



## 多双曲空间传递图解耦表示学习

郑帅, 彭奏章, 朱振峰, 赵耀

引用本文:

郑帅, 彭奏章, 朱振峰, 等. 多双曲空间传递图解耦表示学习[J]. *智能系统学报*, 2025, 20(5): 1243–1255.

ZHENG Shuai, PENG Zouzhang, ZHU Zhenfeng, et al. Graph disentanglement representation learning based on propagation in multiple hyperbolic spaces[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(5): 1243–1255.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202409034>

## 您可能感兴趣的其他文章

### 利用残差密集网络的运动模糊复原方法

Image restoration with residual dense network

智能系统学报. 2021, 16(3): 442–448 <https://dx.doi.org/10.11992/tis.201912002>

### 一种深度自监督聚类集成算法

A deep self-supervised clustering ensemble algorithm

智能系统学报. 2020, 15(6): 1113–1120 <https://dx.doi.org/10.11992/tis.202006050>

### 基于级联宽度学习的多模态材质识别

Cascade broad learning for multi-modal material recognition

智能系统学报. 2020, 15(4): 787–794 <https://dx.doi.org/10.11992/tis.201908021>

### 基于Capsule网络的甲骨文构件识别方法

Recognition of Oracle Radical based on the Capsule network

智能系统学报. 2020, 15(2): 243–254 <https://dx.doi.org/10.11992/tis.201904069>

### 基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

### 视觉同时定位与地图创建综述

A survey of VSLAM

智能系统学报. 2018, 13(1): 97–106 <https://dx.doi.org/10.11992/tis.201703006>

DOI: 10.11992/tis.202409034

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20250623.1427.002>

## 多双曲空间传递图解耦表示学习

郑帅<sup>1,2</sup>, 彭奏章<sup>1,2</sup>, 朱振峰<sup>1,2</sup>, 赵耀<sup>1,2</sup>

(1. 北京交通大学 计算机科学与技术学院, 北京 100044; 2. 现代信息科学与网络技术北京市重点实验室, 北京 100044)

**摘要:** 现有的图表示学习中存在两个问题, 一是缺乏对细粒度邻域建模的考量, 忽略了邻域节点间纠缠的多重语义; 二是图表示学习的空间度量问题, 欧氏空间并非是度量节点表示的最优选择。为解决上述问题, 提出一种多双曲空间下表征传递预测的全新架构, 实现多双曲空间度量下的图解耦表示学习。在该架构下, 通过节点表征将原始拓扑结构映射至双曲空间, 获得双曲空间下的多分辨率传递矩阵。进一步地, 基于混合专家结构设计, 视不同分辨率的双曲标签传递网络为专家网络, 从而发现由不同潜在因素引发的节点连接模式。在多个真实世界的数据集上的实验结果显示, 本文方法在 Squirrel 和 Crocodile 数据集上分别达到 32.3% 和 59.5% 的分类准确率, 可视化实验进一步证明了方法的有效性。

**关键词:** 图表示学习; 图解耦; 双曲空间; 图神经网络; 标签传递; 混合专家系统; 拓扑细化; 多分辨率

**中图分类号:** TP18 **文献标志码:** A **文章编号:** 1673-4785(2025)05-1243-13

中文引用格式: 郑帅, 彭奏章, 朱振峰, 等. 多双曲空间传递图解耦表示学习 [J]. 智能系统学报, 2025, 20(5): 1243-1255.

英文引用格式: ZHENG Shuai, PENG Zouzhang, ZHU Zhenfeng, et al. Graph disentanglement representation learning based on propagation in multiple hyperbolic spaces[J]. CAAI transactions on intelligent systems, 2025, 20(5): 1243-1255.

## Graph disentanglement representation learning based on propagation in multiple hyperbolic spaces

ZHENG Shuai<sup>1,2</sup>, PENG Zouzhang<sup>1,2</sup>, ZHU Zhenfeng<sup>1,2</sup>, ZHAO Yao<sup>1,2</sup>

(1. School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China; 2. Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China)

**Abstract:** There are two salient issues of existing graph representation learning methods. First, there is a dearth of fine-grained neighborhood modeling, which neglects the multifaceted semantic entanglements in the neighborhood structures. Second, the spatial metric employed in graph representation learning presents a significant challenge, since Euclidean space may not constitute the optimal framework for quantifying node representations. To solve these challenges, this study proposes a novel representation propagation and prediction mechanism within multiple hyperbolic spaces, thereby achieving disentangled graph representation learning under multifaceted hyperbolic spatial metrics. Within the proposed framework, the original topological structure is iteratively refined through node representations, yielding propagation matrices embedded in a hyperbolic space. Furthermore, based on a mixture-of-experts design, hyperbolic label propagation networks at different resolutions are treated as expert networks, enabling the discovery of node connection patterns induced by different latent factors. Experimental results on multiple real-world datasets show that the proposed method achieves classification accuracies of 32.3% and 59.5% on the Squirrel and Crocodile datasets, respectively. Additionally, visualization experiments further demonstrate the effectiveness of the proposed approach.

**Keywords:** graph representation learning; graph disentanglement; hyperbolic space; graph neural networks; label propagation; mixture of experts; topology refinement; multiresolution

收稿日期: 2024-09-26. 网络出版日期: 2025-06-23.

基金项目: 中央高校基本科研业务费项目 (2024XKRC088); 国家自然科学基金项目 (62476022); 北京市自然科学基金青年科学基金项目 (4254085).

通信作者: 朱振峰. E-mail: [zhfzhu@bjtu.edu.cn](mailto:zhfzhu@bjtu.edu.cn).

作为一种被用来描述多个实体之间的关系的结构型数据, 图在各类应用场景中广泛存在, 如社交网络、引文网络和分子网络。得益于消息传

递架构发挥出的强大效能,基于图神经网络 (graph neural networks, GNNs) 的图表示学习方法已经成为当下的图数据分析与处理的研究热点<sup>[1]</sup>,并在推荐系统<sup>[2-3]</sup>、医疗及药物预测<sup>[4-6]</sup>等多个方面皆有所建树。图表示学习的主要思想是对原始图数据的拓扑结构和节点特征进行联合建模,进而获得图的嵌入表示,以便于用于下游任务,如节点分类 (node classification)<sup>[7]</sup>、图分类 (graph classification)<sup>[3]</sup> 和连接预测 (link prediction)<sup>[5]</sup>。然而,在当前常用的消息传递神经网络 (message passing neural networks, MPNN)<sup>[7]</sup> 框架下,许多图神经网络模型<sup>[8-9]</sup> 将局部邻域或节点路径作为整体感知对象。值得注意的是,这种局部邻域整体感知的方式缺乏对导致节点连接的多个潜在因子间的纠缠与耦合关系的充分挖掘。这使得模型学习到的节点表征难于刻画与该节点关联的内在多重语义信息,从而限制模型的准确性与泛化性。

为解决上述问题,越来越多的研究工作开始进行图解耦表示学习研究。当前图解耦表示学习有两种常用的解耦方法,一种是基于邻域路由的方法<sup>[10]</sup>,另一种是基于因子图分解的方法<sup>[11]</sup>。上述两类方法均是在欧氏空间中进行解耦表示学习。然而,近期的研究<sup>[12-14]</sup>指出,欧氏空间下的图表示学习会得到扭曲的结果,而双曲空间的指数容纳能力,更为适合以指数级增长子节点个数的图数据进行表示学习<sup>[15-16]</sup>。目前,鲜有工作研究双曲空间下的图解耦表示学习方法。

此外,目前已存在的图解耦表示学习方法并未考虑训练集节点数量的限制。通常来说,在训练集占比较小的情况下,为使得模型捕获全局信息进行节点建模,模型需要使用足够大的感受野来聚合节点信息和长程依赖,而这往往依托于深层图网络。但是,有研究表明<sup>[17]</sup>,在加深网络时,通过使用上述的方式可能导致过平滑问题的出现。为解决这一问题,受启发于 PageRank<sup>[18]</sup> 方法,一些架构解耦的图神经网络方法,如 AP-GCN(adaptive propagation graph convolutional network)<sup>[19]</sup>、APPNP(approximate personalized propagation of neural predictions)<sup>[20]</sup> 等,尝试分离图卷积网络 (graph convolutional network, GCN)<sup>[21]</sup> 中的特征学习与传递步骤,取得了较为不错的效果。尽管这类方法在过平滑问题上表现较好,但仍存在两个问题: 1) 仅使用图数据的非参数化基本拓扑结构进行标签传递,限制了模型的能力,可能导致局部最优解<sup>[22]</sup>; 2) 未能考虑邻域构成的复杂因素,因此直接传递可能受多种纠缠因素的影响,

限制模型的性能。

为了解决上述问题,本文提出了多双曲空间传递预测 (multi-hyperbolic space propagation of neural predictions, MHPP) 模型。具体来说, MHPP 首先通过学习到的节点表征与给定的邻接矩阵,在双曲空间中构建了新的参数化传递矩阵,以此细致化地反映节点之间的关系,对原始的拓扑结构进行一定程度的修正。接着, MHPP 通过多个不同曲率双曲空间来反映导致节点连接的复杂混合潜在因子,使得每个特定曲率的双曲空间对应一个潜在因子,并由不同因素形成各异的传递路径,达到解耦的效果。在此基础上, MHPP 引入混合专家结构 (mixture of experts, MoE), 以便于在强化重要因子及弱化无关因子影响的同时,提升模型泛化性能。本文工作的主要贡献如下:

1) 将图解耦表示学习从欧氏空间拓展到双曲空间,解决了现有解耦方法固有的表征扭曲问题。

2) 针对现有方法仅依赖非参数化拓扑结构的局限,首次提出多双曲空间解耦标签传递方法,结合节点表征与原始拓扑构建多分辨率双曲传递矩阵,以精细化挖掘节点关系的潜在因子;

3) 为挖掘与节点连接相关的直接潜在因子,创新性地引入混合专家结构对这些潜在因子空间进行精细化建模;

4) 在多个常用数据集上的两类实验验证了本文方法的有效性,在 Squirrel 和 Crocodile 数据集上节点分类准确率分别达到 32.3% 和 59.5%。

## 1 相关工作

本文所研究的模型涉及图解耦表示学习、双曲空间中的图表示学习和特征转换-传递解耦架构的图表示学习模型,因此将从上述 3 个方面进行相关工作的阐述。

### 1.1 图解耦表示学习

图解耦表示学习试图探索节点及其邻居中的潜在的节点交互成因及其邻域中存在的影响因素。受胶囊网络<sup>[23]</sup>启发, DisenGCN(disentangled graph convolutional network)<sup>[10]</sup> 将解耦表示学习方法引入至节点级图分类任务中,提出了针对图结构的动态路由方法以聚合邻域信息,并解耦潜在因素。该方法是研究图解耦表示学习方法的先驱。除了节点级解耦方式之外,一些工作从图其他方面的特征 (整图、边等) 对图解耦表示学习进行了研究。例如, FactorGCN(factorizable graph convolutional network)<sup>[11]</sup> 提出了一种等价于多关系解耦的图级解耦表示学习方法,即中心节点可

以聚合来自多个潜在关系中的不同邻域信息。DisGNN<sup>[24]</sup>探索了边级解耦表示学习方法。

在上述 3 种基础图解耦研究方法范式的基础上, 一些工作开展了深入研究来提升模型性能和解耦效果。例如, 在 DisenGCN<sup>[10]</sup> 的基础上, IPG-DN(independence promoted graph disentangled networks)<sup>[25]</sup>优化了图解耦的独立性约束, ADGCN<sup>[26]</sup>则提出了微观解耦和宏观解耦对抗正则, 以提高成分分布之间的可分离性。除此之外, 在异质关系考虑方面, DisenHAN(disentangled heterogeneous graph attention network)<sup>[27]</sup>利用异构图上的注意力机制来学习邻居的权重, 然后利用它们完成分层聚合, 即关系内聚合和关系间聚合。另外, 在自监督框架下, DSSL(decoupled self-supervised learning)<sup>[28]</sup>模型用于解决基于变异推理的潜因和编码参数的解耦学习问题, 其中解耦表示学习是通过解耦相同邻域之间的不同底层语义来实现的。

值得注意的是, 上述各类方法均沿用 DisenGCN 的胶囊网络架构。不同于这些方法, 本文创新地引入了混合专家系统架构, 它能够根据输入动态调整不同因子的重要性, 既保证了模型对关键因素的关注, 又能够抑制噪声因子的干扰, 从而提升模型的泛化能力。

## 1.2 双曲空间中的图表示学习

相较于传统的欧氏空间下的图表示学习, 在双曲空间中进行图表示学习的优势在于, 其指数增长的空间特性与指数增长的图数据相适配, 能够得到更优性能的模式。HGNC(hyperbolic graph convolutional network)<sup>[12]</sup>模型提出了适用于双曲空间的图卷积操作, 在双曲空间中该图卷积操作能够一并学习空间曲率。HGAT(hyperbolic graph attention network)<sup>[14]</sup>方法也重新设计了 GAT(graph attention network)<sup>[29]</sup>的图卷积操作, 并额外设计了双曲空间中的注意力机制计算范式以求得节点权重。HGNN(hyperbolic graph neural network)<sup>[13]</sup>除了设计双曲空间中的图卷积操作外, 还使用了径向基函数来学习整图表征。此外, 由于双曲空间与欧氏空间的特征变换导致双曲图表示学习网络往往有着较高的计算复杂度, 为了加速模型计算, HGCF(hyperbolic graph convolution networks for collaborative filtering)<sup>[30]</sup>在切平面上聚合时, 去除了特征变换和非线性激活来加速模型训练。

进一步地, 在双曲空间下的图表示学习研究中, 文献 [31] 证明单双曲空间对具有环状结构的图数据建模能力不佳。因此, 目前许多方法尝试研究使用多双曲空间对图进行解耦建模。原因在

于不同的双曲空间对图结构建模能力有差异, 通过多双曲空间对多个潜在因子进行建模, 能够结合不同空间的优势, 习得更为出色的节点表征。一些工作进行了多双曲空间图表示学习的研究, 其本质是对不同潜在因子进行解耦建模。例如,  $\kappa$ -GCN( $\kappa$ - graph convolutional network)<sup>[32]</sup>提出了多空间学习框架, 在积空间进行表征学习。GIL(geometry interaction learning)<sup>[33]</sup>同时使用欧氏空间和双曲空间对图进行建模, 并进行特征互补从而学习到更全面的表征。DHGAN(decoupled hyperbolic graph attention network)<sup>[34]</sup>方法通过使用两个双曲空间解耦原始数据中的可替代性(substitutable)关系与补充性(complementary)关系。

本小节介绍的双曲图表示学习方法主要关注双曲空间中的图操作设计, 而未考虑对底层拓扑结构的优化。本文通过节点表征来构建参数化的传递矩阵, 对原始拓扑结构进行优化, 从而更准确地刻画节点间的实际关系, 以辅助实现更有效的表示解耦。

## 1.3 基于特征转换-传递解耦架构图表示学习

在半监督图数据训练任务中, 利用足够大的感受野聚合节点信息和捕获长程依赖关系是提升模型性能的关键步骤。然而, 多项研究表明, 在 GCN 等常用架构中, 通过加深网络层数来捕获高阶邻域信息时, 往往会导致节点特征表示趋于相似, 即所谓的“过平滑”现象, 这种现象严重限制了图神经网络在处理复杂图结构时的表达能力和分类准确性。

因此, 为了避免在进行图表示学习时的过平滑现象, 一些工作将特征转换与传递进行了分离操作。作为此类方法的代表工作之一, APPNP<sup>[20]</sup>将特征平滑问题巧妙地转化成了标签传递问题。具体来说, 该模型通过 MLP(multilayer perceptron)学习特征, 并使用个性化页排名在图上进行多次伪标签传播, 从而实现更深层次的特征传播过程, 得到每个节点平滑稳定的标签, 以用于下游的节点分类任务。与 APPNP 相反, SGC(simplifying graph convolutional network)<sup>[10]</sup>提出先进行多次特征传播, 再进行特征转换。由于特征传播过程不涉及梯度反向传播, SGC 的解耦方式有效提高了模型的可扩展性。这种处理顺序的差异, 反映了不同的架构解耦策略和侧重点, 为后续工作探索提供了有益借鉴。进一步地, DAGNN(deep adaptive graph neural network)<sup>[35]</sup>通过分离特征学习与传递的步骤, 并独立地计算各个感受野保留分数, 来学习最终的节点表示。P&L+C&S(propaga-

tion with correction and smoothness)<sup>[36]</sup> 在数据集上训练一个基础预测器, 然后通过图上的信息传递来纠正它。此外, 部分工作调换了传递与训练的顺序, 比如 PTA(propagation then training adaptively)<sup>[37]</sup>、Meta-PN(meta propagation network)<sup>[38]</sup> 采用标签传递作为教师模型, 生成编码有价值的全局结构知识的伪标签。为进一步提升该架构下的模型性能, 王振胜<sup>[39]</sup> 提出了基于拓扑修改与基于维度划分的特征平滑算法, 在去除对信息传递造成影响的同时使节点特征能融合深层的结构信息。另外, 有工作已经开始尝试采用基于特征转换-传递解耦架构开展表示解耦工作, 如 RAM-CG

(graph relation aware continual graph learning)<sup>[40]</sup> 基于隐藏于边上的关系, 在多个隐因子空间构建因子相关的个性化赋权邻接矩阵实现边级解耦。

本文提出的多双曲空间与传递分离架构结合的图解耦表示学习研究是一个全新的思路, 其能够避免欧氏空间下表征学习的扭曲与失真, 实现精准的特征解耦建模。

## 2 多双曲空间传递预测模型

MHPP 模型架构如图 1 所示, 主要由结构无关的特征学习、多双曲空间解耦标签传递和混合专家引导下标签融合 3 个模块构成。

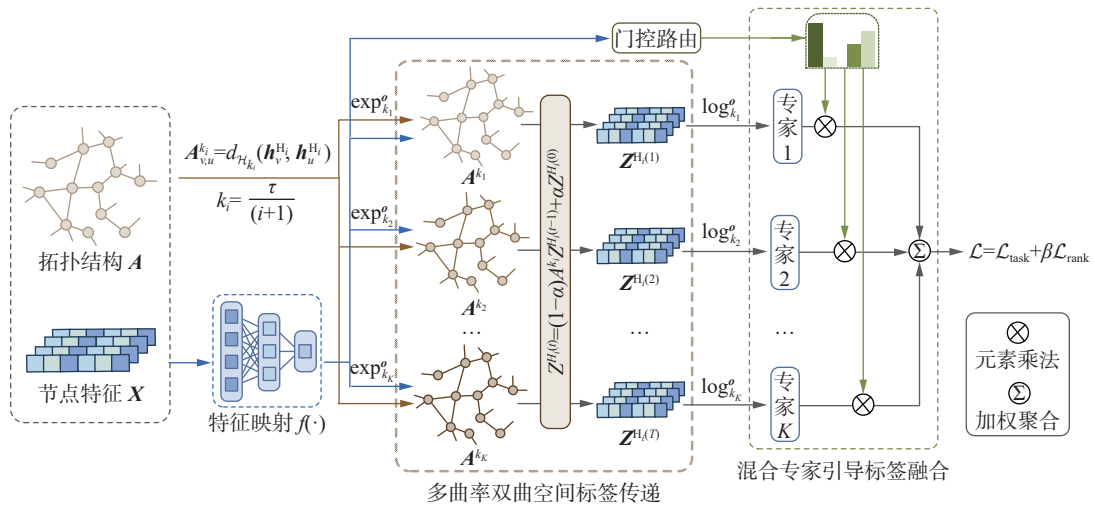


图 1 MHPP 模型架构

Fig. 1 Architecture of MHPP

### 2.1 符号定义

令  $G = (V, E)$  表示无向图, 其中  $V$  是节点集合, 其大小  $|V| = n$ , 并且  $E$  是边集。另外, 使用邻接矩阵  $A$  表示节点之间的连接关系, 当节点  $i$  与节点  $j$  存在边, 即  $(i, j) \in E$  时,  $A_{i,j} = 1$ ; 其他情况下  $A_{i,j} = 0$ 。此外,  $x_u \in \mathbf{R}^f, u = 1, 2, \dots, n$  记作第  $u$  个节点的原始特征, 维度为  $f$ ;  $h_u \in \mathbf{R}^d, u = 1, 2, \dots, n$  表示经过学习得到的第  $u$  个节点的  $d$  维表征。特别地, 在需要区分表征位于欧氏空间或者双曲空间时, 本文中 will 使用上标“E”表明该表征位于欧氏空间, 并且使用上标“H”记作表征处于双曲空间。

### 2.2 必要概念说明

**定义 1** 流形 (manifold)。流形是对曲面概念的推广, 通常使用一个  $n$  维流形  $M$  表示局部近似为  $n$  维欧氏空间  $\mathbf{R}^n$  的拓扑空间。

**定义 2** 某一点的切空间/切平面 (tangent space)。流形  $M$  上关于该点的切空间指在该点处所有切向量组成的向量空间。假设该点为点  $o$ , 那么可以将其切平面记作  $\mathcal{T}_o M$ 。

**定义 3** 指数映射 (exponential map)。用于将点从切空间映射至流形, 即  $\mathcal{T}_o M \rightarrow M$ 。本文将从原点为  $o$  的切空间映射至曲率为  $c$  的流形的指数映射符号记作  $\exp_c^o(\cdot)$ 。

**定义 4** 对数映射 (logarithmic map)。指数映射的逆运算, 用于将点从流形映射至切空间, 即  $M \rightarrow \mathcal{T}_o M$ 。本文将从曲率为  $c$  流形映射至原点为  $v$  的切空间的对数映射符号记作  $\log_c^v(\cdot)$ 。

设双曲空间  $M$  是一个有常数曲率  $c = -1/k, (k > 0)$  的维度为  $d$  的黎曼流形。将  $\langle x, y \rangle_L \rightarrow \mathbf{R}^d$  记作  $x, y \in \mathbf{R}^{d+1}$  两点的洛伦兹内积 (Lorentzian scalar product), 有

$$\langle x, y \rangle_L = -x_0 y_0 + \sum_{i=1}^d x_i y_i$$

这里使用  $\mathcal{H}_c^d$  表示曲率为  $c$  的  $d$  维双曲面流形:

$$\mathcal{H}_c^d = \{x \in \mathbf{R}^{d+1} : \langle x, x \rangle_L = -k, x_0 > 0\}$$

根据定义 2 可以在  $\mathcal{H}_c^d$  上  $x$  处得到一个无限接近  $\mathcal{H}_c^d$  的切空间  $\mathcal{T}_x M$ :

$$\mathcal{T}_x \mathcal{H}^d = \{v \in \mathbf{R}^{d+1} : \langle v, x \rangle_L = 0\}$$

此外, 对于任意两点  $\mathbf{x}, \mathbf{y} \in \mathcal{H}_c^d$ , 距离函数为

$$d_{\mathcal{H}_c^d}(\mathbf{x}, \mathbf{y}) = \sqrt{k} \operatorname{arcosh} \left( -\frac{\langle \mathbf{x}, \mathbf{y} \rangle_L}{k} \right)$$

基于该双曲面  $\mathcal{H}_c^d$  与切平面  $\mathcal{T}_x \mathcal{H}_c^d$ , 可以构造指数映射与对数映射:

$$\begin{aligned} \exp_c^x(\mathbf{y}) &= \sqrt{k} \cosh \left( \frac{\|\mathbf{y}\|_L}{\sqrt{k}} \right) \mathbf{x} + \sqrt{k} \sinh \left( \frac{\|\mathbf{y}\|_L}{\sqrt{k}} \right) \frac{\mathbf{y}}{\|\mathbf{y}\|_L} \\ \log_c^x(\mathbf{y}) &= \sqrt{k} \operatorname{arcosh} \left( -\frac{\langle \mathbf{x}, \mathbf{y} \rangle_L}{k} \right) \frac{\mathbf{y} + \frac{1}{k} \langle \mathbf{x}, \mathbf{y} \rangle_L \mathbf{x}}{\left\| \mathbf{y} + \frac{1}{k} \langle \mathbf{x}, \mathbf{y} \rangle_L \mathbf{x} \right\|_L} \end{aligned}$$

式中  $\|\mathbf{y}\|_L = \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle_L}$  是  $\mathbf{v}$  的洛伦兹模。

### 2.3 多双曲空间解耦标签传递

直觉上来说, 节点连接的构成因素通常是隐

晦而复杂的, 想要深入挖掘和使用这些丰富的潜在信息, 解耦无疑是一个有效的方法。欧氏空间中, 大多数工作往往采用矩阵变换来进行解耦, 如 DisenGCN<sup>[10]</sup> 采用矩阵变换和邻域路由完成解耦。而双曲空间中的图解耦方法却鲜有人研究。以往研究中, DHGAN<sup>[34]</sup> 方法通过使用两个双曲空间解耦原始数据中的可替代性关系与补充性关系。受此启发, 本文通过将数据映射到不同的曲率空间, 从而反映出不同的潜在因子对数据结构的影响。考虑到如表 1 所示的双曲空间的优秀性质, 以及图数据可以较低失真率嵌入到双曲空间的特性<sup>[12-14]</sup>, 本文将节点映射至双曲空间以进行表征度量。

表 1 非欧空间特性分析

Table 1 Analysis of non-Euclidean space characteristics

空间类型	曲率	几何特性	表征优势	表征局限性
双曲空间	负曲率	空间随距离指数级扩张	高效嵌入树状结构和幂律分布网络	对环状和规则网格状结构表征效率较低
球面空间	正曲率	空间随距离收缩	天然保持对称性, 适合球形结构建模	难以表示层次结构; 存在距离上界
黎曼流形	可变曲率	局部结构适应性强	能适应数据内在几何; 表征更精确	模型复杂度和计算成本极高

具体来说, 在双曲空间下, 空间曲率的变化会导致节点之间的表征距离发生改变, 一定程度上反映出节点之间在不同曲率空间下连接的紧密程度, 从而表现出隐因子的存在。受到上述动机的启发, 本文提出了多曲率双曲空间标签传递机制。为了尽可能涵盖多个潜在因子, MHPP 从曲率负倒数最大值为  $\tau$  的情况下选取了  $K$  个双曲空间, 并设定第  $i$  个空间曲率负倒数为  $k_i = \tau / (i + 1)$ ,  $i = 1, 2, \dots, K$ 。

为了使节点特征在双曲空间可运算, 这里首先使用指数映射将特征投影到双曲空间。令  $\mathbf{o} = (\sqrt{K}, 0, \dots, 0) \in \mathcal{H}^d$  记作双曲空间  $\mathcal{H}^d$  的原点。可以通过下式将任意节点  $v \in V$  投影到曲率为  $k_i$  的双曲空间:

$$\mathbf{h}_v^{\mathcal{H}_i} = \exp_{k_i}^{\mathbf{o}}(\mathbf{h}_v^{\mathcal{E}})$$

式中  $\mathbf{h}_v^{\mathcal{E}} = \text{MLP}(\mathbf{x}_v) \in \mathbf{R}^c$  是对节点  $v$  的原始特征  $\mathbf{x}_v$  进行映射变换后在欧氏空间中得到的  $c$  维节点嵌入。

根据经过投影得到的双曲空间下节点表示结合原始的邻接矩阵  $\mathbf{A}$  构建每个双曲空间对应的双曲传递矩阵  $\mathbf{A}^{k_i} \in \mathbf{R}^{n \times n}$ :

$$\mathbf{A}_{v,u}^{k_i} = d_{\mathcal{H}_i}(\mathbf{h}_v^{\mathcal{H}_i}, \mathbf{h}_u^{\mathcal{H}_i}), \mathbf{A}_{v,u} = 1$$

式中  $v, u \in V$ 。在得到因子相关的个性化传递矩阵之后, 为了整合输入节点特征和因子信息, 并降低模型复杂度, MHPP 通过幂迭代来更高效地进

行标签传递:

$$\begin{aligned} \mathbf{z}_v^{\mathcal{H}_i(0)} &= \mathbf{h}_v^{\mathcal{H}_i} \\ \mathbf{z}_v^{\mathcal{H}_i(t)} &= (1 - \alpha) \mathbf{A}^{k_i} \mathbf{z}_v^{\mathcal{H}_i(t-1)} + \alpha \mathbf{z}_v^{\mathcal{H}_i(0)} \end{aligned} \quad (1)$$

式中  $t \in \{1, 2, \dots, T\}$  表示传递次数,  $T$  为总传递次数;  $\alpha$  是用于调节邻域影响大小的超参数。

由于一般在欧氏空间处理下游任务 (如节点分类等), 因此本文所述模型在  $T$  次传递完成后, 将节点表示  $\mathbf{z}_v^{\mathcal{H}_i(T)}$  从曲率为  $k_i$  的双曲空间再次映射回欧氏空间, 记为  $\mathbf{z}_v^{\mathcal{E}_i} \in \mathbf{R}^c$ :

$$\mathbf{z}_v^{\mathcal{E}_i} = \log_{k_i}^{\mathbf{o}}(\mathbf{z}_v^{\mathcal{H}_i(T)})$$

上文对节点为  $v$ 、曲率为  $k_i$  的双曲空间进行了详细论述, 通过相同的流程, 可以在各个双曲空间中都习得所有节点的表征。同样以节点  $v$  为例, 经过多双曲空间下的传递之后, 单个节点的最终结果会包含各空间下的结果, 构成多曲率传递结果矩阵, 即  $\mathbf{z}_v^{\mathcal{E}} = [\mathbf{z}_v^{\mathcal{E}_1}, \mathbf{z}_v^{\mathcal{E}_2}, \dots, \mathbf{z}_v^{\mathcal{E}_K}] \in \mathbf{R}^{K \times c}$ 。

### 2.4 混合专家引导下标签融合

为了提升模型的训练效率, 本文通过分治的思想来处理图解耦表示学习问题, 即通过模块化的多曲率双曲空间传递, 针对每一个潜在因子进行独立的标签传递, 传递彼此之间互不干扰, 仅需考虑该潜在在双曲因子空间内的信息。值得注意的是, 在分别学习到每个因子对应的节点标签之后, 如何将这些标签结合得到最终的节点标签仍

然是该模型所面临的一大问题。目前许多图解耦表示学习方法<sup>[10-11]</sup>采用直接拼接 (concatenate) 方法, 这是一种直观的方法, 但是该方法对不同因子对节点影响的建模准确度不足, 缺乏对各因子重要性的显著性描述。由此, 本研究将通过在因子空间进行精细化加权来解决这一问题。

如果将多个双曲传递网络视为多个专家网络, 并使用一个额外的门网络控制各专家网络的重要性, 再将各专家模块的学习结果聚合到一起, 那么就可以在聚合节点在各空间信息的同时量化各空间对节点的影响大小。因此, 令 MHPP 中 MoE 的  $K$  个专家模块由  $K$  个 MLP 构成, 每个专家网络处理一个特定曲率对应的曲率空间传递结果  $\mathbf{z}_v^E \leftarrow \text{MLP}(\mathbf{z}_v^E)$ , 并构建一个门网络用于学习各专家网络的混合权值, 以捕获节点与因子空间之间的内在关联强度。具体地, 模型基于欧氏空间下的特征规整结果  $\mathbf{h}_v^E$  计算每个专家系统的权值:

$$p_i = \sigma(\mathbf{W}_i \mathbf{h}_v^E)$$

式中:  $\sigma(\cdot)$  是激活函数;  $\mathbf{W}_i \in \mathbf{R}^{1 \times c}$  是每个门网络的参数, 各空间独立但被所有节点共享。

进一步地, 针对每个空间, MPHPP 使用 Softmax 函数将门网络输出的权值归一化为对应的分配概率  $p_i$ , 即相应专家网络的分配权重:

$$p_i = \frac{\exp(p_i)}{\sum_{j=1}^K \exp(p_j)}$$

最后, 模型基于门网络得到的专家网络分配概率向量  $\mathbf{p}_v = [p_1, p_2, \dots, p_K] \in \mathbf{R}^{1 \times K}$  对节点在各个因子空间内的平滑标签进行加权聚合, 得到节点的最后结果  $\mathbf{z}_v^{E*}$ :

$$\mathbf{z}_v^{E*} = \mathbf{p}_v \mathbf{z}_v^E = \sum_{i=1}^K p_i \mathbf{z}_v^{E_i}$$

## 2.5 模型优化

在现实网络中, 同类节点在各个因子空间的重要性分布通常呈现相似模式, 且节点表征往往由少数关键因子主导, 这是网络数据的内在特性。引入门控结构后, 能够通过权值直观地量化并反映这一特性, 实现对因子重要性的精确建模。因此, 为了确保同类节点对相同的因子空间响应趋近一致, 并且对空间响应进行稀疏性约束, 本模型引入了秩约束来控制各节点的门网络权值:

$$\mathcal{L}_{\text{rank}} = (C - \text{rank}(\mathbf{P}))^2 \quad (2)$$

式中:  $C$  是节点类别数量,  $\mathbf{P} \in \mathbf{R}^{n \times K}$  是所有节点的门网络权值矩阵。这种约束不仅能够增强模型对

类内节点相似性的捕获能力, 还能促使模型关注真正重要的因子空间, 减少冗余信息的干扰, 从而提高表征的判别性和模型的泛化能力。通过这种机制, 模型能够更准确地反映现实网络中节点与因子之间的复杂关联模式。特别地, 由于秩约束是非凸且不可微的, 因此, 在具体实现过程中, 采用迹范数  $\|\mathbf{P}\|_{\text{tr}}$  以替代  $\text{rank}(\mathbf{P})$ 。

此外, 在得到节点最终表征矩阵  $\mathbf{Z}^E$  后, 可以通过 MLP 构成的分类器完成由节点表征到节点标签的映射, 并通过交叉熵损失构建分类损失  $\mathcal{L}_{\text{task}}$ :

$$\mathbf{Y}' = \text{MLP}(\mathbf{Z}^E)$$

$$\mathcal{L}_{\text{task}} = \text{CrossEntropy}(\mathbf{Y}, \mathbf{Y}')$$

式中:  $\mathbf{Y} \in \mathbf{R}^{n \times c}$  是节点真实标签,  $\mathbf{Y}'$  为模型的预测结果。从上面论述的整体架构中不难得出模型的总损失:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \beta \mathcal{L}_{\text{rank}}$$

式中  $\beta$  是用于控制秩损失重要性的超参数。

## 3 实验及结果分析

本章首先对实验所用数据集以及实施细节进行介绍, 其次对比本文方法与现有的其他图表示学习方法, 最后对本文方法的网络结构和超参数进行实验分析。

### 3.1 实验设计

#### 3.1.1 数据集

本文使用了 10 个真实数据集, 数据集的详细信息见表 2。其中, 3 个引文图数据集 (Cora、CiteSeer、PubMed) 的节点和边分别表示文章和引用关系。此外, 还引入了几个非同配图, 包括 Texas、Wisconsin、Actor、Squirrel、Chameleon、Cornell 和 Crocodile。

表 2 数据集明细总览  
Table 2 Dataset overview

数据集名称	节点数量	边数量	类别数	特征维度	同配率
Cora	2 708	5 278	7	1 433	0.81
CiteSeer	3 327	4 552	6	3 703	0.74
PubMed	19 717	44 324	3	500	0.80
Texas	183	295	5	1 703	0.11
Wisconsin	251	466	5	1 703	0.21
Actor	7 600	26 752	5	932	0.22
Squirrel	5 201	198 493	5	128	0.22
Chameleon	2 277	31 421	5	128	0.23
Cornell	183	280	5	1 703	0.30
Crocodile	11 631	360 040	5	128	0.24

#### 3.1.2 对比方法

为了评估 MHPP 的性能, 本文选择了多种方

法进行对比: 1) 通用基线方法 MLP; 2) 经典图表示学习基线方法, 包括 GCN<sup>[21]</sup>、GraphSAGE<sup>[41]</sup>、GAT<sup>[29]</sup>、SGC<sup>[9]</sup>、VQGraph<sup>[42]</sup> 和 GraphACL(asymmetric contrastive learning for graphs)<sup>[43]</sup>; 3) 基于特征转换-传递解耦架构图表示学习方法, 包含 APPNP<sup>[20]</sup>、PTA<sup>[37]</sup> 和 GPR-GNN(generalized pagerank graph neural network)<sup>[44]</sup>; 4) 双曲图表示学习方法, HyLa<sup>[45]</sup>、HGCN(hyperbolic graph neural network)<sup>[13]</sup>; 5) 异质图基线方法, H<sub>2</sub>GCN<sup>[46]</sup>、JacobiConv<sup>[47]</sup>。

### 3.1.3 实现细节与参数设置

在实验中, 每个数据集的双曲空间数量被设定为 6。HyperOpt(hyperparameter optimization) 模块被用来当作优化器对其余超参数进行调整, 具体的超参数取值范围见表 3。此外, 对每个超参数的各个取值都进行 300 次的早停法训练, 采用 Adam 作为优化器, 并选择在验证集上的最优结果对应的超参数进行后续实验。

表 3 超参数设置  
Table 3 Hyperparameter setting

参数名称	取值范围
隐藏层嵌入维度	{24,32,40,48,56,64,72,80}
Dropout	[0,1]
学习率	{0.001,0.0025,0.005,0.01,0.025,0.05}
残差权重 $\alpha$	[0,1]
传递次数 $T$	[5,15]
秩损失权重 $\beta$	[e <sup>-10</sup> ,1]

## 3.2 模型应用

### 3.2.1 节点分类效果测试

为进行半监督节点分类实验以验证方法性能, 本研究在 6 个数据集中使用了训练集、验证集和测试集比例为 1:1:8 的随机划分方式。对未使用该划分方式的基线方法, 进行了微调和重训练。对比结果如表 4~5 所示。此外, 为了研究各个模型在高比例训练集上的性能, 对 4 个数据集遵从了 Geom-GCN<sup>[48]</sup> 的划分方式, 实验结果可见表 6。

表 4 同配图半监督节点分类任务准确率对比  
Table 4 Performance comparison of semi-supervised node classification accuracy on homophily datasets %

方法名称	Cora	CiteSeer	PubMed
MLP	72.4	65.9	82.3
GCN <sup>[39]</sup>	81.5	74.6	82.8
APPNP <sup>[20]</sup>	76.7	76.0	82.8

续表 4

方法名称	Cora	CiteSeer	PubMed
SGC <sup>[9]</sup>	74.1	74.6	74.9
PTA <sup>[35]</sup>	83.1	<b>76.8</b>	84.7
GPR-GNN <sup>[44]</sup>	<u>84.4</u>	72.2	86.4
HyLa <sup>[45]</sup>	82.3	68.5	81.1
HGCN <sup>[13]</sup>	78.6	62.8	79.2
JacobiConv <sup>[47]</sup>	<b>85.0</b>	74.1	<u>86.7</u>
H <sub>2</sub> GCN <sup>[46]</sup>	79.2	65.8	<u>86.7</u>
VQGraph <sup>[42]</sup>	82.4	76.3	73.5
GraphACL <sup>[43]</sup>	82.9	73.4	82.8
MHPP	83.7	<u>76.5</u>	<b>86.9</b>

注: 加粗代表本列最优结果, 下划线代表本列次优结果。

表 5 非同配图半监督节点分类任务准确率对比  
Table 5 Performance comparison of semi-supervised node classification accuracy on non-homophily datasets %

方法名称	Squirrel	Chameleon	Crocodile
MLP	24.2	39.8	54.7
GCN <sup>[39]</sup>	26.9	47.4	<u>57.0</u>
APPNP <sup>[20]</sup>	27.2	40.9	55.1
SGC <sup>[9]</sup>	26.0	40.6	51.5
PTA <sup>[35]</sup>	26.5	40.3	55.6
JacobiConv <sup>[47]</sup>	25.4	<b>52.2</b>	—
H <sub>2</sub> GCN <sup>[46]</sup>	<u>27.8</u>	45.0	—
MHPP	<b>32.3</b>	<u>47.8</u>	<b>59.5</b>

注: 加粗代表本列最优结果, 下划线代表本列次优结果, “—”表示数据缺失。

表 6 非同配图(高比例训练集)半监督节点分类任务准确率对比

Table 6 Performance comparison of semi-supervised node classification accuracy on non-homophily datasets (high proportion training sets) %

方法名称	Texas	Wisconsin	Actor	Cornell
GCN <sup>[39]</sup>	59.4	59.8	30.2	57.0
GAT <sup>[41]</sup>	58.3	55.2	26.2	58.9
GraphSAGE <sup>[40]</sup>	82.4	81.1	34.2	75.9
H <sub>2</sub> GCN <sup>[46]</sup>	<u>84.8</u>	<b>86.6</b>	<u>35.8</u>	<b>82.1</b>
DSSL <sup>[28]</sup>	62.1	62.2	28.1	53.1
GraphACL <sup>[43]</sup>	71.0	69.2	30.0	59.3
MHPP	<b>85.0</b>	<u>81.2</u>	<b>35.9</b>	<u>78.1</u>

注: 加粗代表本列最优结果, 下划线代表本列次优结果。

表 4 给出了 3 个同配图数据集上的分类结果。可以发现, MHPP 在降低了训练集比例的同配图, 相较于两种同类方法(即 APPNP 和 PTA), 能够在两种同配率相对高的情况下取得更好的性

能。比如在同配率为 0.81 的 Cora 数据集上, MHPP 分别超过 APPNP 与 PTA 7、0.6 百分点。此外, 对比其他的方法, MHPP 也表现出了一定的竞争力。

表 5 给出了在 3 个非同配图数据集上的分类结果, MHPP 由于加入了解耦模块, 能够处理非同配图这种复杂的数据结构, 因此较同类方法 APPNP 和 PTA 有较强的性能。值得一提的是, MHPP 也在 Squirrel 数据集上大大超过了专门处理异质图的方法 JacobiConv 和 H<sub>2</sub>GCN。

表 6 中, 在 3 个小的非同配图数据集 (Texas、Wisconsin 和 Cornell) 上可以看出, 当训练集数据过多时, MHPP 表现出较差的性能; 但是在大数据集 Actor 上, MPH 却超过已有最好模型 H<sub>2</sub>GCN 0.1 百分点。由此可以判断, 当图较小时, 标签引入得过多会导致已知节点信息较多, 并且较少的传递次数就能完成节点特征的全局的传递, 这可能导致特征中的噪声反复叠加, 进而影响模型学习节点表示的性能。

### 3.2.2 节点聚类效果测试

本研究使用节点聚类任务来进一步评估模型的表示学习能力。具体来说, 使用 K-Means 方法聚类模型学习到的节点表示。表 7 给出了在指标 ACC(accuracy)、NMI(normalized mutual information) 和 ARI(adjusted rand index) 下的节点聚类性能对比。

表 7 节点聚类实验  
Table 7 Results of node clustering %

方法	Cora			CiteSeer		
	ACC	NMI	ARI	ACC	NMI	ARI
GCN <sup>[39]</sup>	63.9	49.8	42.9	52.4	38.7	31.3
HGCN <sup>[14]</sup>	53.8	40.6	31.5	—	—	—
APPNP	<u>72.7</u>	<u>59.7</u>	<u>51.9</u>	<u>61.7</u>	<u>45.1</u>	<u>41.5</u>
MHPP	<b>83.6</b>	<b>67.4</b>	<b>68.3</b>	<b>62.7</b>	<b>47.1</b>	<b>44.0</b>

注: 加粗代表本列最优结果, 下划线代表本列次优结果, “—”表示数据缺失。

本研究采用了经典的 GCN 和双曲空间下的 HGCN, 以及传递分离的 APPNP 与 MHPP 在双曲空间中传递的结果进行对比。每个模型都使用在分类前最后一个隐藏层的节点表示, 其中, 双曲空间下的表征被投影回欧氏空间中进行聚类分析。

从表 7 中可以看出, 相较于传递分离未分开的图卷积网络, 传递分离图网络在学习到的节点表示性能往往更加优秀。举例来说, 在 Cora 数据集上, APPNP 能够在指标 ACC 上超过经典的 GCN 8.8 百分点, 在 NMI 指标上提升达 9.9 百分点。然

而, MHPP 在所选的 2 个数据集上都进一步地超越了 APPNP 的性能, 其中, 在 Cora 数据集上的 ACC 指标提升达到 9.9 百分点。由此可知, MHPP 学习到的节点表征在确保可分类能力足够的前提下, 簇状更加明显, 即类内方差较小, 体现出了足够强的泛化性能。实验结果也进一步说明, 相比于欧氏空间度量, 双曲空间度量更适合描述具有层次性和非欧几何特性的结构化数据, 使得基于双曲度量的传递矩阵可以为模型提供更有效的信息传递。

### 3.2.3 节点嵌入可视化效果测试

本研究使用 Cora 数据集作为示例, 欧氏空间中的表示学习方法直接通过主成分分析 (principal component analysis, PCA) 降维, 双曲空间中的表示学习方法通过映射回欧氏空间后进行主成分分析降维可视化。本研究可视化了经典的欧氏空间图网络 GCN、双曲空间图网络 HGCN, 以及 MHPP 的节点嵌入表示, 结果如图 2 所示。可以看出, 双曲空间下图神经网络学习到的节点表示往往比欧氏空间更加具有区分性, 这直接证明了双曲空间下进行图表示学习的优势。进一步地, 可以看出本研究所提出模型在双曲空间下得到的可视化结果中, 比 HGCN 各类节点簇状更加明显, 即嵌入表示的方差更小, 说明关注节点连接复杂因素并排除不相干的因素, 能够更好地将同类节点与异类节点的嵌入表示分离。这不仅体现了 MHPP 强大的节点嵌入表示学习能力, 更印证模型细化节点嵌入表示学习能力。

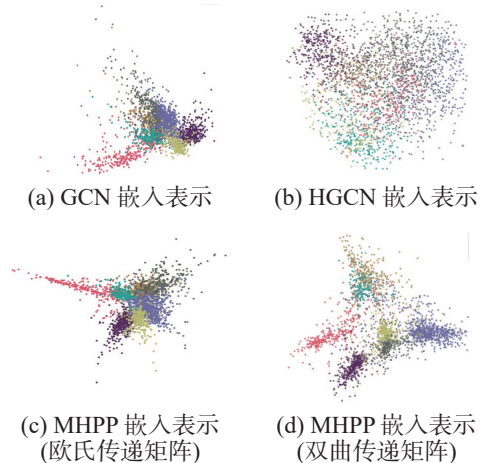
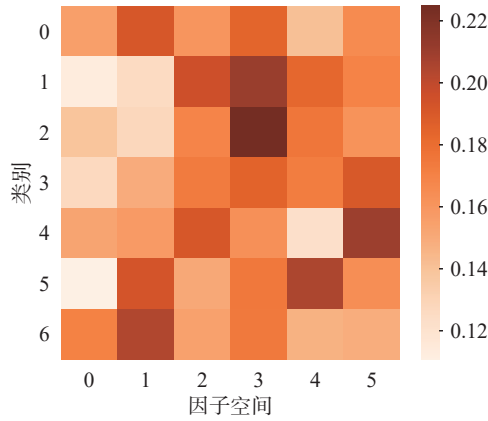


图 2 基于 PCA 降维的节点嵌入可视化  
Fig. 2 Visualization of node embedding based on PCA

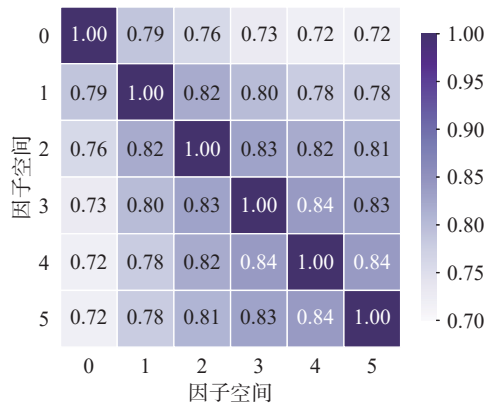
### 3.2.4 解耦有效性分析

为了对 MHPP 的解耦性能进行评估, 本小节对各因子空间的皮尔逊相关系数进行可视化, 并给出了类别相关的门网络权值可视化, 以说明各个隐因子空间的相似性, 结果见图 3。具体来说,

通过对节点在各因子空间中的嵌入进行平均池化 (average pooling) 的方式获得各个因子空间的特征, 并依此进行因子空间之间的皮尔逊相关系数计算。另外, 按照节点类别进行门控权值的平均池化, 并基于此进行门控权值可视化。



(a) 类别平均的因子空间门网络权值可视化



(b) 因子空间皮尔逊相似度可视化

图 3 因子空间解耦可视化

Fig. 3 Visualization of factor space disentanglement

式 (2) 约束同类节点只与少数相关因子连接而非所有因子, 符合节点因子构成假设, 图 3(b) 的可视化结果证明了约束的有效性。为保持模型空间个数设定的高自由度, 因子空间构造时仅预设最大和最小曲率, 根据超参数设定空间数量, 并用线性插值法设定每个空间曲率。图 3 显示相邻空间相关性略高, 远距空间相关性较低, 与模型设定一致。这表明当增大曲率间隔并使用足够多的双曲空间时, 模型能解耦潜在因子, 从多因子角度捕获信息完成图建模。实验结果表明, 当前使用的 6 个因子空间已基本实现多角度因子建模。

### 3.3 消融与超参数实验

#### 3.3.1 消融实验

为了验证模型中每个模块的有效性, 本小节开展消融研究, 在 3 个数据集上分别进行了半监督节点分类实验, 结果如表 8 所示。

表 8 消融分析与性能增益

Table 8 Ablation analysis and performance gain

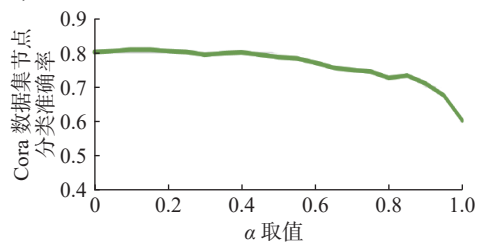
双曲混合专 递矩阵家聚合	Cora/ %	增益/ 百分点	CiteSeer/ %	增益/ 百分点	PubMed/ %	增益/ 百分点	
—	—	76.7	—	76.0	—	82.8	
√	—	79.5	+2.8	76.1	+0.1	85.2	+2.4
—	√	81.4	+4.7	75.9	-0.1	84.6	+1.8
√	√	83.7	+7.0	76.5	+0.5	86.9	+4.1

可以看出, 直接将原始邻接矩阵替换为双曲传递矩阵, 会带给模型较大的性能增益, 如在 PubMed 数据集上从 82.8% 提升到了 85.2%。这能够证明 MHPP 模型在一定程度上优化了非参数化的传递矩阵带来的局部最优解。此外, 在欧氏空间而非双曲空间学习传递矩阵的基础上, 添加混合专家聚合模块, 同样使得模型性能得到提升, 如在 Cora 数据集上准确率提升 4.7 百分点。这一设置可等价为一个多 GCN 的集成模型, 说明本模型能够较好地利用多因子空间进行集成学习, 以增加学习得到的信息。另外, 同时添加这两个模块能进一步提升模型的性能, 这证明了模型所添加额外模块的效果。

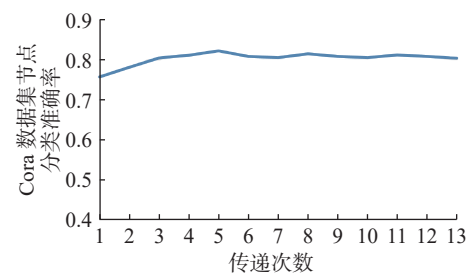
此外, 需要注意的是, 在 CiteSeer 上所添加部分模块的效果不够理想。由本文提出的方法架构可以发现, 这是因为 CiteSeer 数据集中存在孤立节点导致的。对于孤立节点而言, 其难以从标签传递过程中收到来自其他节点信息, 因此, 会在一定程度上影响传递矩阵构建, 进一步影响全局的特征学习。

#### 3.3.2 超参数实验

本小节通过 Cora 数据集上的半监督节点分类实验结果来探讨一些超参数 (包括  $\alpha$  取值和传递次数) 对模型性能的影响, 结果如图 4 所示。



(a)  $\alpha$  取值影响



(b) 传递次数影响

图 4 超参数对模型准确率影响

Fig. 4 Influence of hyper-parameters on ACC

从式 (1) 得知,  $\alpha$  越大, 邻域影响越小。由图 4(a) 可以看出, 当  $\alpha$  为 1 时 (相当于仅通过结构无关的特征学习) 模型性能较差, 证明了图结构在图表征学习中的重要性。通过图 4(b) 可以看出, 本模型使用架构在双曲空间下进行解耦学习时能够缓解过拟合问题, 即较大的传递次数仅会使模型性能逐步收敛而非过拟合。从上述实验可以看出, MHPP 中超参数变换对于模型性能影响较为微小, 这反映出了模型的稳定性。

### 3.4 参数化传递性能分析

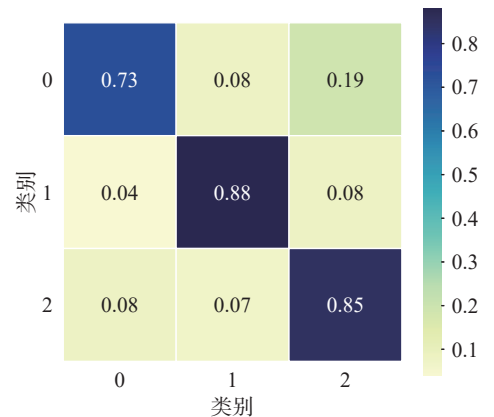
为验证参数化传递矩阵的有效性, 本研究在 3 个常用的引文数据集上对 SGC、APPNP、GPR-GNN 以及 PTA 模型进行了传递矩阵替换实验, 结果见表 9。需说明的是, 原始拓扑指使用归一化邻接矩阵进行传递。由于所选方法均在欧氏空间中进行图表示学习, 因此直接使用了欧氏空间下的节点嵌入参与参数化传递矩阵构建。

表 9 参数传递性能分析  
Table 9 Parametric transfer performance analysis

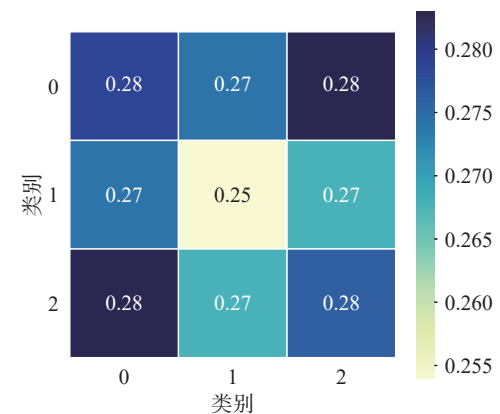
模型	Cora			CiteSeer			PubMed		
	原始拓扑/%	欧氏传递/%	增益/百分点	原始拓扑/%	欧氏传递/%	增益/百分点	原始拓扑/%	欧氏传递/%	增益/百分点
SGC	72.6	74.1	+1.5	74.6	75.1	+0.5	76.2	76.4	+0.2
APPNP	85.1	85.6	+0.5	75.8	76.3	+0.5	77.9	77.3	-0.6
GPRGNN	80.0	79.8	-0.2	67.0	67.0	0	84.2	84.9	+0.7
PTA	85.7	86.3	+0.6	75.7	76.0	+0.3	75.8	75.6	-0.2

从表 9 可以看出, 传递方式对不同方法的影响各异。从架构角度分析: SGC 先传递后特征转换, 能直接捕捉全局结构信息, 因此欧氏传递矩阵为其提供多样化路径后性能一致提升。APPNP 与 PTA 架构相似 (先特征转换后多次传递), 仅优化函数设计不同, 因此在实验中性能变化趋势类似。GPR-GNN 基于 APPNP 架构增加了可学习参数, 虽提高了捕捉复杂关系的能力, 但也可能导致过拟合, 使性能表现不一致。将原始拓扑传递优化为欧氏传递矩阵后, 大多数模型性能都有不同程度提升, 说明对传递矩阵细化能带来增益。然而, 少数情况下模型性能会下降, 这些下降主要出现在 PubMed 数据集上。

为了深入分析原因, 本文比较了 PubMed 数据集在欧氏空间和原始拓扑空间的差异, 计算了类内-类间相似度矩阵和标签兼容性矩阵, 结果如图 5 所示。由图 5 可见, 欧氏空间中各类别对之间的相似度几乎相同, 差异极小; 而在原始拓扑空间中, 类内兼容性极高, 类间兼容性极低。这种结构差异表明: 特征空间中类别边界模糊, 原始拓扑空间中类别界限清晰。因此在 PubMed 数据集中, 原始归一化邻接矩阵已能有效聚合同类节点信息, 而特征区分度不足导致特征空间诱导的传递路径无法发挥预期优势, 甚至产生负面影响。



(a) 节点标签兼容性矩阵



(b) 节点特征相似性矩阵

图 5 PubMed 特性矩阵可视化

Fig. 5 Property matrix visualization of the PubMed dataset

### 3.5 损失有效性分析

为验证秩约束的有效性, 本研究设置不同权

重系数分析其对模型在 3 个数据集上性能的影响, 结果如表 10 所示。当权重为 0 时, 秩约束不起作用, 模型性能可视为基线结果。

**表 10 损失有效性分析**  
**Table 10 Effectiveness analysis of the proposed loss** %

秩损失权重 $\beta$	Cora	CiteSeer	PubMed
0	81.1	72.7	84.8
0.01	82.4	74.5	84.1
0.1	83.7	74.3	83.2
1	78.3	76.5	86.9
10	65.6	59.1	80.8

研究发现, 给予秩约束适当权重时, 模型在 3 个数据集上的性能均有提升, 证明适量引入秩约束有助于提高模型泛化能力。然而, 当权重过大时, 3 个数据集上的性能均大幅下降。这表明适度的秩约束能有效提高模型在不同数据集上的表现, 但过大的权重会导致性能降低。

#### 4 结束语

本研究提出 MHPP 模型, 应对图表示学习中的扭曲和过平滑问题。该模型结合无结构化特征学习、多双曲空间参数化传递矩阵和门控混合专家系统, 实现精确信息传递和因子空间重要性建模。实验证明其有效性, 但多分辨率双曲空间曲率的手动设置限制了泛化能力。未来研究将探索自动曲率学习方法, 如可微代理梯度设计或强化学习。

#### 参考文献:

- [1] 侯磊, 刘金环, 于旭, 等. 图神经网络研究综述[J]. 计算机科学, 2024, 51(6): 282–298.  
HOU Lei, LIU Jinhuan, YU Xu, et al. Review of graph neural networks[J]. Computer science, 2024, 51(6): 282–298.
- [2] 白雪. 基于图神经网络的推荐方法研究[D]. 太原: 山西大学, 2023: 1–46.  
BAI Xue. Research on recommendation method based on graph neural network[D]. Taiyuan: Shanxi University, 2023: 1–46.
- [3] 张舒菡. 基于图表示学习的中医处方推荐方法研究[D]. 北京: 北京交通大学, 2023: 1–64.  
ZHANG Shuhan. Study on TCM prescription recommendation method based on graph representation learning[D]. Beijing: Beijing Jiaotong University, 2023: 1–64.
- [4] 张嘉伟. 基于图结构学习的图神经网络的药物靶标预测应用[D]. 武汉: 华中农业大学, 2023: 1–58.  
ZHANG Jiawei. Application of graph neural network based on graph structure learning to drug target prediction[D]. Wuhan: Huazhong Agricultural University, 2023: 1–58.
- [5] 张宇轩. 基于异构图表示学习的药物—靶蛋白关联关系预测研究[D]. 长春: 东北师范大学, 2023: 1–51.  
ZHANG Yuxuan. Drug-target interaction prediction via heterogeneous graph representation learning[D]. Changchun: Northeast Normal University, 2023: 1–51.
- [6] 邹然, 柳杨, 李聪, 等. 图表示学习综述[J]. 北京师范大学学报(自然科学版), 2023, 59(5): 716–724.  
ZOU Ran, LIU Yang, LI Cong, et al. Graph representation learning: a review[J]. Journal of Beijing Normal University (natural science), 2023, 59(5): 716–724.
- [7] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry[EB/OL]. (2017–06–12)[2024–09–26]. <https://arxiv.org/abs/1704.01212v2>.
- [8] LIN Guangfeng, KANG Xiaobing, LIAO Kaiyang, et al. Deep graph learning for semi-supervised classification[J]. *Pattern recognition*, 2021, 118: 108039.
- [9] WU F, SOUZA A, ZHANG Tianyi, et al. Simplifying graph convolutional networks[C]//International Conference on Machine Learning. Los Angeles: PMLR, 2019: 6861–6871.
- [10] MA Jianxin, CUI Peng, KUANG Kun, et al. Disentangled graph convolutional networks[C]//International Conference on Machine Learning. Los Angeles: PMLR, 2019: 4212–4221.
- [11] YANG Yiding, FENG Zunlei, SONG Mingli, et al. Factorizable graph convolutional networks[J]. *Advances in neural information processing systems*, 2020, 33: 20286–20296.
- [12] CHAMI I, YING R, RÉ C, et al. Hyperbolic graph convolutional neural networks[J]. *Advances in neural information processing systems*, 2019, 32: 4869–4880.
- [13] LIU Qi, NICKEL M, KIELA D. Hyperbolic graph neural networks[J]. *Advances in neural information processing systems*, 2019, 32: 8230–8241.
- [14] ZHANG Yiding, WANG Xiao, SHI Chuan, et al. Hyperbolic graph attention network[J]. *IEEE transactions on big data*, 2022, 8(6): 1690–1701.
- [15] 刘笑梅. 双曲空间中的神经网络图表示学习[D]. 桂林: 广西师范大学, 2019: 1–46.  
LIU Xiaomei. Graph representation learning with neural network in hyperbolic space[D]. Guilin: Guangxi Normal University, 2019: 1–46.
- [16] 周志霞. 基于自编码器与双曲几何的网络表示学习[D]. 石家庄: 河北师范大学, 2023: 1–53.

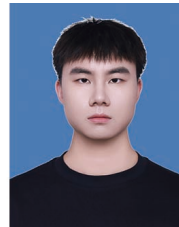
- ZHOU Zhixia. Network representation learning based on auto-encoder and hyperbolic geometry[D]. Shijiazhuang: Hebei Normal University, 2023: 1–53.
- [17] XU Keyulu, HU Weihua, LESKOVEC J, et al. How powerful are graph neural networks?[C]//International Conference on Learning Representations. New Orleans: OpenReview.net, 2019: 1–17.
- [18] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: bring order to the web [C]//International Conference on World Wide Web. Brisbane: Elsevier, 1998: 1–17.
- [19] SPINELLI I, SCARDAPANE S, UNCINI A. Adaptive propagation graph convolutional network[J]. *IEEE transactions on neural networks and learning systems*, 2020, 32(10): 4755–4760.
- [20] GASTEIGER J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: graph neural networks meet personalized pagerank[C]//International Conference on Learning Representations. New Orleans: OpenReview.net, 2019: 1–15.
- [21] KIPF T, WELING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2017–02–22)[2024–05–09]. <https://arxiv.org/abs/1609.02907>.
- [22] LI Guohao, MÜLLER M, GHANEM B, et al. Training graph neural networks with 1000 layers[C]//International Conference on Machine Learning. [S. l.]: PMLR, 2021: 6437–6449.
- [23] SARA S, NICHOLAS F, GEOFFREY H. Dynamic routing between capsules[J]. *Advances in neural information processing systems*, 2017, 30: 3859–3869.
- [24] ZHAO Tianxiang, ZHANG Xiang, WANG Suhang. Exploring edge disentanglement for node classification[C]//Proceedings of the ACM Web Conference 2022. Lyon: ACM, 2022: 1028–1036.
- [25] LIU Yanbei, WANG Xiao, WU Shu, et al. Independence promoted graph disentangled networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 4916–4923.
- [26] ZHENG Shuai, ZHU Zhenfeng, LIU Zhizhe, et al. Adversarial graph disentanglement with component-specific aggregation[J]. *IEEE transactions on artificial intelligence*, 2024, 5(5): 2204–2216.
- [27] WANG Yifan, TANG Suyao, LEI Yuntong, et al. Disentangled heterogeneous graph attention network for recommendation[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. [S. l.]: ACM, 2020: 1605–1614.
- [28] XIAO Teng, CHEN Zhengyu, GUO Zhimeng, et al. Decoupled self-supervised learning for graphs[J]. *Advances in neural information processing systems*, 2022, 35: 620–634.
- [29] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018–02–04) [2024–05–09]. <https://arxiv.org/abs/1710.10903>.
- [30] SUN Jianing, CHENG Zhaoyue, ZUBERI S, et al. HGCF: hyperbolic graph convolution networks for collaborative filtering[C]//Proceedings of the Web Conference 2021. Ljubljana: ACM, 2021: 593–601.
- [31] 张依丁. 基于双曲空间的图表示学习算法研究[D]. 北京: 北京邮电大学, 2022: 1–98.
- ZHANG Yiding. Research on hyperbolic graph representation learning algorithm graph representation learning algorithm based on hyperbolic space[D]. Beijing: Beijing University of Posts and Telecommunications, 2022: 1–98.
- [32] GREGOR B, GARY B, OCTAVIAN G. Constant curvature graph convolutional networks[C]//International Conference on Machine Learning. [S. l.]: PMLR, 2020: 486–496.
- [33] ZHU Shichao, PAN Shirui, ZHOU Chuan, et al. Graph geometry interaction learning[J]. *Advances in neural information processing systems*, 2020, 33: 7548–7558.
- [34] ZHOU Zhiheng, WANG Tao, HOU Linfang, et al. Decoupled hyperbolic graph attention network for modeling substitutable and complementary item relationships[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta: ACM, 2022: 2763–2772.
- [35] LIU Meng, GAO Hongyang, JI Shuiwang. Towards deeper graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. [S. l.]: ACM, 2020: 338–348.
- [36] QIAN Huang, HORACE H, ABHAY S, et al. Combining label propagation and simple models out-performs graph neural networks[C]//International Conference on Learning Representations. [S. l.]: OpenReview.net, 2021: 1–21.
- [37] DONG Hande, CHEN Jiawei, FENG Fuli, et al. On the equivalence of decoupled graph convolution network and label propagation[C]//Proceedings of the Web Conference 2021. Ljubljana: ACM, 2021: 3651–3662.
- [38] DING Kaize, WANG Jianling, CAVERLEE J, et al. Meta propagation networks for graph few-shot semi-supervised learning[C]//AAAI Conference on Artificial Intelligence, 2022, 36(6): 6524–6531.
- [39] 王振胜. 基于特征平滑的图神经网络在节点分类中的应用研究[D]. 北京: 北京化工大学, 2023: 9–41.
- WANG Zhensheng. Research on the application of graph neural networks with feature smoothing in node classification

- ation[D]. Beijing: Beijing University of Chemical Technology, 2023: 9–41.
- [40] SHEN Qinghua, REN Weijieying, QIN Wei. Graph relation aware continual learning [EB/OL]. (2023–08–16) [2024–05–09]. <https://arxiv.org/abs/2308.08259>.
- [41] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[J]. *Advances in neural information processing systems*, 2017, 30: 1–11.
- [42] YANG Ling, TIAN Ye, XU Minkai, et al. VQGraph: rethinking graph representation space for bridging GNNs and MLPs[C]//International Conference on Learning Representations. Vienna: OpenReview.net, 2024: 1–20.
- [43] XIAO Teng, ZHU Huaisheng, CHEN Zhengyu, et al. Simple and asymmetric graph contrastive learning without augmentations[J]. *Advances in neural information processing systems*, 2024, 36: 1–24.
- [44] CHIEN E, PENG Jianhao, LI Pan, et al. Adaptive universal generalized PageRank graph neural network[C]//International Conference on Learning Representations. [S. l. ]: OpenReview.net, 2021: 1–24.
- [45] YU Tao, DE SA C. Random Laplacian features for learning with hyperbolic space[C]//International Conference on Learning Representations. Kigali: OpenReview.net, 2023: 1–23.
- [46] ZHU Jiong, YAN Yujun, ZHAO Lingxiao, et al. Beyond homophily in graph neural networks: current limitations and effective designs[J]. *Advances in neural information processing systems*, 2020, 33: 7793–7804.
- [47] WANG Xiyuan, ZHANG Muhan. How powerful are spectral graph neural networks[C]//International Conference on Machine Learning. Baltimore: PMLR, 2022: 23341–23362.
- [48] PEI Hongbin, WEI Bingzhen, CHANG K, et al. GeomGCN: geometric graph convolutional networks [C]//International Conference on Learning Representations. [S. l. ]: OpenReview.net, 2020: 1–12.

#### 作者简介:



郑帅, 高聘副教授, 博士。主要研究方向为图表示学习、图神经网络和多模态表示学习。发表学术论文 20 余篇。E-mail: [shuaizheng@bjtu.edu.cn](mailto:shuaizheng@bjtu.edu.cn)。



彭奏章, 硕士, 主要研究方向为图分析与解耦图表示学习。E-mail: [pengzouzhang@bjtu.edu.cn](mailto:pengzouzhang@bjtu.edu.cn)。



朱振峰, 教授, 博士生导师, 主要研究方向为机器学习、计算机视觉和动态数据挖掘。主持科技创新 2030—“新一代人工智能”重大项目、国家自然科学基金项目、北京市自然科学基金项目等。获第十三届吴文俊人工智能科学技术科技进步二等奖(排名第 1)。E-mail: [zhfzhu@bjtu.edu.cn](mailto:zhfzhu@bjtu.edu.cn)。