



研究与开发

基于拓扑信息增强的异构图神经网络算法

姚波武^{1,2}, 邓琨^{2,3}, 魏振华¹, 吴桐^{2,3}, 刘星妍⁴

1. 东华理工大学人工智能与信息工程学院, 江西 南昌 330013;
2. 嘉兴大学浙江省全省多模态感知与智能系统重点实验室, 浙江 嘉兴 314001;
3. 嘉兴大学信息科学与工程学院, 浙江 嘉兴 314001;
4. 嘉兴大学人工智能学院, 浙江 嘉兴 314001)

摘要: 现有异构图神经网络算法多侧重于节点属性信息, 利用注意力机制区分节点的重要性; 然而, 此类方法在捕获完整拓扑信息和节点特征多样性方面存在不足。针对以上问题, 提出一种基于拓扑信息增强的异构图神经网络算法。首先, 该算法从局部视角进行聚合, 引入局部拓扑信息; 随后, 结合上下文采样与结构注意力机制, 动态识别并加权聚合关键的高阶拓扑信息; 最后, 通过特征空间叠加技术保留节点的异构信息, 并利用多头自注意力机制跨类型聚合节点属性, 实现复杂语义的捕获。多个公开数据集上的实验结果表明, 该算法在捕获异构图拓扑信息和保留节点异构信息方面表现优异。相较于多种基线方法, 在节点分类任务中, Macro-F1 指标平均提升 0.52%~2.15%; 在聚类任务中, 归一化互信息 (normalized mutual information, NMI) 值平均提升 1.26%~2.65%。

关键词: 异构图; 图神经网络; 图嵌入; 注意力机制

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025208

Heterogeneous graph neural network algorithm based on topology information enhancement

YAO Bowu^{1,2}, DENG Kun^{2,3}, WEI Zhenhua¹, WU Tong^{2,3}, LIU Xingyan⁴

1. College of Artificial Intelligence and Information Engineering, East China University of Technology, Nanchang 330013, China
2. Provincial Key Laboratory of Multimodal Perceiving and Intelligent Systems, Jiaxing University, Jiaxing 314001, China
3. College of Information Science and Engineering, Jiaxing University, Jiaxing 314001, China
4. College of Artificial Intelligence, Jiaxing University, Jiaxing 314001, China

Abstract: Existing heterogeneous graph neural network algorithms primarily focus on node attribute information, uti-

收稿日期: 2025-01-24; 修回日期: 2025-04-12

通信作者: 邓琨, dengkun@hrbeu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62302197); 嘉兴市科技公益项目 (No.2024AY10023)

Foundation Items: The National Natural Science Foundation of China (No.62302197), Jiaxing City Science and Technology Public Welfare Project (No.2024AY10023)

lizing attention mechanisms to distinguish node importance. However, such methods have limitations in capturing complete topological information and node feature diversity. To address these issues, a heterogeneous graph neural network algorithm based on topological information enhancement was proposed. Firstly, local topological information was incorporated, aggregating from a local perspective. Then, context sampling was combined with structural attention mechanisms to dynamically identify and weight aggregate key higher-order topological information. Finally, node heterogeneous information was preserved through feature space stacking technology and multi-head self-attention mechanisms were employed to cross-type aggregate node attributes, achieving the capture of complex semantics. Experimental results on multiple public datasets demonstrate that this method excels in capturing heterogeneous graph topological information and preserving node heterogeneous information. Compared to various baseline methods, in node classification tasks, the Macro-F1 metric is improved by an average of 0.52%~2.15%. In clustering tasks, the normalized mutual information (NMI) value is improved by an average of 1.26%~2.65%.

Key words: heterogeneous graph, graph neural network, graph embedding, attention mechanism

0 引言

异构图是一种包含多种节点类型和连接关系的图结构数据,广泛应用于复杂网络场景,如社交网络^[1]、知识图谱^[2]和引文网络^[3]。异构图节点表示学习旨在将节点映射为低维向量,同时保留其拓扑与语义信息,以提升下游任务性能。目前,研究者主要将图神经网络(graph neural network, GNN)作为核心工具。然而,现有方法主要侧重于节点属性的建模,未能充分挖掘图的拓扑信息。为此,Dong等^[4]提出的异构图嵌入算法(Metapath2vec)利用随机游走策略,从节点邻域中提取拓扑信息,并将其融合至节点嵌入。Jin等^[5]提出了基于属性补全的异构图神经网络(heterogeneous graph neural network via attribute completion, HGNN-AC)利用预训练模型获得的拓扑嵌入来计算无属性节点的权重,以融合拓扑和属性信息。然而,此类方法受预训练信息的限制,难以达到最优效果。

此外,由于信息缺失或隐私保护等原因,异构图中的节点属性可能不完整。为此,LYU等^[6]提出了简单异构图神经网络(simple heterogeneous graph network, Simple-HGN),通过 one-hot 编码方式填补缺失属性;Li等^[7]提出了基于节点属性增强的高阶异构图神经网络(higher order

heterogeneous graph neural network, HOAE),利用基于 Transformer 的自注意力机制填补缺失属性,增强了信息聚合效果。然而,这些方法未能有效表达节点属性的语义信息,且忽略了高阶邻居的语义贡献^[8]。

当前异构图神经网络(heterogeneous graph neural network, HGNN)仍遵循与同构图算法相似的消息传递范式^[9],通过异构感知机制递归聚合相邻节点信息。为了实现节点属性的聚合,Wang等^[10]提出的异构图注意力网络(heterogeneous graph attention network, HAN)和Fu等^[11]提出的元路径聚合图神经网络(metapath aggregated graph neural network, MAGNN)都将不同类型节点的属性强制转换至统一空间。然而,这种强制转换易引发语义混淆,影响节点嵌入的质量^[12]。

针对上述问题,本文提出一种基于拓扑信息增强的异构图神经网络算法(heterogeneous graph neural network algorithm based on topology information enhancement, TEHGNN),引入分层拓扑信息聚合框架,联合建模节点的拓扑结构与语义特征。本文主要贡献包括:(1)构建局部拓扑信息增强模块,将无属性节点作为目标节点的连接参考,捕获局部拓扑信息;(2)构建高阶拓扑信息聚合模块,引入上下文拓扑序列,获取丰富的



高阶拓扑信息；(3) 构建节点异构性聚合模块，采用特征空间叠加机制，在保留节点异构特征的同时提取其语义信息；(4) 在3个基准数据集上的实验结果表明，该算法在节点分类等任务中取得了显著的性能提升。

1 相关工作

为了明确本文方法的研究任务及其与现有工作的关系，本文从图表示学习和注意力机制网络两个方面，回顾了与TEHGNN模型密切相关的研究进展。

1.1 图表示学习

图表示学习能够在保留节点拓扑与语义信息的基础上，得到适用于下游任务的低维向量表示^[13]。在早期研究中，研究者提出了多种以随机游走为核心的浅层模型，能够高效处理小规模图数据^[14-15]。然而，由于模型深度有限，此类方法难以捕捉图中复杂的语义关系，且在应对大规模图数据时性能受限^[16]。为了解决上述问题，研究者提出了多种图神经网络算法，以挖掘图中更深层次的拓扑和语义信息。其中，Kipf和Welling^[17]提出的图卷积网络（graph convolutional network, GCN），首次将卷积运算引入图结构数据的建模中。该模型通过引入自环连接的对称归一化邻接矩阵，实现邻居节点特征的加权聚合。具体而言，每个卷积层将归一化后的邻接矩阵与节点嵌入以及可训练权重相乘，并通过非线性函数进行变换，逐层学习节点的低维表示。其核心卷积计算过程如下：

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (1)$$

其中， $\mathbf{H}^{(l)}$ 为第 l 层的节点嵌入矩阵， $\sigma(\cdot)$ 为激活函数， $\tilde{\mathbf{D}} = \sum_j \tilde{\mathbf{A}}_{ij}$ 是节点对 (i,j) 的度矩阵， $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ 是添加了自连接的邻接矩阵， \mathbf{I} 表示单位矩阵， N 表示节点总数， $\mathbf{W}^{(l)}$ 是第 l 层的可学习权重矩阵。

随后，Schlichtkrull等^[18]提出了关系图卷积网络（modeling relational data with graph convolutional network, R-GCN），对传统图卷积网络的消息传递机制进行扩展。首先，在基本卷积操作的基础上，针对不同类型的边关系，引入关系特定的特征变换矩阵；其次，根据边的语义差异调整邻居节点的贡献；最后，加权聚合邻居节点的互信息，得到节点的低维嵌入表示。信息聚合过程如下：

$$\mathbf{h}_i^{(l+1)} = \sigma \left(\sum_{r \in R_j} \sum_{i \in N_i^r} \frac{1}{c_{i,r}} \mathbf{W}_r^{(l)} \mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)} \mathbf{h}_i^{(l)} \right) \quad (2)$$

其中， $\mathbf{h}_i^{(l+1)}$ 为第 $l+1$ 层节点 i 的嵌入表示， l 为网络层数， $\sigma(\cdot)$ 为激活函数， $r \in R$ 为节点对 (i,j) 间的关系， R 为关系集合， N_i^r 表示关系 r 下节点 i 的邻居索引集合， $c_{i,r}$ 是特定关系 r 的归一化常数， $\mathbf{W}_r^{(l)}$ 是第 l 层关系 r 的可训练权重矩阵。尽管这些方法在捕捉复杂语义方面取得了显著进展，但随着模型层数加深，其卷积操作可能引发梯度消失和语义混淆等问题，从而影响模型训练的稳定性与节点表示的准确性^[19-20]。

1.2 自注意力网络

近年来，自注意力机制在图结构数据建模中展现出巨大潜力，能为输入数据动态分配权重，从而建模节点间的依赖关系，并缓解传统消息传递范式中存在的过度平滑问题。Velickovic等^[21]提出的图注意力网络（graph attention network, GAT）首次将自注意力机制引入图神经网络模型。首先，利用可训练的注意力函数计算邻接节点间的重要性；其次，引入多头机制提升模型稳定性；最后，融合多组注意力头以获得最终嵌入表示。其数学表达式如下：

$$\mathbf{e}_{i,j} = \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_i \parallel \mathbf{W} \mathbf{h}_j]), \alpha_{i,j} = \frac{\exp(\mathbf{e}_{i,j})}{\sum_{j \in N(i)} \exp(\mathbf{e}_{i,j})}$$

$$\mathbf{h}_i^k = \sigma \left(\sum_{j \in N(i)} \alpha_{i,j}^k \mathbf{W}^k \mathbf{h}_j \right), \mathbf{h}_i = \text{Concat}(\mathbf{h}_i^{(1)}, \mathbf{h}_i^{(2)}, \dots, \mathbf{h}_i^{(k)}) \quad (3)$$

其中, $\alpha_{i,j}$ 表示节点对 (i,j) 的注意力系数, $\text{LeakyReLU}(\cdot)$ 表示激活函数, \mathbf{a}^T 表示可学习注意力向量, $N(i)$ 表示节点 i 的邻居节点集合, $||$ 表示级联操作, \mathbf{W} 表示可学习参数矩阵, $\alpha_{i,j}$ 表示归一化后的注意力系数, k 表示多头数, \mathbf{h}_i^k 为节点 i 在第 k 个注意力头下的嵌入表示, $\sigma(\cdot)$ 为激活函数, $\text{Concat}(\cdot)$ 表示多头注意力的拼接操作, \mathbf{h}_i 表示节点 i 的最终嵌入。

在异构图场景下, HAN 进一步将自注意力机制扩展至节点级和语义级两个层面。该方法采用分层结构, 首先, 在节点级视图下计算邻居节点的重要性; 其次, 在语义视图下衡量元路径对最终节点表示的贡献, 最终捕获异构图结构的复杂语义。其核心思路可用以下数学式表达:

$$\alpha_{i,j}^\varphi = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))}{\sum_{j \in N_i^\varphi} \exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))}$$

$$\beta_\varphi = \frac{\exp(\mathbf{q}^T \mathbf{z}_\varphi)}{\sum_{\varphi \in \Phi} \exp(\mathbf{q}^T \mathbf{z}_\varphi)}, \mathbf{z}_\varphi = \frac{1}{|V|} \sum_{i \in V} \mathbf{h}_i^\varphi \quad (4)$$

其中, $\alpha_{i,j}^\varphi$ 表示归一化后元路径 φ 下节点对 (i,j) 的注意力系数, $\varphi \in \Phi$ 表示节点所在的元路径, Φ 表示元路径集合, N_i^φ 表示元路径 φ 下节点 i 的邻居节点集合, \mathbf{a}^T 表示可学习节点级注意力向量, \mathbf{W} 表示可学习参数矩阵, \mathbf{h}_i 表示节点 i 的嵌入表示, $\text{LeakyReLU}(\cdot)$ 表示激活函数, $||$ 表示级联操作, β_φ 表示归一化后元路径 φ 的语义级注意力系数, \mathbf{q}^T 表示可学习语义级注意力向量, \mathbf{z}_φ 为元路径 φ 的嵌入表示, $|V|$ 为异构图的节点总数, \mathbf{h}_i^φ 为融合了节点级互信息的嵌入表示。尽管上述方法有效提升了节点间互信息的融合效果, 但在处理大规模图时仍面临计算复杂度高、训练成本大等问题^[22]。

尽管上述方法在节点表示学习方面取得了显著进展, 但仍主要侧重于节点属性的聚合, 对拓扑结构和节点异构性关注不足^[23]。为此, 本文提出了一种基于拓扑信息增强的异构图神经网络算法 (TEHGNN), 通过局部和高阶拓扑信息的联合聚合, 以及对节点异构特性的建模, 充分保留拓扑和语义信息, 从而学习质量更高的节点表示。

2 基本定义

定义 1 异构网络。异构网络可以表示为 $G=(V,E,F,R,\varphi,\phi)$, 其中, V 表示节点集合, E 表示边集合, F 表示节点类型集合, R 表示边类型集合, 且 $|F|+|R|>2$ 。在异构网络中, 每个节点 $i \in V$ 的类型均由节点类型映射函数 $\varphi:V \rightarrow F$ 决定; 每条边 $e \in E$ 的类型则由边类型映射函数 $\phi:E \rightarrow R$ 决定。异构图示例如图 1 所示, 引文网络是一种典型的异构网络, 由 3 种节点类型 (作者、论文和会议) 以及 2 种边类型 (作者-论文和论文-会议) 组成。

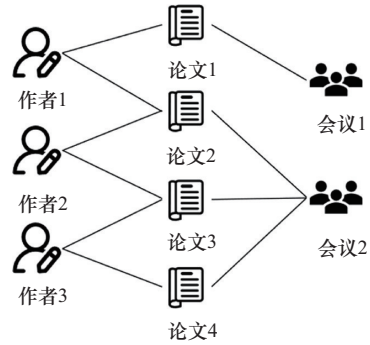


图 1 异构图示例

定义 2 节点上下文拓扑序列。在给定异构网络 $G=(V,E,F,R,\varphi,\phi)$ 中, 对于任意节点 $i \in V$, 可以构建其上下文拓扑序列 $L(i)$ 。该序列包含节点 i 的直接邻居以及多跳邻居。形式化地表示为 $L(i)=\{i_0, i_1, i_2, \dots, i_{\text{seqLen}-1}\}$, 其中元素 i_k 表示节点 i



的 k 跳邻居, 序列长度为 seqLen 。

定义3 异构特征空间。在异构网络中, 由于不同节点类型具有不同的特征属性, 其特征维度通常不一致。设节点类型集合为 F , 节点类型数量为 $|F|$ 。对于任一节点类型 $f \in F$, 其特征表示为 $\mathbf{H}^{(f)} \in \mathbf{R}^{N \times D_f}$, 其中 N 为该类型节点数, D_f 为该类型节点的特征维度。

问题定义 异构网络表示学习: 给定一个异构网络 $G=(V, E, F, R, \phi, \psi)$, 其目标是为任意节点 $i \in V$ 学习一个低维嵌入表示 $\mathbf{h}_i \in \mathbf{R}^d$ 。该过程可以形式化为学习一个映射函数 $f: i \rightarrow \mathbf{h}_i$, 使得节点 i 的低维嵌入表示 \mathbf{h}_i 能够保留其在原异构网络中的拓扑信息与语义信息。

3 基于拓扑信息增强的异构图神经网络算法

本文提出了一种基于拓扑信息增强的异构图神经网络算法 (TEHGNN), 其整体框架如图2所示, 主要由以下3部分构成: (1) 局部拓扑信息增强模块, 聚焦于目标节点的局部结构, 在排除结构噪声的同时, 将预训练的拓扑嵌入和无属性节点作为连接参考, 实现初步的拓扑信息增强; (2) 高阶拓扑信息聚合模块, 通过上下文采样生成高相关度的拓扑序列, 捕获目标节点与高阶邻居之间的潜在关联, 实现高阶信息的递归式聚合; (3) 节点异构性聚合模块, 采用特征映射与特征空间叠加机制, 保留节点异构信息, 最终得到目标节点的低维嵌入表示。

3.1 局部拓扑信息增强模块

由于无属性节点在节点表示学习中仍具有重要作用^[24], 且属性信息与拓扑信息在语义表达上具有相似性。本文设计局部拓扑信息增强模块, 以自注意力机制为基础, 有效聚合目标节点的局部拓扑信息。考虑异构网络中不同类型节点的特征分布存在差异, 为实现拓扑信息的有效融合,

需将所有节点的初始特征映射至统一维度空间。具体映射过程如下:

$$\mathbf{H}^0 = \mathbf{W} \cdot \mathbf{X} + \mathbf{b} \quad (5)$$

其中, $\mathbf{X} \in \mathbf{R}^{N \times d}$ 为初始特征嵌入, N 表示节点总数, d 为特征维度, $\mathbf{H}^0 \in \mathbf{R}^{N \times d}$ 为投影后的特征嵌入, \mathbf{W} 为可学习投影矩阵, \mathbf{b} 为偏置向量。

对于无属性节点, 其初始特征设为零向量, 以降低其对拓扑信息的干扰。同时, 利用预训练模型生成节点的原始拓扑嵌入 \mathbf{s}' 。在此基础上, 结合自注意力机制^[25], 将目标节点 v 的嵌入向量 $\mathbf{s}'_v \in \mathbf{R}^d$ 和邻居节点 u 的嵌入向量 $\mathbf{s}'_u \in \mathbf{R}^d$ 分别映射到高维空间, 并以此计算节点间的结构注意力权重 $\alpha_{v,u}$ 。该计算过程如式(6)所示:

$$\begin{aligned} \mathbf{Q}(v) &= \text{Linear}(\mathbf{s}'_v), \mathbf{K}(u) = \text{Linear}(\mathbf{s}'_u) \\ \alpha_{v,u} &= \frac{\exp(\text{LeakyReLU}(\mathbf{Q}_v^\top \mathbf{K}_u))}{\sum_{u \in N(v)} \exp(\text{LeakyReLU}(\mathbf{Q}_v^\top \mathbf{K}_u))} \end{aligned} \quad (6)$$

其中, $\mathbf{Q}(v) \in \mathbf{R}^d$ 为节点 v 的查询向量, $\mathbf{K}(u) \in \mathbf{R}^d$ 为节点 u 的键向量, d 表示线性映射后的维度, $\text{Linear}(\cdot)$ 表示线性变换函数, $\alpha_{v,u} \in \mathbf{R}^d$ 表示节点对 (v,u) 的结构注意力系数, $\text{LeakyReLU}(\cdot)$ 表示激活函数, $u \in N(v)$ 为目标节点 v 的直接邻居。

在获取邻居节点的注意力权重后, 基于该权重对其特征加权聚合, 提取局部拓扑信息。此外, 由于拓扑信息与语义信息之间存在同质性^[26], 利用结构注意力对无属性节点 u 的特征向量 \mathbf{h}_u^0 聚合, 得到目标节点 v 的局部拓扑矩阵 \mathbf{s}_v^0 , 如式(7)所示:

$$\mathbf{s}_v^0 = \sum_{u \in N(v)} \alpha_{v,u} \cdot \mathbf{h}_u^0 \quad (7)$$

其中, $\mathbf{s}_v^0 \in \mathbf{R}^d$ 表示目标节点 v 的局部拓扑信息, $u \in N(v)$ 表示节点 v 的直接邻居, $N(v)$ 表示节点 v 的直接邻居集合, $\alpha_{v,u} \in \mathbf{R}^d$ 表示节点对 (v,u) 的结构注意力系数, $\mathbf{h}_u^0 \in \mathbf{R}^d$ 表示邻居节点 u 的初始特征。

3.2 高阶拓扑信息聚合模块

在异构网络中, 高阶拓扑信息同样对节点嵌

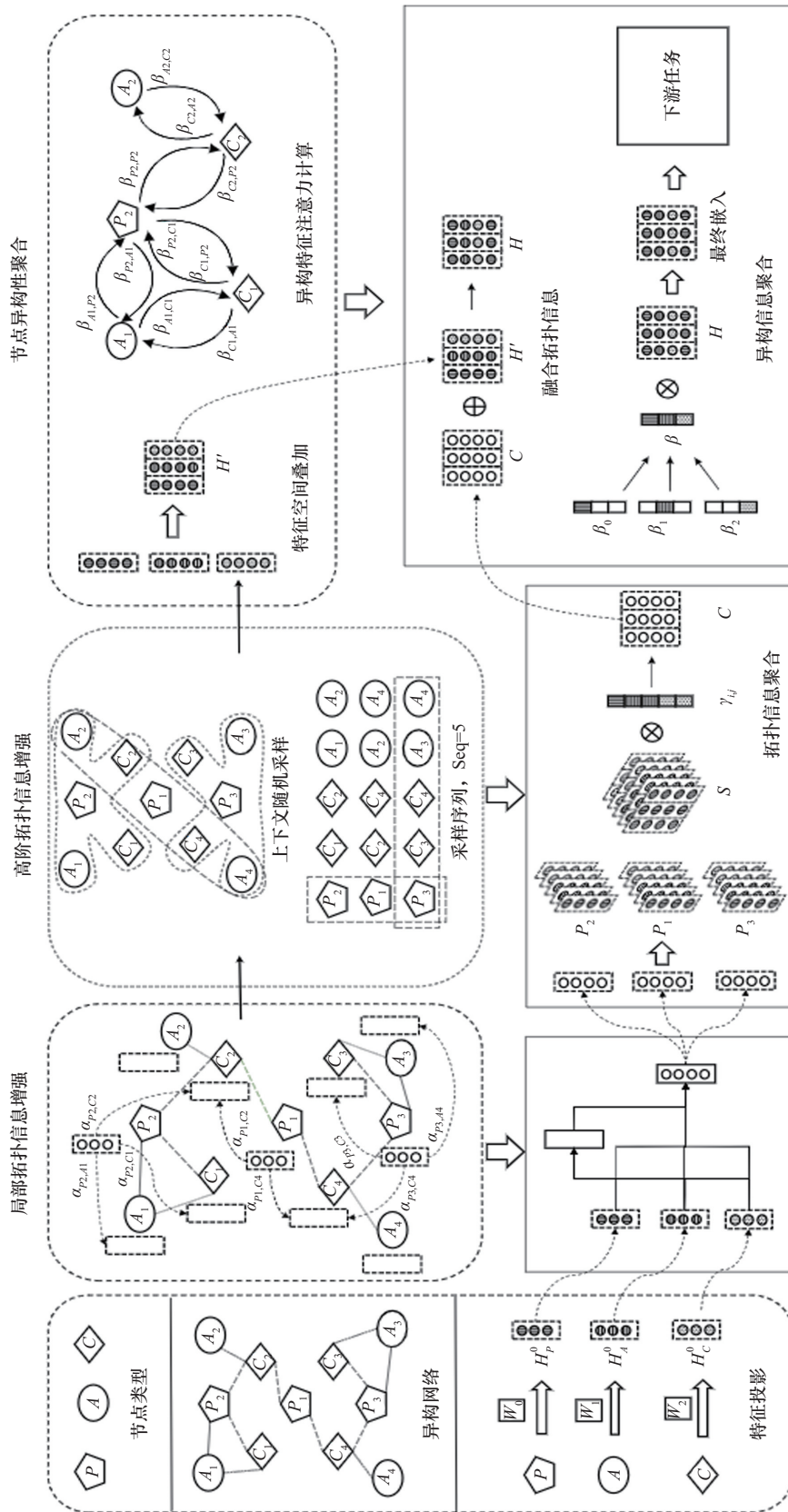


图2 TEHGNN 算法框架



入具有重要影响。为了捕获高阶拓扑信息，本文采用上下文随机采样算法对节点 $v \in V$ 执行多跳邻居采样，生成长度为 seqlen 的拓扑序列 $L(v) = \{u_0, u_1, u_2, \dots, u_{\text{seqlen}-1}\}$, $L(v) \in \mathbf{R}^{\text{seqlen} \times d}$ 。其中, u_i 表示第 i 个邻居节点, 用于捕获高阶拓扑信息并降低远距离噪声节点的干扰^[27]。随后, 对序列 $L(v)$ 进行度归一化和卷积操作, 融合度信息与拓扑信息, 提升模型对拓扑信息的建模能力:

$$s_v^{(l+1)} = \sigma \left(\sum_{u \in L(v)} \frac{1}{\sqrt{d_v d_u}} W^{(l)} s_u^{(l)} \right) \quad (8)$$

其中, $s_v^{(l+1)} \in \mathbf{R}^d$ 表示第 $l+1$ 层网络中节点 v 的特征向量, $W^{(l)} \in \mathbf{R}^d$ 为第 l 层网络中的权重矩阵, $u \in L(v)$ 表示节点 v 的上下文采样节点, $\sigma(\cdot)$ 为激活函数, d_v 和 d_u 为目标节点 v 和邻居节点 u 的度。

为了进一步丰富目标节点的高阶拓扑表示, 将拓扑向量 s_v 填充到序列 $L(v)$, 构建拓扑序列 $P(v)$: $P(v) = \{s_0, s_1, s_2, \dots, s_{\text{seqlen}-1}\}$, $P(v) \in \mathbf{R}^{\text{seqlen} \times d}$ 。为了建模序列中不同邻居对节点的结构性贡献, 引入基于多头自注意力机制进行加权聚合。具体计算过程如下:

$$\bar{Q}(v) = \text{Linear}_l(s_v), \bar{K}(u) = \text{Linear}_r(s_u) \quad (9)$$

$$\begin{aligned} \bar{Q}_v^{(m)} &= \text{Reshape}(\text{Linear}(s_v))_m, \\ \bar{K}_u^{(m)} &= \text{Reshape}(\text{Linear}(s_u))_m \end{aligned} \quad (10)$$

其中, $s_v \in \mathbf{R}^{\text{seqlen} \times d}$ 为节点 v 的嵌入表示, $\text{Linear}(\cdot)$ 为线性映射函数, $\bar{Q}(v) \in \mathbf{R}^{\text{seqlen} \times d}$ 为 v 的查询向量, $\bar{K}(u) \in \mathbf{R}^{\text{seqlen} \times d}$ 表示节点 u 的键向量, seqlen 表示拓扑序列长度, d' 为投影后的嵌入维度, m 表示多头序号, $\text{Reshape}(\cdot)$ 表示维度转换函数, $\bar{Q}_v^{(m)} \in \mathbf{R}^{\text{seqlen} \times k \times d_{\text{head}}}$ 表示第 m 个注意力头下的查询向量, $\bar{K}_u^{(m)} \in \mathbf{R}^{\text{seqlen} \times k \times d_{\text{head}}}$ 表示第 m 个注意力头下的键向量, k 为多头总数, d_{head} 为每个头的维度。

此处引入的多头自注意力机制, 主要用于解决图结构数据方差较大以及节点间噪声干扰严重的问题。在具体实验中, 为了确定多头数的最佳值, 综合考虑嵌入维度、数据集复杂性和模型稳

定性等因素。若多头数过多将压缩每个头的特征维度, 限制表达能力; 而多头数过少则难以捕获不同子空间的互补信息。因此, 在模型训练过程中, 根据目标节点类型的特征维度, 对 4、8、16 等不同的多头数进行多次实验, 确保各注意力头具备足够的表达能力。最后参考实验结果和 HAN 等公开论文^[10] 中的领域经验, 选择 8 作为默认多头数, 充分发挥并行建模的优势。多头自注意力机制框架如图 3 所示。

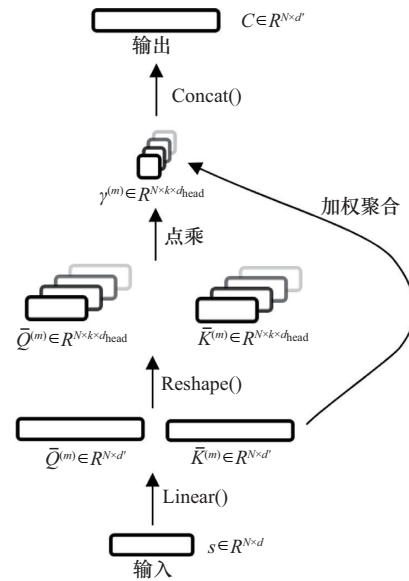


图 3 多头自注意力机制框架

此外, 为了计算高阶节点对 (v, u) 在各个注意力头下的交互强度, 设计了结构注意力系数 $\gamma_{v,u}^{(m)}$, 用于衡量高阶邻居 u 对目标节点 v 的结构性贡献。具体计算如下:

$$\gamma_{v,u}^{(m)} = \text{LeakyReLU}(\text{att}_l^T \bar{Q}_v^{(m)}) + \text{LeakyReLU}(\text{att}_r^T \bar{K}_u^{(m)}) \quad (11)$$

其中, $\gamma_{v,u}^{(m)} \in \mathbf{R}^{\text{seqlen} \times k \times d_{\text{head}}}$ 表示第 m 个注意力头下的结构注意力系数, m 表示多头序号, att_l 和 att_r 为可学习参数, $\text{LeakyReLU}(\cdot)$ 为激活函数。同时, 为了提升跨拓扑维度的可比性, 对 $\gamma_{v,u}^{(m)}$ 进行 Softmax 归一化操作:

$$\Gamma_{v,u}^{(m)} = \frac{\exp(\gamma_{v,u}^{(m)})}{\sum_{u \in N(v)} \exp(\gamma_{v,u}^{(m)})} \quad (12)$$

其中, $\Gamma_{v,u}^{(m)} \in \mathbf{R}^{\text{seqLen} \times k \times d_{\text{head}}}$ 表示归一化后的高阶结构注意力系数。将 $\Gamma_{v,u}^{(m)}$ 用于加权聚合高阶邻居嵌入, 得到每个头的节点表示 $\mathbf{c}_v^{(m)}$ 。最终将 k 个注意力头的输出拼接, 得到节点 v 的整体表示 \mathbf{c}_v , 具体计算如下:

$$\mathbf{c}_v^{(m)} = \sum_{u \in P(v)} \Gamma_{v,u}^{(m)} \bar{\mathbf{K}}_u^{(m)} \quad (13)$$

$$\mathbf{c}_v = \text{Concat}(\mathbf{c}_v^{(1)}, \mathbf{c}_v^{(2)}, \dots, \mathbf{c}_v^{(k)}) \quad (14)$$

其中, $P(v)$ 表示拓扑序列, $\bar{\mathbf{K}}_u^{(m)} \in \mathbf{R}^{\text{seqLen} \times k \times d_{\text{head}}}$ 为邻居节点 u 的嵌入表示, $\mathbf{c}_v^{(m)} \in \mathbf{R}^{\text{seqLen} \times k \times d_{\text{head}}}$ 为第 m 个注意力头下节点 v 的嵌入表示, $\text{Concat}(\cdot)$ 表示多头拼接操作, $\mathbf{c}_v \in \mathbf{R}^{\text{seqLen} \times d}$ 为拼接后的最终表示, k 表示多头数量。拓扑序列中的首个元素可进一步作为节点 v 的高阶拓扑嵌入表示 $\mathbf{c}_v \in \mathbf{R}^d$ 。

为了缓解多层聚合带来的梯度消失问题, 模块采用残差连接与层归一化操作更新节点表示, 提高信息传递的稳定性与模型收敛速度:

$$\mathbf{c}_v^{l+1} = \text{LN}(\mathbf{s}_v^0 + \eta \mathbf{c}_v^l) \quad (15)$$

其中, $\mathbf{c}_v^l \in \mathbf{R}^d$ 为第 l 层节点 v 的嵌入表示, l 为当前网络层数, $\mathbf{s}_v^0 \in \mathbf{R}^d$ 为节点 v 的初始特征, $\text{LN}(\cdot)$ 表示层归一化操作, η 为控制残差连接比例的权重因子。上下文随机采样算法的具体流程如算法 1 所示。

算法 1 上下文随机采样算法

输入 异构图 $G=(V,E,F,R,\varphi,\phi)$, 节点序列长度 seqLen 。

输出 拓扑序列 $P(v)$

- (1) begin
- (2) for $v \in V$ do
- (3) 设定序列起始节点 $P[v, 0] = v$;
- (4) 设计数器 $\text{cnt} = 1$;
- (5) for $\text{cnt} < \text{seqLen}$ do
- (6) 获取节点 v 的高阶邻居集合 $N(v)$;

(7) 随机选择 $N(v)$ 中的节点, 将其添加至 $P[v, \text{cnt}]$;

(8) 更新计数器 $\text{cnt} += 1$;

(9) 若 $(\text{cnt} = \text{seqLen})$, 则跳出循环;

(10) end for

(11) end for

(12) return $P(v)$

3.3 节点异构性聚合模块

在异构网络中, 不同类型节点的特征空间差异显著, 直接聚合可能导致异构信息混淆^[28]。为了解决该问题, 本文引入特征空间叠加机制, 保留节点类型的异构信息并增强模型学习能力。在局部拓扑信息聚合模块中, 所有节点特征已投影至统一特征空间 $\mathbf{H} \in \mathbf{R}^{N \times D}$ 。为了确保不同类型节点特征的独立处理, 本文构建类型感知嵌入表示 $\mathbf{H}' \in \mathbf{R}^{N \times D \times T}$, 其中 T 表示节点类型数。另外, 将各类型节点分配至独立的特征维度区间, 从而避免不同类型特征的混淆。具体而言, 若图中存在 3 种节点类型, 则 t_1 类型节点的初始特征被聚合在 D 至 $2D$ 维度区间, 而其他区域初始化为零:

$$\mathbf{H}'_v = [0 \quad \mathbf{H}_{t_1} \quad 0] \quad (16)$$

在特征聚合过程中, 邻居节点的特征被加权聚合到对应的维度区间。 t_1 和 t_2 类型节点的聚合过程如式 (17) 所示:

$$\mathbf{H}' = [0 \quad \mathbf{H}_{t_1} \quad \lambda \mathbf{H}_{t_2}] \quad (17)$$

其中, $\mathbf{H}' \in \mathbf{R}^{N \times D \times T}$ 为经过特征空间叠加后的嵌入表示, λ 为聚合过程中的特征权重系数。

为了融合节点的拓扑信息与类型信息, 本文引入自适应参数 $\text{struct}_{\text{bt}}$, 用于控制各类型节点在融合拓扑嵌入 \mathbf{c}_v 时的权重比例。该参数的适应过程如下: 首先, 在节点拓扑嵌入与特征嵌入融合前, 为每种节点类型 t 初始化可学习参数向量 $\text{struct}_{\text{bt}}[t_v]$, 其中 t_v 为节点 v 的类型索引; 其次, 利用 $\text{struct}_{\text{bt}}$ 控制各类型节点拓扑嵌入与特征嵌入的融合比例; 最后, 通过随机梯度下降和反向传播优化该



参数。计算过程如式 (18) 所示:

$$\mathbf{H}'_v = \mathbf{H}'_v + \mathbf{H}'_v \cdot \text{struct}_{\text{bt}} \cdot \mathbf{c}_v \quad (18)$$

其中, $\mathbf{H}'_v \in \mathbf{R}^{T \times d}$ 为节点 v 特征空间叠加与拓扑信息聚合后的节点表示, $\text{struct}_{\text{bt}} \in \mathbf{R}^{T \times d}$ 为各节点类型的自适应权重, $\mathbf{c}_v \in \mathbf{R}^{T \times d}$ 为节点 v 的拓扑嵌入表示。

为了计算节点对间的注意力系数 $\beta_{v,u}$, 本文采用自注意力机制, 通过点积形式计算目标节点和邻居节点的相关性, 应用 $\text{Softmax}(\cdot)$ 归一化得到 $\beta_{v,u}$:

$$\beta_{v,u} = \text{Softmax}(\mathbf{q}_v^T \mathbf{k}_u) \quad (19)$$

其中, $\mathbf{q}_v \in \mathbf{R}^{T \times d}$ 和 $\mathbf{k}_u \in \mathbf{R}^{T \times d}$ 分别为目标节点 v 的查询向量和直接邻居 u 的键向量, $\beta_{v,u} \in \mathbf{R}^{T \times d}$ 为节点对 (v,u) 的注意力系数。随后, 将注意力系数 $\beta_{v,u}$ 与邻居节点 u 的嵌入表示 \mathbf{h}_u^l 加权聚合, 得到目标节点 v 最终的特征嵌入:

$$\mathbf{h}_v^{l+1} = \sum_{u \in N(v)} \beta_{v,u} \cdot \mathbf{h}_u^l \quad (20)$$

其中, $\mathbf{h}_u^l \in \mathbf{R}^{T \times d}$ 表示第 l 层节点 u 的特征嵌入, l 表示该模块层数, $\beta_{v,u} \in \mathbf{R}^{T \times d}$ 为节点对 (v,u) 的注意力系数。此外, 为了减小注意力机制计算过程中可能出现的方差波动问题, 引入与高阶拓扑信息聚合模块类似的多头机制以提升计算的稳定性:

$$\mathbf{h}_v^{l+1} = //_{i=1}^K \delta \left(\sum_{u \in N(v)} \beta_{v,u}^i \cdot \mathbf{W}^{(l,i)} \cdot \mathbf{h}_u^l + \mathbf{b}^{(l,i)} \right) \quad (21)$$

其中, $\mathbf{h}_v^{l+1} \in \mathbf{R}^d$ 为仅保留所属类型的最终嵌入, $\mathbf{h}_u^l \in \mathbf{R}^{T \times d}$ 表示第 l 层邻居节点 u 的特征嵌入, $\beta_{v,u} \in \mathbf{R}^{T \times d}$ 为节点对 (v,u) 的注意力系数, K 表示多头数量, i 表示多头序号, $\mathbf{W}^{(l,i)}$ 和 $\mathbf{b}^{(l,i)}$ 表示第 i 个注意力头的权重矩阵和偏置向量, δ 是非线性激活函数, $//$ 表示向量拼接操作。

3.4 模型优化

为了更好地将节点嵌入 \mathbf{h}_v 应用于下游任务中, 首先使用线性函数 $\text{Linear}(\cdot)$ 将其映射至预测维度空间:

$$\tilde{y}_v = \text{Linear}(\mathbf{h}_v; \theta) \quad (22)$$

其中, $\mathbf{h}_v \in \mathbf{R}^d$ 为节点 v 的最终嵌入, $\tilde{y}_v \in \mathbf{R}^{\text{class}}$ 为预测标签, class 为预测标签的类别数, θ 为可学习参数。在半监督分类任务中, 模型使用随机梯度下降和反向传播优化模型权重:

$$\text{Loss} = - \sum_{v \in V_{\text{train}}} \text{CrossEnt}(\tilde{y}_v, y_v) \quad (23)$$

其中, V_{train} 为训练集, \tilde{y}_v 为预测标签, y_v 为真实标签。整体算法描述在算法2中给出。

算法2 TEHGNN 算法

输入 异构图 $G=(V,E,F,R,\phi)$, 采样序列长度 seqLen , 其中 V 为节点集合, E 为边集合, F 为节点类型集合, R 为边类型集合, ϕ 为节点类型映射函数, ϕ 为边类型映射函数。

输出 节点嵌入集合 $\mathbf{H}=\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_{n-1}$ 。

- (1) begin
- (2) for $v \in V$ do
- (3) 将 v 的特征 \mathbf{X}_v 投影到统一特征空间, 得到投影后的特征 \mathbf{h}_v^0 ;
- (4) end for
- (5) for $v \in V$ do
- (6) 计算 v 的邻居结构注意力 $\alpha_{v,u}$;
- (7) 融合局部拓扑信息 \mathbf{s}_v ;
- (8) end for
- (9) for $v \in V$ do
- (10) 使用上下文随机采样算法获取采样序列 $L(v)=\{\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\text{seqLen}-1}\}$;
- (11) 对序列 $L(v)$ 进行图卷积操作增强拓扑信息, 得到拓扑序列 $\mathbf{P}(v)$;
- (12) 对拓扑序列 $\mathbf{P}(v)$ 进行多层聚合, 得到高阶拓扑信息 \mathbf{c}_v ;
- (13) end for
- (14) for $v \in V$ do
- (15) 扩展节点 v 的特征空间;
- (16) 使用自适应参数将高阶拓扑信息 \mathbf{c}_v 与节点特征 \mathbf{H}'_v 进行聚合;
- (17) 将不同类型节点特征映射到对应的类

- 型维度区域，避免特征混淆；
- (18) 对不同类型节点的特征进行加权聚合，得到最终的节点嵌入表示 \mathbf{h}_v ；
- (19) end for
- 返回节点嵌入集合 $\mathbf{H}=\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_{n-1}$ ；
- (20) end

从算法2中可知，代码的第(2)~(4)行对节点的初始特征进行线性映射，将其投影至统一特征空间，生成初步的节点表示；第(5)~(8)行利用节点的拓扑嵌入计算结构注意力系数，以有效捕捉其局部拓扑信息；第(9)~(12)行利用上下文随机采样算法生成拓扑序列，并结合自注意力机制实现对高阶拓扑信息的聚合；第(13)~(18)行表示针对不同类型节点引入特征空间叠加机制，以确保异构信息得到保留。

4 实验与结果分析

4.1 数据集

本文采用3个公开数据集进行实验，数据集描述见表1。

表 1 数据集描述

数据集	节点数量	边数量	目标节点	预测类别数
DBLP	Author: 4 057	239 566	Author	4
	Paper: 14 328			
	Term: 7 723			
	Venue: 20			
ACM	Paper: 3 025	547 872	Paper	3
	Author: 5 959			
	Subject: 56			
	Term: 1 902			
IMDB	Movie: 4 932	86 642	Movie	5
	Director: 2 393			
	Actor: 6 124			
	Keyword: 7 971			

(1) DBLP是计算机科学参考书目网站。节点类型为4种，边类型为6种。节点分类目标为预测作者节点的类标签，包括数据库、数据挖

掘、人工智能和信息检索。其中，训练集与测试集的分类分布较为均衡，各类别比例的差异不超过10%。此外，在DBLP数据集中，仅Author类型节点拥有特征属性，其余节点均为无属性节点，占节点总数的84.48%。因此，补全无属性节点能增强节点表示学习的能力。

(2) ACM是引文网络数据库。节点类型为4种，边类型为8种。节点分类目标为预测论文节点的类标签，包括：数据库、无线通信和数据挖掘。其中，训练集和测试集的分类分布较为均衡，各类别之间的比例差异不超过5%，且两者分布趋势保持一致。此外，在ACM数据集中，Paper类型节点的初始特征维度为1 902，而其余节点均为无属性节点，占节点总数的72.36%。因此，Paper类型节点在补全无属性节点时具备较强的表达能力。

(3) IMDB是关于电影资料的数据库。节点类型为4种，边类型为6种。节点分类目标为预测电影节点的类标签，包括：动作、喜剧、戏剧、爱情和惊悚。其中，训练集和测试集的分类分布较为均衡，各类别的出现频率均处于15%~25%，具有良好的覆盖性与代表性。此外，在IMDB数据集中，Movie类型节点具有丰富的语义特征，其初始特征维度为3 489，而其余节点均为无属性节点，占节点总数的76.98%。因此，Movie节点在异构图中具有显著的语义作用，有助于无属性节点的补全。

4.2 评价指标

节点分类实验使用Macro-F1和Micro-F1作为评价指标，用于衡量模型的性能。其具体定义如式(24)和式(25)所示：

$$P_z = \frac{TP_z}{TP_z + FP_z}, R_z = \frac{TP_z}{TP_z + FN_z}$$

$$\text{Macro-F1} = \frac{1}{Z} \sum_{z=1}^Z \frac{2 \times P_z \cdot R_z}{P_z + R_z} \quad (24)$$



$$\text{Micro-F1} = \frac{2 \cdot \text{TP}_{\text{total}}}{2 \cdot \text{TP}_{\text{total}} + \text{FP}_{\text{total}} + \text{FN}_{\text{total}}} \quad (25)$$

在节点聚类实验中，使用常用评价指标归一化互信息（normalized mutual information, NMI）与调整兰德指数（adjusted Rand index, ARI）来衡量聚类结果与真实标签之间的一致性。其具体定义如式（26）所示：

$$\begin{aligned} \text{NMI}(X, Y) &= \frac{I(X, Y)}{\sqrt{H(X)H(Y)}} \\ I(X, Y) &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ \text{ARI} &= \frac{\text{RI} - E[\text{RI}]}{\max(\text{RI}) - E[\text{RI}]}, \text{RI} = \frac{a + d}{C_2^n} \quad (26) \end{aligned}$$

其中， X 和 Y 分别是聚类结果和真实标签， $I(X, Y)$ 反映了2个变量共享的信息量， $p(x, y)$ 是 X 和 Y 的联合概率分布， $p(x)$ 和 $p(y)$ 分别是 X 和 Y 的边际概率分布。 $H(X)$ 和 $H(Y)$ 分别是 X 和 Y 的熵， a 是2个元素被正确分类到同一组的对数， d 是2个元素被正确分类到不同组的对数， n 是样本总数。

4.3 对比模型

为了全面评估TEHGNN的性能，本文选择多种具有代表性的图嵌入方法，并将其在节点分类和节点聚类任务中的表现进行比较，以突出TEHGNN在解决问题上的优势。具体模型包括如下。

(1) GAT^[21]：一种经典同构图神经网络算法，通过在节点信息传递过程中引入注意力机制，使节点在聚合邻居信息时能够根据邻居的重要性调整权重。与该算法的对比实验，能凸显TEHGNN结构和节点异构性注意力机制的有效性。

(2) GCN^[17]：一种将卷积操作引入图结构数据的同构图神经网络算法，通过图卷积操作对每个节点及其邻居节点的特征进行加权平均或加权聚合，实现特征的学习。与该算法的对比实验，能体现TEHGNN分层拓扑信息聚合的有效性

及捕获节点异构信息的必要性。

(3) GTN（graph transformer network）^[29]：一种自动生成元路径的异构图神经网络算法，利用卷积操作捕获新的图结构和元路径，端到端地学习新图上的节点表示。与该算法的对比实验，能够突出TEHGNN在卷积操作的基础上使用多跳采样序列捕获高阶拓扑信息的优越性。

(4) HAN^[10]：一种拓展了GAT模型的异构图神经网络算法，该方法从节点级和语义级两个层次引入注意力机制，节点级注意力用于捕获局部邻居信息，语义级注意力用于捕获节点异构信息。与该算法的对比实验，能体现TEHGNN捕获图拓扑信息操作的合理性，突出拓扑信息对节点表示学习的重要性。

(5) MAGNN_AC（metapath aggregated graph neural network via attribute completion）^[5]：一种基于属性补全的异构图神经网络框架，将预训练嵌入作为原始嵌入，利用注意力机制补全无属性节点，从而捕获复杂的拓扑信息。与该算法的对比实验，更能突出TEHGNN捕获高阶拓扑信息的必要性。

(6) Simple-HGN^[6]：一种高效的异构图神经网络，利用注意力机制计算节点和边嵌入的重要性，应用残差连接解决梯度消失问题，并通过平均聚合获取节点的最终嵌入。与该算法的对比实验，能验证本文算法使用特征空间叠加机制解决语义混淆问题的能力。

(7) NLA-GNN（non-local information aggregated graph neural network）^[30]：一种聚合非局部信息的异构图神经网络方法，利用局部和非局部聚合优化节点嵌入。该算法首先利用卷积生成元路径，其次通过注意力机制构建多类型节点序列，最后叠加多层卷积进行非局部信息聚合。与该算法的对比实验表明，TEHGNN在捕获高阶拓扑信息和聚合无属性节点拓扑信息方面更为有效。

4.4 实验设置

本文所提 TEHGNN 的运行硬件环境为 Intel(R) Xeon(R) Platinum 8474C CPU+RTX 4090D (24 GB) GPU, 内存为 80 GB, 操作系统为 Ubuntu 18.04。使用的开发工具包括: python3.8、pytorch 1.13.1+cu117、dgl 1.0.0+cu117 以及 cuda 11.7。

本文所提算法 TEHGNN 通过多次实验确定以下参数: 学习率设置为 0.005, 训练轮次为 300, 丢弃率为 0.5, 早停耐心值为 30, 多头自注意力数均为 8。隐藏层的嵌入维度由具体数据集确定, IMDB 为 128, ACM 和 DBLP 为 64。对比实验的参数设置参考原作者给出的推荐参数。

4.5 节点分类

节点分类实验是评估节点嵌入效果的常用实

验。本文采用支持向量机对嵌入结果进行分类, 分别设置 20%、40%、60%、80% 的训练集比例, 并重复实验 10 次取均值作为最终结果。评价指标为 Macro-F1 和 Micro-F1, 节点分类任务实验结果见表 2。表 2 中最优结果以加粗字体表示, 次优结果以下画线标注。

由表 2 可知, TEHGNN 在 3 个不同数据集上表现出色。在 IMDB 数据集上, TEHGNN 的 Macro-F1 和 Micro-F1 指标相较于基线算法平均提升 0.52%~2.04%; 在 ACM 数据集上, 提升幅度为 0.26%~0.39%; 在 DBLP 数据集上, 提升幅度为 0.23%~0.35%。这表明 TEHGNN 在不同异构图场景中均具备较强的鲁棒性。

实验结果表明, 同构图神经网络在处理异构

表 2 节点分类任务实验结果

数据集	评价指标	训练集比例	GAT	GCN	GTN	HAN	MAGNN_AC	Simple-HGN	NLA-GNN	TEHGNN
IMDB	Macro-F1	20.0%	54.8%	50.0%	57.3%	57.5%	59.7%	62.2%	<u>63.6%</u>	63.8%
		40.0%	55.1%	55.4%	57.9%	57.6%	60.2%	62.8%	<u>64.0%</u>	64.5%
		60.0%	55.7%	56.9%	58.0%	57.8%	60.7%	63.6%	<u>64.1%</u>	64.6%
		80.0%	55.4%	57.2%	58.8%	57.8%	60.8%	63.9%	<u>64.2%</u>	64.7%
	Micro-F1	20.0%	55.0%	58.4%	57.1%	57.2%	59.8%	62.2%	<u>63.4%</u>	64.7%
		40.0%	55.6%	58.9%	57.8%	57.6%	60.4%	62.5%	<u>63.5%</u>	65.0%
		60.0%	55.9%	61.6%	57.9%	57.8%	60.8%	63.5%	<u>63.9%</u>	65.9%
		80.0%	55.7%	55.7%	58.7%	58.0%	61.0%	63.9%	<u>64.2%</u>	66.1%
DBLP	Macro-F1	20.0%	66.9%	90.1%	91.7%	91.6%	93.9%	93.8%	<u>94.3%</u>	94.4%
		40.0%	73.2%	90.3%	91.8%	91.6%	94.1%	93.8%	<u>94.5%</u>	94.8%
		60.0%	77.2%	90.5%	92.0%	92.2%	94.0%	93.9%	<u>94.6%</u>	95.2%
		80.0%	78.2%	90.8%	92.2%	92.3%	94.2%	94.3%	<u>94.8%</u>	95.4%
	Micro-F1	20.0%	77.0%	90.5%	92.3%	92.3%	94.3%	94.2%	<u>94.6%</u>	95.2%
		40.0%	79.6%	91.0%	92.5%	92.5%	94.5%	94.6%	<u>94.7%</u>	95.3%
		60.0%	81.6%	91.1%	92.7%	92.7%	94.5%	94.6%	<u>95.0%</u>	95.6%
		80.0%	82.2%	91.2%	92.8%	92.9%	94.6%	94.7%	<u>95.6%</u>	96.0%
ACM	Macro-F1	20.0%	89.6%	90.2%	90.9%	90.2%	91.5%	92.9%	<u>93.0%</u>	93.1%
		40.0%	89.8%	90.8%	91.4%	90.2%	92.8%	93.0%	<u>93.1%</u>	93.2%
		60.0%	89.7%	90.7%	91.7%	91.4%	93.3%	93.2%	<u>93.4%</u>	93.5%
		80.0%	89.4%	91.4%	91.8%	91.6%	93.8%	93.5%	<u>93.9%</u>	94.8%
	Micro-F1	20.0%	89.5%	90.3%	90.8%	89.9%	91.6%	92.7%	<u>92.8%</u>	92.9%
		40.0%	89.7%	90.7%	91.2%	90.4%	92.9%	92.9%	<u>93.0%</u>	93.2%
		60.0%	89.6%	91.3%	91.6%	91.4%	93.2%	93.5%	<u>93.6%</u>	94.2%
		80.0%	89.3%	91.9%	91.8%	91.8%	93.5%	93.8%	<u>93.9%</u>	94.5%



数据时存在明显劣势。这是因为同构图算法未区分节点类型，仅采用单一的聚合方式，难以有效捕获节点的异质性特征。异构图算法相较于同构图算法能捕获节点的异构信息，在3个数据集上有明显的性能提升。具体而言，HAN采用双层注意力机制聚合节点邻居信息，但聚焦于节点属性特征，未能充分考虑拓扑信息。MAGNN_AC通过注意力机制补全无属性节点，利用无属性节点的语义信息，提升了模型对节点属性的捕获能力。Simple-HGN引入边缘注意力机制以建模边缘异构性，进一步提高嵌入表示的准确性。NLA-GNN融合了局部和非局部多层次信息，能够捕获异构节点的复杂语义。然而，这些方法大多只关注拓扑或特征中的一方，忽略了二者之间的关联性与互补性。本文提出的TEHGNN则从拓扑结构与节点特征两个维度出发，实现了两者细粒度的联合建模。在拓扑建模方面，TEHGNN设计了局部与高阶拓扑信息聚合模块，不仅能够捕捉节点的直接邻居关系，还通过上下文采样序列引入远距离节点信息，从而增强对高阶拓扑关系的建模能力；在特征建模方面，TEHGNN引入特征空间叠加机制和自注意力机制，充分提取节点类型信息，保留其异构表达能力。总体而言，TEHGNN实现了拓扑结构与属性特征的互补融合，有效提升了节点表示的准确性与表达力。

4.6 节点聚类

为了进一步比较不同模型所得节点嵌入间的相似性，本文在IMDB、ACM和DBLP 3个数据集上

进行聚类实验。实验使用K-means算法，计算NMI和ARI两项指标。每组实验重复运行10次，取均值作为最终结果，节点聚类任务实验结果见表3。

从表3可见，TEHGNN在节点聚类任务中表现最优，3个数据集上的NMI和ARI指标均优于其他对比算法。这表明TEHGNN生成的节点嵌入在相似性判别与类别区分方面具有更强能力。具体而言，HAN利用双层注意力机制聚合节点特征，但缺乏对高阶拓扑信息的建模；MAGNN_AC通过属性补全增强了无属性节点的表达能力，但对拓扑信息的捕获有所欠缺；Simple-HGN注重边缘异构性，提升了模型对边缘信息的捕获能力，但在高阶拓扑建模方面仍有不足；NLA-GNN融合了局部与非局部的多层次信息，但在处理无属性节点时能力较为有限。相比之下，TEHGNN在特征与拓扑建模的互补融合上表现更为全面，因此在聚类任务中取得了更出色的效果。

4.7 消融实验

为了进一步验证算法的有效性，本节设计多个变体模型，并在相同实验环境和参数设置下进行消融实验。各变体模型的定义如下。

(1) TEHGNN_ls: 表示不使用局部拓扑信息增强模块。

(2) TEHGNN_hs: 表示不进行高阶随机采样，将采样序列长度设置为1。

(3) TEHGNN_s: 表示不聚合拓扑信息，仅利用自注意力机制，从节点属性中捕获拓扑和语义信息。

表3 节点聚类任务实验结果

数据集	评价指标	HAN	MAGNN_AC	Simple-HGN	NLA-GNN	TEHGNN
IMDB	NMI	10.8%	13.7%	13.2%	15.6%	16.9%
	ARI	11.1%	14.4%	14.9%	18.5%	19.4%
DBLP	NMI	77.5%	79.6%	81.2%	81.6%	83.2%
	ARI	78.0%	84.9%	85.0%	86.5%	87.4%
ACM	NMI	77.5%	64.9%	65.3%	72.5%	74.6%
	ARI	78.5%	68.5%	67.8%	74.5%	77.2%

(4) TEHGNN_het: 表示不对节点特征空间进行叠加操作。

TEHGNN的各个变体算法在节点分类任务中的消融实验结果如图4所示。根据结果可以得出以下结论。

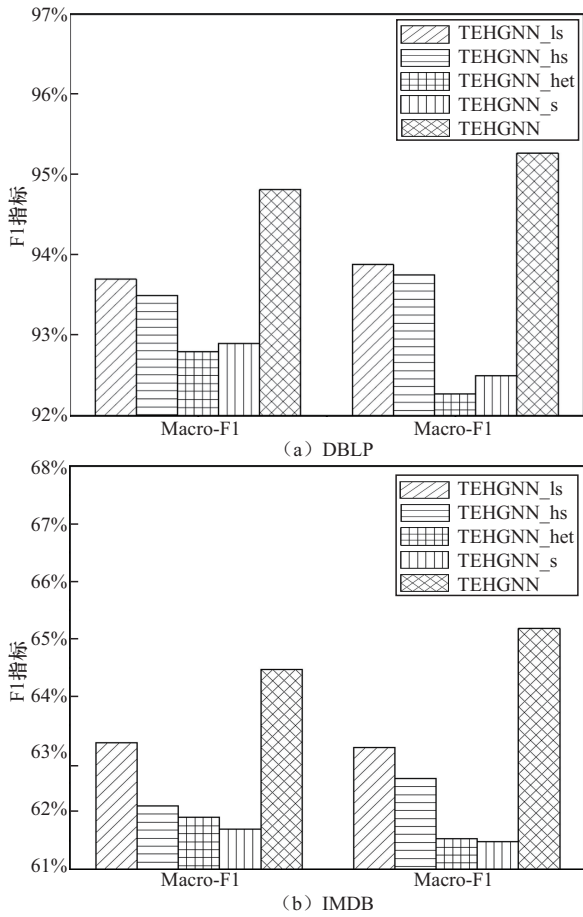


图4 消融实验结果

(1) TEHGNN_ls的性能明显低于TEHGNN, 表明局部拓扑信息增强模块能够有效捕获节点的邻接拓扑信息。对于无属性节点, 该模块能通过邻接注意力机制补全节点特征, 进一步增强局部拓扑信息的捕获能力, 从而增强节点嵌入效果。

(2) TEHGNN_hs的性能低于TEHGNN, 表明高阶随机采样算法能有效挖掘节点的高阶邻居信息, 并通过全局注意力机制加权聚合, 进一步增强拓扑结构特征的捕获能力。

(3) TEHGNN_het在3个数据集上的性能均

显著下降, 表明特征空间叠加机制能保留异构信息, 并有效缓解聚合时的语义混淆, 从而提高节点嵌入的质量。

(4) TEHGNN_s的性能下降最为显著。在DBLP数据集中, Macro-F1和Micro-F1值分别为92.9%和92.5%, 较TEHGNN存在明显差距, 表明拓扑信息在节点表示学习中具有重要意义。该趋势在属性特征稀疏的IMDB数据集中更加明显, 进一步验证了拓扑结构在此类场景中的关键作用。

4.8 参数敏感性实验

为了测试模型在不同参数设置下的性能变化, 本文在DBLP数据集上设计了针对嵌入维度、学习率和丢弃率这三个重要参数的敏感性实验。

4.8.1 嵌入维度

嵌入维度参数分析如图5所示。随着维度增加, 模型性能先升后降。在低维情况下, 增加嵌入维度能提升模型性能; 但当维度超过一定阈值后, 冗余信息增多, 易导致模型过拟合, 从而造成性能下降。在DBLP数据集中, 嵌入维度为64时表现最佳, 表明适度增加维度能提升学习能力, 但过度增加会影响泛化能力。

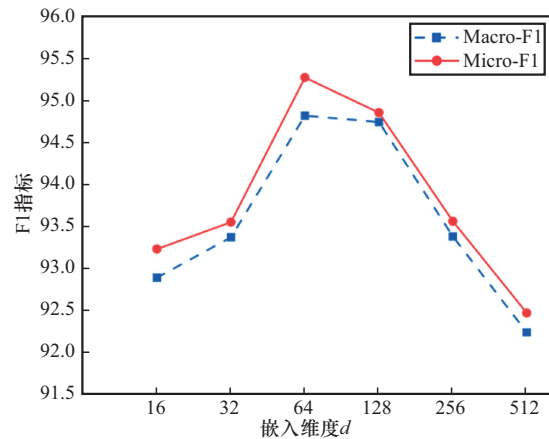


图5 嵌入维度参数分析

4.8.2 学习率

学习率的选择对模型训练过程具有重要影响。较大的学习率可能会使参数更新过快, 进而



导致难以收敛。较小的学习率虽然能稳定训练过程，但可能导致收敛缓慢。学习率参数分析如图6所示，在DBLP数据集上，当学习率设置为0.005时，模型性能达到最佳，且训练过程较为稳定。

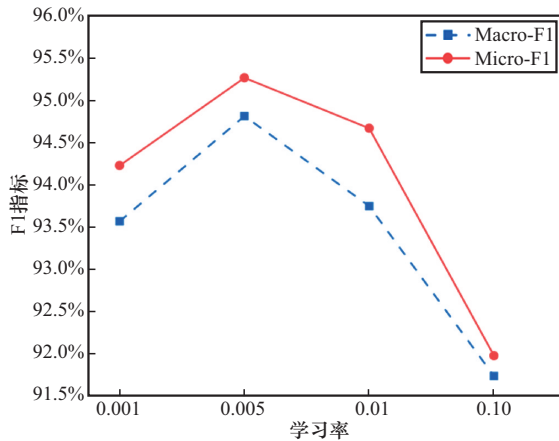


图6 学习率参数分析

4.8.3 丢弃率

丢弃率的设置对模型泛化能力起着关键作用。在训练过程中，低丢弃率可能导致模型对训练数据过度拟合，难以应用于新数据；而较高的丢弃率则可能舍弃过多有效信息，限制模型学习能力。丢弃率参数分析如图7所示，在DBLP数据集上，当丢弃率设为0.5时，模型性能达到最优，展现出较强的抗过拟合能力和良好的泛化效果。

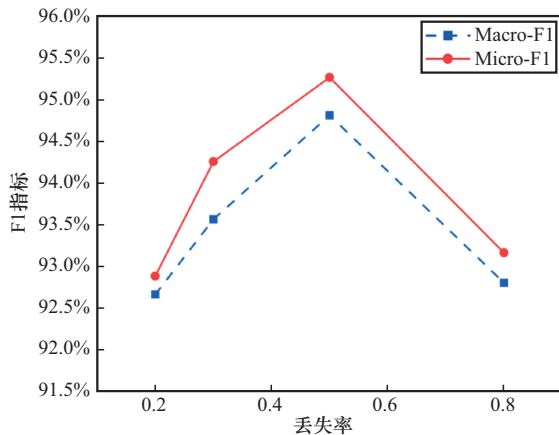


图7 丢弃率参数分析

4.9 有效性分析

为了验证高阶拓扑信息聚合模块的有效性和采样序列长度对模型性能的影响，本文在IMDB数据集上进行了不同采样长度的实验设置。采样长度分析如图8所示，随着采样长度从5增加至50，Macro-F1和Micro-F1指标持续上升，在长度为50时达到最优表现。超过该阈值后，性能开始下降，尤其在采样长度为200时下降最为明显。该趋势表明，采样长度过短会使模块退化为局部拓扑聚合模块，而过长则可能引入噪声信息，削弱高质量邻居对目标节点的语义贡献。此外，该实验同时验证了噪声节点对高阶语义捕获的干扰作用，进一步强调了合理控制采样范围的重要性。

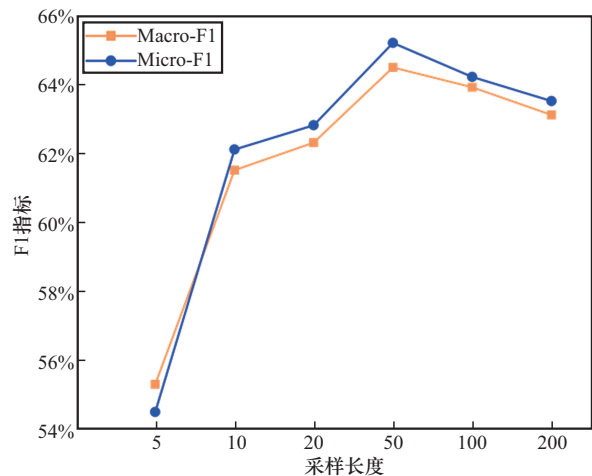


图8 采样长度分析

为了比较4种异构图节点表示算法在节点嵌入上的效果，本文在DBLP数据集上使用T-SNE^[31]方法对节点嵌入进行降维和可视化处理，节点嵌入可视化结果如图9所示。该数据集包含4类目标节点，每类节点使用不同形状区分。实验比较了MAGNN-AC、Simple-HGNN、NLAGNN和TEHGNN的可视化嵌入结果。从可视化结果来看，MAGNN-AC的嵌入分布存在明显的类别重叠和边界模糊现象。这表明其在处理复杂类别关系时区分能力较弱。Simple-HGNN的类别区域相比MAGNN-AC更加分离，每类节点呈紧

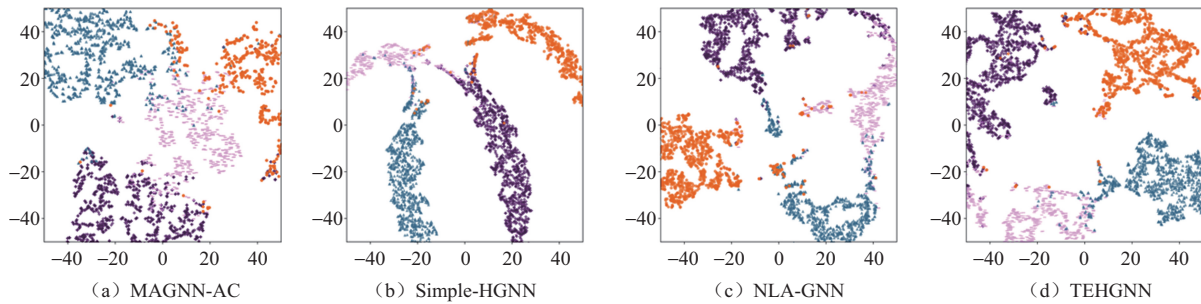


图9 节点嵌入可视化结果

凑的团块状分布，但类别间分隔较窄。这表明其在处理跨类别节点时表现不够理想。

相比之下，NLA-GNN 和 TEHGNN 在节点嵌入上表现优异。两者在可视化图中均呈现出清晰的类簇结构，不同类别的节点分布于图的边缘，形成边界明确、结构紧凑的团块。其中，TEHGNN 的类别间交叉和重叠区域最少，进一步验证了其在节点表示学习方面的优势。TEHGNN 的优势归因于模块设计的协同作用：局部拓扑信息增强模块细粒度地捕获局部拓扑信息，高阶拓扑信息聚合模块通过随机游走和全局注意力机制捕获高阶拓扑信息，而节点异构性聚合模块通过特征空间叠加机制保留了不同类型节点的多样性，显著缓解了语义混淆问题。综上，TEHGNN 不仅在可视化实验中展示出良好的类簇分布特性，也为后续的分类与聚类任务提供了更具判别力的嵌入表示。

5 结束语

为了应对异构图在拓扑信息和节点异构信息聚合上的不足，本文提出了一种基于拓扑信息增强的异构图神经网络算法。该算法通过创新性地融合上下文随机采样算法、注意力机制和特征空间叠加机制等技术，有效整合无属性节点的拓扑信息，并利用节点拓扑信息与属性信息之间的相关性，实现异构图节点信息的全面学习。具体而言，首先，TEHGNN 使用局部拓扑信息增强模块捕获节点的局部拓扑信息；然后，利用高阶拓扑信息聚合模块构建高阶邻

居序列，从而提取多跳邻居的拓扑信息；最后，通过特征空间叠加机制将局部和高阶拓扑信息融入节点特征，并借助注意力机制学习包含复杂语义的节点嵌入。在 3 个公开数据集上的实验结果表明，TEHGNN 在节点分类和节点聚类任务中表现良好，尤其在处理无属性节点和复杂异构网络时展现显著优势，验证了其在实际复杂网络任务中的应用潜力。

未来节点嵌入的研究可以从以下两个方向进行优化和拓展。(1) 针对大规模异构图的处理，可以探索如何更高效地获取拓扑信息，这包括设计适用于大规模网络的节点采样策略和邻居聚合方法，从而更好地适应实际网络场景。(2) 针对异构图中节点和边的多样性，可以研究如何在节点嵌入中更有效地保留节点与边的异构信息，这需要关注不同节点和边类型之间的关系，并设计适用于复杂网络的类型感知策略，以增强模型处理异质性问题的能力。

参考文献：

- [1] WANG X, BO D Y, SHI C, et al. A survey on heterogeneous graph embedding: methods, techniques, applications and sources[J]. IEEE Transactions on Big Data, 2022, 9(2): 415-436.
- [2] 陶涛, 李珍, 王冀彬, 等. 基于图神经网络的权益推荐技术方案研究[J]. 电信科学, 2023, 39(8): 91-101.
TAO T, LI Z, WANG J B, et al. Research on the graphical convolution neural network based benefits recommendation system strategy[J]. Telecommunications Science, 2023, 39(8): 91-101.
- [3] LI S B, GONG J, KE S N, et al. Graph transformer-based heterogeneous graph neural networks enhanced by multiple meta-



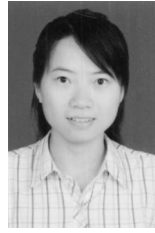
- path adjacency matrices decomposition[J]. *Neurocomputing*, 2025, 629: 129604.
- [4] DONG Y X, CHAWLA N V, SWAMI A. Metapath2vec: scalable representation learning for heterogeneous networks[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2017: 135-144.
- [5] JIN D, HUO C Y, LIANG C D, et al. Heterogeneous graph neural network via attribute completion[C]// *Proceedings of the Web Conference 2021*. New York: ACM Press, 2021: 391-400.
- [6] LYU Q S, DING M, LIU Q, et al. Are we really making much progress?: revisiting, benchmarking and refining heterogeneous graph neural networks[C]// *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. New York: ACM Press, 2021: 1150-1160.
- [7] LI C, FU J H, YAN Y Y, et al. Higher order heterogeneous graph neural network based on node attribute enhancement[J]. *Expert Systems with Applications*, 2024, 238: 122404.
- [8] 贾霄生, 赵中英, 李超, 等. 互信息与多条元路径融合的异质网络表示学习方法[J]. *软件学报*, 2023, 34(7): 3256-3271.
- JIA X S, ZHAO Z Y, LI C, et al. Heterogeneous network representation learning method fusing mutual information and multiple meta-paths[J]. *Journal of Software*, 2023, 34(7): 3256-3271.
- [9] 王瑞琴, 黄熠旻, 纪其顺, 等. 注意力感知的边-节点交换图神经网络模型[J]. *电信科学*, 2024, 40(1): 106-114.
- WANG R Q, HUANG Y M, JI Q S, et al. Attention aware edge-node exchange graph neural network[J]. *Telecommunications Science*, 2024, 40(1): 106-114.
- [10] WANG X, JI H Y, SHI C, et al. Heterogeneous graph attention network[C]// *Proceedings of the World Wide Web Conference*. New York: ACM Press, 2019: 2022-2032.
- [11] FU X Y, ZHANG J N, MENG Z Q, et al. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding[C]// *Proceedings of the Proceedings of The Web Conference 2020*. New York: ACM Press, 2020: 2331-2341.
- [12] LIN W H, CHEN Z L, CHEN Y H, et al. Heterogeneous graph neural network with adaptive relation reconstruction[J]. *Neural Networks*, 2025, 187: 107313.
- [13] NIKOLENTZOS G, MELADIANOS P, VAZIRGIANNIS M. Matching node embeddings for graph similarity[C]// *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. New York: ACM Press, 2017: 2429-2435.
- [14] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: online learning of social representations[C]// *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2014: 701-710.
- [15] TANG J, QU M, WANG M Z, et al. LINE: large-scale information network embedding[C]// *Proceedings of the 24th International Conference on World Wide Web*. New York: ACM Press, 2015: 1067-1077.
- [16] RIBEIRO L F R, SAVERESE P H P, FIGUEIREDO D R. Struc2vec: learning node representations from structural identity[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2017: 385-394.
- [17] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB]. 2016.
- [18] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// *Proceedings of the Semantic Web*. Cham: Springer International Publishing, 2018: 593-607.
- [19] CHEN K J, LU H, LIU Z, et al. Heterogeneous graph convolutional network with local influence[J]. *Knowledge-Based Systems*, 2022, 236: 107699.
- [20] 杨帅, 王瑞琴, 马辉. 基于多通道的边学习图卷积网络[J]. *电信科学*, 2022, 38(9): 95-104.
- YANG S, WANG R Q, MA H. Multi-channel based edge-learning graph convolutional network[J]. *Telecommunications Science*, 2022, 38(9): 95-104.
- [21] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. *Stat*, 2017, 1050(20): 10-48550.
- [22] ZHAO J N, LI C Z, WEN Q L, et al. Gophormer: ego-graph transformer for node classification[EB]. 2021.
- [23] ZHAO Y F, XU S X, DUAN H. HGNN-BRFE: heterogeneous graph neural network model based on region feature extraction[J]. *Electronics*, 2024, 13(22): 4447.
- [24] CHEN Y J, LIU Y. HeGAE-AC: heterogeneous graph auto-encoder for attribute completion[J]. *Knowledge-Based Systems*, 2024, 287: 111436.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017.
- [26] MCPHERSON M, SMITH-LOVIN L, COOK J M. Birds of a feather: homophily in social networks[J]. *Annual Review of Sociology*, 2001, 27: 415-444.
- [27] MAO Q H, LIU Z M, LIU C H, et al. HINormer: representation learning on heterogeneous information networks with graph transformer[C]// *Proceedings of the ACM Web Conference 2023*. New York: ACM Press, 2023: 599-610.
- [28] ZHOU Z A, SHI J M, YANG R C, et al. SlotGAT: slot-based message passing for heterogeneous graph neural network[EB]. 2024.
- [29] YUN S, JEONG M, KIM R, et al. Graph transformer networks[J]. *Advances in Neural Information Processing Systems*, 2019, 32.

- [30] WANG S H, CAO G T, CAO W M, et al. NLA-GNN: non-local information aggregated graph neural network for heterogeneous graph embedding[J]. Pattern Recognition, 2025, 158: 110940.
- [31] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(11).

[作者简介]



姚波武 (2000-), 男, 东华理工大学信息工程学院硕士生, 主要研究方向为异构图神经网络和深度学习等。



魏振华 (1981-), 女, 博士, 东华理工大学副教授, 主要研究方向为智能信息处理和空间信息管理。



吴桐 (1992-), 男, 博士, 嘉兴大学浙江省全省多模态感知与智能系统重点实验室、嘉兴大学信息科学与工程学院讲师, 主要研究方向为太赫兹/毫米波传感、雷达信号处理、纳米光子学等。



邓琨 (1980-), 男, 博士, 嘉兴大学浙江省全省多模态感知与智能系统重点实验室、嘉兴大学信息科学与工程学院副教授、硕士生导师, 主要研究方向为网络结构分析、数据挖掘、异构网络分析等。



刘星妍 (1980-), 女, 嘉兴大学人工智能学院高级工程师, 主要研究方向为数据挖掘、网络结构分析等。