are weight matrices and bias vectors for layer i, respectively. o_i represents the output of layer i, and $ReLU\ (\cdot)$ is the activation function.

The output vector $\vec{o} \in \mathbb{R}^{d_o}$ is transformed into a matrix form $O \in \mathbb{R}^{n_g \times n_{\min}}$, and then the $softmax\ (O_i)$ function is applied to normalize each row of O:

$$A'_i = softmax\ (O_i) = \frac{e^{O_{ij}}}{\sum_{j=1}^{n_{\min}} e^{O_{ij}}}, \quad i = 1, \ldots, n_g \tag{3}$$

Each row of O_i, represents the connection relationship between the generated virtual nodes and the original minority-class nodes. Additionally, each element A'_{ij} represents the normalized weight of the connection between the virtual node $v_i \in V'$ and the original minority node $v_j \in V$, where V' is the set of generated minority nodes, and A' represents the network topological structure information between the generated virtual minority nodes and the original minority nodes.

In order to generate the node attribute feature $F' \in \mathbb{R}^{n_g \times f}$ of the virtual minority nodes, we aggregate the feature attributes of each neighboring node of the virtual minority nodes:

$$F' = A' F_{min} \tag{4}$$

where $F_{min} \in \mathbb{R}^{n_{min} \times f} \subset F$ is the minority class node feature matrix of the original hypergraph X. f is the number of dimensions of the minority class node features of the original hypergraph X. The minority nodes generated by HypergraphGenerator are added to the original hypergraph X to obtain a hypergraph $X_{bal} = \left(V', E', A', F', C', W'\right)$ with a balanced number of each class. where V' is the new node set consisting of the set of student nodes in X and the set of virtual student nodes, E' denotes the new hyperedge set consisting of all hyperedges in X and virtual hyperedges, A' and F' are the new adjacency matrix and feature matrix associated with V', respectively, W' is the expanded the hyperedge weight unit matrix, $C' = \{C_1, C_2\}$, and each node has C_1, C_2 two categories, C_1 is the new class matrix consisting of the class matrix of X and the virtual node class, which indicates the graduation development class of the node, and $C_2 = \{real, fake\}$ indicates whether the node is a virtual node.

The loss function of HypergraphGenerator is as in equation 5:

$$L_{gen} = L_{fake} + L_e + \alpha \left\|\Theta\right\|_2^2 \tag{5}$$

$$L_{fake} = \sum_{i=1}^{n_g} -c_i \log \Pr\left(\hat{y}_i = real | \vec{f}_i\right) \tag{6}$$

$$L_e = \frac{1}{|n_g|} \sum_{i=1}^{n_g} \sum_{j=1}^{n_{\min}} \left\|\vec{f}_i - \vec{f}_j\right\|_2^2 \tag{7}$$

L_{fake} is the confusion discriminator loss on the virtual node, where $c_i \in C_2$ and $\hat{y}_i \in Y$ denote the real label and discriminator output of the node, respectively, and $\vec{f}_i$ is the node embedding vector. L_e is the distance between the embedding vectors of the virtual and real nodes. $\alpha \left\|\Theta\right\|_2^2$ is the regularization term, where Θ is the set of HypergraphGenerator training weights with regularization coefficient α.

2) DISCRIMINATOR(D)

In this paper, we use HGNN [36] as a discriminator, which can distinguish whether a node is a virtual node generated by a generator. Therefore, we use the balanced hypergraph X_{bal}

as the input based on HGNN to obtain the classification result of nodes Y. The calculation procedure of HGNN is shown in equation 8.

$$Y = D_v'^{-\frac{1}{2}} A' W' D_e'^{-1} A'^T D_v'^{-\frac{1}{2}} F' \Theta \tag{8}$$

where Θ is the parameter to be learned during the training process, D_e' and D_v' are the hyperedge degree matrix and vertex degree matrix of the hypergraph X_{bal} generated through the adjacency matrix A', respectively. The degree of hyperedge e_i and the degree of node v_j are calculated as in equation (9,10).

$$D'_e(e_i) = \sum_{j=1}^{n} A'(e_i, v_j) \tag{9}$$

$$D'_v(v_j) = \sum_{i=1}^{n_e} A'(e_i, v_j) \tag{10}$$

The loss function of the discriminator is as in equation 11:

$$L_{dis} = L_{dif} + \beta \|\Omega\|_2^2 \tag{11}$$

$$L_{dif} = \sum_{i=1}^{n'} -\left[c_i \log(p_f) + (1 - c_i)\log(1 - p_f)\right] \tag{12}$$

where $p_f = \Pr(\hat{y}_i = fake|\vec{x}_i)$, L_{dif} is the cross-entropy loss of virtual and real nodes, $\beta \|\Omega\|_2^2$ is the regularization term, and $n' = n_g + n_{\min} + n_{maj}$ is the total number of nodes in the hypergraph X_{bal}, and Ω is the set of training weights for the discriminator with regularization factor β.

The adversarial objective functions of the generator and discriminator are shown in equation 13:

$$\begin{aligned}\min_G \max_D V(D, G) &= E_{x\sim P_{data}(x)}\left[\log D(x) + \beta\|\Omega\|_2^2\right] \\ &\quad + E_{z\sim P_z(z)}\left[\log(1 - D(G(z))) + L_e + \alpha\|\Theta\|_2^2\right]\end{aligned} \tag{13}$$

where x is the real data obeying the distribution p_{data} and z is the noise variable obeying the distribution p_z. Finally, the balanced hypergraph X_{bal} generated by the adversarial is obtained.

D. HYPERGRAPH STRUCTURE CONTRASTIVE LEARNING MODULE

This paper proposes using of hypergraph representation learning to address the adverse implications of long-tailed student social relationships on model performance. Specifically, the proposed approach involves optimizing the hypergraph structure in an end-to-end manner, thereby enhancing the model's efficacy by learning the hypergraph structure of the joint hypergraph X_{bal}, subsequent to imbalanced sampling. This strategy is achieved via node feature masking and hypergraph structure optimization, which ameliorates the negative effects of the long-tailed distribution of student social relationships on the model. This module is divided into two parts: hypergraph structure optimization and node-level contrastive learning.

1) HYPERGRAPH STRUCTURE OPTIMIZATION

In order to learn the hypergraph structure efficiently and improve the model performance, in this paper, two data enhancement schemes, node feature masking and hypergraph structure optimization, are used to simplify the structure and features of the hypergraph.

a: NODE FEATURE MASKING

A mask is used to mask the random node features to optimize the student node features. The masking vector $m^{(f')}$ is first sampled in the achievement feature matrix F', where each element is extracted from a Bernoulli distribution $p^{(f')}$ with probability. Then, the feature vector of each node is masked with $m^{(f')}$.

$$\begin{aligned} F'' &= \Gamma_{mask}(F') \\ &= \left[f'_1 \cdot m^{(f')}, f'_2 \cdot m^{(f')}, \cdots, f'_n \cdot m^{(f')}\right]^T \end{aligned} \tag{14}$$

where F'' is the augmented score feature matrix, $\Gamma_{mask}(F')$ is the feature mask transformation, and f_n' is the transpose of the nth row vector of F'.

b: HYPEREDGE SAMPLING

The balanced hypergraph X_{bal} is first optimized for hypergraph structure to clip task-irrelevant hyperedges. A binary mask vector $m' \in \{0, 1\}^{|E'|}$ is introduced to sample the hyperedges, where $m'_{v'}$ is indicates whether v' in a hyperedge is clipped or not, and $|N'|$, $|E'|$ are the number of nodes and hyperedges in the hypergraph X_{bal}, respectively. $m'_{v'}$ is taken from a Bernoulli distribution with parameter $a^{e'}_{v'}$, $m'_{v'} \sim Bern\left(a^{e'}_{v'}\right)$. $a^{e'} \in [0, 1]^{|E'|}$ is a vector of Bernoulli distribution parameters for the hyperedges, describing the importance of each hyperedge. The smaller the value of $a^{e'}_{v'}$, the more likely it is that v' in a hyperedge is noise and can be removed.

With the sampling mask m', the association matrix A'' of the hypergraph structure after hyperedge sampling can be expressed as:

$$A'' = M^{e'} \odot A', M^{e'} = \Gamma_{bro}(m') \tag{15}$$

$\Gamma_{bro} : R^{|E'|} \rightarrow R^{|N'|\times|E'|}$ denotes the broadcast operator which repeats the mask vector m' $|N'|$ times to construct a binary mask matrix $M^{e'} \in R^{|N'|\times|E'|}$ for superside sampling. $\odot$ denotes the Hadamard product.

2) NODE-LEVEL CONTRASTIVE LEARNING

This paper introduces a SimCLR [37] based contrastive learning framework that will use the optimized association matrix A'' and achievement features F'' to construct an enhanced hypergraph X_{opt} and a balanced hypergraph X_{bal} for node-level contrastive learning to maximize the structural consistency between hypergraphs. The framework consists of the following components:

a: ENCODER

In this paper, we use a two-layer HGNN as a hypergraph encoder $Encoder_\theta(\cdot)$ to extract the node representations of hypergraphs X_{opt} and X_{bal}: $\bar{X}_{opt} = Encoder_\theta(X_{opt})$, $\bar{X}_{bal} = Encoder_\theta(X_{bal})$, where θ is the encoder parameter and $\bar{X}_{opt}$ and $\bar{X}_{bal}$ are the node representation matrices of the enhanced and balanced hypergraphs, respectively.

b: PROJECTION HEAD

The embedding vectors of both encoders are mapped to the same space adopting a two-layer MLP as the projection head $pro_\eta(\cdot)$: $P_{opt} = pro_\eta(\bar{X}_{opt})$, $P_{bal} = pro_\eta(\bar{X}_{bal})$, where n is the parameter of the MLP and P_{opt}, P_{bal} is the projection node representation matrix of the augmented hypergraph and the balanced hypergraph, respectively.

E. LOSS FUNCTION

After mapping the embedding vectors to the same space, contrast loss is used to maximize the consistency between the projections $p_{opt,N}$ and $p_{bal,N}$ of the corresponding nodes v between different views. The contrast loss is as in equation 16

$$L = \frac{1}{2N}\sum_{N=1}^{n}\left[L\left(p_{opt,N}, p_{bal,N}\right) + L\left(p_{bal,N}, p_{opt,N}\right)\right] \quad (16)$$

where $sim(\cdot,\cdot)$ is the cosine similarity function, τ is the temperature parameter, and $L\left(p_{opt,N}, p_{bal,N}\right)$ and $L\left(p_{bal,N}, p_{opt,N}\right)$ are the positive and negative losses of nodal projections $p_{opt,N}$ and $p_{bal,N}$, respectively, calculated as in equations (17,18):

$$L\left(p_{opt,N}, p_{bal,N}\right) = \log \frac{e^{sim\left(p_{opt,N}, p_{bal,N}\right)/\tau}}{\sum_{k=1}^{n} e^{sim\left(p_{opt,N}, p_{bal,k}\right)/\tau}} \quad (17)$$

$$L\left(p_{bal,N}, p_{opt,N}\right) = \log \frac{e^{sim\left(p_{bal,N}, p_{opt,N}\right)/\tau}}{\sum_{k=1}^{n} e^{sim\left(p_{bal,N}, p_{opt,k}\right)/\tau}} \quad (18)$$

IV. EXPERIMENT

In the experimental section, this paper's research questions and experimental dataset are first introduced, followed by a description of the benchmark model, evaluation metrics, and implementation details of the proposed method. To verify the effectiveness of the model in this paper in academic early warning studies, the research questions are as follows:

Q1: Does IS-HGCL outperform existing mainstream graduation development prediction models on four different-year datasets?

Q2: How dose each of the three modules in IS-HGCL affect the model performance?

Q3: How does hyperparameter settings affect IS-HGCL in the experiment?

Q4: How does the convergence of IS-HGCL?

TABLE 1. Student information.

Year	Couse Number	Course Type	Student Number	Class	Student's College	Course Grade
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. Student dormitory information.

Year	Building Number	Dormitory Number	Student Number	College
2014	45	5280	20637	25
2015	45	5361	20722	27
2016	45	5276	20820	14
2017	45	5563	20751	17

TABLE 3. Student graduation development information.

Year	College	Student Number	Graduation Development
2014	25	20428	20428
2015	27	20355	20355
2016	14	20629	20629
2017	19	21423	21423

TABLE 4. The processed student graduation development information.

Year	a	b	c	d
2014	847	214	4567	502
2015	822	270	4320	475
2016	904	213	4670	488
2017	1099	225	4379	376

A. DATASET

In this paper, the original dataset is sourced from a university in China [10], which contains the students' grades dataset, dormitory dataset and graduation development dataset from 2014-2017 in school, etc.

The fields of the student grade dataset include TASK_NO, CUR_NAME, CUR_TYPE, CUR_DEP, CUR_CREDITH, STU_ID, STU_NAME, STU_SEX, STU_CLASS, STU_DEP, GRADE, which represent course number, course name, course type, course affiliation College, student number, name, gender, class, student's college, and course grade. The fields mainly used in this paper include course number, course type, student number, class, student's college, and course grade, and their number distribution is shown in Table 1.

The student dormitory data set includes the student number, name, gender, academic status, nature of enrollment, college, grade, academic system, building number, dormitory number, and dormitory fee rate. The fields mainly used in this paper include student number, college, building number, and dormitory number, and their number distribution is shown in Table 2.

The graduation development dataset includes detailed information on students, such as their college, student number, name, graduation development, type of report card issuance, unit to which the report card is signed, unit code, unit affiliation, location of the signed unit, nature of the unit, industry of the unit, employment status, actual location, organization code (business registration number) of the unit, gender, education, major, academic system, class of difficulty levels, cultivation mode, ethnicity, political affiliation, birth date, birth region, and class. For this study, our analysis primarily focuses on the fields of college, student number, and graduation development. The number distribution is shown in Table 3.

B. DATA PROCESSING

Firstly, preliminary processing of the data set data is carried out, and the main processing steps are the following five steps:

1. Extracting necessary fields: The original dataset consists of many unused fields, such as gender, academic system, dormitory fees, etc. Therefore, all irrelevant data is cleared, and only the fields used to predict the graduates' future are preserved.

2. Handling missing data: Student data with missing essential fields, such as student number, dormitory, and future employment status, are removed. If other missing fields exist, their missing values will be replaced with the median.

3. Merging duplicate data: Duplicate data caused by make-up exams, retakes, and the like is merged by calculating their means.

4. Correcting erroneous data: Values with scores lower than 0 will be modified to 0, and values exceeding 100 will be modified to 100.

5. Feature extraction: The information gained from each feature is calculated to assess their impact on the entire sample, thereby selecting the most influential features.

The distribution of the number of students after the data cleaning is completed is shown in Table 4.

From the student grade dataset, we get the compulsory course grade matrix G and the elective course dataset S. Then, from S and the dormitory dataset, we get the social relationship adjacency matrix A. Then, with the student as the vertex V, we get the edge E with the social relationship adjacency matrix A and initialize the hyperedge weight matrix W. Form the hypergraph $X = (V, E, A, F, C, W)$. Then, we generate nodes by imbalanced sampling to get the balanced hypergraph X_{bal}. Finally, input the hypergraph into the model to get the prediction results. Where a, b, c, and d stand for postgraduate, unemployed, employed, and abroad respectively.

C. EXPERIMENT SETTING

In this paper, we use 70% of the dataset as the training set and 30% as the testing set, exploiting Precision and Recall as evaluation metrics. All experiments are conducted on the PyTorch platform, which uses NVIDIA 3060 GPUs for training. The optimal parameters in the four datasets of this paper are set as follows: Hidder of 3, Learning Rate of 0.005, Loss Rate of 0.4, and Epoch of 1800.

Firstly, in order to verify the effectiveness of the model in this paper, the following sets of existing mainstream graduation development prediction models were designed and analyzed in contrastive with IS-HGCL:

1. GCN [38]: The Graph Convolutional Network (GCN) is the most usual and well-known method for generating balanced network embedding. This method obtains node embedding by aggregating the features of adjacent nodes.

2. GCN-Smote [14]: This method improves the performance of the GCN on imbalanced network embedding issues by incorporating SMOTE technology.

3. GAT [39]: This model leverages multi-head attention mechanisms to dynamically generate aggregation weights for adjacent nodes.

4. SimGRACE [29]: This Graph Contrastive Learning (GCL) model generates graph views by perturbing the model parameters.

5. HGNN [36]: This graph representation learning model employs truncated Chebyshev polynomials to generalize and approximate convolutional operations, thus enhancing representation learning.

6. HGCN [40]: This semi-supervised learning framework is based on spectral theory.

7. HNHN [41]: This HGCN model employs nonlinear activation functions, and includes a normalization model that flexibly adjusts the importance of hyper-edges and nodes.

8. HyperSAGE [42]: This graph representation learning framework learns hypergraph embedding by aggregating messages in a two-stage process.

9. AST [43]: This advanced supervised hypergraph model employs the set functions learned through Deep Sets [44] and Set Transformer [45] for information propagation.

10. TriCL [30]: This advanced unsupervised contrast hypergraph learning model maximizes the mutual information between nodes, hyperedges, and node groups.

In this paper, the following models are designed for contrastive:

1) IS-HGCL: A complete model containing a hypergraph construction module, a hypergraph imbalance sampling module, and a hypergraph structure contrastive learning module.

2) No GO: It does not contain the hypergraph imbalanced sampling module, only the hypergraph construction module and the hypergraph structure contrastive learning module. It takes the hypergraph constructed from the original data as input, without balancing the hypergraph by node sampling.

3) No GS:It does not contain the hypergraph construction module, only the hypergraph imbalanced sampling module, hypergraph structure contrastive learning module, using the unit matrix of student nodes to get the adjacency matrix to construct the hypergraph for input.

4) No CL: No hypergraph structure contrastive learning module is included, only hypergraph construction module, hypergraph imbalanced sampling module is included, and the

TABLE 5. Comparing experimental results.

Dataset	graph representiom	Method	a		b		c		d	
			Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
2014	Graph	GCN	21.15	52.38	35.13	16.67	82.89	23.16	35.80	25.81
		GCN-SMOTE	61.52	78.81	65.88	71.54	78.74	83.17	60.55	72.27
		GAT	20.35	54.76	33.70	18.76	80.77	22.66	36.25	22.58
		SimGRACE	63.64	53.95	76.43	58.79	50.98	90.10	46.72	60.80
	Hypergraph	HGNN	35.49	66.97	34.83	40.07	73.29	34.20	49.09	50.17
		HGCN	45.47	59.49	44.44	36.03	73.20	43.41	59.08	59.95
		HNHN	45.39	65.51	34.63	35.29	74.44	46.25	52.40	52.57
		HyperSAGE	55.85	51.28	44.37	34.19	74.71	42.37	58.63	61.78
		AST	66.26	76.90	74.39	61.99	73.13	71.83	68.46	69.50
		TriCL	76.25	81.68	79.34	78.62	85.18	78.15	78.34	82.67
		IS-HGCL	**82.20**	**84.08**	**84.32**	**81.25**	**87.7**	**82.01**	**83.28**	**84.67**
2015	Graph	GCN	33.46	42.64	20.41	26.18	89.46	25.15	16.49	12.45
		GCN-SMOTE	64.15	76.23	61.21	81.21	85.21	84.25	66.23	65.12
		GAT	26.24	45.54	41.36	16.23	82.67	23.48	26.22	26.21
		SimGRACE	64.14	43.36	77.35	62.23	62.58	81.65	53.34	62.59
	Hypergraph	HGNN	43.58	64.21	38.65	45.66	76.25	42.12	51.56	56.65
		HGCN	48.52	62.22	45.65	34.26	75.24	48.69	63.21	66.45
		HNHN	48.36	62.13	36.47	38.16	77.89	54.45	59.95	59.46
		HyperSAGE	59.74	58.26	48.98	39.46	76.45	48.53	64.55	65.45
		AST	69.86	75.16	76.54	68.41	81.36	76.45	68.16	71.31
		TriCL	83.54	79.21	79.45	80.35	83.84	84.32	83.45	82.32
		IS-HGCL	**86.43**	**84.61**	**83.28**	**84.51**	**86.46**	**87.44**	**85.62**	**86.82**
2016	Graph	GCN	28.49	43.54	34.86	29.84	81.43	26.46	34.35	20.54
		GCN-SMOTE	56.44	64.68	64.31	64.86	73.54	76.15	64.56	66.43
		GAT	26.84	48.46	34.14	25.45	74.53	42.65	43.43	20.84
		SimGRACE	64.81	53.51	74.80	56.64	54.68	81.34	56.46	66.79
	Hypergraph	HGNN	43.43	59.13	34.68	44.61	79.46	34.32	55.62	48.16
		HGCN	44.89	64.65	44.63	34.86	68.43	41.68	61.65	56.86
		HNHN	44.81	64.68	38.46	34.65	71.23	48.46	58.34	43.42
		HyperSAGE	59.61	58.45	48.61	34.68	73.18	48.61	64.62	61.48
		AST	69.23	71.35	76.42	54.63	78.46	73.16	70.86	76.13
		TriCL	80.65	84.61	78.48	76.48	86.42	76.48	79.45	81.61
		IS-HGCL	**86.68**	**85.84**	**82.64**	**80.81**	**88.46**	**81.54**	**81.78**	**84.84**
2017	Graph	GCN	24.34	51.31	34.51	13.54	83.38	21.35	34.46	26.45
		GCN-SMOTE	64.83	76.15	61.68	68.43	76.48	76.15	61.84	73.15
		GAT	26.84	51.31	31.35	14.64	81.34	26.45	34.98	16.81
		SimGRACE	64.35	56.45	76.14	54.89	48.64	86.48	48.61	59.46
	Hypergraph	HGNN	34.82	64.84	30.64	38.45	69.84	34.56	51.86	52.64
		HGCN	44.13	61.85	49.16	40.56	72.15	42.18	61.65	64.82
		HNHN	46.85	68.46	38.46	34.86	76.48	53.10	50.16	56.42
		HyperSAGE	58.64	54.61	46.13	38.94	76.81	43.81	66.14	63.45
		AST	72.10	76.42	73.42	64.82	74.64	73.51	71.16	73.14
		TriCL	79.15	82.53	75.13	79.48	84.61	80.31	79.41	82.31
		IS-HGCL	**85.45**	**86.48**	**83.54**	**82.4**	**86.96**	**82.46**	**82.16**	**83.65**

VOLUME 11, 2023

discriminator HGNN of the imbalanced module is used as a multi-classifier to output the classification results.

Finally, it is discussed that there is a certain degree of influence of the variation of different hyperparameters on the experimental results, which is compared and analyzed, and the influence of the Hidder of the model, the model learning rate, the Epoch count and the loss rate on the model performance.

D. EXPERIMENTAL RESULTS

1) COMPARISON EXPERIMENT (Q1)

To verify the validity of this paper's model in the graduation development prediction study, the IS-HGCL was compared with the four graduation development [35] predicted by the benchmark models mentioned above for the four years of 2014, 2015, 2016, and 2017 on the datasets of postgraduate, unemployed, employed and abroad. The benchmark models are broadly classified into graph-based and hypergraph-based neural network models.

The results of the model comparison analysis experiments are shown in Table 5, where the IS-HGCL model performed better overall than the other models. Among the other models, the GCN and GAT models are affected by imbalanced data, and only the prediction of employed direction students has a higher accuracy rate. The GCN-SMOTE model has a more balanced prediction accuracy rate and significantly improved performance because the SMOTE component oversampled and balanced the data set. The SimGRACE model uses graph contrastive and better learns the data features of long-tailed distribution, improving performance. The HGNN, HGCN, and HNHN models use hypergraphs to dig deeper into the higher-order relationships among students, and their effects are significantly more potent than those of GCN and GAT. TriCL uses contrastive learning on hypergraph-based representations, and its effects are significantly more potent than those of other models except the IS-HGCL model. Moreover, the prediction accuracy of specific targets of the IS-HGCL model is generally higher than that of other models, among which the accuracy of predicting the targets of *a* (postgraduate), *b* (unemployed) and *d* (abroad) is slightly lower than the total accuracy; the prediction result of *c* (employed) target is the best.

Our IS-HGCL model uses a hypergraph to represent the social relationships among students, which effectively makes up for the shortage of adopting graphs to learn students' graduation development in the past. Implicit relationships among students and the final prediction results obtained are significantly higher than those of other models, thus verifying the model's superiority in this paper.

2) COMPONENT ANALYSIS (Q2)

In order to further verify the effectiveness of the method proposed in this paper, IS-HGCL component analysis experiments were conducted on four types of prediction results, and the experimental results are shown in Figure 7:

The data from the component analysis of the model showed that the complete model with graph construction module, graph optimization module, and node-contrast learning module had the best prediction results. The No GL model, which did not use node-level comparative learning, or the No GO model, which did not use the node generator to balance the hypergraph data was clearly less effective than the IS-HGCL model. The No GS model without the graph construction module did not use the students' social relationships for prediction and could have been better. When the graph construction module is not included, the model only considers student performance characteristics and cannot thoroughly learn the social attributes of students. At this time, students only exist in association with themselves, which will lead to a decrease in the accuracy of model detection. When the graph optimization module is not included, virtual nodes cannot be generated to optimize the joint graph, and the model is affected by data imbalance, resulting in lower model accuracy. When the contrastive learning module is not included, the node representation cannot be learned through node-level contrastive learning, the implicit relationship between students cannot be deeply explored, and the model suffers from long-tail distribution interference, leading to lower model performance. The importance of using social relationships and student grades as factors affecting students' graduation development is finally demonstrated. Using a node generator to generate virtual nodes to balance the dataset and node-level contrastive learning to mine implicit relationships among students provides relatively significant enhancements to the model.

3) PARAMETERS SENSITIVITY ANALYSIS(Q3)

Variations of different hyperparameters can influence the experimental results. This paper conducted a comparative analysis of the 2014 dataset to observe the effects of Hidder, model learning rate, Epoch count, and loss rate on the model performance.

The Hidder in Figure 8(a) indicates the number of hidden layers of the hypergraph neural network used in the experiments, set between 1 and 5 with a step size of 1. The model works best when the Hidder is 3, and when the Hidder is too much, the model may be over-smoothed leading to performance degradation.

The learning rate in Figure 8(b) indicates the magnitude of each parameter update in the experiment, and the values are 0.0003, 0.0005, 0.0007, 0.003, 0.005, 0.007, 0.01, 0.03, and 0.05. The model works best when the learning rate is 0.005, and the learning rate is too large to cause the model to fail to converge, which may lead to overfitting.

The loss rate in Figure 8(c) represents the probability of setting neurons to zero in the network, which is used to prevent the network from overfitting, and is set in the range of 0.1 to 0.7 with a step size of 1. The best result is achieved when the loss rate is 0.4. Setting the loss rate too low does not effectively prevent overfitting, and setting it too high will lead to a decrease in test accuracy.

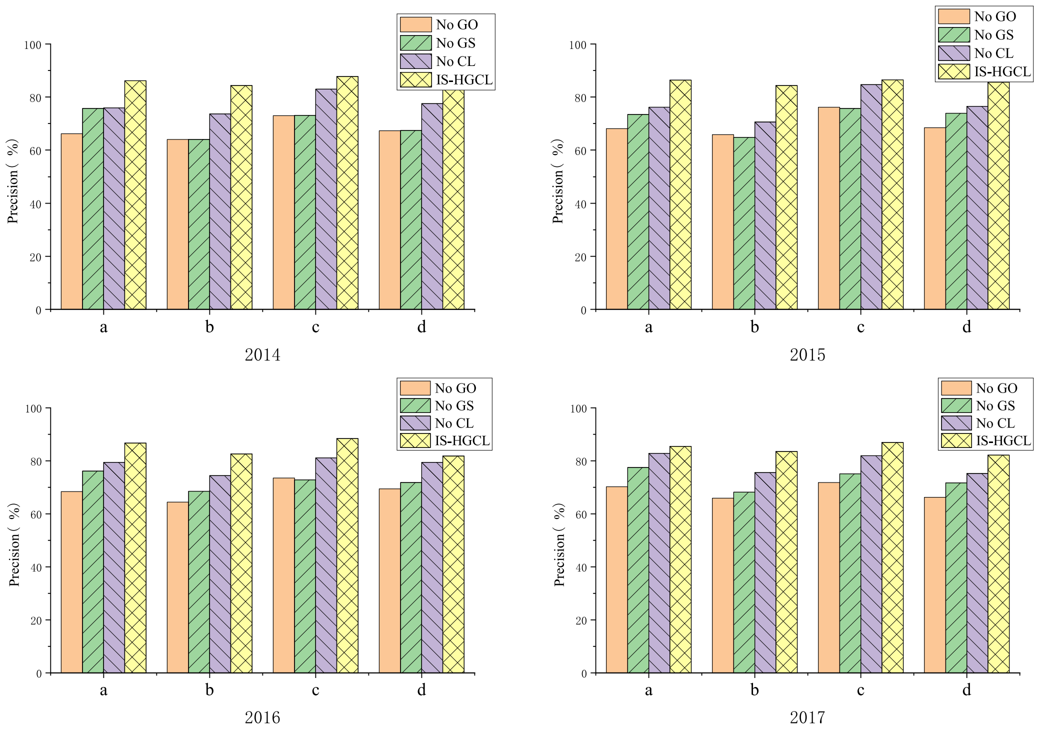

FIGURE 7. Component analysis.

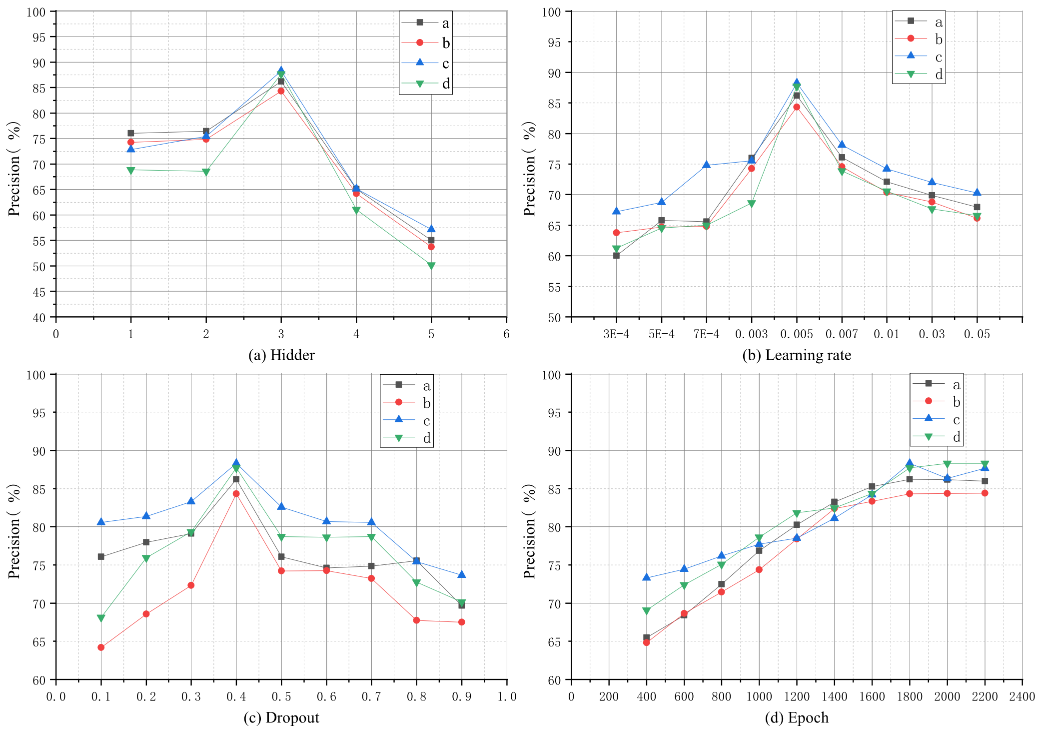

FIGURE 8. Parameter analysis.

In Figure 8(d), Epoch is the training number of the model, which is set from 400 to 2200 with a step size of 200. The model performs best when the Epoch number is 1800, but the model cannot thoroughly learn the data features when the Epoch number is too small and may cause overfitting when the Epoch number is too large.

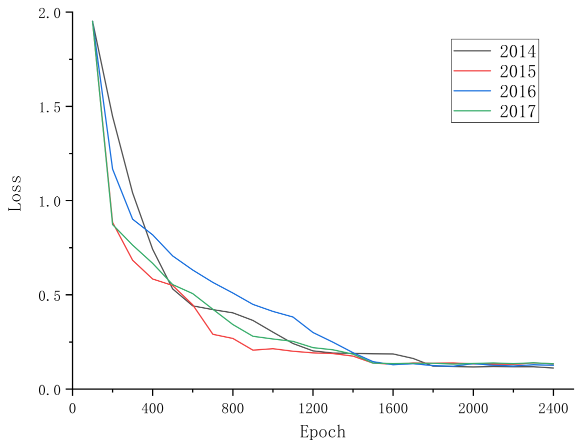

FIGURE 9. Convergence analysis.

We can get the best performance from the experimental results when the Hidder is 3, the model learning rate is 0.005, the Epoch number is 1800, and the loss rate is 0.4.

4) CONVERGENCE ANALYSIS(Q4)

To verify the convergence of the model, we plotted the loss function curves for four data sets. As shown in Figure 9, the loss functions of three data sets converged at 1600 epochs, while the 2014 dataset converged at 1800 epochs. It can be observed that the objective values of our algorithm consistently decreased with each iteration, ultimately reaching convergence. These results validate the convergence of the algorithm.

V. CONCLUSION

Facing the increasingly severe employment problem of graduates, the prediction of students' graduation development is crucial. Since the current work cannot well explore the social relationships among students, this paper characterizes the social relationships among students by hypergraph and proposes a hypergraph contrastive learning graduation development prediction model (IS-HGCL) based on imbalanced sampling. This model generates a few classes of nodes to balance the network by generating an adversarial network, and feature fusion and prediction are performed by comparing and converging students' implicit relationships among nodes. Through experiments on public datasets, it is proved that the IS-HGCL model proposed in this paper is superior to other models in terms of accuracy and can provide more powerful help for graduates' career guidance. However, there are still a small number of students with fewer social activities. Only some of their social relationships can be obtained, which makes it difficult to predict the development of such students, and finding the influence of more social relationships on students' graduation development is a future research direction.

REFERENCES

[1] M. Nie, Z. Xiong, R. Zhong, W. Deng, and G. Yang, "Career choice prediction based on campus big data—Mining the potential behavior of college students," *Appl. Sci.*, vol. 10, no. 8, p. 2841, Apr. 2020.

[2] N. S. Consedine, T.-C. Yu, and J. A. Windsor, "Nursing, pharmacy, or medicine? Disgust sensitivity predicts career interest among trainee health professionals," *Adv. Health Sci. Educ.*, vol. 18, no. 5, pp. 997–1008, Dec. 2013.

[3] B. Khampirat, "The relationship between paternal education, self-esteem, resilience, future orientation, and career aspirations," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0243283.

[4] H. Gao, G. Liang, and H. Chen, "Multi-population enhanced slime mould algorithm and with application to postgraduate employment stability prediction," *Electronics*, vol. 11, no. 2, p. 209, Jan. 2022.

[5] L. Wang, "Big data analysis model for vocational education employment rate prediction," *Sci. Program.*, vol. 2022, Jul. 2022, Art. no. 7576521.

[6] J. P. Shi and H. F. Z. J. Li, "Diversity recommendation approach based on social relationship and user preference (in Chinese)," *Comput. Sci.*, vol. 45, no. S1, pp. 423–427, 2018.

[7] G. Yang, Y. Ouyang, Z. Ye, R. Gao, and Y. Zeng, "Social-path embedding-based transformer for graduation development prediction," *Applied Intelligence*, vol. 52, no. 12, pp. 14119–14136, Sep. 2022.

[8] P. Cheng, "Employment data screening and destination prediction of college students based on deep learning," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–10, Feb. 2022.

[9] M. Li, Y. Zhang, X. Li, L. Cai, and B. Yin, "Multi-view hypergraph neural networks for student academic performance prediction," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105174.

[10] T. Guo, F. Xia, S. Zhen, X. Bai, D. Zhang, Z. Liu, and J. Tang, "Graduate employment prediction with bias," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 670–677.

[11] Y. Ouyang, Y. Zeng, R. Gao, Y. Yu, and C. Wang, "Elective future: The influence factor mining of students' graduation development based on hierarchical attention neural network model with graph," *Appl. Intell.*, vol. 50, no. 10, pp. 3023–3039, Oct. 2020.

[12] Z. Lin, C. Tian, Y. Hou, and W. X. Zhao, "Improving graph collaborative filtering with neighborhood-enriched contrastive learning," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 2320–2329.

[13] A. More, "Survey of resampling techniques for improving classification performance in unbalanced datasets," 2016, *arXiv:1608.06048*.

[14] T. Zhao, X. Zhang, and S. Wang, "GraphSMOTE: Imbalanced node classification on graphs with graph neural networks," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, Mar. 2021, pp. 833–841.

[15] C. Zhao, Y. Xin, X. Li, Y. Yang, and Y. Chen, "A heterogeneous ensemble learning framework for spam detection in social networks with imbalanced data," *Appl. Sci.*, vol. 10, no. 3, p. 936, Jan. 2020.

[16] M. Yu, "Predicament and thinking of college students' employment and entrepreneurship under the background of supply-side reform," *Mobile Inf. Syst.*, vol. 2022, Jun. 2022, Art. no. 1495123.

[17] D. Kumar, C. Verma, P. K. Singh, M. S. Raboaca, R.-A. Felseghi, and K. Z. Ghafoor, "Computational statistics and machine learning techniques for effective decision making on student's employment for real-time," *Mathematics*, vol. 9, no. 11, p. 1166, May 2021.

[18] Z. Wang, G. Liang, and H. Chen, "Tool for predicting college student career decisions: An enhanced support vector machine framework," *Appl. Sci.*, vol. 12, no. 9, p. 4776, May 2022.

[19] M. Peker, H. Gürüler, and B. Sen, "A new fuzzy logic based career guidance system: WEB-CGS," *Tehnički vjesnik*, vol. 24, no. 6, pp. 1863–1868, 2017.

[20] M. Petek Šter, I. Švab, and B. Šter, "Prediction of intended career choice in family medicine using artificial neural networks," *Eur. J. Gen. Pract.*, vol. 21, no. 1, pp. 63–69, Jan. 2015.

[21] X. Li and T. Yang, "Forecast of the employment situation of college graduates based on the LSTM neural network," *Comput. Intell. Neurosci.*, vol. 2021, Sep. 2021, Art. no. 5787355.

[22] Z. Peng, W. Huang, M. Luo, Q. Zheng, Y. Rong, T. Xu, and J. Huang, "Graph representation learning via graphical mutual information maximization," in *Proc. Web Conf.*, Apr. 2020, pp. 259–270.

[23] X. Wang, N. Liu, H. Han, and C. Shi, "Self-supervised heterogeneous graph neural network with co-contrastive learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 1726–1736.

[24] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[26] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 18661–18673.
[27] J. Zhang, J. Zou, Z. Su, J. Tang, Y. Kang, H. Xu, Z. Liu, and S. Fan, "A class-aware supervised contrastive learning framework for imbalanced fault diagnosis," *Knowl.-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109437.
[28] T. Li, P. Cao, Y. Yuan, L. Fan, Y. Yang, R. Feris, P. Indyk, and D. Katabi, "Targeted supervised contrastive learning for long-tailed recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 6908–6918.
[29] J. Xia, L. Wu, J. Chen, B. Hu, and S. Z. Li, "SimGRACE: A simple framework for graph contrastive learning without data augmentation," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 1070–1079.
[30] D. Lee and K. Shin, "I'm me, we're us, and I'm us: Tri-directional contrastive learning on hypergraphs," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 7, pp. 8456–8464.
[31] L. Peng, H. Zhang, Y. Chen, and B. Yang, "Imbalanced traffic identification using an imbalanced data gravitation-based classification model," *Comput. Commun.*, vol. 102, pp. 177–189, Apr. 2017.
[32] Y. Peng, X. He, and J. Zhao, "Object-part attention model for fine-grained image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1487–1500, Mar. 2018.
[33] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
[34] R. Delgado and J. D. Núñez-González, "Bayesian network-based over-sampling method (BOSME) with application to indirect cost-sensitive learning," *Sci. Rep.*, vol. 12, no. 1, p. 8724, May 2022.
[35] Y. Wang, L. Yang, J. Wu, Z. Song, and L. Shi, "Mining campus big data: Prediction of career choice using interpretable machine learning method," *Mathematics*, vol. 10, no. 8, p. 1289, Apr. 2022.
[36] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 3558–3565.
[37] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1597–1607.
[38] N. Heidari and A. Iosifidis, "Progressive graph convolutional networks for semi-supervised node classification," *IEEE Access*, vol. 9, pp. 81957–81968, 2021.
[39] J. Zhang, M. Li, K. Gao, S. Meng, and C. Zhou, "Word and graph attention networks for semi-supervised classification," *Knowl. Inf. Syst.*, vol. 63, no. 11, pp. 2841–2859, Nov. 2021.
[40] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar, "HyperGCN: A new method for training graph convolutional networks on hypergraphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.
[41] Y. Dong, W. Sawin, and Y. Bengio, "HNHN: Hypergraph networks with hyperedge neurons," 2020, *arXiv:2006.12278*.
[42] D. Arya, D. K. Gupta, S. Rudinac, and M. Worring, "HyperSAGE: Generalizing inductive representation learning on hypergraphs," 2020, *arXiv:2010.04558*.
[43] E. Chien, C. Pan, J. Peng, and O. Milenkovic, "You are AllSet: A multiset function framework for hypergraph neural networks," 2021, *arXiv:2106.13264*.
[44] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola, "Deep sets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[45] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teho, "Set Transformer: A framework for attention-based permutation-invariant neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 3744–3753.

YONG OUYANG received the M.S. degree from the Hubei University of Technology, Wuhan, China, in 2007. He is currently an Associate Professor with the School of Computer Science, Hubei University of Technology, where he is also the Director of the Department of Computer Science and Technology, School of Computer Science. His research interests include data mining and intelligent education.

TUO FENG received the M.S. degree from the Hubei University of Technology, Wuhan, China, in 2024. His main research interests include deep learning and data mining.

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

YUBIN XU received the first M.Sc. degree from the University of Glasgow, Glasgow, U.K., in 2020, with a strong interest in the field of education. She is currently pursuing the second master's degree in subject teaching (biology) with South China Normal University, Guangzhou, China. Her research interests include educational technology and AI-assisted teaching.

JINGHANG LIU received the Ph.D. degree from the Wuhan University of Technology, Wuhan, China, in 2018. He is currently an Assistant Professor with the School of Computer Science, Hubei University of Technology, Wuhan. His research interests include machine learning and artificial intelligence.