Multiplex Heterogeneous Graph Neural Networks with Euclidean-Riemannian Mutual Space Synergy

Xiang Li[1], **Yuan Cao**[1], **Zhongying Zhao**[3], **Guoqing Chao**[4], **Yanwei Yu**[1,2*]

[1]Faculty of Information Science and Engineering, Ocean University of China, China
[2]State Key Laboratory of Physical Oceanography, Ocean University of China, China
[3]College of Computer Science and Engineering, Shandong University of Science and Technology, China
[4]School of Computer Science and Technology, Harbin Institute of Technology, China
lixiang1202@stu.ouc.edu.cn, {cy8661, yuyanwei}@ouc.edu.cn, zyzhao@sdust.edu.cn, guoqingchao@hit.edu.cn

Abstract

Multiplex heterogeneous networks are common in real-world scenarios, where entities interact through diverse types of relations across multiple semantic layers. Recent advances in multiplex heterogeneous graph neural networks have achieved remarkable results by incorporating node and relation types into message passing and designing relation-aware architectures. However, most existing methods either decouple relations and risk losing complex semantics or require handcrafted relation patterns, which limit scalability. Moreover, prevailing models are typically restricted to Euclidean space, making it difficult to capture non-Euclidean topologies and to distinguish complex interactions among heterogeneous nodes and relations. Standard GNN message passing, grounded in the homophily assumption, also proves inadequate for the intricate, coupled structures in multiplex heterogeneous graphs. To address these challenges, we propose MRiemGNN, a novel multiplex heterogeneous graph neural network that synergizes Euclidean and Riemannian spaces through a geometry-aware, relation-specific message passing scheme and cross-space mutual learning. Experiments on multiple real-world datasets show that MRiemGNN achieves superior performance, efficiency, and scalability on both node classification and link prediction tasks.

Code — https://github.com/lx970414/MRiemGNN

Introduction

Graphs are widely used to model real-world systems with diverse entities and complex interactions (Wan et al. 2020; Li et al. 2025b). In many domains, such as e-commerce (Fu et al. 2023; Shui et al. 2024; Li et al. 2025c), social networks (Zeng et al. 2024), and knowledge graphs (Wang et al. 2022; Peng et al. 2023), data often naturally form multiplex heterogeneous graphs, where multiple types of nodes and multiple types of relations co-exist and interconnect, reflecting diverse semantics and hierarchical structures. Effectively learning representations from such multiplex heterogeneous graphs is critical for a variety of downstream tasks, including social friendship discovery, recommendation, and so on.

*Yanwei Yu is the corresponding author
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Recently, research on multiplex heterogeneous graph neural networks (MHGNNs) and representation learning has advanced rapidly (Yu et al. 2022; Melton and Krishnan 2023; Fu et al. 2023; Huang et al. 2025). Most approaches aim to extend the success of message passing neural networks by considering node/edge type information, designing relation-aware propagation rules, or leveraging meta-path-based or multi-relational architectures. These models have enabled notable progress in modeling semantic-rich, multi-relational data. However, existing methods are still often constrained by shallow fusion of relational semantics, limited scalability to complex multiplex structures, or an inability to fully capture the coupled, nontrivial dependencies among heterogeneous node and edge types.

Despite substantial progress in graph neural networks (GNNs), existing methods encounter significant challenges when applied to multi-relational heterogeneous graphs. *First, the message passing mechanisms in most GNNs are based on the homophily assumption, which can be inadequate for capturing the intricate, coupled structures present in multiplex heterogeneous graphs*. The assumption of homophily that underlies conventional message passing mechanisms does not hold in heterogeneous graphs, where connected nodes may differ significantly in type, semantics, and behavior. Blind aggregation across node and edge types can obscure meaningful patterns and lead to representation collapse. Many approaches rely on relation-decoupled message passing mechanisms, where each relation is treated independently, often followed by naive aggregation. This strategy neglects the composite semantics emerging from co-occurring relations and fails to capture inter-relation dependencies. Alternative methods attempt to encode handcrafted relational patterns or meta-paths, yet suffer from poor scalability and high design complexity, rendering them difficult to generalize across domains. *Second, most current multiplex graph representation methods operate in Euclidean space, which is ill-suited for encoding non-Euclidean topologies or distinguishing between different node types and interactions*. This geometric limitation makes it difficult to capture hierarchical patterns and nuanced cross-type relations in complex real-world graphs. A vast majority of GNNs operate in Euclidean space, which imposes constraints on capturing the structural richness of real-world graphs. Such

graphs often exhibit a mixture of densely connected clusters and globally tree-like topologies. Euclidean geometry lacks the expressiveness to simultaneously model both. In contrast, non-Euclidean spaces, particularly hyperbolic manifolds, are well-suited for encoding hierarchical or power-law structures due to their exponential growth properties. However, existing geometric GNNs typically operate within a single space, limiting their ability to simultaneously represent local and global patterns, and often failing to integrate the relational heterogeneity of the data.

To address these challenges, we advocate for a unified, geometry-aware approach to relational and structural modeling in multiplex heterogeneous graphs. Specifically, we propose that different types of nodes and relations should be embedded in spaces with adaptive curvature, allowing the model to capture both local and global structures by jointly leveraging Euclidean and hyperbolic (Riemannian) manifolds. Crucially, the learning processes in these geometric spaces should not be independent, but mutually reinforcing, so that information from one manifold can regularize and enrich the representations in the other. In this paper, we present the Euclidean-Riemannian Mutual Space Synergy Heterogeneous Graph Neural Network (MRiemGNN), a novel architecture designed for robust and expressive representation learning on multi-relational, heterogeneous graphs. MRiemGNN incorporates several key innovations: a geometry-aware, relation-specific message passing mechanism utilizing random Fourier feature kernels; a dual-space embedding module that aggregates information from both Euclidean and hyperbolic manifolds; and a cross-space mutual distillation strategy that aligns semantic information via KL divergence. The framework further supports flexible decoding modules for both node classification and link prediction, enhancing its adaptability to a range of downstream applications.

Our main contributions are summarized as follows:

- We address the challenge that standard GNN message passing struggles to model the complex, composite relations in multiplex heterogeneous graphs by innovatively introducing a relation-aware kernelized mechanism that captures both simple and composite relation patterns.
- We propose a Euclidean-Riemannian mutual space synergy mechanism, together with a dual-space embedding aggregation module, to enhance feature complementarity and geometric alignment across manifolds.
- Extensive experiments conducted on five real-world heterogeneous graph datasets demonstrate that MRiemGNN consistently outperforms state-of-the-art baselines, with average improvements of 11.57% and 3.68% on node classification and link prediction tasks, respectively.

Related Work

Heterogeneous Graph Neural Networks

Heterogeneous GNNs (HGNNs) are designed to model node-type and edge-type diversity in real-world graphs. Models such as HAN and HGT employ attention mechanisms for type-aware aggregation, while methods like HetGNN (Zhang et al. 2019) and HGConv (Yang et al. 2021) utilize meta-paths or type-specific encoders to capture complex semantic dependencies. HGT (Hu et al. 2020) further introduces type-dependent parameters and efficient sampling, achieving strong performance at web scale. Despite their success, these methods generally assume a Euclidean latent space, limiting their ability to capture non-Euclidean or hierarchical structures in heterogeneous graphs.

Multi-Relational Graph Neural Networks

Multi-relational GNNs extend traditional models to handle multiple edge types or interaction patterns. R-GCN (Schlichtkrull et al. 2018) uses relation-specific transformations, while CompGCN (Vashishth et al. 2020) combines compositional embeddings with message passing. Some methods, including HAN and HetGNN, leverage meta-paths and attention to integrate higher-order relation patterns, while recent work explores multi-layer or multiplex architectures (Mo et al. 2023a; Huang et al. 2024; Fu et al. 2025). However, most existing methods aggregate relation-specific features via shallow fusion, which can miss the semantics of composite relations and is computationally demanding for large or highly coupled structures.

Geometric and Riemannian Graph Learning

Graph geometry (Iyer et al. 2022; Kang et al. 2023; Li et al. 2025a; Fei et al. 2025) offers a powerful approach for representing structural diversity, especially in hierarchical or scale-free graphs. Hyperbolic GNNs such as HGCN (Liu, Nickel, and Kiela 2019) and HVGNN (Sun et al. 2021) use curved manifolds to capture hierarchy and global structure. Recent work like UHGNN (Law 2021) and $\mathcal{Q}$-GCN (Xiong et al. 2022) further generalize to pseudo-Riemannian manifolds, enabling effective learning on graphs with mixed topologies and surpassing traditional Riemannian models. With advances in Riemannian optimization, some approaches now explore varying curvature, multi-space embeddings, or geometry-aware aggregation (Kan et al. 2023; Sun et al. 2025). However, most geometric GNNs either focus on a single space or treat geometry and relational semantics in isolation, missing their full complementarity.

In contrast, our proposed MRiemGNN jointly captures multi-relational semantics and dual-space geometry. By unifying Euclidean and Riemannian manifolds within a single framework, MRiemGNN enables robust and expressive representation learning on multiplex heterogeneous graphs.

Preliminary

We consider a multiplex heterogeneous network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ is the set of edges, and $\mathbf{X} \in \mathbb{R}^{n \times m}$ is the node feature matrix. Each node $i \in \mathcal{V}$ is assigned a type $\Phi(i) \in \mathcal{O}$, and each edge $e \in \mathcal{E}$ has a type $\psi(e) \in \mathcal{R}$, where $\mathcal{O}$ and $\mathcal{R}$ denote the sets of all node and edge types. The network is *heterogeneous* if $|\mathcal{O}| + |\mathcal{R}| > 2$, and is *multiplex* if multiple edge types exist between the same node pairs.

Distinctively, we define $\mathcal{R}$ as the set of all observed relation patterns—including both single edge types and their

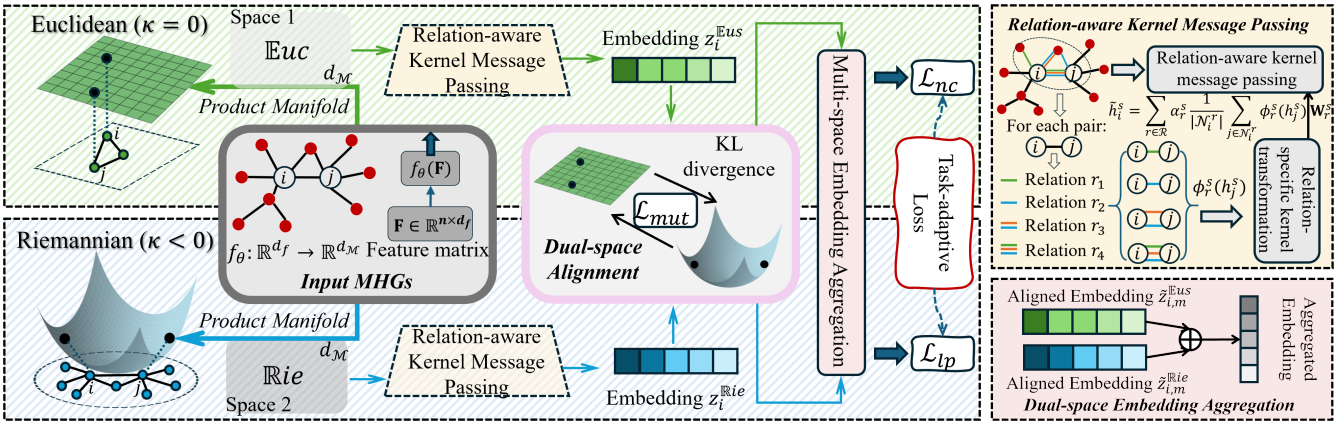


Figure 1: The overview of the proposed MRiemGNN.

co-occurring combinations—enabling the model to capture simple and composite semantics in multiplex heterogeneous graphs. To further encode this structural and relational diversity, we introduce a set of geometric spaces $\mathcal{S} = \{\mathbb{E}uc, \mathbb{R}ie\}$, corresponding to Euclidean and hyperbolic (Riemannian) manifolds. For each node $i \in \mathcal{V}$, separate embeddings z_i^s are learned in each space $s \in \mathcal{S}$, which are then aligned and aggregated for robust representation.

Our goal is to learn expressive node representations z_i that fully exploit both relational patterns and geometric diversity, supporting downstream tasks such as node classification and link prediction.

Methodology

MRiemGNN is a unified geometric neural framework designed for representation learning on multiplex heterogeneous graphs (MHGs), where nodes participate in multiple types of semantic relations and complex structural patterns co-exist. To capture the rich diversity and structural complexity inherent in such graphs, MRiemGNN integrates relation-aware message propagation in both Euclidean and Riemannian (hyperbolic) manifolds and bridges the two views through a mutual distillation mechanism. The final node representation is constructed as a weighted combination of both geometric spaces, serving as a universal embedding for downstream tasks. The overall workflow is illustrated in Fig. 1.

Overview of Architecture

The MRiemGNN consists of four tightly coupled modules: (1) a ***Relation-aware Kernel Message Passing*** module that encodes diverse inter-node semantics in a geometry-specific fashion, (2) a ***Euclidean-Riemannian Mutual Space Synergy*** module for promoting knowledge transfer and alignment between different geometric spaces, (3) a ***Dual-space Embedding Aggregation*** module for fusing dual-space representations, and (4) a ***Task-adaptive Decoder*** module supporting both node-level and edge-level inference.

Relation-aware Kernel Message Passing

Traditional GNN message passing relies on the homophily assumption that connected nodes are likely to be similar, which often fails in multiplex heterogeneous graphs, where neighbors can belong to different types and connect via diverse, potentially composite, relations. As a result, conventional message passing struggles to capture the intricate and coupled semantics of MHGs, leading to less effective node representations.

To overcome this, we propose a relation-aware kernel message passing mechanism. Specifically, we relabel not only each individual relation, but also every observed combination of relations between node pairs. For example, if a MHG contains three relation types a, b, and c, we enumerate all present combinations such as $\{r_a, r_b, r_c, r_{ab}, r_{ac}, r_{bc}, r_{abc}\}$ and assign them to a set of abstract relation labels $\{r_1, r_2, \dots\}$. This relabeling allows the model to recognize both simple and composite semantic patterns, ensuring that the set $\mathcal{R}$ comprehensively reflects all unique relation structures in the dataset.

For each geometric space $s \in \mathcal{S} = \{\mathbb{E}uc, \mathbb{R}ie\}$ and each relation $r \in \mathcal{R}$, we then define a space- and relation-specific kernel transformation via random Fourier features (RFF), which efficiently approximate nonlinear kernels:

$$\phi_r^s(h_j^s) = \sqrt{\frac{2}{D}} \cdot \cos(\omega_r^s \cdot h_j^s + b_r^s), \tag{1}$$

where $h_j^s \in \mathbb{R}^{d_{\mathcal{M}}}$ is the embedding of node j in space s, $\omega_r^s \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ is a random projection matrix, $b_r^s \sim \mathcal{U}(0, 2\pi)$ is a random phase, and D controls the kernel approximation fidelity. The message aggregated to node i in space s is then obtained as follows:

$$\tilde{h}_i^s = \sum_{r \in \mathcal{R}} \alpha_r^s \cdot \frac{1}{|\mathcal{N}_i^r|} \sum_{j \in \mathcal{N}_i^r} \phi_r^s(h_j^s) \cdot \mathbf{W}_r^s, \tag{2}$$

where $\mathcal{N}_i^r$ denotes the neighbors of node i under relation r, α_r^s is a learnable importance coefficient reflecting the contribution of relation r in space s, and $\mathbf{W}_r^s \in \mathbb{R}^{D \times d_{\mathcal{M}}}$ is a

trainable transformation matrix that projects kernel features back to the latent space.

After aggregation, the node embedding is updated through a nonlinear activation:

$$h_i^{(l+1),s} = \sigma\left(\tilde{h}_i^{(l),s}\right), \tag{3}$$

where $\sigma(\cdot)$ is an element-wise activation function such as ReLU. This process is operated for L message passing layers in each geometric space, allowing the model to progressively capture higher-order, relation-specific structural patterns.

Euclidean-Riemannian Mutual Space Synergy

While the above mechanism enables each space to independently learn from relational structure, it is also critical to promote semantic consistency and knowledge transfer between the Euclidean and Riemannian representations. To this end, we introduce a symmetric mutual distillation loss. After L layers, the final-layer embeddings for node i in the Euclidean and Riemannian spaces are denoted as $z_i^{\mathbb{E}uc} = h_i^{(L),\mathbb{E}uc}$ and $z_i^{\mathbb{R}ie} = h_i^{(L),\mathbb{R}ie}$, respectively.

We compute temperature-scaled softmax distributions over each embedding:

$$q_i^s = \text{softmax}\left(\frac{z_i^s}{\tau}\right), \quad s \in \mathcal{S}, \tag{4}$$

where τ is a temperature parameter. The mutual distillation loss is then defined as:

$$\mathcal{L}_{\text{mut}} = \sum_{i \in V} \text{KL}\left(q_i^{\mathbb{E}uc} \,\|\, q_i^{\mathbb{R}ie}\right) + \text{KL}\left(q_i^{\mathbb{R}ie} \,\|\, q_i^{\mathbb{E}uc}\right), \tag{5}$$

where $\text{KL}(\cdot\|\cdot)$ is the KL divergence. This regularization aligns high-order semantic distributions between spaces while allowing each to preserve its unique geometric inductive bias. After mutual learning, we can obtain the aligned Euclidean and Riemannian embeddings $\tilde{z}_i^{\mathbb{E}uc}$ and $\tilde{z}_i^{\mathbb{R}ie}$.

Dual-space Embedding Aggregation

To obtain a universal node embedding that fully exploits both geometric priors, we perform a weighted aggregation of the two space-specific embeddings:

$$z_i = \beta \cdot \tilde{z}_i^{\mathbb{E}uc} + (1-\beta) \cdot \tilde{z}_i^{\mathbb{R}ie}, \tag{6}$$

where $\beta \in [0,1]$ is a learnable or tunable hyperparameter. This fusion enables the final embedding to flexibly integrate the strengths of both Euclidean geometry (effective for dense, locally connected patterns) and Riemannian geometry (powerful for hierarchical and non-Euclidean topology).

Task-adaptive Decoder

The fused embedding z_i serves as a unified feature for downstream node classification and link prediction tasks.

Node Classification. The predicted class probabilities for node i are given by:

$$p_i = \text{softmax}(\mathbf{W}_{\text{cls}} z_i + b_{\text{cls}}), \tag{7}$$

where $\mathbf{W}_{\text{cls}}$ and $\mathbf{b}_{\text{cls}}$ are trainable parameters. The node classification loss over labeled nodes is:

$$\mathcal{L}_{\text{cls}} = -\sum_{i \in \mathcal{V}_{\text{train}}} \log p_{i,y_i}. \tag{8}$$

The total objective for node classification is:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{mut}}, \tag{9}$$

where λ_1 balances the mutual distillation loss.

Link Prediction. For a node pair (i,j), we construct a composite feature vector:

$$\mathbf{g}_{ij} = [\mathbf{z}_i \| \mathbf{z}_j \| \mathbf{z}_i * \mathbf{z}_j \| |\mathbf{z}_i - \mathbf{z}_j|], \tag{10}$$

which is input to an MLP decoder to compute the predicted edge score:

$$s_{ij} = \text{MLP}(\mathbf{g}_{ij}). \tag{11}$$

The link prediction loss is binary cross-entropy over positive and negative edges:

$$\mathcal{L}_{\text{lp}} = -\sum_{(i,j)\in E^+} \log \sigma(s_{ij}) - \sum_{(i,j)\in E^-} \log(1-\sigma(s_{ij})), \tag{12}$$

where E^+, E^- denote positive and negative edge sets. The total link prediction objective is:

$$\mathcal{L} = \mathcal{L}_{\text{lp}} + \lambda_2 \mathcal{L}_{\text{mut}}, \tag{13}$$

where λ_2 balances the auxiliary mutual distillation loss.

In summary, MRiemGNN achieves robust and expressive graph representations by tightly coupling geometry-aware, relation-specific message passing, multi-space embedding synergy, and flexible, task-adaptive decoding. All modules are jointly optimized end-to-end.

Experiment

In this section, we evaluate the performance of our proposed MRiemGNN through extensive experiments and answer the following questions:

- **(RQ1)** How does MRiemGNN effectively perform in the node classification task?
- **(RQ2)** How does MRiemGNN effectively perform in the link prediction task?
- **(RQ3)** What are the effects of different modules in MRiemGNN on performance?
- **(RQ4)** How does MRiemGNN perform in terms of efficiency compared to baselines on multiplex graphs?
- **(RQ5)** How do different hyperparameter settings affect the performance?

Dataset	#nodes	#edges	#n-types	#e-types	#feat.	#labels	Mult.
IMDB	11,616	17,106	3	2	19	3	✗
Alibaba	15,218	27,036	2	3	18	5	✓
Amazon	10,166	148,865	1	2	1,156	–	✓
Taobao	21,318	41,676	2	4	19	3	✓
Mag	100,000	1,831,462	1	3	128	120	✓

Table 1: Statistical summaries of datasets.

	Method	IMDB		Alibaba		Taobao		Mag	
		Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F11	Macro-F1	Micro-F1
Euc.	FAME [CIKM'20]	0.5564	0.5646	0.2412	0.2536	0.3128	0.4035	0.3927	0.4806
	MHGCN [KDD'22]	0.6136	0.6197	0.4303	0.4774	0.3156	0.4136	0.4034	0.4967
	DMG [ICML'23]	0.6303	0.6297	0.4964	0.5450	0.3388	0.4295	0.4828	0.5479
	MGDCR [AAAI'23]	0.6357	0.6339	0.4869	0.5345	0.3385	0.4290	0.4788	0.5403
	CoCoMG [MM'23]	0.6415	0.6398	0.4854	0.5360	0.3392	0.4297	0.4717	0.5444
	BPHGNN [KDD'23]	0.6447	0.6468	0.4972	0.5455	0.3393	0.4304	0.4732	0.5474
	MGHC [AAAI'25]	0.6435	0.6457	0.4976	0.5460	0.3401	0.4299	0.4905	0.5517
Rie.	HGCN [NIPS'19]	0.5525	0.5607	0.4231	0.4774	0.2925	0.3747	0.4129	0.4828
	SELFMGNN [AAAI'22]	0.5957	0.6116	0.4521	0.5039	0.3129	0.4026	0.4321	0.5016
	HIE [ICML'23]	0.6363	0.6441	0.4697	0.5203	0.3198	0.4091	0.4518	0.5260
	RRN [NIPS'23]	0.6478	0.6562	0.4769	0.5254	0.3253	0.4148	0.4538	0.5275
	MotifRGC [AAAI'22]	0.6415	0.6709	0.4788	0.5266	0.3290	0.4188	0.4569	0.5302
	CUSP [ICLR'25]	0.6547	0.6848	0.4842	0.5347	0.3367	0.4274	0.4623	0.5366
	GraphMoRE [AAAI'25]	0.6635	0.6844	0.4851	0.5360	0.3351	0.4259	0.4705	0.5417
	H-EDML [WWW'25]	0.6636	0.6975	0.4825	0.5340	0.3328	0.4211	0.4707	0.5432
	MRiemGNN	**0.7657***	**0.8428***	**0.5151***	**0.6402***	**0.3653***	**0.4521***	**0.5574***	**0.6039***
	Improvement	*15.39%*	*20.83%*	*3.52%*	*17.25%*	*7.41%*	*5.04%*	*13.64%*	*9.46%*

Table 2: Node classification performance comparison of different methods on four datasets. Marker * indicates the result is statistically significant (t-test with p-value < 0.01).

Experimental Settings

Datasets To assess the efficacy of MRiemGNN, we conduct node classification and link prediction experiments on eight publicly available datasets, including IMDB, Alibaba, Amazon, Taobao, and Mag (adapted from ogbn-mag). The summary of dataset statistics is presented in Table 1.

Baselines To demonstrate the effectiveness, we compare our proposed MRiemGNN with the following two categories of methods: (1) Euclidean GNN methods: **FAME** (Liu et al. 2020), **MHGCN** (Yu et al. 2022), **DMG** (Mo et al. 2023b), **MGDCR** (Mo et al. 2023c), **CoCoMG** (Peng, Wang, and Zhu 2023), **BPHGNN** (Fu et al. 2023), and **MGHC** (Huang et al. 2025); (2) Riemannian GNN methods: **HGCN** (Chami et al. 2019), **SELFMGNN** (Sun et al. 2022), **HIE** (Yang et al. 2023), **RRN** (Katsman et al. 2023), **MotifRGC** (Sun et al. 2024), **CUSP** (Grover et al. 2025), **GraphMoRE** (Guo et al. 2025), and **H-EDML** (Cao et al. 2025) for comparison to evaluate the node classification and link prediction performance.

Evaluation Metrics and Experimental Settings We use Macro-F1 and Micro-F1 metrics for evaluating the node classification task, and AUROC and AUPR metrics for the link prediction task. We employ the Adam optimizer for the entirety of model training, initializing all trainable parameters using Xavier. We systematically explore several key hyperparameters: The number of GNN message passing layers ranges from {1, 2, 3, 4, 5}. The weight decay ranges from {1e-4, 5e-4, 1e-3, 5e-3, 1e-2}, and the learning rate is in a range of {0.0001,0.001,0.01,0.1}. In the embedding layer, we consider different vector sizes, selecting from {16, 32, 64, 128, 256}, and the dimension of hidden embedding is searched in {8, 16, 32, 128, 256, 512}. Additionally, we implement early stopping to mitigate the risk of overfitting. The decay ratio, denoted as θ, is set to 0.5 to further fine-tune training dynamics. We use early stopping with a 50-epoch patience, halting training if validation performance does not improve. The complete model implementation is carried out using the Deep Graph Library and PyTorch, harnessing the computational power of an Intel Xeon Gold 5320T CPU (2.3GHz) with a NVIDIA RTX 3090 GPU.

Node Classification (RQ1)

We first evaluate our model against state-of-the-art (SOTA) baselines on the node classification task, as shown in Table 2. The top seven methods are Euclidean-based GNNs, while the remainder are Riemannian-based.

Our MRiemGNN consistently outperforms all baselines on both Macro-F1 and Micro-F1 across all evaluated heterogeneous networks. Specifically, MRiemGNN achieves average improvements of 10.01% (Macro-F1) and 13.15% (Micro-F1) over the strongest competitors on four datasets. The gains are especially pronounced on multiplex heterogeneous networks such as Alibaba, Taobao, and Mag. For example, MRiemGNN surpasses the SOTA MGHC by 17.25% in Micro-F1 on Alibaba. These improvements are largely due to our model's ability to capture rich multiplex structures via dual-space embeddings and to align Euclidean and Riemannian representations through mutual learning. MRiemGNN also demonstrates strong performance on the general heterogeneous IMDB dataset, achieving 15.39% (Macro-F1) and 20.83% (Micro-F1) improvements. Even in settings without explicit multiplex structures, the Riemannian component effectively captures heterogeneity among node types and interactions, enabling more informative node

	Method	IMDB		Alibaba		Amazon		Taobao		Mag	
		AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
Euc.	FAME [CIKM'20]	0.8635	0.8593	0.6255	0.6194	0.9507	0.9515	0.9942	0.9913	0.7327	0.7275
	MHGCN [KDD'22]	0.8924	0.8567	0.9024	0.8592	0.9718	0.9744	0.9942	0.9953	0.8493	0.8272
	DMG [ICML'23]	0.8998	0.8802	0.9055	0.8597	0.9732	0.9749	0.9940	0.9948	0.8501	0.8311
	MGDCR [AAAI'23]	0.9055	0.8843	0.9139	0.8748	0.9527	0.9514	0.9915	0.9898	0.8898	0.8735
	CoCoMG [MM'23]	0.9163	0.8952	0.9246	0.8847	0.9669	0.9658	0.9935	0.9920	0.9008	0.8829
	BPHGNN [KDD'23]	0.9223	0.8982	0.9364	0.8941	0.9735	0.9739	0.9955	0.9934	0.9017	0.8840
	MGHC [AAAI'25]	0.9175	0.9010	0.9289	0.8910	0.9723	0.9710	0.9947	0.9955	0.9108	0.8987
Rie.	HGCN [NIPS'19]	0.8231	0.8127	0.8159	0.7948	0.8987	0.8824	0.9342	0.9428	0.8035	0.7891
	SELFMGNN [AAAI'22]	0.8540	0.8486	0.8484	0.8289	0.9251	0.9149	0.9795	0.9756	0.8298	0.8103
	HIE [ICML'23]	0.8823	0.8456	0.8924	0.8595	0.9420	0.9344	0.9892	0.9910	0.8480	0.8275
	RRN [NIPS'23]	0.8966	0.8754	0.8939	0.8611	0.9459	0.9376	0.9890	0.9898	0.8514	0.8313
	MotifRGC [AAAI'22]	0.9037	0.8803	0.9095	0.8757	0.9538	0.9494	0.9904	0.9875	0.8802	0.8653
	CUSP [ICLR'25]	0.9109	0.8921	0.9207	0.8997	0.9614	0.9600	0.9937	0.9915	0.8958	0.8779
	GraphMoRE [AAAI'25]	0.9205	0.8974	0.9289	0.8851	0.9670	0.9652	0.9950	0.9930	0.8988	0.8811
	H-EDML [WWW'25]	0.9172	0.9010	0.9365	0.8940	0.9695	0.9680	0.9955	0.9952	0.9087	0.8955
	MRiemGNN	**0.9506***	**0.9536***	**0.9647***	**0.9308***	**0.9839***	**0.9862***	**0.9957**	**0.9960**	**0.9643***	**0.9445***
	Improvement	*3.07%*	*5.84%*	*3.02%*	*4.10%*	*1.07%*	*1.26%*	*0.02%*	*0.05%*	*5.87%*	*5.10%*

Table 3: Link prediction performance comparison of different methods on five datasets. Marker * indicates the result is statistically significant (t-test with p-value < 0.01).

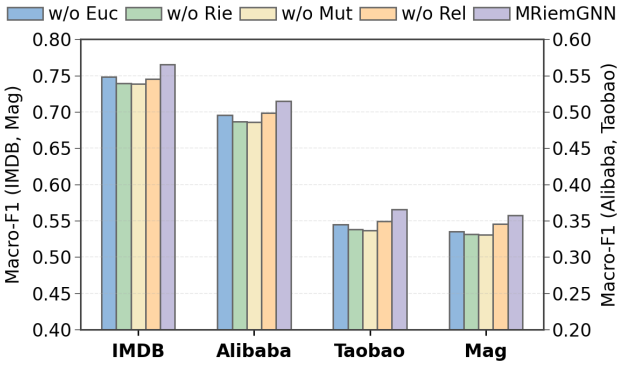


Figure 2: Performance comparison of different variants.

representations than previous methods.

Link Prediction (RQ2)

We further evaluate MRiemGNN on unsupervised link prediction against SOTA baselines, as shown in Table 3. The first seven methods are Euclidean-based GNNs, and the remaining are Riemannian-based.

MRiemGNN consistently achieves SOTA results across all metrics and datasets. On average, it improves AUROC and AUPR by 2.6% and 3.27% over the best baselines on five datasets. Notably, many strong baselines already achieve over 90% accuracy, making further improvements highly challenging. Even so, MRiemGNN delivers the best results—for instance, on Taobao dataset, 11 baselines exceed 99% AUROC, yet MRiemGNN still leads. The model also demonstrates significant gains on other datasets, highlighting its robustness. Specifically, on the large-scale dataset Mag, MRiemGNN achieves average improvements of 5.87% (AUROC) and 5.10% (AUPR), benefiting from its sparse coding and efficient computation strategy. This enables MRiemGNN to effectively capture global patterns while maintaining manageable computational complexity.

Ablation Study (RQ3)

To evaluate the effectiveness of each component in our MRiemGNN, we further conduct an ablation study on different MRiemGNN variations. We report the results of the ablation study on four datasets for node classification in Table 4. Specifically, we generate four variants:

- ***w/o Euc*** removes the learned embeddings from the Euclidean space module and uses only the Riemannian ones.
- ***w/o Rie*** removes the learned embeddings from the Riemannian space module and uses only the Euclidean ones.
- ***w/o Mut*** removes the Euclidean-Riemannian mutual space synergy module.
- ***w/o Rel*** excludes the relation-aware kernel message passing module.

Based on the ablation results, we make the following observations: (1) Removing either the Euclidean or Riemannian space (***w/o Euc*** or ***w/o Rie***) leads to a notable performance drop across all multiplex networks, highlighting the effectiveness of our dual-space embedding strategy. (2) The model without mutual space synergy (***w/o Mut***) performs worst, even compared to the single-space variants, underscoring the importance of cross-space mutual learning. Modeling in Euclidean or Riemannian space alone is limited—Euclidean space struggles with non-Euclidean topologies and complex interactions, while Riemannian space may

Method	IMDB		Alibaba		Taobao		Mag	
	Each	Total	Each	Total	Each	Total	Each	Total
BPHGNN	0.52	52.11	0.08	8.40	0.18	17.94	8.73	873.32
MGHC	1.51	181.32	1.10	132.12	1.15	138.01	20.58	2,469.13
GraphMoRE	1.75	175.32	0.98	98.33	1.23	122.85	23.55	2,354.60
H-EDML	0.92	101.31	0.65	71.83	0.43	47.74	13.15	1,578.12
MRiemGNN	0.35	35.44	0.05	5.83	0.12	12.22	2.87	286.96

Table 4: Runtime (s) of several best-performing baselines.

overemphasize relation differentiation at the expense of capturing detailed node interactions. The mutual learning mechanism aligns and complements these strengths, enabling the model to distinguish both relation types and structural features more effectively. (3) Removing relation-aware kernel message passing (***w/o Rel***) consistently degrades performance, but less so than other ablations, confirming that this module significantly enhances the capacity of GNNs to model multiplex relational patterns and supports more effective node representation learning.

Efficiency Analysis (RQ4)

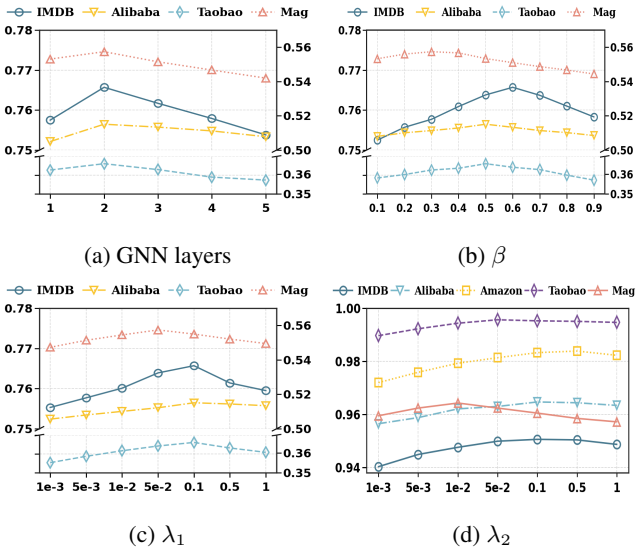
Table 4 shows a runtime comparison between our proposed MRiemGNN and several strong baselines on four benchmark datasets. MRiemGNN consistently achieves the lowest per-epoch and total runtime across all datasets. For instance, on the large-scale Mag dataset, MRiemGNN completes training in just 286.96 seconds, offering a $3\times$ to $8\times$ speedup over the next best baseline. On smaller graphs such as Alibaba, it also outperforms all competitors, with per-epoch runtime less than half that of the fastest baseline.

This efficiency stems from two key factors: the kernelized message passing with random Fourier features enables expressive relational modeling without heavy meta-path or attention computations, while the weighted fusion of Euclidean and Riemannian embeddings reduces overhead compared to multi-branch frameworks. These design choices allow MRiemGNN to deliver strong performance with exceptional scalability, making it especially suitable for large-scale, real-world multiplex heterogeneous graphs.

Parameter Sensitivity (RQ5)

We finally evaluate the sensitivity of our MRiemGNN with respect to three main hyperparameters (*i.e.*, the number of aggregation layers, parameter A, and the number of training rounds). The Macro-F1 score on node classification with different settings on five datasets is depicted in Figure 4. Notice that the performance on three multiplex datasets refers to the right-ordinate axis, and the performance on IMDB w.r.t. also refers to the right-ordinate axis.

Effect of message passing Layers L. As shown in Fig. 3a, model performance improves as the number of GNN layers increases from 1 to 2, reaching optimal results across datasets. Beyond two layers, further increases lead to saturation or a slight decline in accuracy, likely due to over-smoothing or noise accumulation. This suggests that a moderate depth (*e.g.*, 2 layers) achieves the best balance between feature propagation and model stability in MHGs.



(a) GNN layers (b) β (c) λ_1 (d) λ_2

Figure 3: Hyperparameter impact *w.r.t* the number of GNN layers L, balanced weight β, loss weight λ_1 and λ_2.

Effect of hyperparameter β. We also examine the impact of the fusion coefficient β in the embedding fusion process of Euclidean and Riemannian space, which controls the balance between Euclidean and Riemannian embeddings. As illustrated in Fig. 3b, performance rises with increasing β up to around 0.6, then gradually decreases. This indicates that moderately integrating information from both geometric spaces leads to optimal node representations, while overemphasizing either space reduces model effectiveness.

Effect of hyperparameters λ_1 and λ_2. We further investigate the sensitivity of the model to the mutual learning loss weights λ_1 (node classification) and λ_2 (link prediction). As shown in Fig.3c and Fig.3d, model performance generally improves as λ_1 and λ_2 increase from very small values, peaking around 0.1 to 0.5. Setting either parameter too low underutilizes the benefits of cross-space mutual learning, while excessively large values may lead to over-regularization and slight performance drops. This pattern is consistent across all datasets, confirming that moderate mutual learning strength is key to maximizing the advantages of dual-space representation.

Conclusion

We proposed MRiemGNN, a novel multiplex Riemannian GNN for multiplex heterogeneous graph representation. It unifies Euclidean and Riemannian geometries, incorporates relation-aware kernelized message passing, and leverages dual-space mutual learning to boost representational capacity. The framework flexibly supports various downstream tasks via adaptive decoders. Extensive experiments on real-world heterogeneous graph datasets demonstrate the effectiveness and efficiency of our approach.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under grant Nos. 62176243, 62472263, and 62276079, and the Fundamental Research Funds for the Central Universities under grant No. 202442005.

References

Cao, H.; Wang, Y.; Li, J.; Zhu, P.; and Hu, Q. 2025. Hyperbolic-Euclidean Deep Mutual Learning. In *Proceedings of the ACM on Web Conference 2025*, 3073–3083.

Chami, I.; Ying, Z.; Ré, C.; and Leskovec, J. 2019. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32.

Fei, Y.; Liu, Y.; Jia, C.; Li, Z.; Wei, X.; and Chen, M. 2025. A survey of geometric optimization for deep learning: from Euclidean space to Riemannian manifold. *ACM Computing Surveys*, 57(5): 1–37.

Fu, C.; Zheng, G.; Huang, C.; Yu, Y.; and Dong, J. 2023. Multiplex heterogeneous graph neural network with behavior pattern modeling. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, 482–494.

Fu, S.; Peng, Q.; He, Y.; Du, B.; Zou, B.; Jing, X.-Y.; and You, X. 2025. Unsupervised multiplex graph diffusion networks with multi-level canonical correlation analysis for multiplex graph representation learning. *Science China Information Sciences*, 68(3): 132102.

Grover, K.; Yu, H.; Song, X.; Zhu, Q.; Xie, H.; Ioannidis, V. N.; and Faloutsos, C. 2025. Spectro-Riemannian graph neural networks. *arXiv preprint arXiv:2502.00401*.

Guo, Z.; Sun, Q.; Yuan, H.; Fu, X.; Zhou, M.; Gao, Y.; and Li, J. 2025. GraphMoRE: Mitigating Topological Heterogeneity via Mixture of Riemannian Experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11754–11762.

Hu, Z.; Dong, Y.; Wang, K.; and Sun, Y. 2020. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, 2704–2710.

Huang, Y.; Mo, Y.; Liu, Y.; Nie, C.; Wen, G.; and Zhu, X. 2024. Multiplex graph representation learning via bi-level optimization. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2081–2089.

Huang, Y.; Nie, C.; He, H.; Mo, Y.; Zhu, Y.; Wen, G.; and Zhu, X. 2025. Multiplex Graph Representation Learning with Homophily and Consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11835–11842.

Iyer, R. G.; Bai, Y.; Wang, W.; and Sun, Y. 2022. Dual-geometric space embedding model for two-view knowledge graphs. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 676–686.

Kan, X.; Li, Z.; Cui, H.; Yu, Y.; Xu, R.; Yu, S.; Zhang, Z.; Guo, Y.; and Yang, C. 2023. R-mixup: Riemannian mixup for biological networks. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, 1073–1085.

Kang, Q.; Zhao, K.; Song, Y.; Wang, S.; and Tay, W. P. 2023. Node embedding from neural Hamiltonian orbits in graph neural networks. In *International conference on machine learning*, 15786–15808. PMLR.

Katsman, I.; Chen, E.; Holalkere, S.; Asch, A.; Lou, A.; Lim, S. N.; and De Sa, C. M. 2023. Riemannian residual neural networks. *Advances in Neural Information Processing Systems*, 36: 63502–63514.

Law, M. 2021. Ultrahyperbolic neural networks. *Advances in neural information processing systems*, 34: 22058–22069.

Li, L.; Jin, Z.; Zhang, X.; Duan, H.; Wang, J.; Tao, Z.; Zhao, H.; and Zhu, X. 2025a. Multi-view riemannian manifolds fusion enhancement for knowledge graph completion. *IEEE Transactions on Knowledge and Data Engineering*.

Li, X.; Qi, J.; Liu, H.; Cao, Y.; Chao, G.; Zhao, Z.; Dong, J.; Liu, X.; and Yu, Y. 2025b. ScaleGNN: Towards Scalable Graph Neural Networks via Adaptive High-order Neighboring Feature Fusion. *arXiv preprint arXiv:2504.15920*.

Li, X.; Qi, J.; Zhao, Z.; Zheng, G.; Cao, L.; Dong, J.; and Yu, Y. 2025c. UMGAD: Unsupervised Multiplex Graph Anomaly Detection. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, 3724–3737. IEEE Computer Society.

Liu, Q.; Nickel, M.; and Kiela, D. 2019. Hyperbolic graph neural networks. *Advances in neural information processing systems*, 32.

Liu, Z.; Huang, C.; Yu, Y.; Fan, B.; and Dong, J. 2020. Fast attributed multiplex heterogeneous network embedding. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 995–1004.

Melton, J.; and Krishnan, S. 2023. muxGNN: Multiplex graph neural network for heterogeneous graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 11067–11078.

Mo, Y.; Chen, Y.; Lei, Y.; Peng, L.; Shi, X.; Yuan, C.; and Zhu, X. 2023a. Multiplex graph representation learning via dual correlation reduction. *IEEE Transactions on Knowledge and Data Engineering*, 35(12): 12814–12827.

Mo, Y.; Lei, Y.; Shen, J.; Shi, X.; Shen, H. T.; and Zhu, X. 2023b. Disentangled multiplex graph representation learning. In *International conference on machine learning*, 24983–25005. PMLR.

Mo, Y.; Wu, Z.; Chen, Y.; Shi, X.; Shen, H. T.; and Zhu, X. 2023c. Multiplex graph representation learning via common and private information mining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 9217–9225.

Peng, L.; Mo, Y.; Xu, J.; Shen, J.; Shi, X.; Li, X.; Shen, H. T.; and Zhu, X. 2023. GRLC: Graph representation learning with constraints. *IEEE transactions on neural networks and learning systems*, 35(6): 8609–8622.

Peng, L.; Wang, X.; and Zhu, X. 2023. Unsupervised multiplex graph learning with complementary and consistent information. In *Proceedings of the 31st ACM International Conference on Multimedia*, 454–462.

Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, 593–607. Springer.

Shui, C.; Li, X.; Qi, J.; Jiang, G.; and Yu, Y. 2024. Hierarchical graph contrastive learning for review-enhanced recommendation. In *Joint European conference on machine learning and knowledge discovery in databases*, 423–440. Springer.

Sun, L.; Huang, Z.; Wang, Z.; Wang, F.; Peng, H.; and Yu, P. S. 2024. Motif-aware riemannian graph neural network with generative-contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 9044–9052.

Sun, L.; Huang, Z.; Zhou, S.; Wan, Q.; Peng, H.; and Yu, P. 2025. Riemanngfm: Learning a graph foundation model from riemannian geometry. In *Proceedings of the ACM on Web Conference 2025*, 1154–1165.

Sun, L.; Zhang, Z.; Ye, J.; Peng, H.; Zhang, J.; Su, S.; and Yu, P. S. 2022. A self-supervised mixed-curvature graph neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4146–4155.

Sun, L.; Zhang, Z.; Zhang, J.; Wang, F.; Peng, H.; Su, S.; and Yu, P. S. 2021. Hyperbolic variational graph neural network for modeling dynamic graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4375–4383.

Vashishth, S.; Sanyal, S.; Nitin, V.; and Talukdar, P. P. 2020. Composition-based Multi-Relational Graph Convolutional Networks. In *ICLR*.

Wan, G.; Du, B.; Pan, S.; and Haffari, G. 2020. Reinforcement learning based meta-path discovery in large-scale heterogeneous information networks. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 34, 6094–6101.

Wang, C.; Zhou, S.; Yu, K.; Chen, D.; Li, B.; Feng, Y.; and Chen, C. 2022. Collaborative knowledge distillation for heterogeneous information network embedding. In *Proceedings of the ACM web conference 2022*, 1631–1639.

Xiong, B.; Zhu, S.; Potyka, N.; Pan, S.; Zhou, C.; and Staab, S. 2022. Pseudo-riemannian graph convolutional networks. *Advances in neural information processing systems*, 35: 3488–3501.

Yang, M.; Zhou, M.; Ying, R.; Chen, Y.; and King, I. 2023. Hyperbolic representation learning: Revisiting and advancing. In *International conference on machine learning*, 39639–39659. PMLR.

Yang, Y.; Guan, Z.; Li, J.; Zhao, W.; Cui, J.; and Wang, Q. 2021. Interpretable and efficient heterogeneous graph convolutional network. *IEEE Transactions on Knowledge and Data Engineering*, 35(2): 1637–1650.

Yu, P.; Fu, C.; Yu, Y.; Huang, C.; Zhao, Z.; and Dong, J. 2022. Multiplex heterogeneous graph convolutional network. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 2377–2387.

Zeng, Z.; Luo, M.; Kong, X.; Liu, H.; Guo, H.; Yang, H.; Ma, Z.; and Zhao, X. 2024. Mitigating world biases: A multimodal multi-view debiasing framework for fake news video detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6492–6500.

Zhang, C.; Song, D.; Huang, C.; Swami, A.; and Chawla, N. V. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD conference on knowledge discovery and data mining*, 793–803.