DGCD: An Adaptive Denoising GNN for Group-level Cognitive Diagnosis

Haiping Ma[1], **Siyu Song**[1,2], **Chuan Qin**[2,3*], **Xiaoshan Yu**[4],
Limiao Zhang[1*], **Xingyi Zhang**[5] and **Hengshu Zhu**[2]
[1]Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University, Hefei, Anhui, China
[2]Career Science Lab, Boss Zhipin, Beijing, China
[3]PBC School of Finance, Tsinghua University, Beijing China
[4]School of Artificaial Intelligence, Anhui University, Hefei, Anhui, China
[5]School of Computer Science and Technology, Anhui University, Hefei, Anhui, China
hpma@ahu.edu.cn, siyusong00@gmail.com, chuanqin0426@gmail.com, yxsleo@gmail.com, zhanglm@ahu.edu.cn, xyzhanghust@gmail.com, zhuhengshu@gmail.com

Abstract

Group-level cognitive diagnosis, pivotal in intelligent education, aims to effectively assess group-level knowledge proficiency by modeling the learning behaviors of individuals within the group. Existing methods typically conceptualize the group as an abstract entity or aggregate the knowledge levels of all members to represent the group's overall ability. However, these methods neglect the high-order connectivity among groups, students, and exercises within the context of group learning activities, along with the noise present in their interactions, resulting in less robust and suboptimal diagnosis performance. To this end, in this paper, we propose **DGCD**, an adaptive **D**enoising graph neural network for realizing effective **G**roup-level **C**ognitive **D**iagnosis. Specifically, we first construct a **g**roup-**s**tudent-**e**xercise (GSE) graph to explicitly model higher-order connectivity among groups, students, and exercises, contributing to the acquisition of informative representations. Then, we carefully design an adaptive denoising module, integrated into the graph neural network, to model the reliability distribution of student-exercise edges for mining purer interaction features. In particular, edges of lower reliability are more prone to exclusion, thereby reducing the impact of noisy interactions. Furthermore, recognizing the relational imbalance in the GSE graph, which could potentially introduce bias during message passing, we propose an entropy-weighted balance module to mitigate such bias. Finally, extensive experiments conducted on four real-world educational datasets clearly demonstrate the effectiveness of our proposed DGCD model. The code is available at https://github.com/BIMK/Intelligent-Education/tree/main/DGCD.

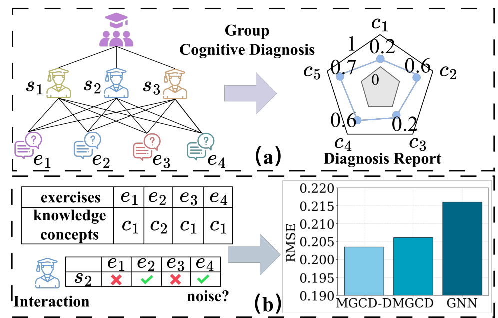

Figure 1: (a) The process of group-level cognitive diagnosis. (b) The interaction between student s_2 and exercise e_4 might be a noisy one. We eliminate similar noisy interactions and conduct the experiment, denoted as MGCD-D. Additionally, experiments are performed on the raw data employing MGCD and GNN.

1 Introduction

Cognitive diagnosis endeavors to assess students' mastery of knowledge concepts by predictively analyzing their accuracy in answering exercises, stemming from their interactions with exercises. With the advancements in deep learning [Qin *et al.*, 2024; Lin *et al.*, 2017; Wang *et al.*, 2021; Sun *et al.*, 2019], cognitive diagnostics research has received increasing attention. Abundant researches have focused on modeling the abilities of individual students and diagnosing their understanding of various concepts [De La Torre, 2011; Wu *et al.*, 2015] through methods such as Item Response Theory (IRT) [Embretson and Reise, 2013], multidimensional IRT (MIRT) [Ackerman *et al.*, 2003], matrix factorization (MF) [Ackerman *et al.*, 2003], and neural cognitive diagnosis (NeuralCD) frameworks [Wang *et al.*, 2020a]. However, students often participate in collaborative learning activities, including tests, exams, and group assignments, which offer additional benefits[Hammar Chiriac, 2014]. Therefore, evaluating group abilities is of paramount importance.

Currently, insufficient research has been undertaken in the realm of group-level cognitive diagnosis. Figure 1(a) illustrates the process of group-level cognitive diagnosis.

[*]Corresponding Authors

Traditional approaches consider the group as a unified entity and then apply individual cognitive diagnosis directly to the group, overlooking the intricate relationship between the group and the individual. The other approaches derive group competence by aggregating the competencies of individuals. For instance, the latest group diagnosis framework (MGCD) [Huang *et al.*, 2021] utilizes deep neural networks to handle the group-level cognitive diagnosis task and attention mechanisms to learn group representation. However, previous methods failed to consider higher-order connectivity among groups, students, and exercises, resulting in suboptimal outcomes. To address this limitation, we construct the group-student-exercise graph to establish higher-order connectivity among groups, students, and exercises, and employ GNN to learn higher-order representations of groups.

In the real world, the data about student-exercise interactions exhibits noise [Zhang *et al.*, 2023; Ma *et al.*, 2024]. Previous models neglected to consider the impact of this noise on the overall performance of the model, rendering them less robust. As illustrated in Figure1(b), the interaction data involving student s_2 is noisy. The RMSE for MGCD-D is less than that of MGCD, indicating the adverse influence of noise on the model's effectiveness. Meanwhile, although GNN can model higher-order relationships between groups, students, and exercises, it is susceptible to interaction noise due to its reliance on message passing mechanisms [Wu *et al.*, 2021]. As evident in Figure1(b), MGCD exhibits superior performance compared to GNN in the presence of noisy data. To enhance the higher-order representation of the group, it is imperative to mitigate the impact of noisy interactions on GNN. Furthermore, the existence of relational imbalance in the GSE graph introduces bias in different relation information during message passing, further undermining GNN's effectiveness. For instance, in the message passing, a student node has a group neighbor node and multiple exercise neighbor nodes. The relationship between student and group differs from that between student and exercise.

To this end, in this paper, we propose an adaptive denoising GNN for group-level cognitive diagnosis, namely DGCD. Specifically, we first construct the GSE graph and employ GNN to model higher-order connectivity among groups, students, and exercises. Subsequently, to address the influence of noisy interactions, we introduce an adaptive denoising method to drop the noisy edges based on the reliability of the student-exercise edges, thus alleviating the impact of the noisy edges on the learning of subsequent representations. Additionally, to tackle the bias introduced by the unbalanced relations in the GSE graph during the message passing, we propose an entropy-weighted balance module to reweight the information of unbalanced relations, balancing the information aggregation across various relations. Finally, Extensive experiments conducted on four real-world datasets demonstrate the effectiveness of our DGCD.

2 Related Work

2.1 Cognitive Diagnosis

In the domain of educational psychology, diverse cognitive diagnosis models (CDMs) have been formulated to provide comprehensive insights into students' cognitive abilities [DiBello *et al.*, 2006; Li *et al.*, 2020; Yang *et al.*, 2024; Zhang *et al.*, 2024]. Previous research mainly focuses on the individual cognitive diagnosis, with DINA [De La Torre, 2011; Junker and Sijtsma, 2001] and IRT [Embretson and Reise, 2013]standing out as notable contributions that profile students based on inherent attributes. Additionally, MIRT [Reckase, 2009], an extension of IRT, enables the representation of students' cognitive abilities through multidimensional latent traits. Moreover, certain methodologies utilize matrix factorization (MF) to reveal underlying eigenvectors of students and exercises via score matrix decomposition [Toscher and Jahrer, 2010; Thai-Nghe and Schmidt-Thieme, 2015]. Diverging from traditional approaches representing student-exercise interactions using linear functions, the neural cognitive diagnostic framework (NeuralCD) [Wang *et al.*, 2020a] employs neural networks to capture complex interactions between students and exercises. ReliCD [Zhang *et al.*, 2023] proposes a Bayesian method to explicitly estimate the state uncertainty of different concepts of knowledge for students and enable the confidence quantification of diagnostic feedback.

In recent years, group-level cognitive diagnosis has gained significant prominence across diverse domains [Liu *et al.*, 2023; Huang *et al.*, 2021; Yu *et al.*, 2024]. The initial strategy involves extending traditional CDMs and adapting them to assess group abilities. Traditional group-level cognitive diagnostic models primarily build upon Group IRT (GIRT) [Reise *et al.*, 2006; Mislevy, 1983; Birenbaum *et al.*, 2004], an extension of IRT tailored for group-level analysis. The second approach focuses on initially modeling individual abilities and subsequently assumes that a group's collective abilities are equivalent to the average proficiency of its members [Agrawal *et al.*, 2014; Liu *et al.*, 2016]. The recently introduced approach MGCD [Huang *et al.*, 2021] utilizes an attention mechanism to calculate individual student weights and aggregates student ability states into group ability states. Nevertheless, existing methods fail to consider higher-order connectivity among groups, students, and exercises, while disregarding noise in interaction data. In this paper, we construct the GSE graph to establish higher-order connectivity between groups, students, and exercises, and we propose an adaptive denoising method to alleviate the impact of noisy interactions on model performance.

2.2 GNN-based Representation Learning

GNN-based representation learning methods have gained increased attention in recent years. In contrast to conventional representation learning methods, GNN-based approaches leverage graph structures to model node representations. Current GNN-based representation learning methods fall into two main categories: homomorphic graph-based [Kipf and Welling, 2016; Hamilton *et al.*, 2017; Brody *et al.*, 2021] and heteromorphic graph-based [Schlichtkrull *et al.*, 2018; Wang *et al.*, 2019; Hu *et al.*, 2020]. Homogeneous graph-based representation learning methods do not account for the distinctions among node and edge types during the aggregation of information from neighboring nodes. Heterogeneous graph-based representation learning methods [Yang *et al.*, 2021;

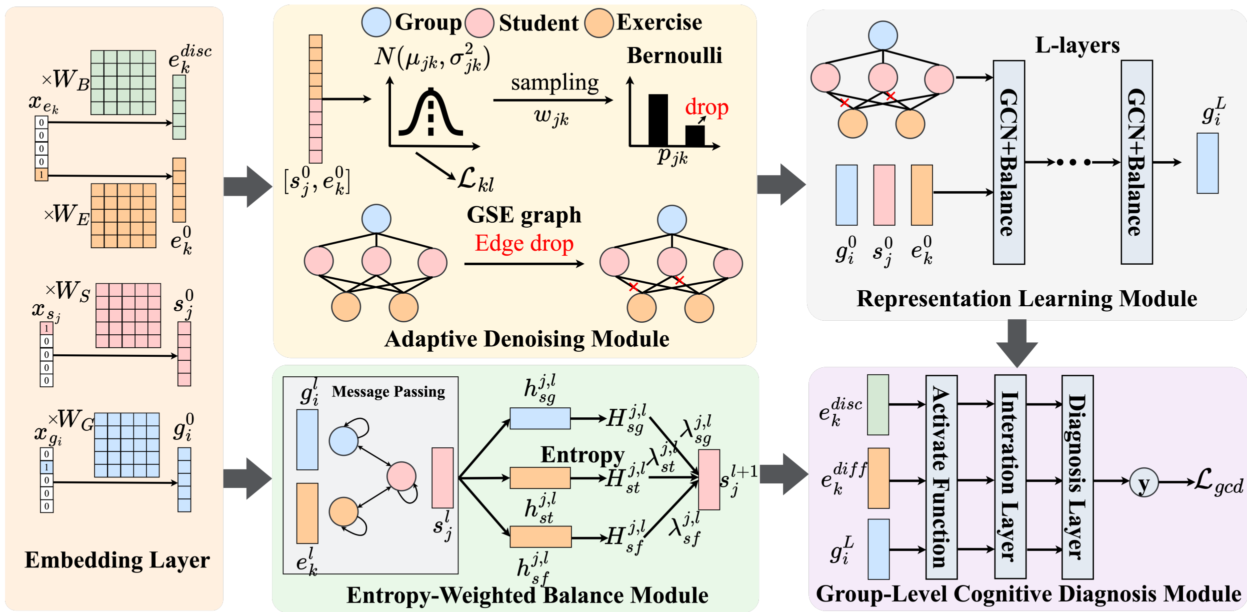

Figure 2: Overall framework of DGCD. The Adaptive Denoising Module illustrates the denoising procedure for the GSE graph, the Entropy-Weighted Balance Module showcases the information aggregation process for various relationships when updating student nodes. The Group-Level Cognitive Diagnostics Module is utilized to predict the scores of the groups in responding to the exercises.

Yang *et al.*, 2023; Liu *et al.*, 2021; Wu *et al.*, 2022] take into account various node types and relationship types during the aggregation of neighbor information, allowing for the capture of diverse and complex relationships between nodes.

Although GNN-based representation learning [Qin *et al.*, 2023; Shen *et al.*, 2021a] has demonstrated notable success, its neighbor aggregation scheme enlarges the impact of interactions on representation learning, rendering the process more susceptible to interaction noises. Furthermore, GNN may be influenced by relation unbalance. Several methods [Qiu *et al.*, 2020; You *et al.*, 2020; Wu *et al.*, 2021; Ye *et al.*, 2023] alleviate the impact of interaction noise through self-supervision, incorporating data augmentation and contrast learning to enhance node representation learning. However, these methods cannot be directly applied to the group-level cognitive diagnosis. In this paper, we introduce a novel denoising method to eliminate the noise edges in the GSE graph. Furthermore, to address the relational imbalance, we propose an information entropy-based approach that aims to balance the information of different relations in the message passing.

3 Problem Formulation

We define the group-level cognitive diagnosis task and establish the group-student-exercise graph to utilize GNN within the broader context of group-level cognitive diagnosis.

Suppose there are m students $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$, n group $\mathcal{G} = \{g_1, g_2, \ldots, g_n\}$, h exercises $\mathcal{E} = \{e_1, e_2, \ldots, e_h\}$ and t knowledge concepts $\mathcal{K} = \{k_1, k_2, \ldots, k_t\}$. The l-th group $g_l \in \mathcal{G}$ consists of a set of students, i.e., group members with student indexes $\mathcal{K}_l = \{k_{l,1}, k_{l,2}, \ldots, k_{l,|g_l|}\}$, where $s_{k_{l,*}} \in \mathcal{S}$ and $|g_l|$ is the size of the group. In addition, there is an exercise-concept correlation matrix $\mathbf{Q} = \{Q_{ij}\}_{m\times n}$. If exercise e_i requires knowledge concept k_j then Q_{ij} = 1. We employ a set of triplet (g_i, e_j, y_{ij}) to denote the response logs H, where $g_i \in \mathcal{G}$, $e_j \in \mathcal{E}$ and y_{ij} is the correct rate that group g_i got on exercise e_j.

Group-Student-Exercise Graph To establish higher-order connectivity between groups, students, and exercises, we construct the group-student-exercise (GSE) graph, denoted as G_r = (V, R). Here, $V = \{V_g, V_s, V_e\}$ represents the set of three node types, with $V_g \subseteq \mathcal{G}$, $V_s \subseteq \mathcal{S}$, and $V_e \subseteq \mathcal{E}$, where $\mathcal{G}$, $\mathcal{S}$, and $\mathcal{E}$ denote the sets of groups, students, and exercises, respectively. The set of relationships among the nodes is denoted as $R = \{r_1, r_2, r_3\}$, where r_1 indicates the inclusion relationship between groups and students, r_2 and r_3 denote the correct and incorrect relationships between students and exercises. Each graph comprises a group node, all student nodes within the group, and the exercise nodes the group has interacted with. We specify that edges between nodes are categorized into different types based on distinct relationships.

Problem Definition. *Given group-exercise response records H, student-exercise response records F (the student-exercise response records in F here are the student-exercise interaction records contained in the group-exercise interaction records H), group-student-exercise (GSE) graphs, and exercise-concept correlation matrix Q, our task aims to utilize the GSE graphs to conduct group-level cognitive diagnosis and obtain more accurate and robust diagnostic results.*

4 Methodology

This paper introduces an adaptive denoising graph neural network approach for group-level cognitive diagnosis. The following sections provide a concise overview of our method and a detailed presentation of individual modules.

Figure 2 illustrates our method comprising four primary modules: the representation learning module, the adaptive denoising module, the entropy-weighted balance module, and the group-level cognitive diagnosis module. The representation learning module captures higher-order connectivity among groups, students, and exercises. The adaptive denoising module seeks to alleviate the impact of noisy interactions. The entropy-weighted balance module aims to equalize information aggregation across various relationships and the group-level cognitive diagnosis module is employed to predict scores for interactions between the group and the exercises. The following will introduce their details.

4.1 Representation Learning Module

In practical settings, students within a group collaboratively interact with exercises, leading to intricate relationships among groups, students, and exercises. To capture intricate higher-order connectivity, we construct the group-student-exercise graph. Each group, along with its associated students and exercises they interact with collectively, serves as nodes, while the connections between them represent the edges. We then utilize a graph neural network to learn the higher-order representation of the group. First, this embedding layer takes the group one-hot vector $\boldsymbol{x}_{g_i}$, student one-hot vector $\boldsymbol{x}_{s_j}$ and exercise one-hot vector $\boldsymbol{x}_{e_k}$ as its inputs, and obtains the corresponding initial embedding vectors $\boldsymbol{g}_i^0 \in \mathbb{R}^{1\times d}$, $\boldsymbol{s}_j^0 \in \mathbb{R}^{1\times d}$ and $\boldsymbol{e}_k^0 \in \mathbb{R}^{1\times d}$ as follows:

$$\boldsymbol{g}_i^0 = \boldsymbol{x}_{g_i}\boldsymbol{W}_G, \boldsymbol{s}_j^0 = \boldsymbol{x}_{s_j}\boldsymbol{W}_S, \boldsymbol{e}_k^0 = \boldsymbol{x}_{e_k}\boldsymbol{W}_E, \tag{1}$$

where $\boldsymbol{W}_G \in \mathbb{R}^{n\times d}$, $\boldsymbol{W}_S \in \mathbb{R}^{m\times d}$, and $\boldsymbol{W}_E \in \mathbb{R}^{h\times d}$ are trainable matrices, n, m, and h denote the number of groups, students, and exercises, while d denotes the embedding size.

Afterward, we categorize groups, students, and exercises as three distinct types of nodes, classifying the edges between them into three different types based on their respective relationships. Then, we aggregate information from neighboring nodes based on different node types and relationship edges, respectively. We denote the embedding of node i (it can denote either a group node, a student node, or an exercise node.) at the l-th GNN layer as $\boldsymbol{o}_i^l$. We formally define the message passing process from the $(l-1)$-th layer to the l-th layer as:

$$\boldsymbol{o}_i^l = \boldsymbol{W}_o^{l-1}\boldsymbol{o}_i^{l-1} + \sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_r(i)} \boldsymbol{W}_r^{l-1}\boldsymbol{o}_j^{l-1}, \tag{2}$$

where $\boldsymbol{W}_o^{l-1}$ and $\boldsymbol{W}_r^{l-1} \in \mathbb{R}^{d\times d}$ are trainable matrices, $\mathcal{R}$ denotes the set of relationships between groups, students, and exercises. $\mathcal{N}_r(i)$ denotes the set of neighboring nodes that have relationship r with node i. After stacking L GNN layers, the layer-aggregation mechanism is adapted to generate the final representation $\boldsymbol{o}_i^L$.

4.2 Adaptive Denoising Module

In the GSE graph, the relationship between groups and students is fixed, resulting in no noise in the group-student edge. However, noise is present in the edges between students and exercises. This noise is amplified during the message passing of the GNN, subsequently affecting the learning of the representation. To mitigate the impact of noise on representation learning, we propose an adaptive denoising strategy. This strategy aims to model the reliability of student-exercise edges and calculate the sampling probability p for each edge which denotes the retention probability of an edge.

Specially, we model that the reliability of edge v_{jk} between student s_j and exercise e_k as a Gaussian distribution. To obtain the unique reliability distribution for edge v_{jk}, We concatenate the vector representations of student s_j and exercise e_k and subsequently apply distinct transformation matrices to acquire the mean and variance parameters, respectively, i.e.,

$$\boldsymbol{\mu}_{jk} = \boldsymbol{W}_\mu([\boldsymbol{s}_j^0, \boldsymbol{e}_k^0]), log\boldsymbol{\sigma}_{jk}^2 = \boldsymbol{W}_\sigma([\boldsymbol{s}_j^0, \boldsymbol{e}_k^0]), \tag{3}$$

$$q(z_{jk}|s_j, e_k) = \mathcal{N}(\boldsymbol{\mu}_{jk}, \boldsymbol{\sigma}_{jk}^2), \tag{4}$$

where $\boldsymbol{\mu}_{jk}$ and $\boldsymbol{\sigma}_{jk} \in \mathbb{R}^1$ represent mean and variance parameters for the reliability of edge v_{jk}, respectively. $\boldsymbol{W}_\mu$ and $\boldsymbol{W}_\sigma \in \mathbb{R}^{1\times 2d}$ are different trainable matrices.

We proceed to sample the reliability distribution of edge v_{jk} to obtain the reliability w_{jk} of edge v_{jk}. To facilitate gradient backpropagation during training, we employ the resampling technique as follows:

$$w_{jk} = \mu_{jk} + \epsilon \odot \sigma_{jk}, \tag{5}$$

where random variable $\epsilon \sim \mathcal{N}(0, 1)$. A larger w_{jk} indicates that the edge v_{jk} is more important and should be retained, while a smaller w_{jk} suggests that the edge v_{jk} is more likely to be a noisy edge. Therefore, we drop the edge v_{jk} based on w_{jk}. To make the edge dropping procedure differentiable and enable an end-to-end optimization process, we apply a Relaxed Bernoulli distribution to obtain edge retention probability p_{jk} through sampling. Specifically, the probability p_{jk} is calculated by:

$$p_{jk} = sigmoid(\frac{1}{t}(log(\frac{\rho(w_{jk})}{1-\rho(w_{jk})}) + log(\frac{u}{1-u}))), \tag{6}$$

where ρ is a sigmoid activation function, t is temperature parameter of relaxation, and $u \sim$ Unif[0,1]. Then, we can obtain a new denoising view based on the retention probabilities of the edges. In the denoising view, the student-exercise correct answer adjacency matrix A_t^d and the student-exercise incorrect answer adjacency matrix A_f^d as follows:

$$\begin{cases} a_{jk}^t = p_{jk}, if \;\; s_j \;\; correctly \;\; answered \;\; e_k \\ a_{jk}^f = p_{jk}, if \;\; s_j \;\; incorrectly \;\; answered \;\; e_k \end{cases}, \tag{7}$$

where a_{jk}^t, a_{jk}^f are the elements of A_t^d and A_f^d, respectively.

4.3 Entropy-Weighted Balance Module

The number of neighboring nodes varies for nodes in the GSE Graph across different relationships. Unbalanced message passing can adversely affect the update of node representations. For instance, each student has only one group node neighbor and multiple exercise node neighbors, diminishing the influence of the group's representation in updating the student's representation during message passing. Hence, it is essential to balance information from different relations during the message passing.

Specifically, we employ the entropy weighting approach to balance information from different relationships by reweighting the information during the message passing. First, we normalize the aggregation of neighboring nodes for different relationships, and subsequently, we calculate the information entropy corresponding to different relationships using Eq.(9). The specific process is as follows:

$$h_{st}^{j,l} = \sum_{k \in N_e^t} \frac{a_{jk}^t}{d_j^{st} d_k^{et}} W_{r_2}^{l-1} e_k^{l-1}, \tag{8}$$

where $h_{st}^{j,l}$ represents the normalized embedding regarding the exercise neighbors correctly answered by student s_j in the l-th GNN layer. d_j^{st} is the sum of the j-th row of A_t^d,and d_k^{et} is the sum the k-th column of A_t^d. Similarly we can obtain $h_{sf}^{j,l}$, $h_{sg}^{j,l}$, $h_{et}^{k,l}$ and $h_{ef}^{k,l}$. After normalization, we computed the information entropy for each relation:

$$F_H = -\frac{1}{\sqrt{d}} \sum_i h_i \log_2(h_i), \tag{9}$$

$$H_{st}^{j,l} = F_H(h_{st}^{j,l}), \tag{10}$$

where h_i denotes the value of the i-th position in the vector h, and d is the dimension of the vector h. $H_{st}^{j,l}$ is the information entropy of $h_{st}^{j,l}$. Similarly we can obtain $H_{sf}^{j,l}$, $H_{sg}^{j,l}$, $H_{et}^{k,l}$ and $H_{ef}^{k,l}$. A higher information entropy indicates a smaller amount of information. Therefore, to balance the information from different relationships, we define the weights of information corresponding to different relationships as:

$$\lambda_{st}^{j,l} = \frac{H_{st}^{j,l}}{H_{st}^{j,l} + H_{sf}^{j,l} + H_{sg}^{j,l}}. \tag{11}$$

Then we can get $\lambda_{sf}^{j,l}$, $\lambda_{sg}^{j,l}$, $\lambda_{et}^{k,l}$ and $\lambda_{ef}^{k,l}$. Using the weights derived from information entropy, we redefine the message passing as follows:

$$\boldsymbol{g}_i^l = \boldsymbol{W}_g^{l-1} \boldsymbol{g}_i^{l-1} + \sum_{j \in N_i^s} \frac{1}{|N_i^s|} \boldsymbol{W}_{r_1}^{l-1} \boldsymbol{s}_j^{l-1}, \tag{12}$$

$$\begin{aligned} \boldsymbol{s}_j^l = {} & \boldsymbol{W}_s^{l-1} \boldsymbol{s}_j^{l-1} + \sum_{k \in N_j^{st}} \frac{\lambda_{st}^{j,l} a_{jk}^t}{d_j^{st} d_k^{et}} \boldsymbol{W}_{r_2}^{l-1} \boldsymbol{e}_k^{l-1} \\ & + \sum_{h \in N_j^{sf}} \frac{\lambda_{sf}^{j,l} a_{jh}^f}{d_j^{sf} d_h^{ef}} \boldsymbol{W}_{r_3}^{l-1} \boldsymbol{e}_h^{l-1} + \frac{\lambda_{sg}^{j,l}}{|N_i^s|} \boldsymbol{W}_{r_1}^{l-1} \boldsymbol{g}_i^{l-1}, \end{aligned} \tag{13}$$

$$\begin{aligned} \boldsymbol{e}_k^l = {} & \boldsymbol{W}_e^{l-1} \boldsymbol{e}_k^{l-1} + \sum_{j \in N_k^{et}} \frac{\lambda_{et}^{k,l} a_{jk}^t}{d_j^{st} d_k^{et}} \boldsymbol{W}_{r_2}^{l-1} \boldsymbol{s}_j^{l-1} \\ & + \sum_{q \in N_k^{ef}} \frac{\lambda_{ef}^{k,l} a_{qk}^f}{d_q^{sf} d_k^{ef}} \boldsymbol{W}_{r_3}^{l-1} \boldsymbol{s}_q^{l-1}, \end{aligned} \tag{14}$$

where N_i^s denotes the set of students in group g_i. N_j^{st} and N_j^{sf} represent the sets of exercises that student s_j answered correctly and incorrectly, respectively. N_k^{et} and N_k^{ef} represent the sets of students that answered exercise e_k correctly and incorrectly, respectively.

Dataset	ASSIST2012	NIPS_Edu	SLPbio	SLPmath
# Students	1,802	2,113	3,922	4,152
# Groups	114	150	145	153
# Exercises	707	688	120	138
# Knowledge concepts	122	77	21	39
# AVG. group size	19.16	14.78	27.05	27.14
# AVG. responses for a group	7.48	31.96	70.49	72.56

Table 1: Statistics of all datasets.

4.4 Group-Level Cognitive Diagnosis Module

After obtaining a group's representation $\boldsymbol{g}_i^L$ through the L-layer of GNNs, we can diagnose the cognitive state of the group through the diagnosis layer. Here, we utilize a multi-layer perceptron (MLP) as a diagnostic layer. The first layer of the interaction layer is represented as:

$$\boldsymbol{e}_k^{disc} = \phi\left(\boldsymbol{x}_{e_k} \boldsymbol{W}_B\right), \boldsymbol{e}_k^{diff} = \phi\left(\boldsymbol{e}_k^0\right), \boldsymbol{g}_i^{state} = \phi\left(\boldsymbol{g}_i^L\right), \tag{15}$$

$$\boldsymbol{x}_g = \boldsymbol{Q}_e \left(\boldsymbol{g}_i^{state} - \boldsymbol{e}_k^{diff}\right) \boldsymbol{e}_k^{disc}, \tag{16}$$

where $\boldsymbol{W}_B \in \mathbb{R}^{h \times d}$ is a trainable matrix, the $\boldsymbol{e}_k^{disc}$ is the differentiation of exercises, $\boldsymbol{Q}_e$ is a matrix of knowledge concepts, and ϕ is an activate function.

Then we can obtain the correct rate of the group g_i's answer on exercise e_k through multiple linear layers:

$$\begin{cases} \boldsymbol{z}_1 = \phi\left(\boldsymbol{W}_1 \boldsymbol{x}_g + b_1\right) \\ \boldsymbol{z}_2 = \phi\left(\boldsymbol{W}_2 \boldsymbol{z}_1 + b_2\right) \\ \qquad \cdots \\ \boldsymbol{z}_h = \phi\left(\boldsymbol{W}_h \boldsymbol{z}_{h-1} + b_h\right) \end{cases}. \tag{17}$$

Finally, we can obtain the prediction score $\hat{y}_{ik}$ via:

$$\hat{y}_{ik} = \phi\left(\boldsymbol{W}_{h+1} \boldsymbol{z}_h + b_{h+1}\right). \tag{18}$$

To optimize the model parameters and obtain the ability level of all groups, we maximize the likelihood $p(Y;\Theta)$, which indicates the correct probability that all groups answer the exercises. We optimize the parameters with the evidence lower bound (ELBO):

$$logp(Y;\Theta) \geq \mathbb{E}_{q(w|X)}[logp(Y|w)] - KL(q(w|X)||p(w)), \tag{19}$$

where $p(w)$ is the prior distribution for the reliability of the edge. $q(w|X)$ is the posterior distribution we constructed for the edge. X is the set of representations, $X = \{S, E\}$. $logp(Y|w)$ measures the likelihood that groups answer exercises correctly based on the reliability w of the student-exercise edge. We follow the variational autoencoder (VAE) [Kingma and Welling, 2014; Zha *et al.*, 2023; Qin *et al.*, 2022; Shen *et al.*, 2021b] and calculate the two terms in Equation (19) separately:

$$\mathcal{L}_{gcd} = \mathbb{E}_{q(w|X)}[logp(Y|w)] = \sum_{i=1}^{n} \sum_{k \in H_i^e} \left(\hat{y}_{ik} - y_{ik}\right)^2, \tag{20}$$

$$\begin{aligned} \mathcal{L}_{kl} &= KL(q(w|X)||p(w)) \\ &= \sum_{i=1}^{n} \sum_{j \in H_i^s} \sum_{k \in H_i^e} \frac{1}{2}(\mu_{jk}^2 + \sigma_{jk}^2 - ln\sigma_{jk}^2 - 1), \end{aligned} \tag{21}$$

Model	ASSIST2012		NIPS_Edu		SLPbio		SLPmath	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
IRT	$0.3214_{\pm 0.0029}$	$0.2772_{\pm 0.0035}$	$0.2414_{\pm 0.0011}$	$0.1919_{\pm 0.0012}$	$0.1336_{\pm 0.0035}$	$0.1016_{\pm 0.0024}$	$0.1316_{\pm 0.0067}$	$0.0953_{\pm 0.0059}$
MIRT	$0.2304_{\pm 0.0012}$	$0.1827_{\pm 0.0009}$	$0.2145_{\pm 0.0020}$	$0.1674_{\pm 0.0016}$	$0.1250_{\pm 0.0002}$	$0.0953_{\pm 0.0003}$	$0.1261_{\pm 0.0002}$	$0.0911_{\pm 0.0002}$
NCD	$0.2074_{\pm 0.0012}$	$0.1531_{\pm 0.0009}$	$0.2036_{\pm 0.0004}$	$0.1568_{\pm 0.0011}$	$0.1264_{\pm 0.0008}$	$0.0955_{\pm 0.0007}$	$0.1320_{\pm 0.0010}$	$0.0936_{\pm 0.0004}$
ReliCD	$0.2088_{\pm 0.0011}$	$0.1561_{\pm 0.0019}$	$\underline{0.1997}_{\pm 0.0005}$	$0.1548_{\pm 0.0007}$	$0.1278_{\pm 0.0007}$	$0.0959_{\pm 0.0006}$	$0.1308_{\pm 0.0006}$	$0.0919_{\pm 0.0006}$
MGCD	$\underline{0.2061}_{\pm 0.0012}$	$\underline{0.1524}_{\pm 0.0016}$	$0.2022_{\pm 0.0017}$	$0.1554_{\pm 0.0022}$	$\underline{0.1212}_{\pm 0.0012}$	$\underline{0.0927}_{\pm 0.0006}$	$0.1192_{\pm 0.0013}$	$0.0856_{\pm 0.0007}$
RGCN	$0.2145_{\pm 0.0005}$	$0.1622_{\pm 0.0027}$	$0.2035_{\pm 0.0011}$	$0.1565_{\pm 0.0011}$	$0.1260_{\pm 0.0011}$	$0.0957_{\pm 0.0010}$	$0.1242_{\pm 0.0005}$	$0.0893_{\pm 0.0006}$
GATv2	$0.2123_{\pm 0.0046}$	$0.1594_{\pm 0.0018}$	$0.2026_{\pm 0.0018}$	$\underline{0.1542}_{\pm 0.0014}$	$0.1236_{\pm 0.0026}$	$0.0940_{\pm 0.0023}$	$\underline{0.1153}_{\pm 0.0005}$	$\underline{0.0829}_{\pm 0.0014}$
SGL	$0.2072_{\pm 0.0011}$	$0.1578_{\pm 0.0021}$	$0.2092_{\pm 0.0006}$	$0.1582_{\pm 0.0008}$	$0.1320_{\pm 0.0006}$	$0.1004_{\pm 0.0004}$	$0.1384_{\pm 0.0011}$	$0.0992_{\pm 0.0009}$
RocSE	$0.2076_{\pm 0.0012}$	$0.1543_{\pm 0.0016}$	$0.2066_{\pm 0.0005}$	$0.1555_{\pm 0.0008}$	$0.1312_{\pm 0.0009}$	$0.1005_{\pm 0.0007}$	$0.1342_{\pm 0.0005}$	$0.0959_{\pm 0.0004}$
DGCD	$\mathbf{0.1977_{\pm 0.0023}}$	$\mathbf{0.1437_{\pm 0.0022}}$	$\mathbf{0.1929_{\pm 0.0009}}$	$\mathbf{0.1495_{\pm 0.0014}}$	$\mathbf{0.1127_{\pm 0.0007}}$	$\mathbf{0.0859_{\pm 0.0005}}$	$\mathbf{0.1072_{\pm 0.0010}}$	$\mathbf{0.0762_{\pm 0.0010}}$

Table 2: Experimental results on group performance prediction. The **bold** indicates the best result and the second-best results are underlined. We conduct five experiments and compute the average to derive the final result.

where n represents the number of group, and H_i^s, H_i^e denote all students of group g_i and exercises responded by group g_i.

Finally, we define the final loss function as follows:

$$\mathcal{L} = \mathcal{L}_{gcd} + \lambda_{kl}\mathcal{L}_{kl} + \lambda_{reg}\|\Theta\|_2^2, \tag{22}$$

where Θ represents all the learnable parameters, and λ_{kl}, λ_{reg} are hyper-parameters controlling the effect strength of $\mathcal{L}_{kl}$ and the regularization.

5 Experiments

5.1 Dataset Description

We conduct the experiments on four public education benchmarks, including ASSIST12 [Feng *et al.*, 2009], NIPS_Edu [Wang *et al.*, 2020b], SLPbio [Lu *et al.*, 2021], and SLPmath [Lu *et al.*, 2021]. All of these datasets contain group labels, with students from the same group belonging to the same class. Each dataset contains common interaction records for groups: group-exercise responses. Specifically, for each group-exercise response, we calculated the percentage of students in the group who answered the exercise correctly as the response outcome, and we used the student responses from the group-exercise response as the student-exercise responses. Detailed statistics of these datasets are presented in Table 1.

5.2 Experimental Setup

To verify the effectiveness of our DGCD, each dataset of group-exercise responses is divided randomly into two subsets: 80% for training and 20% for testing. As group-level cognitive diagnosis is a regression task to predict the correct rate of group responses to exercises, we selected root mean square error (RMSE) and the mean absolute error (MAE), to evaluate model performance.

In our DGCD, we set the dimension d of the vector to be the number of knowledge concepts. The number of the GNN layers in the representation learning module is set to 2. The number of diagnostic layers is set to 3. Additionally, the hyper-parameter λ_{kl} was searched in [1,1e-1,1e-2,1e-3,1e-4,1e-5,1e-6,1e-7]. And, t was optimized over the values [0.57,0.67,0.77,0.87].

5.3 Baseline Approaches

To validate the effectiveness of DGCD, we compare it against two baseline types. Initially, we assess its performance in contrast to cognitive diagnosis-based approaches as follows: **IRT** [Embretson and Reise, 2013], **MIRT** [Reckase, 2009], **NeuralCDM** [Wang *et al.*, 2020a], **MGCD** [Huang *et al.*, 2021] and **Relicd** [Zhang *et al.*, 2023]. Then, we compared the graph-based approaches as follows: **RGCN** [Schlichtkrull *et al.*, 2018], **GATv2** [Brody *et al.*, 2021], **SGL** [Wu *et al.*, 2021] and **RocSE** [Ye *et al.*, 2023].

5.4 Performance Comparison

To validate the effectiveness of DGCD, we conducted a comparative analysis utilizing the above four datasets. Table 2 presents the experimental results of our DGCD and the baseline models. We can find that our model consistently surpasses all baseline models by a substantial margin. Compared with IRT, MIRT, NCD ReliCD, and MGCD, our model shows better performance. This implies that the modeling of high-order connectivity among groups, students, and exercises using graph neural networks results in a more accurate group ability level. Our model exhibits noteworthy improvement across all four datasets compared to GAT, RGCN, SGL, and RocSE. This implies that our model outperforms in identifying noisy edges in student-exercise interactions, facilitated by the adaptive denoising module. Furthermore, in terms of the reliability of randomly deleting edges (for SGL) and calculating edges based on similarity (for RocSE), our approach demonstrates greater flexibility and adaptation to diagnostic tasks. Additionally, our entropy-weighted balance module ensures the equilibrium of information related to distinct relationships during aggregation. This mitigates bias in the aggregated information resulting from variations in the number of neighbors associated with different relationships.

5.5 Ablation Analysis

To study the impact of our key components, we introduced different variants of DGCD from two perspectives and analyzed their effects. “w/o AD” and “w/o EB” means removing the adaptive denoising module and the entropy-weighted balance module, respectively. As shown in Figure 3, our DGCD consistently achieves the best performance, demonstrating the contribution of each component. Meanwhile, it

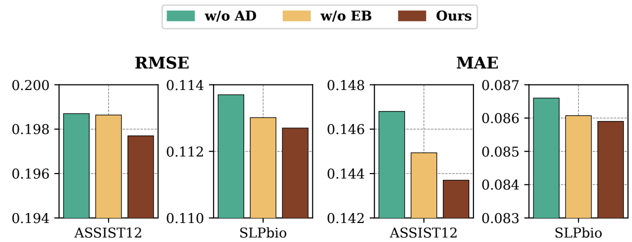

Figure 3: Comparison of DCGD and its ablated variants.

is evident that the adaptive denoising module significantly enhances model performance. This improvement can be attributed to the adaptive module's ability to eliminate noise edges, thereby mitigating the impact of noise on the model.

5.6 Effect of Hyperparameters

We study the impact of the hyperparameters t and λ_{kl} in Figure 4. t is a temperature parameter that affects the loose Bernoulli sampling process of the edge dropout. λ_{kl} is the weight of the $\mathcal{L}_{kl}$ that affects the modeling of student-exercise edge reliability. It has been observed that varying values of t yield different outcomes. Specifically, the model achieves optimal performance on ASSIT12 when t is equal to 0.67 and on SLPbio when t is equal to 0.77. Notably, different datasets are associated with distinct values of t. Similarly, optimal performance is observed for the model on ASSIST12 when λ_{kl} is set to 1e-3 and on SLPbio when λ_{kl} is set to 1e-6.

5.7 Case Study

To assess our model's capability against noise, we systematically introduced noisy data by replacing a certain percentage (5%, 10%) of population-exercise interactions in the training set, while keeping the test set unchanged. Figure 5(a) shows the result on the ASSIST12 dataset. We observed that the performance of MGCD, RocSE, and DGCD is influenced as the percentage of noise increases. In particular, DGCD consistently surpasses MGCD and RocSE, indicating its proficiency in mitigating the impact of noise interactions through adaptive denoising. Additionally, after introducing 10% noise to the ASSIST12 dataset, the performance of DGCD exceeds that of MGCD on the noise-free dataset. This further supports the superior performance and robustness of our DGCD in comparison to MGCD.

Meanwhile, we analyzed the retention probability of the edges learned by the adaptive denoising module on ASSIST12. As shown in Figure 5(b), student s_{547} answered incorrectly on exercises e_7 and e_{22} but correctly on e_{20}. Exercises e_7, e_{20}, and e_{22} share identical knowledge concepts and difficulty levels. Consequently, the interaction between s_{547} and e_{20} is deemed potentially noisy data, with our model learning a retention probability of 0.2956 for the edges connecting s_{547} and e_{20}. This value is lower than the retention probability for the edges of s_{547} with e_7 and e_{22}. This observation demonstrates our model's ability to identify noisy edges and mitigate their impact on representation learning.

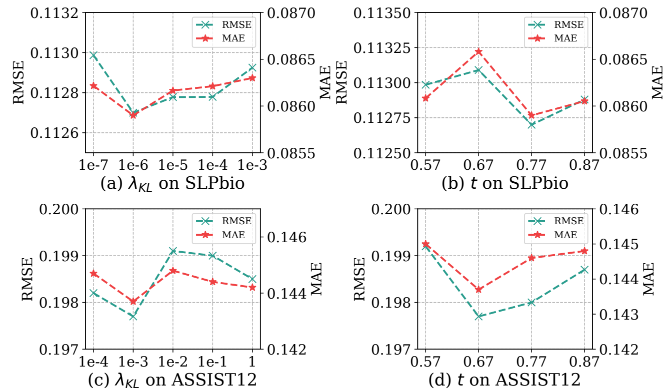

Figure 4: Hyper-parameter sensitivity analysis.

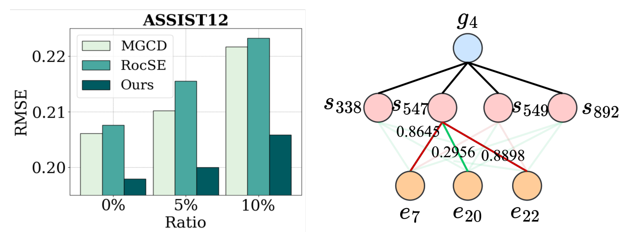

Figure 5: Left: Model performance w.r.t. noise ratio. Right: Retention probability of edges. Red edges symbolize incorrect answers, while green edges denote correct answers.

6 Conclusion

In this paper, we introduced a denoising graph neural network approach for group-level cognitive diagnosis, namely DGCD. The method considers higher-order connectivity among groups, students, and exercises, effectively alleviating the impact of noisy interactions on diagnostic results. To be specific, We initially constructed the group-student-exercise graph and subsequently utilized Graph Neural Networks to capture intricate higher-order connectivity among groups, students, and exercises, thereby obtaining advanced representations of the groups. Then, we introduced an adaptive denoising module to address the impact of noise interactions on diagnostic results. This module selectively removes noisy edges based on their reliability, resulting in a denoised view for subsequent representation learning. Furthermore, we designed an entropy-weighted balance module that utilizes information entropy reweighting to alleviate information bias resulting from the relational imbalance in the message passing. Finally, we conducted extensive experiments on four public education benchmarks and the experimental results demonstrate the effectiveness of our DGCD.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62107001, 62302009 and 62202006.

References

[Ackerman *et al.*, 2003] Terry A Ackerman, Mark J Gierl, and Cindy M Walker. Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22(3):37–51, 2003.

[Agrawal *et al.*, 2014] Rakesh Agrawal, Behzad Golshan, and Evimaria Terzi. Grouping students in educational settings. In *ACM SIGKDD*, pages 1017–1026, 2014.

[Birenbaum *et al.*, 2004] Menucha Birenbaum, Curtis Tatsuoka, and Tomoko Yamada. Diagnostic assessment in timss-r: Between-countries and within-country comparisons of eighth graders' mathematics performance. *Studies in Educational Evaluation*, 30(2):151–173, 2004.

[Brody *et al.*, 2021] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *ICLR*, 2021.

[De La Torre, 2011] Jimmy De La Torre. The generalized dina model framework. *Psychometrika*, 76:179–199, 2011.

[DiBello *et al.*, 2006] Louis V DiBello, Louis A Roussos, and William Stout. 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics*, 26:979–1030, 2006.

[Embretson and Reise, 2013] Susan E Embretson and Steven P Reise. *Item response theory*. Psychology Press, 2013.

[Feng *et al.*, 2009] Mingyu Feng, Neil Heffernan, and Kenneth Koedinger. Addressing the assessment challenge with an online system that tutors as it assesses. *User modeling and user-adapted interaction*, 19(3):243–266, 2009.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[Hammar Chiriac, 2014] Eva Hammar Chiriac. Group work as an incentive for learning–students' experiences of group work. *Frontiers in psychology*, 5:558, 2014.

[Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *WWW*, pages 2704–2710, 2020.

[Huang *et al.*, 2021] Jie Huang, Qi Liu, Fei Wang, Zhenya Huang, Songtao Fang, Runze Wu, Enhong Chen, Yu Su, and Shijin Wang. Group-level cognitive diagnosis: A multi-task learning perspective. In *IEEE ICDM*, pages 210–219. IEEE, 2021.

[Junker and Sijtsma, 2001] Brian W Junker and Klaas Sijtsma. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272, 2001.

[Kingma and Welling, 2014] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *stat*, 1050:1, 2014.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2016.

[Li *et al.*, 2020] Xiaomin Li, Wen-Chung Wang, and Qin Xie. Cognitive diagnostic models for rater effects. *Frontiers in psychology*, 11:525, 2020.

[Lin *et al.*, 2017] Hao Lin, Hengshu Zhu, Yuan Zuo, Chen Zhu, Junjie Wu, and Hui Xiong. Collaborative company profiling: Insights from an employee's perspective. In *AAAI*, volume 31, 2017.

[Liu *et al.*, 2016] Yuping Liu, Qi Liu, Runze Wu, Enhong Chen, Yu Su, Zhigang Chen, and Guoping Hu. Collaborative learning team formation: a cognitive modeling perspective. In *DASFAA*, pages 383–400. Springer, 2016.

[Liu *et al.*, 2021] Mengfan Liu, Pengyang Shao, and Kun Zhang. Graph-based exercise-and knowledge-aware learning network for student performance prediction. In *Artificial Intelligence: First CAAI International Conference, CICAI 2021, Hangzhou, China, June 5–6, 2021, Proceedings, Part I 1*, pages 27–38. Springer, 2021.

[Liu *et al.*, 2023] Shuhuan Liu, Xiaoshan Yu, Haiping Ma, Ziwen Wang, Chuan Qin, and Xingyi Zhang. Homogeneous cohort-aware group cognitive diagnosis: A multi-grained modeling perspective. In *CIKM*, pages 4094–4098, 2023.

[Lu *et al.*, 2021] Yu Lu, Yang Pian, Ziding SHEN, Penghe CHEN, and Xiaoqing Li. Slp: A multi-dimensional and consecutive dataset from k-12 education. In *Proceedings of the 29th International Conference on Computers in Education*, pages 261–266, 2021.

[Ma *et al.*, 2024] Haiping Ma, Yong Yang, Chuan Qin, Xiaoshan Yu, Shangshang Yang, Xingyi Zhang, and Hengshu Zhu. Hd-kt: Advancing robust knowledge tracing via anomalous learning interaction detection. In *WWW*, 2024.

[Mislevy, 1983] Robert J Mislevy. Item response models for grouped data. *Journal of Educational Statistics*, 8(4):271–288, 1983.

[Qin *et al.*, 2022] Chuan Qin, Kaichun Yao, Hengshu Zhu, Tong Xu, Dazhong Shen, Enhong Chen, and Hui Xiong. Towards automatic job description generation with capability-aware neural networks. *IEEE TKDE*, 35(5):5341–5355, 2022.

[Qin *et al.*, 2023] Chuan Qin, Hengshu Zhu, Dazhong Shen, Ying Sun, Kaichun Yao, Peng Wang, and Hui Xiong. Automatic skill-oriented question generation and recommendation for intelligent job interviews. *ACM TOIS*, 42(1):1–32, 2023.

[Qin *et al.*, 2024] Chuan Qin, Le Zhang, Yihang Cheng, Rui Zha, Dazhong Shen, Qi Zhang, Xi Chen, Ying Sun, Chen Zhu, Hengshu Zhu, and Hui Xiong. A comprehensive survey of artificial intelligence techniques for talent analytics. *arXiv preprint arXiv:2307.03195v2*, 2024.

[Qiu *et al.*, 2020] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD*, pages 1150–1160, 2020.

[Reckase, 2009] Mark D. Reckase. *Multidimensional Item Response Theory*. Springer New York, 2009.

[Reise *et al.*, 2006] Steven P Reise, Rob R Meijer, Andrew T Ainsworth, Leo S Morales, and Ron D Hays. Application of group-level item response models in the evaluation of consumer reports about health plan quality. *Multivariate behavioral research*, 41(1):85–102, 2006.

[Schlichtkrull *et al.*, 2018] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *ESWC*, pages 593–607. Springer, 2018.

[Shen *et al.*, 2021a] Dazhong Shen, Chuan Qin, Chao Wang, Zheng Dong, Hengshu Zhu, and Hui Xiong. Topic modeling revisited: A document graph-based neural network perspective. *NeurIPS*, 34:14681–14693, 2021.

[Shen *et al.*, 2021b] Dazhong Shen, Chuan Qin, Chao Wang, Hengshu Zhu, Enhong Chen, and Hui Xiong. Regularizing variational autoencoder with diversity and uncertainty awareness. In *IJCAI*, pages 2964–2970. International Joint Conferences on Artificial Intelligence, 2021.

[Sun *et al.*, 2019] Ying Sun, Fuzhen Zhuang, Hengshu Zhu, Xin Song, Qing He, and Hui Xiong. The impact of person-organization fit on talent management: A structure-aware convolutional neural network approach. In *Proceedings of the 25th ACM SIGKDD*, pages 1625–1633, 2019.

[Thai-Nghe and Schmidt-Thieme, 2015] Nguyen Thai-Nghe and Lars Schmidt-Thieme. Multi-relational factorization models for student modeling in intelligent tutoring systems. In *KSE*, pages 61–66. IEEE, 2015.

[Toscher and Jahrer, 2010] Andreas Toscher and Michael Jahrer. Collaborative filtering applied to educational data mining. *KDD cup*, 2010.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *WWW*, pages 2022–2032, 2019.

[Wang *et al.*, 2020a] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. Neural cognitive diagnosis for intelligent education systems. In *AAAI*, volume 34, pages 6153–6161, 2020.

[Wang *et al.*, 2020b] Zichao Wang, Angus Lamb, Evgeny Saveliev, Pashmina Cameron, Yordan Zaykov, José Miguel Hernández-Lobato, Richard E Turner, Richard G Baraniuk, Craig Barton, Simon Peyton Jones, et al. Instructions and guide for diagnostic questions: The neurips 2020 education challenge. *arXiv preprint arXiv:2007.12061*, 2020.

[Wang *et al.*, 2021] Chao Wang, Hengshu Zhu, Peng Wang, Chen Zhu, Xi Zhang, Enhong Chen, and Hui Xiong. Personalized and explainable employee training course recommendations: A bayesian variational approach. *ACM TOIS*, 40(4):1–32, 2021.

[Wu *et al.*, 2015] Runze Wu, Qi Liu, Yuping Liu, Enhong Chen, Yu Su, Zhigang Chen, and Guoping Hu. Cognitive modelling for predicting examinee performance. In *IJCAI*, 2015.

[Wu *et al.*, 2021] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. Self-supervised graph learning for recommendation. In *ACM SIGIR*, pages 726–735, 2021.

[Wu *et al.*, 2022] Kaifang Wu, Yonghui Yang, Kun Zhang, Le Wu, Jing Liu, and Xin Li. Multi-relational cognitive diagnosis for intelligent education. In *CAAI*, pages 425–437. Springer, 2022.

[Yang *et al.*, 2021] Yang Yang, Jia-Qi Yang, Ran Bao, De-Chuan Zhan, Hengshu Zhu, Xiao-Ru Gao, Hui Xiong, and Jian Yang. Corporate relative valuation using heterogeneous multi-modal graph neural network. *IEEE TKDE*, 35(1):211–224, 2021.

[Yang *et al.*, 2023] Yang Yang, Chubing Zhang, Xin Song, Zheng Dong, Hengshu Zhu, and Wenjie Li. Contextualized knowledge graph embedding for explainable talent training course recommendation. *ACM Transactions on Information Systems*, 42(2):1–27, 2023.

[Yang *et al.*, 2024] Shangshang Yang, Xiaoshan Yu, Ye Tian, Xueming Yan, Haiping Ma, and Xingyi Zhang. Evolutionary neural architecture search for transformer in knowledge tracing. *NeurIPS*, 36, 2024.

[Ye *et al.*, 2023] Haibo Ye, Xinjie Li, Yuan Yao, and Hanghang Tong. Towards robust neural graph collaborative filtering via structure denoising and embedding perturbation. *ACM TOIS*, 41(3):1–28, 2023.

[You *et al.*, 2020] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.

[Yu *et al.*, 2024] Xiaoshan Yu, Chuan Qin, Dazhong Shen, Haiping Ma, Le Zhang, Xingyi Zhang, Hengshu Zhu, and Hui Xiong. Rdgt: Enhancing group cognitive diagnosis with relation-guided dual-side graph transformer. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Zha *et al.*, 2023] Rui Zha, Chuan Qin, Le Zhang, Dazhong Shen, Tong Xu, Hengshu Zhu, and Enhong Chen. Career mobility analysis with uncertainty-aware graph autoencoders: A job title transition perspective. *IEEE Transactions on Computational Social Systems*, 2023.

[Zhang *et al.*, 2023] Yunfei Zhang, Chuan Qin, Dazhong Shen, Haiping Ma, Le Zhang, Xingyi Zhang, and Hengshu Zhu. Relicd: A reliable cognitive diagnosis framework with confidence awareness. In *IEEE ICDM*, 2023.

[Zhang *et al.*, 2024] Zheng Zhang, Le Wu, Qi Liu, Jiayu Liu, Zhenya Huang, Yu Yin, Yan Zhuang, Weibo Gao, and Enhong Chen. Understanding and improving fairness in cognitive diagnosis. *SCIENCE CHINA Information Sciences*, 67(5):152106–, 2024.