PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

HiCD: Hyperbolic Insight through Decomposed Educational Graphs for Long-Tailed Cognitive Diagnosis

Shengwei Ji[1,3] , **Wenli Wang**[1] , **Yongqiang Xie**[1] , **Fei Liu**[2,3] * , **Yonghui Yang**[4]
[1]School of Artificial Intelligence and Big Data, Hefei University, Hefei, China
[2]Innovation School of Artificial Intelligence, Hefei University of Technology, Hefei, China
[3]Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, Hefei, China
[4]NExT++ Research Center, National University of Singapore, Singapore
swji@mail.hfut.edu.cn, wangwenli@stu.hfuu.edu.cn, xieyongqiang@stu.hfuu.edu.cn, feiliu@mail.hfut.edu.cn, yyh.hfut@gmail.com

Abstract

Cognitive diagnosis (CD) aims to infer students' mastery of knowledge concepts from their response behaviors and constitutes a core component of intelligent education and personalized learning. However, existing graph-based CD models struggle to handle the pronounced long-tail distributions in educational data, where most students and concepts interact with only a limited number of exercises, resulting in suboptimal representation learning and poor generalization to low-frequency instances. To address this challenge, we propose **HiCD** (**H**yperbolic **i**nsight for **C**ognitive **D**iagnosis), a novel hyperbolic model that embeds students, exercises, and concepts into non-Euclidean space. By exploiting the exponential representational capacity of hyperbolic geometry, HiCD naturally captures hierarchical and sparse structures, effectively alleviating long-tail bias while enhancing embedding expressiveness. A key contribution of HiCD is a hyperbolic diagnostic function that operates directly on the manifold, avoiding Euclidean approximations and preserving geometric consistency. Moreover, HiCD decomposes the educational graph into three semantically distinct subgraphs and assigns each a dedicated curvature, enabling adaptive geometric modeling of heterogeneous relations. Extensive experiments on multiple benchmark datasets demonstrate that HiCD consistently improves diagnostic accuracy and robustness, particularly under severe long-tail scenarios. The source code is available at: https://github.com/CyberXie/HiCD.

1 Introduction

Cognitive diagnosis (CD) aims to infer students' mastery of latent knowledge concepts from their response behaviors and predict performance on unseen exercises. It serves as a fundamental component of intelligent education and personalized learning, as illustrated in Fig. 1. Early CD models, including traditional approaches [Cai *et al.*, 2016; Templin and Henson, 2006] and subsequent deep learning-based methods [Wang *et al.*, 2020; Pandey and Karypis, 2019], primarily rely on manually designed diagnostic functions or neural networks to model student responses and estimate concept mastery. Despite their success, these methods often model student–exercise interactions independently, limiting their ability to capture complex relationships among students, exercises, and concepts. To overcome these limitations, recent studies have introduced graph-based approaches into CD to model structural dependencies among students, exercises, and knowledge concepts. Early methods, such as KaNCD [Wang *et al.*, 2022] and RCD [Gao *et al.*, 2021], leverage educational graphs to enhance representation learning by exploiting relational structures. Subsequent work, including SCD [Wang *et al.*, 2023] and SVGCD [Xia *et al.*, 2025], further incorporates self-supervised learning and semantic-aware modeling to alleviate data sparsity and improve diagnostic performance. Overall, graph-based CD methods significantly advance cognitive diagnosis by inte-

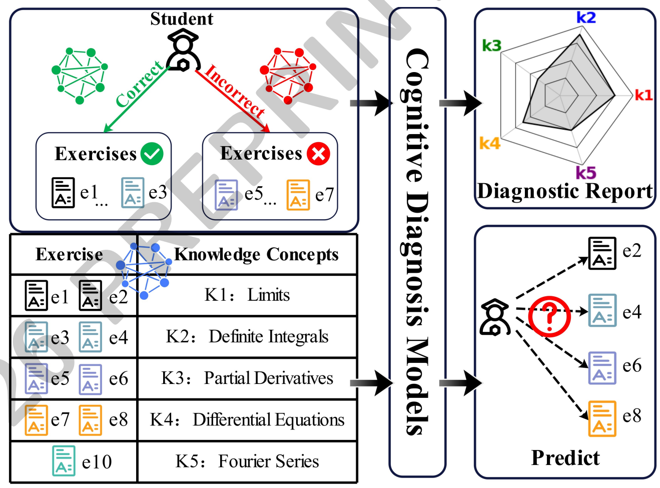


Figure 1: The process of cognitive diagnosis (CD)

*Corresponding author

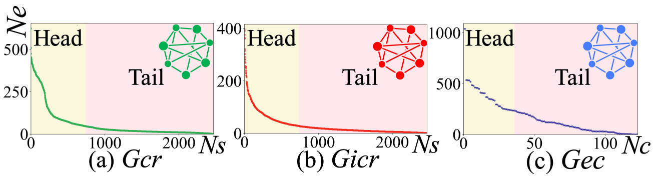


Figure 2: Long-tail distributions of student–exercise responses ((a) correct and (b) incorrect) and (c) concept–exercise relations on the ASSIST-0910 dataset. Here, N_s, N_e, and N_c denote the numbers of students, exercises, and concepts, respectively.

grating structural reasoning with semantic representation.

However, a challenge persists in graph-based cognitive diagnosis: **the long-tail distribution** in educational graphs. As illustrated in Fig. 2, most students and concepts interact with few exercises, resulting in long-tail sparsity. Such sparsity constrains representation learning, making it difficult to obtain discriminative embeddings for low-frequency students and concepts, degrading diagnostic accuracy. Although existing methods, such as SCD [Wang *et al.*, 2023] and EIRS [Yao *et al.*, 2023], attempt to alleviate this issue through degree balancing or response generation, they remain confined to Euclidean space. Due to the flat geometry of Euclidean manifolds, these models are limited in modeling sparse, hierarchical, and scale-free structures observed in educational graphs, and thus struggle to capture subtle cognitive differences among long-tail nodes.

To address the above limitations, we propose **HiCD** (**H**yperbolic **i**nsight for **C**ognitive **D**iagnosis), a hyperbolic cognitive diagnosis model. HiCD introduces hyperbolic geometry into CD modeling, leveraging the exponential growth of metric volume to accommodate power-law long-tail distributions observed in educational data. Unlike existing hyperbolic graph models that perform computation via Euclidean tangent spaces [Liu *et al.*, 2019; Chami *et al.*, 2019], HiCD conducts diagnostic modeling directly on the hyperbolic manifold, avoiding repeated exponential and logarithmic mappings and improving numerical stability [Sun *et al.*, 2024]. Moreover, HiCD decomposes the educational graph into three semantically distinct subgraphs and adopts a curvature-aware mechanism to assign each subgraph an appropriate hyperbolic space, following advances in curvature estimation [Cruceru *et al.*, 2021; Shu *et al.*, 2023; Sun *et al.*, 2023]. A multi-level fusion strategy integrates information across subgraphs. By jointly addressing geometric mismatch and structural heterogeneity, HiCD mitigates the adverse effects of long-tail distributions in cognitive diagnosis. Our contributions are summarized as follows:

- We propose HiCD, a hyperbolic cognitive diagnosis model for long-tail distributions in educational graphs.
- We design manifold-based hyperbolic diagnostic functions and introduce a curvature-aware mechanism to adaptively model semantically distinct subgraphs.
- Extensive experiments on multiple benchmark datasets show that HiCD consistently outperforms strong baselines, especially on sparse long-tail interactions.

2 Preliminaries

2.1 Hyperbolic Geometry

Hyperbolic space is a Riemannian manifold with constant negative curvature, whose metric volume grows exponentially with the radius r. This property makes it well suited for modeling hierarchical and long-tailed distributions commonly observed in educational graphs [Nickel and Kiela, 2017; Sarkar, 2011]. In this work, we adopt the Lorentz (hyperboloid) model [Nickel and Kiela, 2018; Chami *et al.*, 2019] due to its favorable numerical stability. For curvature $k > 0$, the d-dimensional Lorentz model is defined as

$$\mathbb{H}_k^d = \{x \in \mathbb{R}^{d+1} \mid \langle x, x \rangle_H = -k,\ x_0 > 0\}, \tag{1}$$

where the Lorentzian inner product is given by $\langle x, y \rangle_H = -x_0 y_0 + \sum_{i=1}^{d} x_i y_i$. The tangent space at a point $o \in \mathbb{H}_k^d$ is defined as $T_o\mathbb{H}_k^d = \{v \in \mathbb{R}^{d+1} \mid \langle v, o \rangle_H = 0\}$, which provides a locally Euclidean approximation of the manifold [Ganea *et al.*, 2018]. Exponential and logarithmic maps establish smooth mappings between the tangent space and the hyperbolic manifold; their explicit forms are provided in Appendix A.1. These mappings enable gradient-based optimization while preserving the underlying hyperbolic geometry.

2.2 Problem Statement

In cognitive diagnosis, we consider three types of entities: students S ($|S| = N$), exercises E ($|E| = M$), and knowledge concepts C ($|C| = K$). The exercise–concept association matrix $\mathbf{Q} = \{q_{ec}\}_{M \times K}$ is binary, where $q_{ec} = 1$ indicates that exercise e involves concept c, and $q_{ec} = 0$ otherwise. Student response logs are represented as $R = \{(s, e, r_{se}) \mid s \in S, e \in E, r_{se} \in \{0, 1\}\}$, where $r_{se} = 1$ denotes a correct response. Based on R, we construct an educational bipartite graph $\boldsymbol{G} = (S \cup E, A)$, where A is the adjacency matrix induced by the observed responses. The objective of cognitive diagnosis is to predict, for any student–exercise pair (s, e), the probability that student s correctly answers exercise e:

$$\hat{y}_{se} = P(r_{se} = 1 \mid s, e), \qquad \hat{y}_{se} \in [0, 1]. \tag{2}$$

These predictions further enable the inference of students' mastery levels over knowledge concepts.

3 HiCD Model

This section presents the architecture of the proposed HiCD model. As illustrated in Fig. 3, HiCD consists of three core components: graph decomposition, hyperbolic representation learning, and hyperbolic diagnosis. Specifically, the educational graph is decomposed into three semantically distinct subgraphs, whose representations are learned in hyperbolic spaces and subsequently fused. The algorithmic details are presented in Appendix E. Based on the fused representations, diagnostic functions operating directly on the hyperbolic manifold are employed to predict student responses.

3.1 Graph Decomposition

The educational graph exhibits both semantic and structural heterogeneity: correct and incorrect responses convey opposite semantics, while student–exercise and exercise–concept

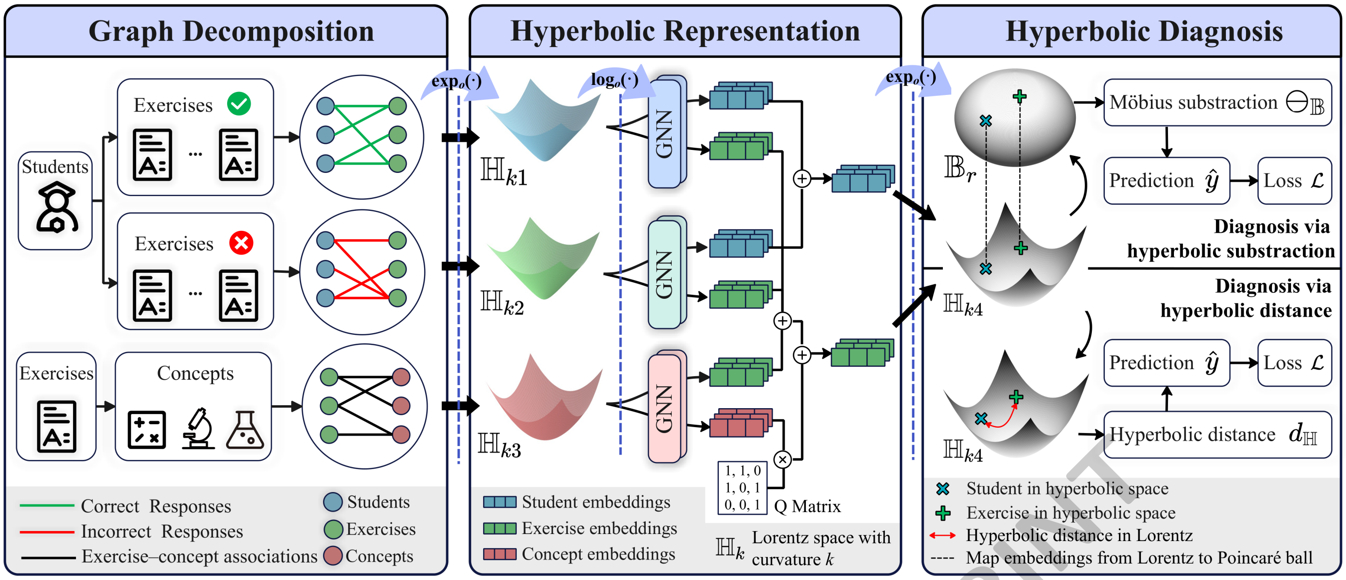


Figure 3: Overall framework of the proposed HiCD model, consisting of (a) graph decomposition into semantically distinct subgraphs, (b) hyperbolic representation learning with curvature-aware modeling and multi-level fusion, and (c) hyperbolic diagnostic functions defined directly on the manifold for student response prediction.

relations differ substantially in connectivity patterns and degree distributions (see Fig. 2). Accordingly, we decompose the educational graph into three semantic subgraphs based on relation types: the correct-response graph G_{cr}, the incorrect-response graph G_{icr}, and the exercise–concept association graph G_{ec}. These subgraphs respectively model student mastery, non-mastery behaviors, and concept associations. For each subgraph, we initialize learnable node embeddings $\boldsymbol{h}^{\mathbb{R}}_{cr}$, $\boldsymbol{h}^{\mathbb{R}}_{icr}$, $\boldsymbol{h}^{\mathbb{R}}_{ec} \in \mathbb{R}^d$ in Euclidean space, which are optimized end-to-end via backpropagation. These embeddings serve as the initial representations for students, exercises, and concepts, and are subsequently refined through hyperbolic representation learning to encode student knowledge states, exercise difficulty, and concept semantics.

3.2 Hyperbolic Representation

Following the graph decomposition, each semantic subgraph is mapped to a dedicated hyperbolic space with curvature-specific geometry. Curvature characterizes the geometric properties of hyperbolic space and determines how distances and relational structures are encoded. In HiCD, three subgraphs with distinct structural and semantic characteristics—the student correct-response graph, the student incorrect-response graph, and the exercise–concept association graph—are embedded into hyperbolic spaces with potentially different curvatures. To estimate suitable curvatures, we adopt an efficient lower-bound approximation of Ollivier–Ricci curvature based on local graph statistics [Ollivier, 2009; Shu *et al.*, 2023]. Edge curvatures are first computed and aggregated to node-level curvatures by averaging over incident edges [Weber *et al.*, 2017; Leal *et al.*, 2021], after which each subgraph curvature is defined as the mean over its nodes [Sun *et al.*, 2023; Cruceru *et al.*, 2021]. This adaptive curvature assignment strategy facilitates geometry–semantics alignment, mapping each semantic relation to an appropriately scaled hyperbolic manifold with curvature parameter $k \in \{k_1, k_2, k_3\}$ (see Appendix A.2 for details).

Building upon the curvature-specific hyperbolic manifolds, we perform hyperbolic representation learning for each subgraph. Node embeddings are mapped between Euclidean space and the corresponding hyperbolic manifold via exponential and logarithmic maps:

$$\boldsymbol{h}^{\mathbb{H}} = \exp_o^k(\boldsymbol{h}^{\mathbb{R}}), \quad \boldsymbol{h}^{\mathcal{T}} = \log_o^k(\boldsymbol{h}^{\mathbb{H}}), \tag{3}$$

where $\exp_o^k(\cdot)$ and $\log_o^k(\cdot)$ denote the exponential and logarithmic maps at the origin, respectively, and $\boldsymbol{h}^{\mathcal{T}}$ lies in the tangent space. Message passing and aggregation are then performed in the tangent space using independent GraphSAGE encoders [Hamilton *et al.*, 2017] for each subgraph (see Appendix A.3):

$$\boldsymbol{h}_i^{(l+1)} = \sigma\left(\mathbf{W}^{(l)} \cdot \left[\boldsymbol{h}_i^{(l)} \,\middle\|\, \underset{j \in \mathcal{N}(i)}{\mathrm{AGGREGATE}}\left(\boldsymbol{h}_j^{(l)}\right)\right] + \boldsymbol{b}^{(l)}\right). \tag{4}$$

The refined representations from different subgraphs are then integrated through a multi-level fusion strategy to support subsequent hyperbolic diagnosis. For a student node, its representations from the correct-response (G_{cr}) and incorrect-response (G_{icr}) subgraphs are fused, as both directly model student behavior. For an exercise node, its representations from all three subgraphs are integrated, incorporating information from student interactions (G_{cr}, G_{icr}) and conceptual associations (G_{ec}). The Q-matrix is used to select the relevant concept embeddings for each exercise.This yields the final hyperbolic student representation $\boldsymbol{h}(s)$ and exercise representation $\boldsymbol{h}(e)$ for the subsequent diagnostic layer (see Appendix A.3 for details).

3.3 Hyperbolic Diagnosis

We perform cognitive diagnosis directly in hyperbolic space by defining geometry-aware diagnostic functions. Since Euclidean operations are not intrinsically compatible with hyperbolic geometry, we design two hyperbolic diagnostic functions—*hyperbolic subtraction* and *hyperbolic distance*—to model the interaction between student ability and exercise difficulty.

Diagnosis via Hyperbolic Subtraction

In conventional cognitive diagnosis models, the difference between student and exercise embeddings is often used to estimate the likelihood of a correct response. To generalize this operation to hyperbolic space, we adopt Möbius subtraction to capture the relative relationship between student ability and exercise difficulty. Specifically, embeddings are first mapped from the Lorentz model to the Poincaré ball via a Lorentz-to-Poincaré transformation [Ganea *et al.*, 2018] (see Appendix A.4), where Möbius operations are well defined.The Möbius subtraction is given by

$$\boldsymbol{h}(s) \ominus_c \boldsymbol{h}(e) := \boldsymbol{h}(s) \oplus_c \big(-\boldsymbol{h}(e)\big), \tag{5}$$

and the resulting discrepancy is measured as

$$d_B(\boldsymbol{h}(s), \boldsymbol{h}(e)) = \big\|\boldsymbol{h}(s) \ominus_c \boldsymbol{h}(e)\big\|_2. \tag{6}$$

Diagnosis via Hyperbolic Distance

To avoid numerical instability caused by repeated manifold mappings and Möbius operations, we alternatively compute the hyperbolic distance directly in the Lorentz model. The geodesic distance between student and exercise embeddings is defined as

$$d_H(\boldsymbol{h}(s), \boldsymbol{h}(e)) = \sqrt{k}\ \operatorname{arcosh}\left(-\frac{\langle \boldsymbol{h}(s), \boldsymbol{h}(e)\rangle_{\mathbb{H}}}{k} \right), \tag{7}$$

where k denotes the curvature of the corresponding hyperbolic space and $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ is the Lorentzian inner product. This formulation computes the geodesic distance directly on the manifold, ensuring geometric consistency and improved numerical stability [Sun *et al.*, 2024].

Unified Decoding via Fermi–Dirac Function

Both diagnostic functions yield a distance measure $d(\boldsymbol{h}(s), \boldsymbol{h}(e))$, which is converted into the probability of a correct response using a Fermi–Dirac decoder [Chami *et al.*, 2019]:

$$\hat{y}_{se} = \left[\exp\left(\frac{d(\boldsymbol{h}(s), \boldsymbol{h}(e)) - r}{t} \right) + 1 \right]^{-1}, \tag{8}$$

where r and t denote the threshold and temperature parameters, respectively.

3.4 Model Optimization

We optimize HiCD by minimizing the Binary Cross-Entropy (BCE) loss over observed student–exercise interactions. Given the predicted probability $\hat{y}_{se}$ produced by the hyperbolic diagnostic decoder, the loss function is defined as

$$\mathcal{L} = - \sum_{(s,e)\in R} \left[r_{se} \log \hat{y}_{se} + (1 - r_{se}) \log(1 - \hat{y}_{se}) \right], \tag{9}$$

where $r_{se} \in \{0, 1\}$ denotes the observed response label. The model parameters are optimized end-to-end by minimizing $\mathcal{L}$ via gradient-based optimization.

Datasets	ASSIST-0910	Junyi
# Students	2,493	10,000
# Exercises	17,676	734
# Knowledge concepts	123	734
# Response logs	267,423	408,057
# Avg logs per student	107.240	40.806
# Sparsity in student-exercise logs	99.394%	94.441%

Table 1: Statistics of two real-world datasets.

4 Experiments

This section evaluates the proposed HiCD model through extensive experiments. We first describe the experimental setup, including datasets, baselines, and implementation details, and then analyze the performance of HiCD through a series of research questions. Two model variants are considered: **HiCD-sub**, which performs diagnosis via hyperbolic subtraction, and **HiCD-dist**, which adopts hyperbolic distance as the diagnostic function. Both variants share the same graph decomposition strategy and curvature-aware mechanism.

The experiments are designed to answer the following research questions: **RQ1:** How does HiCD compare with existing baselines on public benchmark datasets? **RQ2:** How does each design component contribute to the overall performance? **RQ3:** How does HiCD perform under different data distribution settings? **RQ4:** How sensitive is HiCD to key hyperparameters? **RQ5:** Can HiCD capture and distinguish different long-tailed structures in hyperbolic space?

4.1 Experimental Settings

Implementation Details: All experiments are conducted on the ASSIST-0910 and Junyi datasets (see Table 1). Models are implemented in PyTorch with Xavier normal initialization. The embedding dimension is set to 128, and training is performed using a learning rate of 1×10^{-4}, weight decay of 1×10^{-2}, and a Riemannian optimizer. Each model is run five times with different random seeds on an NVIDIA RTX 4090 GPU, and we report the average AUC, RMSE, and ACC. Additional implementation details are provided in Appendix B.

Baselines: We compare HiCD with representative cognitive diagnosis methods from two categories: (1) *traditional models*, including IRT [Cai *et al.*, 2016] and NeuralCD [Wang *et al.*, 2020]; and (2) *graph-based models*, including KaNCD [Wang *et al.*, 2022], RCD [Gao *et al.*, 2021], SCD [Wang *et al.*, 2023], and SVGCD [Xia *et al.*, 2025].

We also compare with a unified hyperbolic GNN baseline (see Appendix D.2). All baselines are reproduced and tuned following their original implementations, and performance is evaluated using AUC, RMSE, and ACC.

4.2 Overall Performance Comparison (RQ1)

Table 2 reports the overall performance of different methods on the ASSIST-0910 and Junyi datasets. Across various train–test split ratios, HiCD achieves consistently strong performance and outperforms competing baselines on most evaluation metrics. In particular, both HiCD-sub and HiCD-dist obtain higher AUC and ACC compared with existing meth-

(a) ASSIST-0910

Train:Test	5:5			6:4			7:3			8:2		
Methods	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑
IRT	0.7225	0.4472	0.7062	0.7333	0.4434	0.7100	0.7415	0.4412	0.7153	0.7480	0.4390	0.7175
NeuralCD	0.7340	0.4440	0.7131	0.7413	0.4408	0.7155	0.7508	0.4386	0.7151	0.7558	0.4372	0.7186
KaNCD	0.7412	0.4468	0.7126	0.7499	0.4399	0.7214	0.7584	0.4340	0.7269	0.7640	0.4304	0.7315
RCD	0.7563	0.4287	0.7256	0.7612	**0.4257**	0.7293	0.7644	0.4255	0.7305	0.7685	**0.4231**	0.7330
SCD	0.7543	0.4304	0.7222	0.7602	0.4276	0.7262	0.7668	0.4245	0.7330	0.7701	0.4233	0.7321
SVGCD	0.7591	0.4357	0.7200	0.7658	0.4320	0.7237	0.7699	0.4291	0.7352	0.7698	0.4268	0.7346
HiCD-sub	0.7602	**0.4286**	0.7242	0.7629	0.4248	0.7316	0.7659	**0.4235**	0.7319	0.7662	0.4230	0.7323
HiCD-dist	**0.7613**	0.4338	**0.7261**	**0.7706**	0.4364	**0.7317**	**0.7787**	0.4364	**0.7403**	**0.7972**	0.4303	**0.7531**

(b) Junyi

Train:Test	5:5			6:4			7:3			8:2		
Methods	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑
IRT	0.7867	0.4171	0.7281	0.7901	0.4131	0.7390	0.7927	0.4103	0.7454	0.7858	0.4075	0.7508
NeuralCD	0.7803	0.4129	0.7473	0.7804	0.4125	0.7479	0.7806	0.4122	0.7476	0.7811	0.4126	0.7473
KaNCD	0.7850	0.4092	0.7504	0.7859	0.4091	0.7509	0.7869	0.4089	0.7516	0.7885	0.4083	0.7556
RCD	0.8035	0.3997	0.7666	0.8056	0.3998	0.7675	0.8066	0.3984	0.7673	0.8087	0.3972	0.7689
SCD	0.8026	0.4003	0.7646	0.8073	0.3981	0.7676	0.8099	0.3968	0.7687	0.8134	0.3949	0.7726
SVGCD	0.8056	0.4027	0.7627	0.8089	**0.3953**	0.7687	0.8095	**0.3932**	0.7689	0.8037	0.4033	0.7664
HiCD-sub	0.8114	**0.3959**	0.7675	0.8108	0.3964	0.7613	0.8103	0.3943	0.7684	0.8155	**0.3943**	0.7731
HiCD-dist	**0.8142**	0.4078	**0.7682**	**0.8154**	0.4043	**0.7696**	**0.8123**	0.4039	**0.7693**	**0.8207**	0.3961	**0.7734**

Table 2: Performance comparison under different train–test split ratios. Best results are in bold and runner-up results are underlined. ↑ (↓) indicates that higher (lower) values are better. (a) Results on ASSIST-0910; (b) Results on Junyi.

Datasets	**ASSIST-0910**			**Junyi**		
Metrics	AUC↑	RMSE↓	ACC↑	AUC↑	RMSE↓	ACC↑
w/o SubG-Dec	0.7637	0.4361	0.7138	0.8093	0.4113	0.7520
w/o Fusion	0.7648	0.4357	0.7130	0.8078	0.4120	0.7508
w/o Q Matrix	0.7725	0.4352	0.7164	0.8073	0.4121	0.7505
w/o Hyperbolic	0.7513	0.4428	0.6589	0.7852	0.4248	0.6876
w/o Hyp-Diag_1	0.7787	0.4422	0.7053	0.8049	0.4220	0.7377
w/o Hyp-Diag_2	0.7224	0.4614	0.6920	0.7843	0.4306	0.7257
w/o C-A	0.7697	0.4413	0.7044	0.8105	0.4127	0.7528
HiCD-dist	**0.7972**	**0.4369**	**0.7531**	**0.8195**	**0.4034**	**0.7604**

Table 3: Ablation results of HiCD on two benchmark datasets. Best results are highlighted in bold. ↑ (↓) indicates that higher (lower) values are better.

ods, indicating the effectiveness of hyperbolic modeling for cognitive diagnosis under long-tailed data distributions.

Between the two variants, HiCD-dist generally yields better overall performance, especially under the 8:2 train–test split. Compared with HiCD-sub, it achieves higher AUC and more stable ACC across multiple runs, suggesting that distance-based modeling provides a more consistent diagnostic signal. This observation highlights the advantage of performing diagnosis directly based on hyperbolic distances, particularly in the presence of sparse and long-tailed interaction patterns.

4.3 Ablation Study (RQ2)

To assess the contribution of individual components in HiCD, we conduct ablation studies on the ASSIST-0910 and Junyi datasets. All variants are derived from HiCD-dist, with each variant removing or modifying a specific component. The results are reported in Table 3.

Impact of Graph Structure and Fusion Mechanisms. Removing graph decomposition (w/o SubG-Dec), which merges all relations into a single heterogeneous graph, leads to noticeable drops in AUC, indicating the importance of separating semantic relations. Disabling the multi-level fusion module (w/o Fusion) degrades performance, suggesting that integrating information across subgraphs contributes to improved diagnostic accuracy. When the Q-matrix enhancement is removed (w/o Q Matrix), performance decreases, highlighting its role in aligning exercise semantics with concept associations.

Impact of Hyperbolic Geometric Mechanisms. Replacing hyperbolic operations with Euclidean counterparts (w/o Hyperbolic) results in the largest performance degradation among all variants, underscoring the effectiveness of hyperbolic modeling for sparse and long-tailed educational data. Using a shared curvature across subgraphs (w/o C-A) instead of subgraph-specific curvatures also leads to performance drops, indicating the benefit of curvature-aware modeling. Moreover, variants employing Euclidean diagnostic functions (w/o Hyp-Diag_1 and w/o Hyp-Diag_2) consistently underperform HiCD-dist, suggesting that hyperbolic diagnostic functions better capture the geometric relationships required for accurate cognitive diagnosis. Detailed ablation settings are provided in Appendix C.

4.4 Long-tail Grouped Experiments (RQ3)

To evaluate HiCD under long-tailed interaction distributions, we partition students into six groups (S1–S6) according to their interaction frequencies. We additionally conduct exercise- and concept-side long-tail analyses, which show consistent trends (see Appendix D.1). Specifically, S1–S2 correspond to low-interaction (sparse) students, S3–S4 to medium-activity students, and S5–S6 to high-interaction (dense) students. As shown in Fig. 4, HiCD consistently outperforms competing methods across all groups, with more pronounced performance gains observed for sparse students (S1–S2). This indicates that HiCD is more robust to data sparsity and can generalize more effectively from limited interaction records. For medium- and high-activity groups, HiCD also achieves competitive and stable performance improvements over baselines. These results suggest that hyperbolic modeling in HiCD maintains balanced performance across different interaction regimes, effectively alleviating the impact of long-tailed interaction distributions.

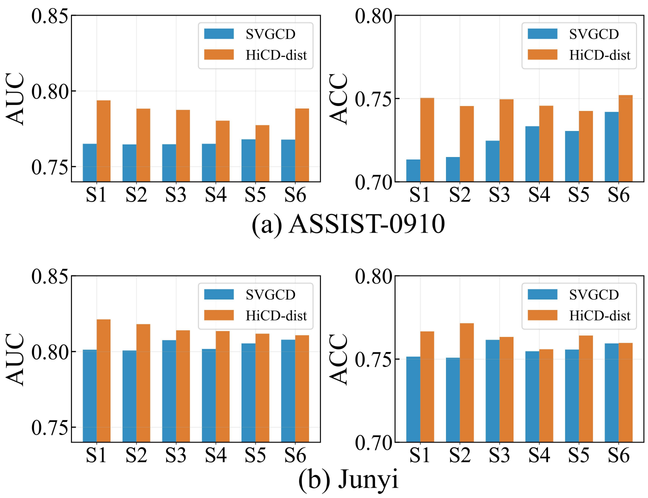


Figure 4: Performance across different student groups on the ASSIST-0910 and Junyi datasets.

4.5 Hyperparameter Sensitivity Analysis (RQ4)

To assess the robustness of HiCD with respect to hyperparameter choices, we conduct sensitivity analyses on both datasets by varying three key factors: the curvature parameter controlling manifold geometry, the decoder parameters governing probabilistic mapping, and the number of GNN layers.

Sensitivity to the curvature parameter k. HiCD assigns subgraph-specific curvatures while keeping the diagnostic space curvature learnable. To examine the effect of curvature, we fix k_1, k_2, and k_3 to the same constant and vary $k \in \{0, 0.5, 1.0, 2.0, 3.0, 4.0\}$. As shown in Fig. 5, performance first improves and then degrades as curvature increases. The adaptive curvature setting consistently yields the best AUC and ACC, indicating its effectiveness in balancing local structural preservation and global geometric smoothness.

Sensitivity to decoder parameters r and t. We evaluate fixed decoder configurations with $r = t \in \{1, 2, 3, 4\}$ and compare them with a learnable decoder. As shown in Fig. 6(a), larger values of r and t lead to a smoother transition region and reduced discrimination, whereas the learnable decoder consistently achieves better performance across all metrics, demonstrating greater flexibility in adapting to different data distributions.

Sensitivity to the number of GNN layers. We vary the number of GNN layers $L \in \{1, 2, 3, 4, 5\}$ to examine the effect of encoder depth. As shown in Fig. 6(b), performance

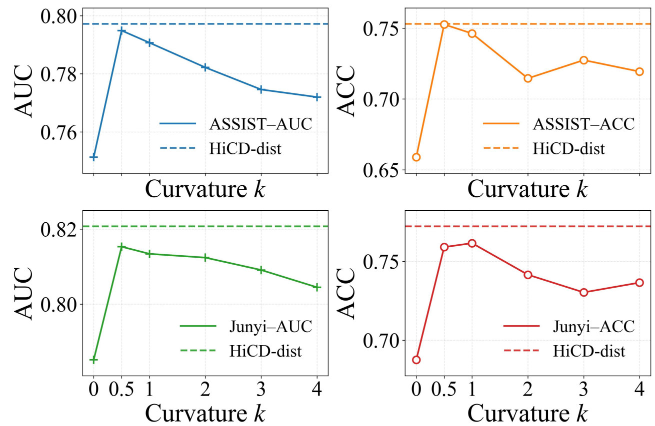


Figure 5: Sensitivity analysis of curvature on the ASSIST-0910 and Junyi datasets.The dashed line denotes HiCD with adaptive curvature, while solid lines correspond to fixed curvature values.

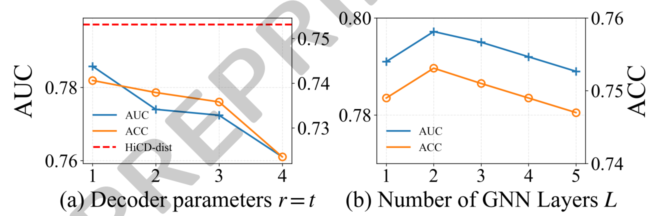


Figure 6: Sensitivity analysis on the ASSIST-0910 dataset. (a) Effects of the Fermi–Dirac decoder parameters r and t, where solid lines correspond to fixed parameter settings and the dashed line denotes HiCD with a learnable decoder. (b) Effects of the number of GNN layers on model performance.

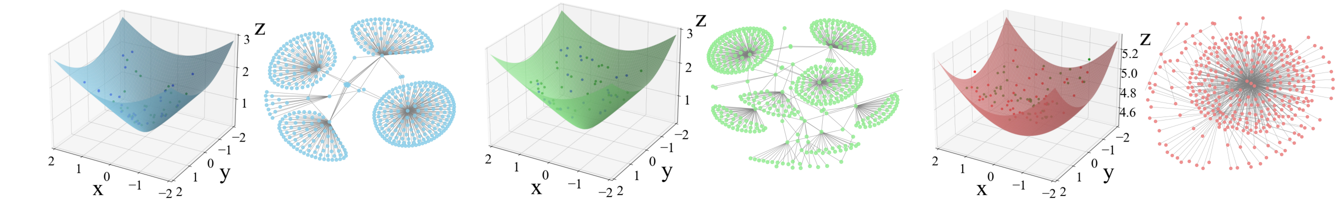
improves with increasing depth up to $L = 2$ and then declines as deeper layers are added. This suggests that a moderate depth effectively captures multi-hop dependencies while avoiding over-smoothing, consistent with common observations in graph neural networks.

Overall, the results indicate that HiCD exhibits stable performance across a wide range of hyperparameter settings, and that adaptive curvature modeling, a learnable decoder, and an appropriate GNN depth contribute to its robustness.

4.6 Embedding Visualization Analysis (RQ5)

To qualitatively examine how HiCD captures long-tailed structures across different semantic subgraphs, we visualize the learned embeddings on the ASSIST-0910 dataset, as shown in Fig. 7. We apply t-SNE [Maaten and Hinton, 2008] to project high-dimensional embeddings into a two-dimensional space for visualization. Student and exercise/concept embeddings learned from different subgraphs are visualized separately, corresponding to hyperbolic spaces with curvatures k_1, k_2, and k_3.

As observed in Fig. 7, embeddings from the correct-response and incorrect-response subgraphs (H_{k_1} and H_{k_2}) exhibit more dispersed and hierarchical patterns, reflecting the separation of mastery and non-mastery behaviors. In contrast, embeddings from the exercise–concept subgraph (H_{k_3})



(a) G_{cr} in Hyperbolic space $\mathbb{H}_{k1}$ (b) G_{cr} Visualization (c) G_{icr} in Hyperbolic space $\mathbb{H}_{k2}$ (d) G_{icr} Visualization (e) G_{ec} in Hyperbolic space $\mathbb{H}_{k3}$ (f) G_{ec} Visualization

Figure 7: Hyperbolic embedding visualization of students, exercises, and concepts on the ASSIST-0910 dataset.

form comparatively more compact structures, consistent with denser semantic associations across concepts. These visual patterns suggest that HiCD learns subgraph-specific representations that align with their underlying structural properties, facilitating effective modeling of long-tailed educational interactions.

5 Related Work

5.1 Cognitive Diagnosis Models

Cognitive diagnosis (CD) aims to infer students' mastery of latent knowledge concepts from their observed exercise responses. Traditional CD models, such as IRT [Cai *et al.*, 2016], DINA [De La Torre, 2009], and DINO [Templin and Henson, 2006], rely on expert-defined Q-matrices and probabilistic assumptions to model student ability and item difficulty. Although effective in controlled settings, these methods exhibit limited expressiveness and struggle to scale to complex, large-scale educational data.

With the rise of deep learning, neural CD models have significantly improved representation capacity. NCD [Wang *et al.*, 2020] introduced neural networks for modeling nonlinear student–exercise interactions, while DKT [Piech *et al.*, 2015] incorporated sequential modeling to capture temporal knowledge dynamics. Recent neural CD methods also explore richer representation learning to better capture latent cognitive characteristics and interaction patterns. However, most neural approaches treat interactions independently and do not explicitly model the relational structure among students, exercises, and concepts.

To address this limitation, graph-based CD models have emerged as a dominant direction. KaNCD [Wang *et al.*, 2022] leveraged exercise–concept graphs to enhance knowledge propagation, and RCD [Gao *et al.*, 2021] unified students, exercises, and concepts within a multi-relational graph. More recent works incorporate self-supervised and semantic-aware learning to improve robustness under data sparsity and long-tailed distributions, including SCD [Wang *et al.*, 2023] and SVGCD [Xia *et al.*, 2025].

Despite these advances, existing methods primarily operate in Euclidean space, which is ill-suited for modeling long-tailed interaction patterns, and often overlook the distinct structural characteristics of different semantic subgraphs.

5.2 Hyperbolic Graph Representation Learning

Graph representation learning has become a fundamental approach for modeling relational data by capturing structural dependencies among entities. Graph neural networks (GNNs), including GCN [Kipf and Welling, 2017], GraphSAGE [Hamilton *et al.*, 2017], and GAT [Wang *et al.*, 2019], enable message passing and have achieved strong performance in learning node representations.

Hyperbolic geometry offers an alternative to Euclidean space for modeling hierarchical and scale-free structures due to exponential volume growth [Nickel and Kiela, 2017; Sarkar, 2011]. Nickel and Kiela introduced Poincaré embeddings for hierarchical relations, while subsequent works extended graph message passing to hyperbolic space [Chami *et al.*, 2019]. The Lorentz model [Nickel and Kiela, 2018; Chami *et al.*, 2019] further improved numerical stability and has been widely adopted in hyperbolic graph learning.

Recent studies further investigate curvature-aware message passing and geometry-adaptive representation learning for heterogeneous graph structures. Trainable curvature parameters [Nickel and Kiela, 2018] and curvature estimation methods such as Forman–Ricci and Ollivier–Ricci curvature [Weber *et al.*, 2017; Leal *et al.*, 2021; Ollivier, 2009; Shu *et al.*, 2023] enable more flexible geometric modeling.

Despite these advances, most hyperbolic graph models rely on a single global curvature, which limits their expressiveness for heterogeneous or multi-relational graphs. Educational graphs, in particular, contain multiple semantic relations with distinct structural characteristics, motivating the need for more fine-grained geometric modeling.

6 Conclusion

In this paper, we propose HiCD, a hyperbolic cognitive diagnosis model designed to address data sparsity and long-tailed interaction distributions in educational settings. By directly defining cognitive diagnostic functions on the hyperbolic manifold, HiCD enables geometry-consistent modeling of student–exercise interactions, while leveraging subgraph decomposition and adaptive-curvature representations for heterogeneous educational relations. Extensive experiments on the ASSIST-0910 and Junyi datasets demonstrate that HiCD achieves strong diagnostic performance, particularly under sparse and long-tailed conditions. Future work will explore more fine-grained cognitive structures and extend hyperbolic modeling to broader educational scenarios.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62306100 and Grant 62406096; in part by Anhui Educational Committee through the Natural Science Research Project of under Grant 2025AHGXZK40026; in part by the Key Laboratory of Knowledge Engineering with Big Data (Ministry of Education of China) under Grant BigKEOpen2025-01; and in part by Hefei University Arithmetic Platform.

References

[Cai *et al.*, 2016] Li Cai, Kilchan Choi, Mark Hansen, and Lauren Harrell. Item response theory. *Annual Review of Statistics and Its Application*, 3(1):297–321, 2016.

[Chami *et al.*, 2019] Ines Chami, Rex Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *Proceedings of the Advances in neural information processing systems*, page 4869, 2019.

[Cruceru *et al.*, 2021] Calin Cruceru, Gary Bécigneul, and Octavian-Eugen Ganea. Computationally tractable riemannian manifolds for graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7133–7141, 2021.

[De La Torre, 2009] Jimmy De La Torre. Dina model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, 34(1):115–130, 2009.

[Ganea *et al.*, 2018] Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 5350–5360, 2018.

[Gao *et al.*, 2021] Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu-Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. Rcd: Relation map driven cognitive diagnosis for intelligent education systems. In *Proceedings of the international ACM SIGIR conference on research and development in information retrieval*, pages 501–510, 2021.

[Hamilton *et al.*, 2017] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 1025–1035, 2017.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[Leal *et al.*, 2021] Wilmer Leal, Guillermo Restrepo, Peter F Stadler, and Jürgen Jost. Forman–ricci curvature for hypergraphs. *Advances in Complex Systems*, 24(01):2150003, 2021.

[Liu *et al.*, 2019] Qi Liu, Maximilian Nickel, and Douwe Kiela. Hyperbolic graph neural networks. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 8230–8241, 2019.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[Nickel and Kiela, 2017] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Proceedings of the Advances in neural information processing systems*, pages 6341–6350, 2017.

[Nickel and Kiela, 2018] Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the International conference on machine learning*, pages 3779–3788, 2018.

[Ollivier, 2009] Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.

[Pandey and Karypis, 2019] Shalini Pandey and George Karypis. A self-attentive model for knowledge tracing. In *Proceedings of the International Conference on Educational Data Mining, EDM 2019*, pages 384–389, 2019.

[Piech *et al.*, 2015] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas Guibas, and Jascha Sohl-Dickstein. Deep knowledge tracing. In *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, pages 505–513, 2015.

[Sarkar, 2011] Rik Sarkar. Low distortion delaunay embedding of trees in hyperbolic plane. pages 355–366, 2011.

[Shu *et al.*, 2023] Dong Wook Shu, Youjin Kim, and Junseok Kwon. Localized curvature-based combinatorial subgraph sampling for large-scale graphs. *Pattern Recognition*, 139:109475, 2023.

[Sun *et al.*, 2023] Li Sun, Junda Ye, Hao Peng, Feiyang Wang, and Philip S Yu. Self-supervised continual graph learning in adaptive riemannian spaces. In *Proceedings of the AAAI conference on artificial intelligence*, pages 4633–4642, 2023.

[Sun *et al.*, 2024] Li Sun, Zhenhao Huang, Zixi Wang, Feiyang Wang, Hao Peng, and Philip S Yu. Motif-aware riemannian graph neural network with generative-contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9044–9052, 2024.

[Templin and Henson, 2006] Jonathan L Templin and Robert A Henson. Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, 11(3):287, 2006.

[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 *Proceedings of the The world wide web conference*, pages 2022–2032, 2019.

[Wang *et al.*, 2020] 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 *Proceedings of the AAAI conference on artificial intelligence*, pages 6153–6161, 2020.

[Wang *et al.*, 2022] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yu Yin, Shijin Wang, and Yu Su. Neuralcd: a general framework for cognitive diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):8312–8327, 2022.

[Wang *et al.*, 2023] Shanshan Wang, Zhen Zeng, Xun Yang, and Xingyi Zhang. Self-supervised graph learning for long-tailed cognitive diagnosis. In *Proceedings of the AAAI conference on artificial intelligence*, pages 110–118, 2023.

[Weber *et al.*, 2017] Melanie Weber, Emil Saucan, and Jürgen Jost. Characterizing complex networks with forman-ricci curvature and associated geometric flows. *Journal of Complex Networks*, 5(4):527–550, 2017.

[Xia *et al.*, 2025] Chenao Xia, Fei Liu, Zihan Wang, Zhuangzhuang He, Pengyang Shao, Haowei Zhou, and Yonghui Yang. Semantic-tailored variational-contrastive graph learning for cognitive diagnosis. In *Proceedings of the ACM on Web Conference*, pages 1946–1952, 2025.

[Yao *et al.*, 2023] Fangzhou Yao, Qi Liu, Min Hou, Shiwei Tong, Zhenya Huang, Enhong Chen, Jing Sha, and Shijin Wang. Exploiting non-interactive exercises in cognitive diagnosis. *Interaction*, 100(200):300, 2023.