

双空间令牌化与对比增强的异质图 Transformer

焦鹏飞¹, 范子暘², 范浩杨², 鲁逸凡²

¹(杭州电子科技大学 网络空间安全学院, 杭州 310018)

²(杭州电子科技大学 卓越学院, 杭州 310018)

E-mail: fzy@hdu.edu.cn

摘要:近年来,异质图神经网络在处理多类型节点和边的复杂关系方面展现出强大能力,但基于消息传递的架构仍面临表达能力受限、过平滑和过挤压等问题.本文提出基于Transformer架构的CHGormer模型,通过融合对比学习与异构关系编码的令牌生成机制,创新性地解决了异质图中局部异构关系与全局语义依赖的整合难题.具体而言,CHGormer设计了双空间令牌生成策略,在属性与拓扑特征空间中分别采样正负令牌序列,并通过类型感知的邻域聚合生成异构关系表征,将其作为注意力偏置引入全局交互.此外,基于对比学习的跨序列优化进一步增强了节点表示的判别性.在DBLP、Freebase和AMiner 3个基准数据集上的实验表明,CHGormer在节点分类任务中表现优异.本研究为异质图表示学习提供了新思路,并在社交推荐和知识推理等场景中展现出应用潜力.

关键词:图神经网络;Transformer;对比学习;异质图;节点分类

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2026)05-1070-09

Contrastive-enhanced Heterogeneous Graph Transformer with Dual-space Tokenization

JIAO Pengfei¹, FAN Ziyang², FAN Haoyang², LU Yifan²

¹(School of Cyberspace Hangzhou Dianzi University, Hangzhou 310018, China)

²(Zhuoyue Honors College Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: In recent years, Heterogeneous Graph Neural Networks have demonstrated strong capabilities in modeling complex relationships involving multiple types of nodes and edges. However, message-passing-based architectures still face limitations such as restricted expressive power, over-smoothing, and over-squashing. This paper proposes a novel Transformer architecture, CHGormer, which innovatively addresses the challenge of integrating local heterogeneous relations with global semantic dependencies in heterogeneous graphs by combining contrastive learning with a token generation mechanism for heterogeneous relation encoding. Specifically, CHGormer introduces a dual-space token generation strategy, where positive and negative token sequences are sampled separately from the attribute and topology feature spaces. These are then aggregated through type-aware neighborhood aggregation to form heterogeneous relational representations, which are incorporated into global interactions as attention biases. Moreover, a contrastive learning-based cross-sequence optimization is employed to further enhance the discriminative power of node representations.

Keywords: graph neural networks; Transformer; contrastive learning; heterogeneous graph; node classification

0 引言

近年来,异质图因其在学术网络^[1-5], 社交网络^[6-9] 和知识图谱^[10-15] 等复杂现实系统中对多类型节点与边关系的强大建模能力而受到广泛研究,成为图表示学习领域的重要研究方向^[16-21]. 与同质图不同,异质图中节点和边蕴含丰富的语义信息,例如学术网络中作者、论文和会议之间的多元关联,典型表现为多位作者在某一会议上合作发表论文.因此,作为研究异质图的工具,异质图神经网络受到了极大的关注.异质图神经网络通常需要处理两类关键信息:1) 节点类型、边关系等局部异构结构特征;2) 节点的语义关联等全局依赖关系.然而,如何有效融合这两类信息,同时避免传统图神经网络(Graph Neural Network, GNN)的过平滑和过挤压问题,

仍是当前研究的核心挑战.

现有的异质图表示学习方法主要分为两类^[22-28]:一类基于消息传递机制(如HetGNN^[29], RGCN^[30]),通过类型感知的邻居聚合捕捉局部异质性,但由于单层GNN只能聚合直接邻居的信息,且在实际应用中,为了避免高层数导致过平滑^[31,32],因此GNN层数一般限制为2~4层,因此其有限的感受野难以建模长距离依赖;另一类基于图Transformer(如PHGT^[33], HINormer^[34]),通过全局注意力机制突破局部视野限制,并引入异质关系建模模块,来捕获节点间的语义关联.然而,现有图Transformer在异质图上的探索仍存在不足:一方面,传统方法过度依赖全局节点注意力,导致局部异质关系的捕获能力下降;另一方面,虽然tokenization方法通过将复杂数据分割为具有独立语义的单元^[35],缓解了对全局信息的

收稿日期:2025-06-16 收修改稿日期:2025-07-17 基金项目:国家自然科学基金项目(62372146)资助;浙江省自然科学基金项目(LDT23F01015F01)资助. 作者简介:焦鹏飞(通信作者),男,1990年生,博士,教授,CCF会员,研究方向为复杂网络分析、图机器学习、数据智能和网络空间安全;范子暘,男,2004年生,研究方向为异质图神经网络;范浩杨,男,2004年生,研究方向为异质图神经网络;鲁逸凡,男,2002年生,研究方向为异质图神经网络.

过度依赖,但现有方法倾向于选择与目标节点相似性高的节点,使得低相似性节点的信息被弱化^[36],导致模型无法捕获到潜在的关键的语义联系.简单异质图如图 1 所示.

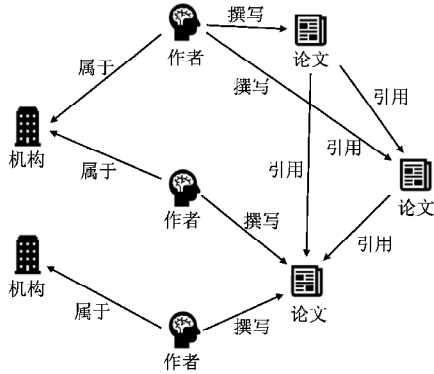


图 1 简单异质图

Fig. 1 Simple heterogeneous graph

针对上述问题,本文提出一种新型异质图 Transformer 模型 CHGormer,通过融合对比增强与异构关系编码的令牌生成机制,实现异质信息的全面建模和内在图属性的捕获.具体而言,首先,本文将原始特征经过结构信息编码,并且设计了异构关系编码器,通过类型感知的邻域聚合生成节点语义关系表征,并将其作为注意力偏置引入全局交互;其次,本文提出双空间令牌生成策略,在属性与拓扑特征空间中分别采样正负令牌序列,捕捉节点间属性与拓扑的共性与差异.此外,基于对比学习的跨序列优化进一步推动节点表示在异质空间中的判别性.在多个异质图数据集上进行的大量实验验证了 CHGormer 的有效性.本文的贡献可总结为:

1) 本文深入剖析了异质图 Transformer 中局部异构关系与全局语义依赖的整合挑战,并开发了基于异构关系感知的对比学习框架以提供系统性解决方案;

2) 本文设计了双空间动态令牌生成策略,通过属性-拓扑双视角对比采样和关系编码偏置注意力实现高效异质语义建模;

3) 在多个异质图数据集上进行了实验,验证了 CHGormer 的有效性.

本研究为异质图表示学习提供了新思路,并在社交推荐^[37-42]、知识推理等场景中展现出应用潜力.在现实世界中,社交推荐系统作为连接社交网络结构与用户兴趣行为的重要桥梁,已在电商、短视频平台等领域广泛部署.传统推荐系统主要依赖用户的历史行为或协同过滤策略,但在异质社交网络中,用户之间存在丰富的多类型关系(如关注、评论、转发),这些异构交互蕴含深层的社会影响机制.异质图神经网络为建模此类复杂社交关系提供了结构化表示框架,能够将用户、内容、标签等多种类型节点与边融合进统一语义空间中进行表征学习.例如,在微博或小红书平台上,用户的兴趣偏好不仅受到其直接关注对象的影响,还受到间接社交路径中多类型实体(如共同话题、转发链路)的共同作用.因此,有效地融合结构信息与语义依赖,实现精准的社交推荐,是异质图学习模型的重要应用方向.本文所提出的 CHGormer 模型,通过属性与拓扑空间下的双视角 Token 建模策略,以及对比增

强机制,在处理此类任务中具有天然优势,能够挖掘潜在的社交兴趣群体并提升推荐准确性.

1 相关工作

1.1 图 Transformer (Graph Transformer, GT)

图 Transformer 通过将 Transformer 架构引入图表示学习,不仅克服了传统 GNN 在长程依赖建模上的局限性,还通过全局自注意力机制实现了全图节点的交互建模,显著提升了对跨层级语义关联的捕获能力;其并行化计算特性大幅优化了训练效率,同时通过类型感知的注意力权重动态调整,能够有效处理异质图中多类型节点和边的复杂语义关系,在节点分类、链接预测等任务中展现出比传统 GNN 模型更高的鲁棒性和泛化性能.现有研究主要分为两类:

1) 同质图上的 GT: 在同质图上,GT 主要通过全局注意力机制来学习节点之间的重要性,部分模型还进一步引入图结构信息来弥补 Transformer 在图数据上的局限性.代表性的模型如 GraphGPS^[43]、GTN^[44]和 Graphormer^[45]. GraphGPS 结合了 GNN 的局部聚合能力和 Transformer 的全局依赖建模能力,通过局部采样减少 Transformer 的计算成本,使其更适用于大规模图数据. GTN 提出结构增强的位置编码机制,通过拉普拉斯特征向量和节点度数构建位置编码,使 Transformer 能够感知图拓扑的层次化特征.此外,GTN 设计动态边权重生成模块,将邻接矩阵作为注意力偏置项,实现图结构信息与全局注意力的协同优化. Graphormer 通过度中心性编码和最短路径编码显式建模图的结构语义.将节点度数作为注意力机制的偏置项,强化中心节点的影响力;利用节点对间的最短路径长度和边特征生成路径感知的注意力权重.

2) 异质图上的 GTs: 在异质图上,GTs 需要学习不同类型的节点和边的关系,因此需要做关系感知的注意力建模.代表性的模型如 HGT^[46]、PHGT 和 HINormer. HGT 提出异构注意力机制,根据边类型动态调整注意力权重,引入元路径来区分不同关系的注意力计算. PHGT 通过引入多类型 token 结果传统 GNN 和现有 GTs 的局限性,核心是 3 种 token: 节点 token、语义 token 和全局 token. PHGT 通过结合 GCN^[47]提取局部结构特征,并通过多类型 token 的协同注意力机制,同时建模局部结构、复杂语义和长程依赖关系. HINormer 的核心是局部结构编码器:用于提取节点邻域的局部结构信息,和异构关系编码器. HINormer 采用 k-hop 上下文采样策略生成固定长度的节点序列输入,并利用改进的 GATv2 注意力机制降低计算复杂度.

1.2 异质图神经网络 (Heterogeneous Graph Neural Networks, HGNNs)

异质图神经网络专注于处理包含多类型节点和边的复杂图数据,其核心在于建模异构语义信息.现有方法可分为两类:

1) 基于元路径的方法: 如 HAN^[48]、MAGNN^[49]、Ro-He^[50]、MECCH^[51]和 HGMAE^[52],通过预定义或自动生成的元路径 (Meta-path) 聚合高阶语义信息. HAN 利用节点级和语义级双重注意力,动态融合不同元路径下的邻域特征.此类方法依赖元路径的质量,且计算复杂度随元路径数量增长而显著增加. MAGNN 通过多层级语义聚合框架实现高阶异质

关系的建模,首先利用 RNN 对预定义元路径进行编码,显式捕获路径中节点的语义依赖,其次设计 3 种注意力机制,依次在节点、元路径和语义级进行动态权重分配,从而兼顾局部结构和全局语义的协同优化. RoHe 通过引入注意力净化器(attention purifier),结合元路径转移概率先验与特征相似性,动态剪枝恶意邻居边,提升异质图神经网络在对抗攻击下的鲁棒性. MECCH 通过三阶段框架(元路径上下文构建、编码器、卷积融合)将元路径转换为无冗余子图结构,避免随机游走的高开销,实现无损节点信息聚合. HGMAE 通过元路径掩码与重建技术强化异质图语义建模:基于预定义元路径随机掩蔽部分语义边,结合语义级注意力机制动态加权不同元路径的重要性,强制模型通过剩余结构重建被掩蔽的高阶关系;同时联合元路径位置特征预测节点拓扑角色,并采用缩放余弦误差精准恢复元路径隐含的复杂依赖.

2) 基于邻域聚合的方法:如 RSHN^[53]、SimpleHGN^[54]、GNNMutation^[55] 和 PNA^[56],将类型感知嵌入到消息传递过程中.例如,SimpleHGN 通过边类型嵌入增强 GAT 的异构建模能力.尽管无需元路径,此类方法仍受限于消息传递的局部性,难以捕捉长程依赖. RSHN 通过对实体邻居进行结构建模以及构建一个以关系为节点的“关系图”进行关系建模,实现了层次化的信息聚合.相比传统的 GCN, RSHN 不仅考虑了实体的局部结构,还引入了关系间的上下文关联,能够更有效地处理多种关系下的信息传递. GNNMutation 构建包含患者与蛋白质双节点的异质图,通过蛋白质互作边(Eppi)和突变导向边(Emut)连接,并采用 BM25-tf-rf 特征加权法量化基因突变重要性. PNA 通过集成多类聚合器(均值、最大值、最小值、标准差)并设计对数度缩放器解决节点度分布偏差问题.

1.3 节点分类

节点分类是图表示学习中的基础任务之一,旨在根据节点的结构特征和属性信息预测其所属类别.在同质图中,节点分类任务通常借助 GCN、GAT 等模型完成,通过局部邻域信息聚合获得判别性表示.然而,在异质图中,节点所属的类型与其邻居类型通常不同,关系多样性使得节点表示的学习更加复杂.

为此,研究者提出了多种适用于异质图的节点分类方法.例如, RGCN 通过为不同关系类型引入独立的变换矩阵,使得模型能够区分不同边的语义;HAN 采用元路径引导的双层注意力机制,分别在节点层与语义层建模多视角信息;MAGNN 在此基础上引入基于元路径的上下文序列编码,进一步提升语义建模能力;Simple-HGN 则通过边类型嵌入强化异构图的建模能力,同时避免了手工设计元路径的需求.近年来,Transformer 结构也被引入节点分类任务中,如 HGT 和 HI-Normer,通过全局注意力机制捕捉跨类型长程依赖信息,提升节点嵌入表达能力.

2 预备知识

2.1 异质图

给定一个图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,其中 \mathcal{V} 是节点集合, \mathcal{E} 是边集合,二者通过类型映射函数区分多样性:节点类型映射函数: $\phi: \mathcal{V} \rightarrow \mathcal{A}$,其中 \mathcal{A} 为预定义的节点类型集合;边类型映射函数: $\psi: \mathcal{E} \rightarrow \mathcal{R}$,其中 \mathcal{R} 为预定义的边类型集合;满足 $|\mathcal{A}| + |\mathcal{R}| > 2$,即图中至少包含两种类型的节点和边.

$\rightarrow \mathcal{A}$,其中 \mathcal{A} 为预定义的节点类型集合;边类型映射函数: $\psi: \mathcal{E} \rightarrow \mathcal{R}$,其中 \mathcal{R} 为预定义的边类型集合;满足 $|\mathcal{A}| + |\mathcal{R}| > 2$,即图中至少包含两种类型的节点和边.

2.2 Transformer

Transformer 是一种基于自注意力机制的深度学习框架,它依赖于多头自注意力(Multi-Head Self-Attention)和前馈神经网络(Feedforward Neural Network)来处理输入数据,使其在捕捉长距离依赖关系方面具有显著优势.对于节点表示矩阵 $\mathbf{H} \in \mathbb{R}^{n \times d}$,自注意力计算可以表示为:

$$\text{Attention}(\mathbf{H}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (1)$$

其中 $\mathbf{Q} = \mathbf{H}\mathbf{W}_Q, \mathbf{K} = \mathbf{H}\mathbf{W}_K, \mathbf{V} = \mathbf{H}\mathbf{W}_V$,分别为查询、键、值矩阵, $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ 为可学习的参数.

多头注意力通过并行多个独立的注意力头,捕获不同子空间的特征,增强模型表达能力:

$$\text{MultiHead}(\mathbf{H}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}_O \quad (2)$$

前馈神经网络由两个线性层和中间的非线性激活函数构成,用于对每个位置的特征进行单独处理,可以表示如下:

$$\text{FFN}(\mathbf{H}) = \text{ReLU}(\mathbf{H}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \quad (3)$$

其中 $\mathbf{W}_1 \in \mathbb{R}^d \times d_{\text{hid}}, \mathbf{W}_2 \in \mathbb{R}^{d_{\text{hid}} \times d}$ 为可学习参数矩阵, d_{hid} 为中间隐藏层维度, ReLU 为激活函数.

2.3 问题定义

节点分类是图学习中的核心任务之一,旨在根据图结构和节点特征预测未标注节点的类别.每个节点 $v \in \mathcal{V}$ 对应一个标 $y_v \in \{1, \dots, C\}$,其中 C 为类别总数.对于异质图,由于存在多种不同类型的节点和关系,因此需要考虑不同类型节点的语义信息和关系依赖.

给定一个异质图 $G = (\mathcal{V}, \mathcal{E}, \mathcal{T}_V, \mathcal{T}_E)$,其中部分节点 $V_L \in \mathcal{V}$ 具有已知类别标签 Y_L ,目标是学习一个分类函数 $f: \mathcal{V} \rightarrow \mathcal{C}$,其中 \mathcal{C} 为类别集合,使得未标注节点 $V_U = \mathcal{V} \setminus V_L$,可以准确预测其类别标签 Y_U .

2.4 对比学习

对比学习^[57-61]是一种自监督表示学习方法,其核心思想是通过拉近语义相似样本(正样本对)的表示,同时推远语义不相似样本(负样本对)的表示,从而学习具有强判别性的特征空间.在图表示学习中,对比学习被广泛用于增强节点嵌入的质量,克服数据稀疏性,并捕捉复杂的图结构信息.

给定一个锚节点 v_i ,对比学习的目标是学习一个编码器 $f(\cdot)$,使其嵌入 $h_i = f(v_i)$ 满足:锚节点 v_i 与其正样本 v_i^+ 的嵌入 h_i 和 h_i^+ 在潜在空间中距离相近,与其负样本 $\{v_{i,j}^-\}_{j=1}^N$ 的嵌入 h_i 和 $\{h_{i,j}^-\}_{j=1}^N$ 在潜在空间距离较远.

一种广泛使用的对比损失是 InfoNCE 损失的变体.对于一个锚节点 v_i ,正样本 v_i^+ 和 $\{v_{i,j}^-\}_{j=1}^N$,损失函数定义为:

$$\mathcal{L}_{\text{ci}}^{(i)} = -\log \frac{\exp(\text{sim}(h_i, h_i^+)/\tau)}{\exp(\text{sim}(h_i, h_i^+)/\tau) + \sum_{j=1}^N \exp(\text{sim}(h_i, h_{i,j}^-)/\tau)} \quad (4)$$

其中, $\text{sim}(u, v)$ 是衡量嵌入向量相似度的函数,例如余弦相似度; $\tau > 0$ 是一个温度参数,用于调节对负样本的关注程度;分子部分最大化正样本对的相似度,分母部分最小化正样本相似度的同时,最大化锚节点与负样本的区分度.

3 方法

在本节中,本文将具体介绍 CHGormer,系统模型图如图 2 所示.首先,本文引入 token 生成策略,为节点生成属性和拓扑特征空间上的正负 token.其次,本文引入结构编码,用于对节点特征进行邻域聚合,保留图的结构特征;引入异构关系编码,用于捕获异构语义关系并增强注意力机制的类型感知能力.最后,本文引入对比学习,用于增强节点表示.

3.1 Token 生成 (Tokenization)

针对异质图的复杂语义,本文设计了双空间 token 生成

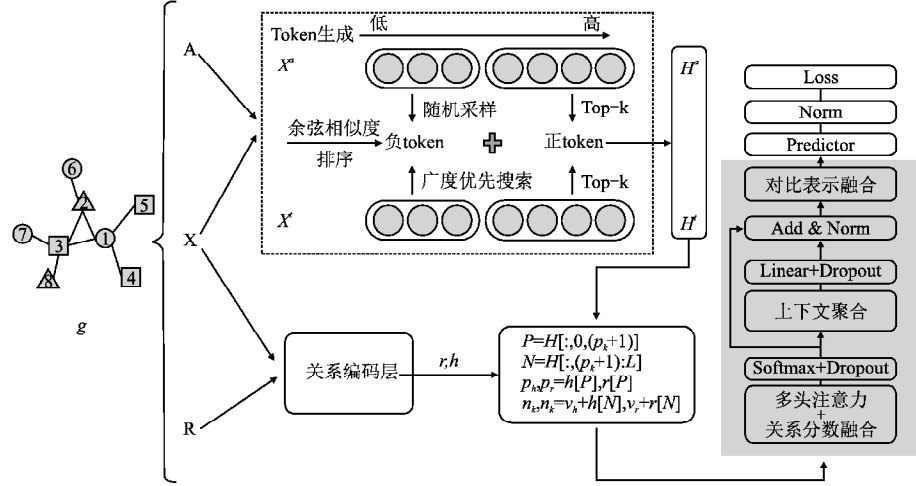


图 2 系统模型图

Fig. 2 System model

为目标节点 v_i 选择 p_k 个相似度最高的节点,构建正 token 序列 $V_i^{a,p}$.从剩余节点中随机采样 n_k 个低相似度节点,构建负 token 序列 $V_i^{a,n}$.

3.1.2 基于拓扑特征空间的 token

为了保留节点之间的复杂关系,本文基于 X^t 生成特征矩阵: $X^t = \hat{A}^k X^a$,捕捉 k -跳邻域信息.与属性特征空间类似,计算节点间余弦相似度矩阵 S^t ,用于捕获节点在结构上的相似性.

给定一对节点 (v_i, v_j) ,相似度计算如下:

$$S_{i,j}^t = \frac{X_i^t \cdot X_j^{tT}}{|X_i^t| |X_j^t|} \quad (6)$$

基于结构相似性,执行广度优先搜索为目标节点 v_i 采样 p_k 个邻近节点作为正 token 序列 $V_i^{t,p}$,在剩余节点中随机选择 n_k 个非邻近节点作为负 token 序列 $V_i^{t,n}$.

3.1.3 token 序列构建

将正负 token 序列与目标节点自身拼接,形成输入序列:

$$H^{a,i^0} = \text{Concat}(V_i^{a,p}, V_i^{a,n}) \in \mathbb{R}^{(1+p_k+n_k) \times d} \quad (7)$$

$$H^{t,i^0} = \text{Concat}(V_i^{t,p}, V_i^{t,n}) \in \mathbb{R}^{(1+p_k+n_k) \times d} \quad (8)$$

其中 p_k, n_k 为超参数.

3.2 提出的模型:CHGormer

为了捕获异构语义关系并增强注意力机制的类型感知能力,首先,本文设计了关系编码层,通过类型感知权重调整节点特征,并进行图卷积聚合,生成异构感知的关系编码.其次,本文设计了关系感知注意力层,通过多头注意力机制融合节

策略,分别基于属性特征空间和拓扑特征空间,构建正负 token 序列进行对比学习.

3.1.1 基于属性特征空间 X^a 的 token

为了建模节点本身的语义相似性,本文基于加入自环的归一化邻接矩阵 \hat{A} 传播节点属性特征矩阵 X^a ,基于 X^a ,计算节点间余弦相似度矩阵 S^a ,用于捕获节点在特征空间中的语义相似性.给定一对节点 (v_i, v_j) ,相似度计算如下:

$$S_{i,j}^a = \frac{X_i^a \cdot X_j^{aT}}{|X_i^a| |X_j^a|} \quad (5)$$

点特征和关系编码,生成类型感知的上下文表示.

3.2.1 关系编码层

在关系编码层中,使用节点类型的独热编码初始化关系编码,通过多层邻域聚合更新关系编码,引入可学习的类型感知权重调整不同的类型节点的影响:

$$r_v^t = \sum_{u \in \mathcal{N}(v)} w_{\phi(u)}^{t-1} f(r_u^{t-1}; \theta_h^t) \quad (9)$$

其中 $w_{\phi(u)}$ 是节点类型相关的权重, $f(\cdot; \theta_h^t)$ 是由参数化权重定义的可学习聚合函数.

3.2.2 关系感知注意力层

在关系感知注意力层中,本文首先计算标准注意力得分:

$$a_{i,j} = a_i^T \text{LeakyReLU}(Q_i) + a_j^T \text{LeakyReLU}(K_j) \quad (10)$$

其中 a_i, a_j 是可学习的注意力参数; Q_i, K_j 分别表示第 i 个查询和第 j 个键的特征向量.

其次,关系编码 r 被投影为查询和键,并作为偏置项调整注意力分数:

$$q_i^R = W_{Q_R} r_i, k_j^R = W_{K_R} r_j \quad (11)$$

进一步,注意力机制可以表示为:

$$\hat{\alpha}_{i,j} = \alpha_{i,j} + \beta \cdot q_i^R k_j^R \quad (12)$$

其中 β 是超参数.

3.2.3 Transformer Layer

对于属性空间和拓扑空间的 token 序列 H ,处理流程相同.首先,CHGormer 通过特征矩阵 X 和关系编码 r ,经过结构编码得到处理后的特征和关系编码:

$$H = \text{GCN}(X), r = \text{RELayers}(r) \quad (13)$$

之后将 token 分割为正负样本:

$$\mathbf{P} = \mathbf{H}[:, 0:(pk+1)], \mathbf{N} = \mathbf{H}[:, (pk+1):L] \quad (14)$$

为了方便从负 token 中提取到有用的信息,因此本文在负 token 的第 1 位添加了一个具有可学习特征的虚拟 token. 由此得到正、负 token 的特征和关系编码 $\mathbf{p}_h, \mathbf{p}_r, \mathbf{n}_h, \mathbf{n}_r$:

$$\mathbf{p}_h, \mathbf{p}_r = \mathbf{h}[\mathbf{P}], \mathbf{r}[\mathbf{P}] \quad (15)$$

$$\mathbf{n}_h, \mathbf{n}_r = \mathbf{v}_h + \mathbf{h}[\mathbf{N}], \mathbf{v}_r + \mathbf{r}[\mathbf{N}] \quad (16)$$

其中 $\mathbf{v}_h, \mathbf{v}_r$ 是虚拟 token. 之后通过多层 Transformer 层处理正负样本得到增强后的正负样本表示: $\mathbf{p}_h, \mathbf{n}_h$.

3.3 节点嵌入获取

为了解决异质图的数据稀疏性和复杂关系,提升多模态信息融合能力,本文在模型当中加入了对比学习. 根据增强后的正负样本表示,得到属性对比表示 z_a 和拓扑对比表示 z_t :

$$z_a = \mathbf{p}_h^a - \delta \cdot \mathbf{n}_h^a, z_t = \mathbf{p}_h^t - \delta \cdot \mathbf{n}_h^t \quad (17)$$

其中 δ 是超参数,控制差异强度.

根据两个模态的对比表示,得到融合后的表示:

$$\mathbf{z} = \alpha \cdot z_a + (1 - \alpha) \cdot z_t \quad (18)$$

其中 α 是超参数,控制拓扑对比表示的融合权重.

算法 1. 模型 CHGormer

输入: 节点特征: X_a, X_t , 类型嵌入: type_emb , 节点类型索引: node_type , token 序列: $\text{token_a}, \text{token_t}$, batch 索引 idx , GCN 层数 num_gnns , GTs 层数 num_layers .

输出: 最终得分和 token 表示.

1. 定义 $\text{inner}(\text{tag} \in \{ "a", "t" \})$:
2. $\mathbf{h} = X_{\text{tag}}; \mathbf{r} = \text{type_emb}[\text{node_type}]$
3. $\text{token} = \text{token_tag}[\text{idx}]; \mathbf{v}_h, \mathbf{v}_r = \text{v_tag}[\text{idx}]$
4. **for** l in $l, \text{num_gnns}$: $\#h, r$ 经过 GCN 和 REConv 层
 $\mathbf{h} = \text{GCN}^l(\mathbf{h}), \mathbf{r} = \text{RE}^l(\mathbf{r})$
5. $\text{p_token} = \text{token}[:, :l + p_k], \text{p_token} = \text{token}[:, l + p_k:]$ #token 切分
6. $\mathbf{p}_h = \mathbf{h}[\text{p_token}], \mathbf{p}_r = \mathbf{r}[\text{p_token}]$ #正负样本
7. $\mathbf{n}_h = \text{concat}(\mathbf{v}_h, \mathbf{h}[\text{n_token}]), \mathbf{n}_r = \text{concat}(\mathbf{v}_r, \mathbf{r}[\text{n_token}])$
8. **for** l in $l, \text{num_layers}$: $\#\text{Transformer}$
 $\mathbf{p}_h = \text{AGT}^l(\mathbf{p}_h, \mathbf{p}_r), \mathbf{n}_h = \text{AGT}^l(\mathbf{n}_h, \mathbf{n}_r)$
9. **return** $z_{\text{tag}} = \mathbf{p}_h[:, :l], \mathbf{p}_h[:, l:] - \delta \times \mathbf{n}_h[:, :l], \mathbf{p}_h[:, l:]$
10. $z_a, z_t = \text{inner}("a"), \text{inner}("t")$
11. $\mathbf{z} = \alpha \times z_a + (1 - \alpha) \times z_t$
12. $\text{output} = \text{predictor}(\mathbf{z})$
13. 若 $\text{norm} = \text{True}$, 则执行归一化
14. 返回 output 以及所有 token 表示: $\mathbf{p}_a, \mathbf{n}_a, \mathbf{p}_t, \mathbf{n}_t$

3.4 训练目标

在 CHGormer 中,为了充分利用负 token 增强节点表示,本文引入对比损失:

$$\mathcal{L}_{cl} = -\log \frac{\exp(\mathbf{P}_0^{a,i} \cdot \hat{\mathbf{P}}^{t,iT} / \tau)}{\sum_{j=1}^{n_k} \exp(\mathbf{P}_0^{a,i} \cdot \mathbf{N}_j^{a,iT} / \tau)} - \log \frac{\exp(\mathbf{P}_0^{t,i} \cdot \hat{\mathbf{P}}^{a,iT} / \tau)}{\sum_{j=1}^{n_k} \exp(\mathbf{P}_0^{t,i} \cdot \mathbf{N}_j^{t,iT} / \tau)} \quad (19)$$

其中 $\mathbf{P}_0^{a,i}, \mathbf{N}_0^{a,i}$ 表示从正、负 token 序列得到的信息; $\hat{\mathbf{P}}^{a,i} = \frac{1}{p_k}$

$\sum_{j=1}^{p_k} \mathbf{P}_j^{a,i}, \hat{\mathbf{P}}^{t,i} = \frac{1}{p_k} \sum_{j=1}^{p_k} \mathbf{P}_j^{t,i}, \tau$ 是超参数.

此外,本文引入分类损失:标准交叉熵损失监督节点分类:

$$\mathcal{L}_{ce} = -\sum_{i \in \mathcal{V}_i} \mathbf{Y}_i \ln \hat{\mathbf{Y}}_i, \hat{\mathbf{Y}}_i \quad (20)$$

结合对比损失和分类损失:

$$\mathcal{L} = \mathcal{L}_{cl} + \gamma \cdot \mathcal{L}_{ce} \quad (21)$$

其中 γ 是超参数.

3.5 时间复杂度分析

在这一节中,本文总结了 CHGormer 各个部分的计算开销,Token 生成部分的时间复杂度是 $O(N)$,关系编码层的时间复杂度是 $O(|E|)$,Transformer 层的时间复杂度是 $O(|E| + N)$,因此总体的时间复杂度是 $O(|E| + N)$.

4 实验

4.1 数据集介绍

DBLP 是计算机科学领域的书目数据集,涵盖 1994 至 2014 年间 4 个研究领域(数据库、数据挖掘、机器学习和信息检索)在 20 个顶级会议上发表的论文. 包含作者(A)、论文(P)、研究术语(T)和会议地点(V)4 类节点,构成学术关系网络.

AMiner 是由清华大学构建的学术信息平台,本文采用其包含 4 类论文(数据库、数据挖掘、机器学习和信息检索)的子图数据. 节点类型包括论文(P)、作者(A)和参考文献(R),体现学术成果的关联结构.

Freebase 知识图谱的子集聚焦影视领域,涵盖电影(M)、演员(A)、导演(D)和编剧(W)4 类实体节点,通过语义关系描述影视产业的创作关系网络.

数据集的详情如表 1 所示.

表 1 数据集参数

Table 1 Statistics of datasets

数据集	节点	边	目标
DBLP	26182	239566	Author
Freebase	43854	151034	Movie
AMiner	55783	153676	Paper

4.2 比较基线

为了全面评估本文提出的 CHGormer 与先进的方法,本文从以下两个主要类别中考虑了一系列比较基线:异质图神经网络模型(Heterogeneous GNNs, 包括 RGCN, HAN, HetGNN, MAGNN 和 SimpleHGN),基于异质图 Transformer(包括 HGT, GTN, PHGT 和 HINormer).

- RGCN: RGCN 扩展了传统 GCN 以处理多关系图数据,通过为每种边类型分配独立权重矩阵,在消息传递中区分不同关系. 其核心创新是引入高斯分布表示节点特征,并基于方差设计注意力权重,以抑制噪声传播. 训练时添加正则化损失,强制首层输出服从高斯分布并约束参数,显著提升对抗环境下的鲁棒性,

- HAN: HAN 通过在节点类型和元路径级别引入双层注意力机制,自适应地聚合不同类型节点及关系的信息,以学习异质图中节点的高质量表示.

- HetGNN: HetGNN 通过 3 阶段框架解决异构图节点表示问题: 1) 异构邻居采样: 基于重启随机游走按节点类型分层采样,确保覆盖冷启动节点; 2) 多模态特征融合: 使用双向

LSTM 聚合文本、图像等异构属性;3) 类型感知聚合:注意力机制为不同邻居类型分配权重。

- **MAGNN**:MAGNN 针对元路径的中间节点信息丢失问题,提出 3 组件架构:1) 节点特征对齐:线性变换将不同类型节点特征映射到统一空间;2) 元路径内部编码:使用 RNN 或关系旋转编码捕获路径中间节点序列;3) 元路径间聚合:语义注意力加权融合多路径信息。

- **SimpleHGN**:SimpleHGN 证明轻量化设计可高效处理异构图,其核心是扩展 GAT 的 3 要素:1) 可学习边类型嵌入:将边类型信息融入注意力分数计算;2) 预激活残差连接:跨层传递节点表示以缓解过平滑;3) 输出嵌入 L2 归一化:提升相似度计算稳定性。

- **HGT**:HGT 将 Transformer 适配到动态异构图,提出三元组注意力机制:依据源节点类型、边类型、目标节点类型动态生成 Query/Key/Value 权重。引入相对时间编码 RTE 处理任意时间跨度的边,并设计异构子图采样算法 HGSampling 平衡节点类型分布以支持超大规模图训练。

- **GTN**:GTN 的核心是自动生成元路径通道:通过堆叠可学习矩阵的“图变换层”,软选择边类型组合成新邻接矩阵

(如融合“用户-点击”与“商品-类别”生成高阶路径“用户-兴趣类别”)。每条生成路径分配通道注意力权重,替代人工设计元路径。

- **PHGT**:PHGT 通过引入 3 类 token(节点 token、语义 token 和全局 token)在 Transformer 框架下同时建模异质图中的局部结构,高阶语义关系与全局长程依赖从而实现高效且语义感知的异质图表示学习。

- **HINormer**:HINormer 用全局注意力替代消息传递,解决过平滑与长程依赖问题:1) 局部结构编码器:GCN 聚合邻居特征捕获局部拓扑;2) 异构关系编码器:学习类型感知位置编码调整注意力;3) D-hop 上下文采样:为每个节点生成定制化邻居序列输入 Transformer。采用轻量 GATv2 注意力减少参数,实现全局交互与结构感知的统一。

4.3 参数设置

在实验中,本文使用 Micro-F1 和 Macro-F1 作为评价指标。学习率设为 0.0001,结构编码器和异质编码器的层数设为 4,隐藏层维度设为 256,Transformer 层数设为 2,节点序列的长度设为 200,注意力分数的温度设为 1.0,对比损失中的权重 τ 为 35,控制拓扑对比表示的融合权重 α 设为 0.5,控制分类损

表 2 节点分类表现

Table 2 Node classification performance

Methods	DBLP		Freebase		AMiner	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
RGCN	92.07 ± 0.50	91.52 ± 0.50	60.82 ± 1.23	59.08 ± 1.44	81.58 ± 1.44	62.53 ± 2.31
HAN	92.05 ± 0.62	91.67 ± 0.49	61.42 ± 3.56	57.05 ± 2.06	81.90 ± 1.51	64.67 ± 2.21
HetGNN	92.33 ± 0.41	91.76 ± 0.43	62.99 ± 2.31	58.44 ± 1.99	72.34 ± 1.42	55.42 ± 1.45
MAGNN	93.76 ± 0.45	93.28 ± 0.51	64.43 ± 0.73	58.18 ± 3.87	82.64 ± 1.59	68.60 ± 2.04
Simple-HGN	94.46 ± 0.22	94.01 ± 0.24	67.49 ± 0.97	62.49 ± 1.69	86.44 ± 0.48	75.73 ± 0.97
HGT	93.49 ± 0.25	93.01 ± 0.23	66.43 ± 1.88	60.03 ± 2.21	85.74 ± 1.24	74.98 ± 1.61
GTN	93.97 ± 0.54	93.52 ± 0.55	OOM	OOM	OOM	OOM
PHGT	95.33 ± 0.18	94.96 ± 0.17	<u>68.74 ± 1.42</u>	61.73 ± 1.86	OOM	OOM
HINormer	<u>94.94 ± 0.21</u>	94.57 ± 0.23	67.78 ± 0.39	62.76 ± 1.10	<u>88.04 ± 0.12</u>	79.88 ± 0.24
CHGormer	94.82 ± 0.31	<u>94.65 ± 0.32</u>	68.79 ± 0.23	63.91 ± 0.31	88.53 ± 0.18	<u>78.81 ± 0.37</u>

其中,粗体代表最高的得分,下划线代表第二高的得分,“OOM”代表显卡显存不足。

失影响的权重 γ 设为 0.8,控制关系分数融合的权重 β 设为 1.0,控制正负样本差异的权重 δ 设为 1,正、负 token 的长度 p_k, n_k 均为 25。

4.4 性能比较

本文在 3 个数据集上进行了节点分类实验,并且本文的 CHGormer 取得了有竞争力的表现。CHGormer 在 Freebase 数据集以及 AMiner 的 Micro-F1 得分上取得了第一,在 AMiner 的 Macro-F1 上取得了第二,在 DBLP 数据集的 Macro-F1 上也取得了第二。这得益于模型对异构关系的动态建模能力,本文引入 token 生成策略,为节点生成属性和拓扑特征空间上的正负 token。通过类型敏感的注意力权重强化全局交互中的关键连接,结合结构编码与 Transformer 的多层级特征融合策略,并引入对比学习,有效平衡了复杂异构数据的特性。

此外,本文发现基于 Transformer 的模型在节点分类任务中显著优于传统非 Transformer 方法。例如,PHGT 在 DBLP 上远超 RGCN,而 HINormer 在 AMiner 中大幅领先 MAGNN。这一优势源于 Transformer 的全局自注意力机制,能够直接建模全图节点的长距离依赖,而传统 GNN 受限于局部消息传

递的瓶颈。同时,Transformer 的多头注意力动态分配异构关系权重,例如 PHGT 对边类型的自适应增强,并利用并行计算提升大规模图训练效率,从而在复杂异构场景中避免性能下降或内存溢出风险。

在 3 个异质图数据集的实验中,本文观察到 CHGormer 在 Freebase 数据集上取得了最优表现,这主要归因于 Freebase 异质结构更强,节点和边类型丰富,关系语义明确,有利于模型中基于属性与拓扑双空间的 token 生成策略和关系感知注意力机制的发挥。而在 AMiner 和 DBLP 数据集上,尽管 CHGormer 仍具有竞争力,但相较于 Freebase 优势相对减弱,原因在于这些学术网络结构更为稠密、异质性较低,token 所构建的对比信息量相对有限。总体来看,CHGormer 在节点类型丰富、关系异构性显著的图结构中具有显著优势,而在结构较集中或关系单一的图中,其表现相对与其他模型接近。

4.5 消融实验

为了验证模型中不同组件对性能的影响,本文进行了消融实验。实验中,本文移除了模型的 3 个关键组件,并观察了模型性能的变化。这 3 个组件分别是:1) 负样本机制;2) 异

质类型采样; 3) 图卷积网络(GCN)结构函数. 实验结果如图3所示. 本文在3个不同的数据集上进行了消融实验, 对于每个数据集, 本文计算了模型在移除特定组件后的 Micro-F1 和 Macro-F1 分数. 在图中分别用 w/o negative; w/o hete; w/o GCN 表示.

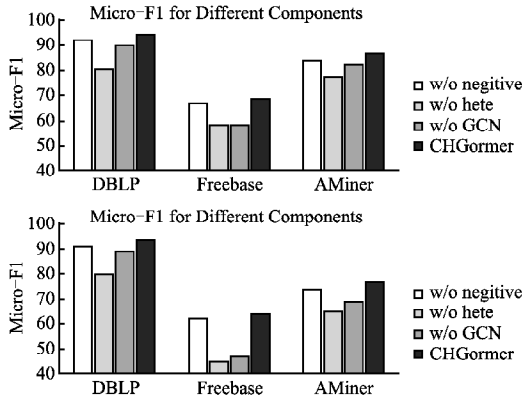


图3 消融实验
Fig. 3 Ablation study

1) 从图3中可以看出, 移除负样本机制后, 模型在所有数据集上的 Micro-F1 和 Macro-F1 分数均显著下降. 这表明负样本机制对于提高模型的区分能力和泛化能力至关重要, 它有助于模型学习区分正负样本的特征表示.

2) 异质类型采样是处理异构图数据的关键技术之一, 它允许模型捕捉不同类型节点之间的关系和交互. 移除异质类型采样后, 模型在 DBLP 和 AMiner 数据集上的性能有所下降, 而在 Freebase 数据集上的性能下降更为显著. 这表明异质类型采样对于捕捉异构图中的复杂关系和多样性信息非常重要.

3) 图卷积网络(GCN)结构函数在模型中用于捕捉图结构信息, 这对于理解节点之间的连接模式和传播信息至关重要. 移除 GCN 结构函数后, 模型在 AMiner 数据集上的性能下降最为明显, 而在 DBLP 和 Freebase 数据集上的性能也有所下降. 这表明 GCN 结构函数在捕捉图结构信息方面发挥了重要作用.

4.6 参数敏感性分析

在这一部分中, 本文进行了以下4个参数的敏感性分析: GCN 和关系编码层的层数 num_gnns, 关系感知注意力层的层数 num_layers, 对比损失中的权重, 以及控制拓扑对比表示的融合权重. 实验结果如图4所示.

对于 num_gnns, 一个适中的值, 例如3能使模型拥有更好的性能, 因为过大的层数会引发过平滑和过挤压问题, 导致远距离节点信息在多层传播中被压缩丢失.

对于 num_layers, 一个较小的值, 例如2能够达到一个更好的分数, 因为过高的 Transformer 层数会导致过拟合, 较小的值可以减少噪声累积.

对于 τ , 当 τ 增大时, 分子和分母中的指数项会变得更加接近, 从而使得损失函数更加平滑. 这种平滑的损失函数更容易优化, 并且能够更好地反映样本之间的全局相似性. 由于它因此较大的值更加适合模型.

对于 α , 由于在异构图中, 节点的属性信息和拓扑信息都

是重要的特征, 因此一个适中的值, 例如0.5, 能够平衡这两

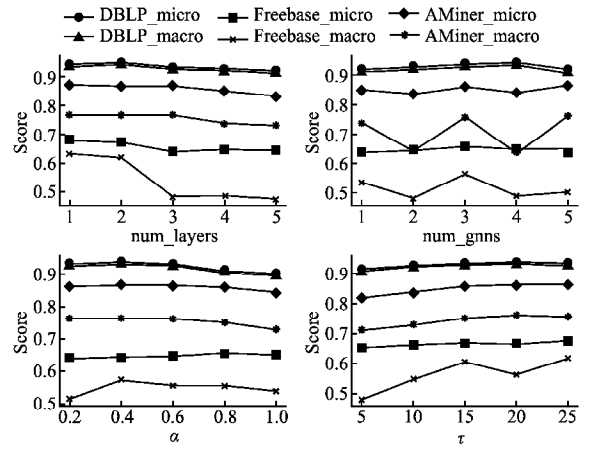


图4 参数敏感性
Fig. 4 Parameter sensitivity

种信息, 使得模型能够同时利用节点本身的特征和节点之间的关系来做出更准确的预测.

5 结论

本文针对异质图表示学习中局部异构关系与全局语义依赖的整合挑战, 提出了 CHGormer 模型. 通过异构关系感知的对比学习框架和双空间动态令牌生成策略, CHGormer 实现了高效的异质语义建模. 实验结果表明, CHGormer 在所有数据集上取得了有竞争力的性能, 验证了其设计的有效性. 进一步分析揭示了 Transformer 架构在异质图任务中的显著优势——全局自注意力机制能够直接建模长距离依赖, 而类型感知的权重分配则增强了异构关系的动态捕获能力. 未来工作将探索更轻量化的稀疏注意力设计, 以进一步提升模型在大规模异质图上的可扩展性. 本研究为复杂图数据的表示学习提供了新的技术路径, 具有重要的理论价值和应用前景.

References:

[1] Sun Q, Peng H, Li J, et al. Pairwise learning for name disambiguation in large-scale heterogeneous academic networks[C]//International Conference on Data Mining, 2020; 511-520.
 [2] Kong X, Shi Y, Yu S, et al. Academic social networks: modeling, analysis, mining and applications[J]. Journal of Network and Computer Applications, 2019, 132: 86-103, doi: 10. 1016/j. jnca. 2019. 01. 029.
 [3] Pan L, Dai X, Huang S, et al. Academic paper recommendation based on heterogeneous graph[C]//China National Conference on Chinese Computational Linguistics, 2025; 381-392.
 [4] Tang J, Zhang J, Yao L, et al. Arnetminer: extraction and mining of academic social networks[C]//14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008: 990-998.
 [5] Van den Brink M, Benschop Y. Gender in academic networking: the role of gatekeepers in professorial recruitment[J]. Journal of Management Studies, 2014, 51(3): 460-492.
 [6] Salamat A, Luo X, Jafari A. HeteroGraphRec: a heterogeneous graph-based neural networks for social recommendations [J].

- Knowledge-Based Systems, 2021, 217:106817, doi:10.1016/j.knosys.2021.106817.
- [7] Wang Y, Sun H, Zhao Y, et al. A heterogeneous graph embedding framework for location-based social network analysis in smart cities [J]. IEEE Transactions on Industrial Informatics, 2019, 16(4): 2747-2755.
- [8] Cai D, Shao Z, He X, et al. Mining hidden community in heterogeneous social networks [C]//3rd International Workshop on Link Discovery, 2005:58-65.
- [9] Milroy L, Llamas C. Social networks [J]. The Handbook of Language Variation and Change, 2013: 407-427, doi: 10.1002/9780470756591.
- [10] Li Z, Liu H, Zhang Z, et al. Learning knowledge graph embedding with heterogeneous relation attention networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(8): 3961-3973.
- [11] Asprino L, Daga E, Gangemi A, et al. Knowledge graph construction with a façade: a unified method to access heterogeneous data sources on the web [J]. ACM Transactions on Internet Technology, 2023, 23(1):1-31.
- [12] Chen X, Jia S, Xiang Y. A review: knowledge reasoning over knowledge graph [J]. Expert Systems with Applications, 2020, 141:112948, doi:10.1016/j.eswa.2019.112948.
- [13] Hogan A, Blomqvist E, Cochez M, et al. Knowledge graphs [J]. ACM Computing Surveys, 2021, 54(4):1-37.
- [14] Fensel D, Şimşiek U, Angele K, et al. Introduction: what is a knowledge graph? [J]. Knowledge Graphs: Methodology, Tools and Selected Use Cases, 2020:1-10, doi:10.1007/978-3-030-37439-6_1.
- [15] Chen Z, Wang Y, Zhao B, et al. Knowledge graph completion: a review [J]. IEEE Access, 2020, 8: 192435-192456, doi: 10.1109/ACCESS.2020.3030076.
- [16] Yu L, Sun L, Du B, et al. Heterogeneous graph representation learning with relation awareness [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(6):5935-5947.
- [17] Chan T H, Cendra F J, Ma L, et al. Histopathology whole slide image analysis with heterogeneous graph representation learning [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023:15661-15670.
- [18] Shao Z, Xu Y, Wei W, et al. Heterogeneous graph neural network with multi-view representation learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(11):11476-11488.
- [19] Yang X, Yan M, Pan S, et al. Simple and efficient heterogeneous graph neural network [C]//AAAI Conference on Artificial Intelligence, 2023:10816-10824.
- [20] Zhao J, Wang X, Shi C, et al. Heterogeneous graph structure learning for graph neural networks [C]//AAAI Conference on Artificial Intelligence, 2021:4697-4705.
- [21] Chen M, Huang C, Xia L, et al. Heterogeneous graph contrastive learning for recommendation [C]//16th ACM International Conference on Web Search and Data Mining, 2023:544-552.
- [22] Wang X, Bo D, 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.
- [23] Yang C, Xiao Y, Zhang Y, et al. Heterogeneous network representation learning: a unified framework with survey and benchmark [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(10):4854-4873.
- [24] Zhang K, Wang W, Zhang H, et al. Learning to represent programs with heterogeneous graphs [C]//30th IEEE/ACM International Conference on Program Comprehension, 2022:378-389.
- [25] Dong Y, Hu Z, Wang K, et al. Heterogeneous network representation learning [C]//29th International Joint Conference on Artificial Intelligence, 2020:4861-4867.
- [26] Chang Y, Chen C, Hu W, et al. Megnn: meta-path extracted graph neural network for heterogeneous graph representation learning [J]. Knowledge-Based Systems, 2022, 235:107611, doi:10.1016/j.knosys.2021.107611.
- [27] Zhang L, Guo J, Bai Q, et al. Dynamic heterogeneous graph representation learning with neighborhood type modeling [J]. Neurocomputing, 2023, 533:46-60, doi:10.1016/j.neucom.2023.02.060.
- [28] Zhao B W, Hu L, You Z H, et al. HINGRL: predicting drug-disease associations with graph representation learning on heterogeneous information networks [J]. Briefings in Bioinformatics, 2022, 23(1):bbab515, doi:10.1093/bib/bbab515.
- [29] Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network [C]//25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019:793-803.
- [30] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks [C]//The Semantic Web: 15th International Conference ESWC, 2018:593-607.
- [31] Liu Y, Zheng Y, Zhang D, et al. Beyond smoothing: unsupervised graph representation learning with edge heterophily discriminating [C]//AAAI Conference on Artificial Intelligence, 2023:4516-4524.
- [32] Keriven N. Not too little, not too much: a theoretical analysis of graph(over) smoothing [J]. Advances in Neural Information Processing Systems, 2022, 35: 2268-2281, doi: 10.5555/3600270.3600435.
- [33] Lu Z, Fang Y, Yang C, et al. Heterogeneous graph transformer with poly-tokenization [C]//33rd International Joint Conferences on Artificial Intelligence, 2024:2234-2242.
- [34] Mao Q, Liu Z, Liu C, et al. Hinormer: representation learning on heterogeneous information networks with graph transformer [C]//ACM Web Conference, 2023:599-610.
- [35] Tang J, Yang Y, Wei W, et al. Higtpt: Heterogeneous graph language model [C]//30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024:2842-2853.
- [36] Chen J, Liu H, Hopcroft J, et al. Leveraging contrastive learning for enhanced node representations in tokenized graph transformers [J]. Advances in Neural Information Processing Systems, 2024, 37: 85824-85845, doi:10.52202/079017-2725.
- [37] Zhang Y, Gao S, Pei J, et al. Improving social network embedding via new second-order continuous graph neural networks [C]//28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022:2515-2523.
- [38] Kumar S, Mallik A, Khetarpal A, et al. Influence maximization in social networks using graph embedding and graph neural network [J]. Information Sciences, 2022, 607:1617-1636, doi:10.1016/j.ins.2022.06.075.
- [39] Jain L, Katarya R, Sachdeva S. Opinion leaders for information diffusion using graph neural network in online social networks [J].

- ACM Transactions on the Web, 2023, 17(2):1-37.
- [40] Yang L, Wang S, Tao Y, et al. Dgrec: graph neural network for recommendation with diversified embedding generation [C]//16th ACM International Conference on Web Search and Data Mining, 2023:661-669.
- [41] Wu Z, Zhou J, Zhang J, et al. A deep prediction framework for multi-source information via heterogeneous GNN[C]//30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024:3460-3471.
- [42] Lu Y, Gao M, Liu H, et al. Neighborhood overlap-aware heterogeneous hypergraph neural network for link prediction[J]. Pattern Recognition, 2023, 144:109818, doi:10.1016/j.patcog.2023.109818.
- [43] Rampúšek L, Galkin M, Dwivedi V P, et al. Recipe for a general, powerful, scalable graph transformer[C]//36th International Conference on Neural Information Processing Systems, 2022:14501-14515.
- [44] Yun S, Jeong M, Kim R, et al. Graph transformer networks[C]//33rd International Conference on Neural Information Processing Systems, 2019:11983-11993.
- [45] Ying C, Cai T, Luo S, et al. Do transformers really perform badly for graph representation? [J]. Advances in Neural Information Processing Systems, 2021, 34: 28877-28888, doi: 10.5555/3540261.354273.
- [46] Hu Z, Dong Y, Wang K, et al. Heterogeneous graph transformer [C]//Proceedings of the Web Conference, 2020:2704-2710.
- [47] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C]//International Conference on Learning Representations, 2017:1-14, doi:10.48550/arXiv.1609.02907.
- [48] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network [C]//World Wide Web Conference, 2019:2022-2032.
- [49] Fu X, Zhang J, Meng Z, et al. Magnn: metapath aggregated graph neural network for heterogeneous graph embedding[C]//Proceedings of the Web Conference, 2020:2331-2341.
- [50] Zhang M, Wang X, Zhu M, et al. Robust heterogeneous graph neural networks against adversarial attacks[C]//AAAI Conference on Artificial Intelligence, 2022:4363-4370.
- [51] Fu X, King I. MECCH: metapath context convolution-based heterogeneous graph neural networks[J]. Neural Networks, 2024, 170: 266-275, doi:10.1016/j.neunet.2023.11.030.
- [52] Tian Y, Dong K, Zhang C, et al. Heterogeneous graph masked autoencoders [C]//AAAI Conference on Artificial Intelligence, 2023:9997-10005.
- [53] Zhu S, Zhou C, Pan S, et al. Relation structure-aware heterogeneous graph neural network[C]//International Conference on Data Mining, 2019:1534-1539.
- [54] Lv Q, Ding M, Liu Q, et al. Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks [C]//27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021:1150-1160.
- [55] Özcan Şimşek N Ö, Özgür A, Gürgen F. GNNMutation: a heterogeneous graph-based framework for cancer detection[J]. BMC Bioinformatics, 2025, 26(1):1-17.
- [56] Corso G, Cavalleri L, Beaini D, et al. Principal neighbourhood aggregation for graph nets[J]. Advances in Neural Information Processing Systems, 2020, 33:13260-13271, doi:10.5555/3495724.3496836.
- [57] Jiao P, Yu K, Bao Q, et al. Graph contrastive learning with node-level accurate difference[J]. Fundamental Research, 2025, 5(2): 818-829.
- [58] Guo X, Jiao P, Shi D, et al. Learning node representations via sketching the generative process with events benefits link prediction on heterogeneous networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(5):3961-3974.
- [59] Wang X, Liu N, Han H, et al. Self-supervised heterogeneous graph neural network with co-contrastive learning [C]//27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021:1726-1736.
- [60] Lin T, Zhou C, Li Q, et al. Multi-level disentangled contrastive learning on heterogeneous graphs[C]//16th International Conference on Machine Learning and Computing, 2024:628-634.
- [61] Gui H, Liu J, Tao F, et al. Embedding learning with events in heterogeneous information networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(11):2428-2441.