

# 一种属性偏移感知的图不变表示学习方法

曹付元<sup>1,2)</sup> 赵亮<sup>1)</sup> 陈庆强<sup>1)</sup>

<sup>1)</sup>(山西大学计算机与信息技术学院 太原 030006)

<sup>2)</sup>(计算智能与中文信息处理教育部重点实验室(山西大学) 太原 030006)

**摘要** 图表示学习通过将图数据嵌入至低维向量空间,为图分类任务生成关键表征,并在生物分子性质预测与社区分类等应用中发挥重要作用。图神经网络通过消息传递机制提取图结构的全局表示,已成为图表示学习的核心技术。近年来,为提升图神经网络在分布外场景下的泛化能力,图不变表示学习受到广泛关注。然而,现有方法大多侧重于挖掘结构不变子图,忽略了图中节点属性的分布偏移,导致模型易受节点伪相关特征干扰,削弱模型的表征能力。针对上述问题,本文提出一种属性偏移感知的图不变表示学习方法,通过自适应特征加权与Beta混合增强的联合优化方式,提升模型在分布外场景下的泛化性能。具体而言,首先使用不变子图提取器,以提取结构不变表示;随后,通过自适应特征维度加权机制对节点属性进行动态权重分配,从而弱化属性偏移的影响;最后,基于Beta分布的Mixup对加权表示进行线性插值,生成跨分布的虚拟混合样本,并在原始与混合表示上分别施加一致性和正交约束损失,进一步强化模型对稳定特征的聚焦能力。在合成数据和真实数据集上的实验表明,所提方法在SPMOTIF数据集上相较于ERM基线准确率平均提升了19.6%,并在五个真实任务中全面优于所有对比方法,进一步验证了其有效性。

**关键词** 图表示学习;图神经网络;分布外泛化;节点属性偏移;特征加权;混合数据增强  
中图分类号 TP181 DOI号 10.11897/SP.J.1016.2026.01342

## An Attribute-Shift-Aware Graph Invariant Representation Learning Method

CAO Fu-Yuan<sup>1,2)</sup> ZHAO Liang<sup>1)</sup> CHEN Qing-Qiang<sup>1)</sup>

<sup>1)</sup>(School of Computer and Information Technology, Shanxi University, Taiyuan 030006)

<sup>2)</sup>(Key Laboratory of Computational Intelligence and Chinese Information Processing(Shanxi University),  
Ministry of Education, Taiyuan 030006)

**Abstract** Graph representation learning plays a crucial role in graph classification tasks, as it enables the embedding of graph data into low-dimensional vector spaces, thereby generating meaningful representations that are applied in various domains such as biomolecular property prediction, social network analysis, and community detection. Among the many approaches, Graph Neural Networks (GNNs) have emerged as a core technology in graph representation learning, owing to their ability to extract global representations of graph structures through effective message-passing mechanisms. Despite their success, one of the key challenges that remain is enhancing the generalization ability of GNNs in out-of-distribution (OOD) scenarios. To address this challenge, the concept of graph-invariant representation learning has gained increasing attention in recent years. However, most existing methods primarily focus on

收稿日期:2025-07-19;在线发布日期:2026-01-30。本课题得到国家自然科学基金联合基金项目(U24A20323)、国家自然科学基金面上项目(62376145)、山西省科技创新人才团队专项(202204051002016)、太行山西省实验室技术攻关专项资助项目(THYF-JSZX-24010700)、太原市“双百攻关行动”关键核心技术攻关“揭榜挂帅”项目(2024TYJB0127)、山西省基础研究计划项目(202503021212095)资助。曹付元(通信作者),教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为数据挖掘、机器学习。E-mail: cfy@sxu.edu.cn。赵亮,硕士研究生,中国计算机学会(CCF)学生会员,主要研究领域为图机器学习、分布外泛化。陈庆强,博士,讲师,中国计算机学会(CCF)会员,主要研究领域为数据挖掘、机器学习。

extracting structural invariant subgraphs and overlook the issue of distribution shifts in node attributes, which are often prevalent in real-world data. This oversight leads to the interference of spurious correlations in node features, weakening the representational power of the learned model and making it vulnerable to distribution shifts in test data. In response to these challenges, we introduce a novel attribute-shift-aware graph-invariant representation learning method. Our proposed approach aims to improve the generalization performance of GNNs in OOD environments by addressing both structural and attribute distribution shifts. Our method utilizes a joint optimization framework that integrates adaptive feature weighting and Beta-mixed augmentation techniques. Specifically, our method begins by applying an invariant subgraph extractor, which is designed to capture structural invariant representations of the graph. This allows our model to preserve essential structural information, regardless of shifts in the distribution of node attributes. Following this, an adaptive feature weighting mechanism is employed to dynamically assign weights to node attributes based on their relevance, thus reducing the influence of attribute shifts and minimizing the risk of overfitting to spurious correlations. Moreover, our method employs Beta-distribution-based Mixup, a data augmentation technique, to generate virtual mixed samples by performing linear interpolation on the weighted representations. This process creates samples that lie in-between different distributions, which helps our model learn more robust and generalized representations that can better handle OOD scenarios. To further refine the learning process and ensure that our model focuses on stable and invariant features, consistency and orthogonal loss constraints are applied on both the original and mixed representations. This encourages our model to prioritize stable patterns while avoiding overfitting to noise or spurious correlations in the data. The proposed method is evaluated using both synthetic and real-world datasets to assess its effectiveness. In experiments conducted on the SPMOTIF dataset, our method achieves an average accuracy improvement of 19.6% over the ERM baseline. Furthermore, it demonstrates superior performance in five real-world tasks, outperforming all comparison methods in terms of generalization and robustness to attribute shifts. These results highlight the significant advantages of incorporating attribute-shift-aware learning into graph-invariant representation methods. Our findings strongly underscore the importance of addressing both structural and attribute distribution shifts in order to achieve robust, reliable, and generalizable graph representations. Ultimately, our proposed approach offers a promising and effective solution for significantly improving the performance of GNNs in real-world applications, where the presence of attribute shifts is common and challenging.

**Keywords** graph representation learning; graph neural networks; out-of-distribution generalization; attribute shift; feature reweighting; mixup data augmentation

## 1 引言

图表示学习作为图机器学习领域的基础,通过将复杂图结构数据嵌入低维向量空间,为图分类任务生成判别性图表征,已在社交网络的社区分类<sup>[1]</sup>、生物化学的分子属性预测<sup>[2]</sup>等应用中发挥关键作用。近年来,图神经网络(Graph Neural Networks, GNNs)凭借层次化消息传递机制<sup>[3-5]</sup>,通过多轮邻域特征聚合,

有效捕捉图的局部与全局拓扑信息并生成高质量的图级表示,已成为端到端图表示学习的核心技术。

尽管如此,现有大多数GNN方法均假设训练与测试数据满足独立同分布条件<sup>[6]</sup>,在真实场景中,这一假设往往难以成立,图数据常因生成机制、环境因素或采样策略的差异,在图规模<sup>[7-8]</sup>、节点度分布<sup>[9]</sup>、子结构模式<sup>[10]</sup>(如分子支架)及节点属性分布<sup>[11]</sup>等方面产生偏移,导致模型易于捕捉到数据中的伪相关特征,进而在分布外(Out-Of-Distribution,

OOD<sup>[12-14]</sup>测试上出现性能骤降。以引文网络中的论文主题分类为例,模型训练阶段以引文结构(如引用关系)和文本内容作为输入,基于历史标注数据学习分类模型。由于早期高被引论文多出自知名机构且行文规范,模型易将作者机构(结构属性)与文本风格(节点属性)作为判别依据。测试阶段若结构或

属性分布发生偏移,如新兴机构论文增多、语言风格口语化增强,伪相关性难以泛化,导致分类性能下降。如图1所示,在SPMOTIF合成数据集上的实证研究进一步验证了这一现象:当前主流GNN方法在分布外图分类中的准确率显著下降,尤其在属性分布偏移下,模型泛化能力受到严重影响。

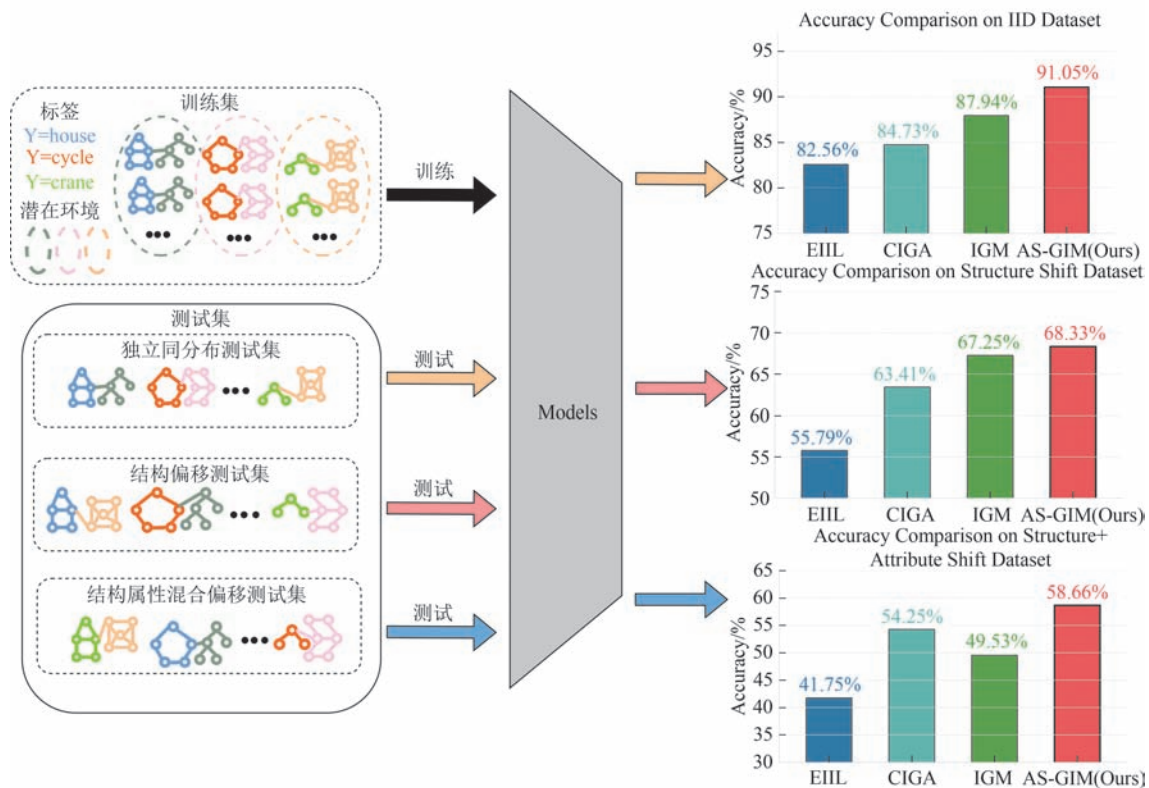


图1 分布转移对模型性能的影响

为应对上述挑战,已有研究尝试从不同角度提升图神经网络在OOD场景下的泛化能力。一类方法主要基于对抗性训练和域自适应技术<sup>[15-17]</sup>,例如域对抗训练(Domain-Adversarial Training)以及多任务学习方法,旨在通过缩小不同域之间的特征分布差异来增强泛化能力。然而,这类方法通常依赖大量的环境标签,而在图数据中获取此类标签往往较为困难,限制了其实际应用。

与此不同,另一类方法侧重于利用预定义环境划分来学习不变表示。代表性方法包括 Arjovsky 等人提出的 Invariant Risk Minimization (IRM) 方法<sup>[18]</sup>,其通过在多个已知环境中约束预测器共享以学习环境不变表示;Krueger 等人提出的 V-REx 方法<sup>[19]</sup>则通过最小化不同环境间的经验风险方差以提升模型在分布变化下的泛化能力。然而,这类方法对环境划分质量高度敏感,且依赖人工先验,因而在

环境信息不足或划分不准确时,其性能易受限制。

为进一步缓解对环境标签的依赖,还有一些研究则尝试通过自动环境划分或伪环境构造实现不变学习。例如,Creager 等人提出的 EIIIL 方法<sup>[20]</sup>,通过在训练过程中基于样本损失分布或预测差异等信号,动态自适应地推断环境划分,从而有效减轻伪相关性对模型的干扰;为进一步提升环境多样性, Jia 等人提出了 IGM 方法<sup>[21]</sup>,基于提取的不变子图与环境子图,通过联合环境 Mixup(混合环境相关子图生成伪环境)与不变 Mixup(混合不变子图)策略,学习具有稳定结构特征的表示,从而在无需真实环境标签的情况下增强模型对分布变化的鲁棒性。此外,还有部分研究将信息瓶颈原理引入图 OOD 泛化任务,例如 Chen 等人提出的 CIGA 方法<sup>[22]</sup>和 Miao 等人提出的 GSAT 方法<sup>[23]</sup>,通过最大化标签与不变表示之间的互信息,同时压缩输入中与标签无关的信

息,以提升模型的泛化能力。然而,上述方法大多聚焦于图结构层面的不变性,对于节点属性分布偏移问题考虑不足。例如,如图1所示,当引入节点属性分布偏移后,EIIL、CIGA以及IGM算法的泛化性能都会明显下降。

为克服现有方法在应对属性偏移方面的局限,本文提出一种属性偏移感知的图不变表示学习方法(Attribute-Shift-Aware Graph Invariant Modeling, AS-GIM)。该方法在保持图结构不变性优势的基础上,有效增强模型对节点属性扰动的适应能力,从而提升其在分布外(OOD)场景下的泛化性能与鲁棒性。

具体而言,本文从结构信息与属性信息两个关键维度出发,设计了三项核心机制以提升模型对分布变化的应对能力。首先,构建不变子图提取模块,结合图神经网络编码器与可微分采样技术,从图中提取具有稳定结构特征的子图表示,以捕捉与任务无关的干扰信息之外的结构不变性。其次,为应对属性偏移带来的伪相关性干扰,引入基于自适应学习的特征维度加权方法。该方法通过学习一个特征维度加权向量,自动识别并压制受属性扰动影响较大、对任务贡献较低的噪声维度,同时强化与标签具有因果关系的判别性维度,从而提升模型对关键特征的感知能力。最后,在上述加权后得到的不变子图嵌入上应用Beta分布Mixup增强策略,以线性组合生成跨分布的虚拟训练样本,并针对原始加权表示与Mixup混合表示分别施加一致性损失和正交约束损失,引导模型学习更加稳健的不变特征,进一步提升其对属性偏移的适应性。

本文的贡献主要体现在以下三个方面:

(1)提出了一种全新的视角,从节点属性偏移问题出发,对现有不变表示学习方法进行了补充和改进;

(2)构建了特征维度加权机制与不变混合数据增强策略两个关键模块,提升模型对属性扰动的鲁棒性,并分析了其合理性和必要性;

(3)在多个标准数据集上进行了广泛实验,结果表明本文方法在缓解节点属性偏移影响、提高OOD泛化性能方面均显著优于现有方法。

## 2 相关工作

近年来,图神经网络在图表示学习中取得显著进展,其通过消息传递机制聚合邻居信息,并利用图

级读出函数整合节点表示以支撑图分类任务,展现出强大性能。然而,传统GNN模型普遍假设训练与测试数据服从相同分布,这一假设在实际存在分布外问题的场景下往往难以成立,导致模型性能急剧下降。

早期关于GNN泛化能力的研究主要关注于模型架构设计和算法优化,而对数据分布偏移对图表示的影响考虑不足。近年来,随着OOD问题在各领域的重要性日益凸显<sup>[24-27]</sup>,相关研究开始从因果推断与不变表示学习的角度出发,尝试将图划分为不变子图与环境子图。其基本思想在于:不变子图包含跨不同数据分布均保持稳定的结构-标签关系,而环境子图则主要携带易受环境因素影响的噪声信息,通过仅利用不变子图进行预测,可望提升模型在跨域场景下的鲁棒性。

在相关工作中,Wu等人率先提出了DIR方法<sup>[28]</sup>,构建了首个基于图神经网络的无监督因果图发现框架,通过可微掩码机制自动识别与标签存在因果关联的子图结构,无需依赖显式环境划分;然而,其“子图因果性仅由拓扑结构决定”的隐含假设在节点属性分布出现显著偏移时容易失效,从而误将噪声属性识别为因果特征。Li等人提出的GIL方法<sup>[29]</sup>利用训练数据中分布差异构造虚拟环境,并采用环境间对比学习区分环境与不变子图,但该方法要求已知分布差异的边界条件,缺乏自适应性。Zhu等人提出的MARIO方法<sup>[30]</sup>基于信息瓶颈原理构建不变表示,通过最大化子图与标签间的互信息并同时最小化子图与原始输入之间的互信息,以期过滤冗余信号;然而,其单阶段优化过程难以平衡上述目标,存在可能过度压缩有用结构信息的风险。GSAT<sup>[23]</sup>则在节点级注意力中引入可学习的随机性,通过对比不同随机掩码生成的子图视角来增强抗干扰能力,但随机掩码可能破坏关键结构连续性,尤其在属性噪声存在的情况下会加剧预测不稳定性。最近,Piao等人<sup>[31]</sup>提出利用图粗化构建层次化语义环境,并通过跨粒度子图对比提升不变性识别精度;Sun等人<sup>[32]</sup>则设计了图元网络以显式建模多环境相关性,旨在强化模型在复合分布偏移下的泛化性能。

尽管上述方法在图结构层面的泛化性能上取得了一定进展,但在应对节点属性的跨环境分布变化时仍存在挑战。部分方法在实现过程中已尝试利用节点属性信息以增强模型的泛化性能。例如,Sui等人提出的AdvCA方法<sup>[33]</sup>,通过在节点属性和边特征

上施加对抗性扰动生成新的环境,以扩展环境多样性并保持因果相关特征的稳定性,从而在一定程度上缓解了属性分布偏移的影响。然而,该方法生成的增强样本依赖于训练数据的潜在分布,难以覆盖真实测试环境中更复杂的属性差异。随后,Sui等人提出的AIA方法<sup>[34]</sup>,采用对抗性不变增强策略在节点属性层面生成新的环境,并利用判别器区分稳定特征与环境特征,从而约束模型聚焦于稳定特征。尽管该方法能够缓解因属性扰动导致的性能下降,但其效果高度依赖于对稳定特征与环境特征的准确划分,而这一假设在实际图数据中往往难以满足。相比之下,本文提出的属性偏移感知图不变表示学习方法,通过自适应特征加权直接在表示学习过程中建模节点属性分布差异,并结合跨分布混合增强与一致性、正交约束优化,从根本上减弱了属性偏移带来的干扰,从而在分布外场景下表现出更强的泛化能力。

### 3 基础知识

设图数据集为  $\mathcal{G} = \{G_i, Y_i\}_{i=1}^N$ , 其中每个图  $G_i = (V_i, E_i, X_i)$  由节点集合  $V_i$ 、边集合  $E_i$  以及节点属性矩阵  $X_i \in \mathbb{R}^{|V_i| \times D}$  组成,  $D$  表示属性维度,  $Y_i \in \mathcal{Y}$  为对应标签。根据 Bengio 等人<sup>[35]</sup> 与 Arjovsky 等人<sup>[18]</sup> 提出的分布偏移研究范式, 假设训练集与测试集的分布不相同, 即  $P_{train}(G_i) \neq P_{test}(G_i)$ 。进一步, 假设每个图  $G_i$  可分解为两个部分: 不变子图  $G_i^I = (V_i^I, E_i^I, X_i^I)$  和受环境影响 (Env) 的环境子图  $G_i^{Env} = (V_i^{Env}, E_i^{Env}, X_i^{Env})$ , 其中  $G_i^{Env}$  为  $G_i^I$  的补集。相应地, 定义不变子图集合为  $\mathcal{G}_I = \{G_i^I, Y_i\}_{i=1}^N$  及环境子图集合为  $\mathcal{G}_{Env} = \{G_i^{Env}, Y_i\}_{i=1}^N$ 。此外, 引入特征维度加权向量  $R = \{\omega_k\}_{k=1}^D \subset [0, 1]^D$ , 用于衡量各属性维度对分布偏移的敏感性, 其中当  $\omega_k \rightarrow 0$  时, 表明对该维度的噪声信息实现了有效抑制。定义不变混合表示为  $Z_{Mix} = \lambda Z_i + (1 - \lambda) Z_j$ , 其中混合系数  $\lambda \sim \text{Beta}(\alpha, \alpha)$ ; 且  $Z_i = f_\theta(G_i^I \odot R)$ ,  $\odot$  表示逐元素乘法。

在本研究中, 假设每个图  $G_i$  可分解为不变子图  $G_i^I$  与环境子图  $G_i^{Env}$ , 其中不变子图  $G_i^I$  承载与标签  $Y_i$  稳定相关的核心信息, 而环境子图  $G_i^{Env}$  可能引入与  $Y_i$  存在虚假相关性的干扰信息。这一假设的提出基于以下考虑。首先, 从因果不变性理论出发, 决定预测目标的因果机制在不同环境下应保持稳定, 因

此与标签因果相关的结构对应于不变子图<sup>[36]</sup>; 其次, 这一假设在实际任务中也有直观体现。例如, 在分子性质预测任务中, 功能基团通常决定分子的关键性质, 可视为不变子图, 而分子量或背景结构的差异更多反映环境偏移, 易形成干扰信息。类似的理论假设已在分布外泛化研究中被广泛应用, 如 Arjovsky 等人提出的 Invariant Risk Minimization (IRM) 理论<sup>[18]</sup>, 因而具有坚实的理论基础与实践支撑。

基于上述合理性分析, 本文进一步假设不变子图不仅在不同环境下保持稳定, 还包含决定图标签的核心信息, 即  $P(Y|G_i^I) = P(Y|G_i^I, Env)$ , 从而保证在不同环境下通过提取  $G_i^I$  可获得对未知测试数据具有良好泛化能力的表示。

为进一步缓解节点属性分布偏移带来的影响, 本文引入特征维度加权向量  $R$ , 用于衡量并抑制那些受环境影响较大的噪声维度, 同时强化与任务相关的高信息量特征。基于此, 模型通过  $GNN_\theta$  从加权后不变子图  $G_i^I \odot R$  中提取不变嵌入  $Z_i$ 。接着, 为增强模型在跨环境场景下的鲁棒性, 采用基于 Beta 分布的混合策略构造混合表示  $Z_{Mix} = \lambda Z_i + (1 - \lambda) Z_j$ , 其中  $\lambda \sim \text{Beta}(\alpha, \alpha)$  且  $Z_j$  为另一图的不变嵌入。最终, 旨在学习一个预测函数  $f = \omega \circ h$ , 使得在面对不同环境下的分布偏移时, 模型依然能够保持稳定的预测性能。

## 4 方法设计

为应对图数据中结构及节点属性分布偏移对泛化能力的影响, 本文提出一种属性偏移感知的不变表示学习方法 (见图 2)。该方法无需预定义环境划分, 融合混合增强 (Mixup)<sup>[37]</sup> 与不变学习思想, 旨在捕捉真实的不变模式, 从而有效抑制噪声。具体而言, 此方法由不变子图提取、特征维度加权和不变子图混合增强三大模块构成, 并辅以  $(\mathcal{L}_{RR})$  正则化损失, 以进一步降低冗余噪声对模型的干扰并提高其鲁棒性。

### 4.1 整体框架

在图神经网络的 OOD 泛化研究中, 除了结构偏移, 节点属性分布的偏移同样是导致模型泛化性能下降的关键因素。然而, 现有不变表示学习方法多依赖人工划分的环境, 且聚焦结构不变性, 几乎忽视了属性偏移带来的伪相关风险。针对这一挑战,

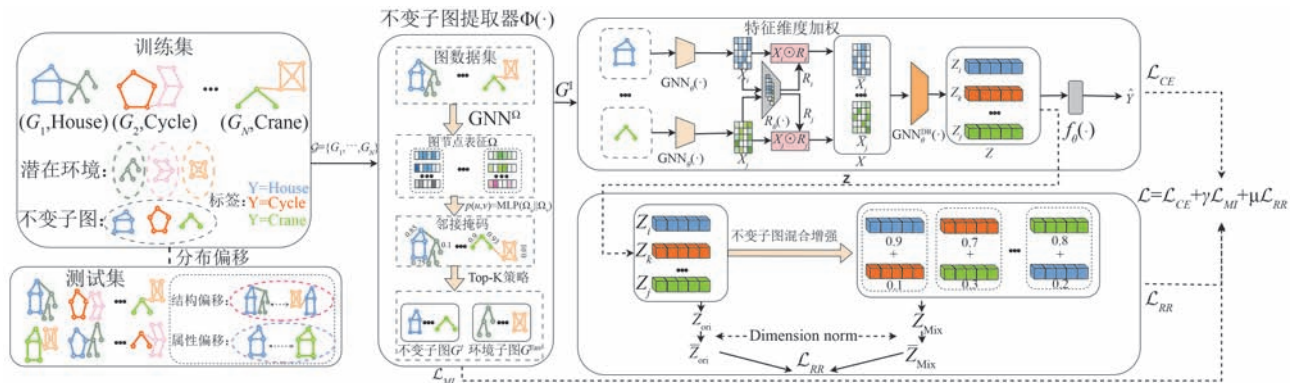


图2 模型框架图

本文提出了一种全新的框架,该框架无需外部环境标签,能够从原始图数据中挖掘与任务标签具有因果关联的稳定子图。在此基础上,通过动态加权节点属性来削弱伪相关特征的影响,并在图嵌入空间中利用Beta分布Mixup策略生成多样化的混合样本,从而引导模型聚焦于与标签具有因果关联的稳定特征,有效提升其在分布外数据上的鲁棒性。

具体而言,所提出的框架包含三个核心模块:首先,不变子图提取模块利用图神经网络编码器和可微分采样方法,从每个图 $G_i$ 中分别提取与标签 $Y_i$ 紧密相关的核心结构(即不变子图 $G_i'$ )以及可能受分布偏移影响的伪相关部分(环境子图 $G_i^{Env}$ ),并通过最大化 $I(G_i'; Y_i)$ 的互信息约束,强化不变子图的稳定性;其次,特征维度加权模块采用基于注意力机制的方法学习权重向量 $R = \{\omega_k\}_{k=1}^D$ (其中 $D$ 为属性维度),以动态调控各特征在表示中的贡献,从而抑制因分布偏移而带来的特征干扰;最后,不变子图混合增强模块在获得稳定的不变表示后,在图嵌入空间中应用混合增强策略,通过线性插值生成具有跨分布的虚拟训练样本,并借助正则化损失进一步抑制噪声与伪相关信息。这三个模块通过协同训练,实现了分类损失、互信息损失和正则化损失的联合最小化,从而确保模型在不同分布下具有鲁棒的预测性能。

#### 4.2 不变子图提取

在图数据中,并非所有结构信息均对预测任务有利,只有部分子图能够稳定地与标签建立因果关联。为此,本文提出一种不变子图提取器 $\Phi(\cdot)$ ,用于从输入图 $G_i$ 中筛选出对预测任务具有稳定贡献的不变子图,形式化表示为

$$G_i' = \Phi(G_i) \quad (1)$$

在理想状态下,该模块应满足不变性要求,即对于任意环境 $Env \in \mathcal{E}$ ,满足:

$$P(Y|G_i') = P(Y|G_i', Env) \quad (2)$$

这表明无论数据采样自哪个环境,利用不变子图所得到的条件分布均保持一致,从而确保提取的不变表示具备跨环境的一致性。

在具体实现上,本文首先采用图神经网络编码器(例如GAT或GCN)将图 $G_i$ 中的各节点映射至低维嵌入空间:

$$\Omega = GNN_{enc}(G_i) \quad (3)$$

其中, $\Omega \in \mathbb{R}^{V \times d}$ 表示图 $G_i$ 中所有节点的嵌入表示, $d$ 为嵌入维度。对于图中的每一条边 $(u, v)$ ,本文将对对应节点 $u$ 和 $v$ 的嵌入向量进行拼接,得到该边的特征向量:

$$\phi(u, v) = \Omega_u \parallel \Omega_v \quad (4)$$

其中, $\parallel$ 表示向量拼接。在此基础上,将 $\phi(u, v)$ 输入多层感知器(MLP)以计算边的重要性得分:

$$p(u, v) = MLP_{enc}(\phi(u, v)) \quad (5)$$

其中, $p(u, v)$ 表示边 $(u, v)$ 被选入不变子图的概率。

为了构造不变子图 $G_i'$ ,本文采用稀疏Top-K采样策略,从所有边中选取得分最高的前 $k\%$ 边。由于直接采样过程不可微,使用Gumbel-Softmax<sup>[38]</sup>技术对采样过程进行可微近似,从而可以在端到端训练中直接优化该模块。

为进一步增强不变子图与标签之间的关联性,本文引入互信息约束损失,其目标是最大化不变子图表示 $\Phi(G_i')$ 与真实标签 $Y_i$ 之间的互信息,形式化表示为

$$\mathcal{L}_{MI} = -I(\Phi(G_i'); Y_i) \quad (6)$$

该约束有助于增强所提取子图的任务相关性,从而提升模型在跨环境场景中的泛化能力。

#### 4.3 特征维度加权

尽管不变子图提取器在一定程度上能够过滤无关信息,但由于节点属性本身存在分布偏移,部分属

性维度在不同环境下可能表现出显著差异,从而对模型稳定性产生负面影响。为缓解该问题,本文设计了特征维度加权模块,通过学习权重向量:

$$\mathbf{R} = \{\omega_k\}_{k=1}^D \subset [0, 1]^D \quad (7)$$

对每个不变子图  $G_i^l$  中的节点特征矩阵  $\mathbf{X}_i \in \mathbb{R}^{V_i \times d}$  进行逐维加权。具体而言,本文采用基于注意力机制的方法,通过交叉熵损失  $\mathcal{L}_{CE}$  对各属性维度的梯度敏感性进行动态调整。对于第  $k$  个维度,其权重按下式计算:

$$\omega_k = 1 - \frac{1}{N} \sum_{i=1}^N \left| \frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{X}_{i,k}} \right| \quad (8)$$

其中,  $\mathbf{X}_{i,k}$  表示第  $i$  个样本中节点特征矩阵  $\mathbf{X}_i$  的第  $k$  列,即第  $k$  个属性。梯度绝对值越大,表明该维度对损失的贡献越显著,且更容易受到环境变化的干扰,因此对应权重会降低。

在此基础上,对于每个由不变子图提取器提取的不变子图  $G_i^l$ ,采用逐元素乘法对节点特征进行更新:

$$\bar{\mathbf{X}}_i = \mathbf{X}_i \odot \mathbf{R} \quad (9)$$

其中,  $\odot$  表示逐元素相乘操作,  $\bar{\mathbf{X}}_i$  表示加权后的节点特征矩阵。随后,采用全局池化操作(例如平均池化或最大池化)对不变子图  $G_i^l$  中所有节点的加权特征进行聚合,构造出图级表示:

$$\boldsymbol{\psi}_i = \text{Pooling}(\{\bar{\mathbf{x}}_v | v \in V_i\}) \quad (10)$$

其中,  $\bar{\mathbf{x}}_v$  表示图  $G_i^l$  中单个节点  $v$  的加权特征表示。该图级表示  $\boldsymbol{\psi}_i$  融合动态加权节点信息,提升对图整体结构及关键特征的表征能力。最后,将图表示  $\boldsymbol{\psi}_i$  输入到后续分类器中,该分类器由全连接层和 SoftMax 层构成,以生成最终预测分布:

$$\hat{Y}_i = \text{SoftMax}(\text{MLP}(\boldsymbol{\psi}_i)) \quad (11)$$

在训练阶段,模型通过最小化交叉熵损失函数  $\mathcal{L}_{CE}$  对预测分布  $\hat{Y}_i$  与真实标签  $Y_i$  进行监督学习:

$$\mathcal{L}_{CE} = - \sum_{i=1}^N Y_i \log(\hat{Y}_i) \quad (12)$$

该设计不仅在特征层面实现了对节点属性的有效降噪与加权,同时通过池化操作构建图级表示,确保后续分类器能够基于鲁棒的表示进行准确预测。

#### 4.4 不变子图混合增强

为进一步提升模型的泛化能力,避免其过于依赖单一分布信息,本文在图嵌入空间中引入了不变混合增强策略。具体而言,在经过不变子图提取和特征加权之后,每个图将得到一个稳定的表示  $Z_i = \text{GNN}_\theta(G_i^l \odot \mathbf{R})$ 。本文采用 Mixup 方法<sup>[37]</sup>,对来自不

同样本的不变表示进行线性插值,生成新的合成样本:

$$Z_{Mix} = \lambda Z_i + (1 - \lambda) Z_j, \lambda \sim \text{Beta}(\alpha, \alpha) \quad (13)$$

为了平衡各特征维度的贡献并消除尺度差异,本文对嵌入表示  $Z_i$  以及  $Z_{Mix}$  进行标准化处理进行实例-维度归一化处理,使每个特征维度均具有零均值和  $1/\sqrt{N}$  的标准差:

$$\bar{Z} = \frac{Z - \mu(Z)}{\sigma(Z) * \sqrt{N}} \quad (14)$$

其中,  $Z$  表示原始嵌入表示,  $\mu(Z)$  与  $\sigma(Z)$  分别表示其均值和标准差,  $N$  为样本总数。

该混合增强策略能够扩展训练数据的分布空间,使模型在训练过程中见识更多跨环境样本,从而迫使其学习到真正的、不受噪声干扰的稳定特征。为进一步保证混合样本与原始样本在语义上的一致性,本文引入了  $\mathcal{L}_{RR}$  正则化损失。该损失由一致性损失和正交约束损失组成:

$$\mathcal{L}_{RR} = \mathcal{L}_{consist} + \mathcal{L}_{ortho} \quad (15)$$

其中,一致性损失定义为

$$\mathcal{L}_{consist} = \left\| \bar{Z}_i - \bar{Z}_{Mix} \right\|_F^2 \quad (16)$$

用于测量原始不变表示  $\bar{Z}_i$  与混合表示  $\bar{Z}_{Mix}$  之间的差异;正交约束损失为

$$\mathcal{L}_{ortho} = \left\| \bar{Z}_i^T \bar{Z}_i - I \right\|_F^2 + \left\| \bar{Z}_{Mix}^T \bar{Z}_{Mix} - I \right\|_F^2 \quad (17)$$

其中,  $\|\cdot\|_F^2$  表示 Frobenius 范数,  $I$  为单位矩阵。该损失项的引入旨在减少嵌入空间中冗余信息的干扰,从而提升生成表示的辨识度和鲁棒性。

将上述各模块整合后,本文设计的整体训练目标为联合最小化多项损失函数,从而高效提取图数据中的不变特征信息。具体目标函数为:  $\mathcal{L} = \mathcal{L}_{CE} + \gamma \mathcal{L}_{MI} + \mu \mathcal{L}_{RR}$ , 其中  $\gamma$  和  $\mu$  分别为控制互信息损失和正则化损失权重的超参数。本方法的整体流程如算法 1 所示。

**算法 1.** 一种属性偏移感知的图不变表示学习方法

输入: 图数据集  $\mathcal{G} = \{G_1, G_2, \dots, G_N\}$ , 每个图  $G = (V, E, X, y)$ , 训练轮数  $T$ , 采样边数  $k$ , Beta 参数  $\alpha$ , 超参数  $\gamma$  (互信息权重),  $\mu$  (正则化权重)。

输出: 图结构数据的不变表示  $Z$ , 分类结果。

1. 初始化: 图编码器  $\text{GNN}(\cdot)$ , 边评分器  $\text{MLP}(\cdot)$ , 特征加权模块  $\text{R}(\cdot)$ , 分类器  $f(\cdot)$ ;
2. FOR epoch = 1 TO  $T$  DO
3. 将数据集划分为小  $D = \{B_1, B_2, \dots, B_m\}$ ;

4. FOR 每个批次  $B \in DDO$
5. FOR 每个图  $G = (V, E, X, y) \in BDO$
6. 根据公式(3)~(5)计算图中每条边的权重  $p(u, v)$ , 并通过 Top-K 策略选取权重最高的  $k$  条边, 构造不变子图  $G_i^l$ ;
7.  $Z_i = GNN_{enc}(G_i^l)$ ;
8.  $\mathcal{L}_{MI} = MutualInfo(Z_i, y_i)$ ;
9. 根据公式(7) - (11)对不变子图  $G_i^l$  中节点属性进行动态权重分配, 使用公式(12)计算  $\mathcal{L}_{CE}$  损失;
10. 根据公式(13) - (17)对加权后不变表示执行混合增强, 并计算一致性与正交正则化损失  $\mathcal{L}_{RR}$ ;
11. END FOR
12.  $\mathcal{L} = \mathcal{L}_{CE} + \gamma \mathcal{L}_{MI} + \mu \mathcal{L}_{RR}$ ;
13. 执行反向传播并更新  $\theta_{GNN}, \theta_{MLP}, \theta_R, \theta_f$ ;
14. END FOR
15. END FOR

## 5 实验与结果分析

为验证所提出的属性偏移感知图不变表示学习方法在图分类任务中缓解节点属性分布偏移、提升OOD泛化能力的有效性, 本文在合成数据集及多个真实数据集上进行了系列实验。实验旨在评估模型在不同分布偏移场景下的泛化能力, 并对各模块的贡献及超参数的敏感性进行详细分析。

### 5.1 实验设置

#### 5.1.1 数据集

在实验中, 本文选取多个数据集, 以系统评估模型在不同类型分布偏移下的泛化性能。

SPMotif合成数据集用于研究模型在受控偏移下的泛化能力。为模拟现实场景中可能出现的结构与属性偏移, 本文设计了三类扰动: 首先, 结构扰动通过随机替换或删除图中部分边, 引入节点度分布变化, 从而模拟结构相关偏移; 其次, 属性扰动在节点特征的不同维度上添加偏置或随机噪声, 以模拟属性偏移; 最后, 混合扰动同时注入结构与属性偏移, 并通过参数  $\text{bias} \in \{0.33, 0.6, 0.9\}$  调节干扰强度, 以生成不同程度的环境变化, 用于验证模型在伪相关信息干扰下捕捉不变特征的能力。

在验证模型在真实图数据上的泛化能力方面, Graph-SST5<sup>[39]</sup>文本图数据集通过调整句子长度分布, 使不同环境中图的节点数量存在系统性差异, 并改变节点的依存边连接数, 引入结构偏移, 从而测试

模型在节点度分布偏移下的稳健性。DrugOOD<sup>[10]</sup>分子数据集则包含天然存在的属性偏移, 可反映真实化学空间中的分布差异, 其主要偏移类型包括: Assay(实验条件差异导致的分子活性测量偏移)、Scaffold(分子骨架结构差异导致的化学空间分布偏移)和Size(分子大小分布差异)。上述数据集可用于评估模型在属性级别分布变化下的泛化能力。

此外, 本文还使用ColoredMNIST改造的图数据集来模拟图属性偏移场景。该数据集通过将彩色信息引入MNIST手写数字的图结构表示中, 构造类别标签与颜色属性之间的虚假相关性, 从而形成不同环境, 用于测试模型在属性级别的伪相关干扰下是否能够学习到真正与类别标签相关的不变特征。

#### 5.1.2 评价指标

为了全面评估模型性能, 本文针对不同数据集选择了与其任务特性相匹配的评价指标。在Graph-SST5和ColoredMNIST数据集上, 由于类别分布较为均衡, 采用分类准确率(Accuracy, ACC)作为主要指标, 以衡量整体分类能力; 在SPMotif合成数据集上, 同样使用Accuracy来评估模型在合成环境下的泛化表现; 而在DrugOOD数据集上, 由于其为类别高度不平衡的二分类任务数据集, 采用ROC-AUC指标<sup>[10]</sup>, 以更合理地反映模型在属性偏移条件下对正负样本的区分能力。所有实验均重复5次, 并报告平均值与标准差, 以确保结果的稳定性与可靠性。

#### 5.1.3 基线方法

除了采用经验风险最小化(ERM)<sup>[6]</sup>作为基线之外, 实验中还比较了多种主流的图OOD泛化方法, 以验证本文提出的结构-属性协同优化策略在消除伪相关信息和提升分布外泛化能力方面的优势。基线方法主要分为三类, 第一类为不变性学习方法。该类方法涵盖依赖人工环境标签的算法(如IRM<sup>[18]</sup>、V-Rex<sup>[19]</sup>), 其核心机制是在预定义环境下强制模型提取一致特征以抑制噪声; 同时也包括无需外部标注的算法(如EIL<sup>[20]</sup>), 该算法通过自适应聚类推断环境来学习不变表示。为进一步增强不变性, AIA<sup>[34]</sup>引入了对抗性不变增强策略, 通过在节点属性层面生成新的环境, 并利用判别器引导模型聚焦于稳定特征, 从而强化模型的泛化能力。第二类方法基于信息瓶颈原理, 例如CIGA<sup>[22]</sup>和GSAT<sup>[23]</sup>, 通过最大化标签与不变特征的互信息并压缩任务无关信息来提升泛化能力。第三类为内部环境划分方

法,如DIR<sup>[28]</sup>,其在单图内部构造多个环境子图并施加不变性约束。本文还对比了IGM<sup>[21]</sup>与EQuAD<sup>[40]</sup>:前者通过环境Mixup与不变性Mixup的联合策略生成多样化环境并提炼稳定子图;后者基于Infomax原理解耦不变特征与伪相关特征,以构建更稳健的归纳偏置。上述方法分别从数据增强、不变表示提取和信息解耦的角度出发,构成了充分且多样化的对比实验体系,为验证所提方法在复杂分布偏移场景下的有效性提供了坚实支撑。

#### 5.1.4 超参数设置

为保证模型在不同数据集和环境下的性能稳定性,本文采用交叉验证与网格搜索相结合的方法来确定关键超参数。具体而言,每个数据集首先划分为训练集和测试集,训练集进一步进行5折交叉验证。在每次交叉验证中,通过网格搜索在预设的参数范围内训练模型并评估性能,推荐的参数范围为: $\gamma \in [0.3, 0.8]$ 、 $\mu \in [0.3, 0.6]$ 和 $\alpha \in [0.2, 0.6]$ 。该方法能够在不同数据分布和环境下为模型提供合理的超参数初始化,从而有效提升泛化性能。

## 5.2 实验结果

### 5.2.1 在合成数据上的实验

在合成数据集SPMOTIF-MIXED上的实验结果(见表1)表明,随着虚假相关性bias从0.33、0.60到0.90逐渐增强,传统的ERM方法以及部分现有的不变表示学习方法(如IRM、V-REx和EIL)均出现显著性能下降,其准确率分别从约58.18%降至41.36%。与之相比,本文所提出的方法在各个bias条件下均实现了明显的提升:在bias=0.33时,本文方法的准确率达到75.83%;在bias=0.60时为73.25%;在bias=0.90时为58.66%。与ERM相比,本文方法在低、中、高三个场景下的准确率分别提高约17.7%、23.9%和17.3%,平均提升近19.6%。此外,相较于当前最先进的不变表示学习方法,如AIA、IGM和EQuAD,本文方法在各实验设置下分别实现了平均3.4%、5.1%和2.9%的性能提升,进一步验证了其在复杂属性偏移环境下的优势表现。这些结果充分证明,通过引入特征维度加权和不变混合增强策略,本文方法能够有效过滤节点属性偏移带来的噪声,准确捕捉图中真正的因果相关性,从而显著提升模型在分布外数据上的泛化性能。

### 5.2.2 在真实数据上的实验

如表2所示,AS-GIM在五个真实OOD基准任务

表1 在合成数据集上的图分类结果

数据集	SPMOTIF-MIXED		
	BIAS=0.33	BIAS=0.60	BIAS=0.90
评价指标	ACC指标		
ERM	58.18±4.30	49.29±8.17	41.36±3.29
IRM	58.20±1.97	49.29±3.67	40.73±1.93
V-REX	57.82±5.93	48.25±2.79	43.27±1.32
EIL	53.91±3.15	48.41±5.53	41.75±4.97
DIR	67.28±4.06	51.66±14.1	38.58±5.88
CIGA	72.42±4.80	70.83±7.54	54.25±5.38
GSAT	70.22±3.26	69.65±4.23	50.29±3.81
AIA	71.35±4.37	71.53±3.81	54.77±2.59
IGM	70.47±3.66	72.44±6.38	49.53±4.29
EQuAD	73.06±3.58	71.32±6.55	53.68±3.74
<b>AS-GIM</b>	<b>75.83±3.47</b>	<b>73.25±4.28</b>	<b>58.66±3.54</b>

中均表现出显著优势。在DRUG-ASSAY(73.42%准确率)、DRUG-SCA(70.31%)及DRUG-SIZE(68.25%)三项分子预测任务中,AS-GIM分别较ERM提升约1.5%,并小幅领先于CIGA、IGM与AIA,证明了其在分子属性分布偏移场景下的稳健性。在风格扰动最为剧烈的CMNIST-SP任务中,AS-GIM达到50.43%的准确率,显著超越ERM与V-REx(10%~14%)及次优IGM(47.93%)与AIA(47.85%),充分说明其对结构与属性伪相关的抑制效果。在Graph-SST5情感分析任务上,AS-GIM取得45.78%的分类准确率,同样优于ERM(43.89%)及其他基线方法。总体来看,AS-GIM在各类OOD场景下均超越现有方法,且在属性与结构偏移最剧烈的环境中优势最为突出,验证了其特征维度加权和不变性混合增强策略能够有效降低伪相关性干扰,促进因果特征的学习,从而显著提升分布外泛化能力。

### 5.3 算法复杂度分析

为了进一步验证方法的可扩展性,本文对算法的时间与空间复杂度进行了分析。整体来看,算法的计算开销来自于边权重估计、Top-K边筛选以及节点属性加权过程。在单图规模为 $N$ 个节点、 $M$ 条边、特征维度为 $d$ 的情况下,时间复杂度为 $O(M \cdot d + M \log k + N \cdot d^2)$ ,其中 $k$ 为选取的边数。结合批次大小 $B$ 与训练轮数 $T$ ,总体复杂度为 $O(T \cdot B \cdot (M \cdot d + M \log k + N \cdot d^2))$ ,与现有基于子图选择的图不变表示方法保持同一量级。在空间开销方面,除图结构与节点属性存储外,模型参数仅需占用少量额外空间,整体空间复杂度为 $O(N \cdot d + M)$ 。

表 2 面向对象设计(OOD)在真实世界图的复杂分布变化上的泛化性能

数据集	DRUG-ASSAY	DRUG-SCA	DRUG-SIZE	CMNIST-SP	GRAPH-SST5
评价指标	AUC 指标			ACC 指标	
ERM	71.79±0.27	68.85±0.62	66.70±1.08	13.96±5.48	43.89±1.73
IRM	72.12±0.49	68.69±0.65	66.54±0.42	31.58±9.52	43.69±1.26
V-REX	72.05±1.25	68.92±0.98	66.33±0.74	10.29±0.46	43.28±0.52
EIIL	72.60±0.47	68.45±0.53	66.38±0.66	30.04±10.9	42.98±1.03
DIR	68.25±1.40	63.91±1.36	60.40±1.42	15.50±8.65	41.12±1.96
CIGA	73.17±0.39	69.70±0.27	67.78±0.76	44.91±4.31	45.25±1.27
GSAT	70.59±0.43	66.45±0.50	66.70±0.37	43.38±5.44	45.09±1.34
AIA	72.15±0.46	68.93±0.91	66.92±1.12	47.85±3.71	44.13±1.67
IGM	73.07±0.89	68.32±0.48	67.99±1.26	47.93±3.62	43.79±1.49
EQuAD	72.98±0.32	69.78±0.35	68.06±0.18	48.23±4.15	44.76±0.89
<b>AS-GIM</b>	<b>73.42±0.41</b>	<b>70.31±0.19</b>	<b>68.25±0.69</b>	<b>50.43±6.83</b>	<b>45.78±1.03</b>

#### 5.4 模型鲁棒性分析

为验证所提方法在节点属性偏移下的鲁棒性,本文在 Graph-SST5 与 Drug-Assay 数据集上设计了属性扰动实验。具体地,通过随机替换节点属性的比例  $p = \{0.1, 0.3, 0.5\}$  模拟不同程度的噪声与分布偏移,并比较 AS-GIM 与 CIGA、IGM、EQUAD 的性能。实验结果(图 3)显示,随着扰动增加,各方法性能均下降,但 AS-GIM 下降幅度明显较小。在

Graph-SST5 数据集上,当  $p=0.5$  时,CIGA、IGM 与 EQUAD 准确率分别下降约 6.53%、4.78% 和 6.93%,而 AS-GIM 仅下降约 2.53%,仍保持最高准确率 43.25%;在 Drug-Assay 数据集上,AS-GIM 的 AUC 下降约 2.17%,其他方法均超过 4%。结果表明,本文方法在缓解节点属性分布偏移方面更稳健,有效验证了自适应特征加权机制与 Beta 混合增强策略在强化模型稳定表征能力上的作用。

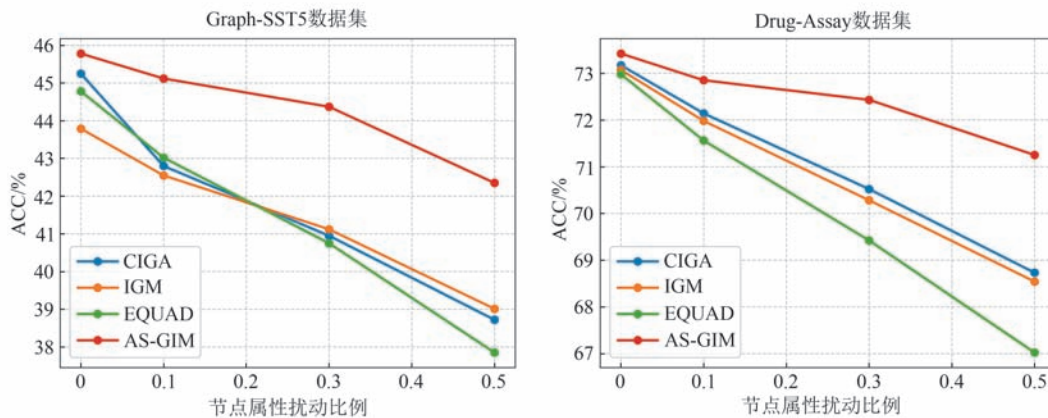


图 3 鲁棒性分析

#### 5.5 消融实验

如图 4 所示,在消融实验中,本文分别在 DrugOOD 数据集和 Graph-SST5 数据集上评估了模型中两个关键模块(特征维度加权(Feature Dimension Weighting, FDW)和不变混合增强(Invariant Mixup Augmentation, IMA))的作用及其协同效应。实验结果表明,在 DrugOOD 数据集中,固定不变子图提取器后,仅保留 FDW 模块配置相比完整模型,ROCAUC 平均下降约 2%;仅采用 IMA 模块时,ROCAUC 下降约 3%。在 Graph-

SST5 数据集中,移除 FDW 模块后,分类准确率下降约 1.6%;去除 IMA 模块则下降约 2.5%。同时移除两模块时,模型性能在两个数据集中均进一步下降。结果表明,FDW 模块在动态调控节点特征、抑制属性偏移带来的噪声信息方面至关重要,而 IMA 模块则有助于扩充训练数据的多样性并提升模型鲁棒性。综合实验结果表明,两个模块的协同作用显著提高了模型的 OOD 泛化能力,完整模型在应对分布外变化时表现出更强的泛化性能。

为了进一步验证所设计损失函数在模型优化中

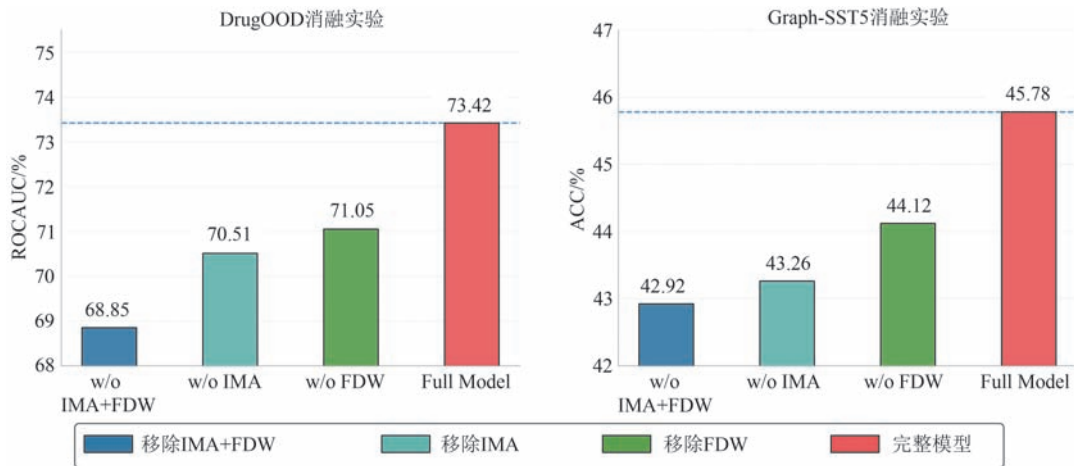


图 4 消融实验

的作用,本文在 Graph-SST5 与 DrugOOD 数据集上进行了额外的消融实验,结果如表 3 所示。可以观察到,完整的 AS-GIM 模型在两个数据集上均取得最佳性能。去除互信息约束会导致模型在 Graph-SST5 上的准确率下降约 3.38%,在 DrugOOD 上的 ROC-AUC 下降约 3.43%,说明互信息约束在不变子图提取中发挥了关键作用,有助于模型捕获因果相关特征并减轻属性偏移干扰。相比之下,去除一致性损失或正交约束损失会分别导致模型在两数据集上的性能下降 1%~2%,表明它们在防止过拟合和实现特征解耦方面具有辅助作用。当去除所有损失函数时,模型性能进一步下降至最低水平,这验证了联合约束机制对提升整体模型性能的必要性。

### 5.6 超参数敏感性分析

图 5、图 6 中,本实验分别对 SPMotif-Mixed (bias=0.9) 和 Graph-SST5 两个数据集的关键超参数  $\gamma$  与  $\mu$  进行了敏感性分析。图 5(a) 展示了在 SPMotif-Mixed 上固定  $\mu=0.3$  时,不同  $\gamma$  值对模型

表 3 损失函数消融对实验影响

配置\数据集	Graph-SST5	DrugOOD
完整模型	45.78	73.42
去互信息损失	42.40	69.99
去一致性损失	44.51	71.14
去正交损失	44.06	71.42
去所有损失	42.13	70.97

准确率的影响:当  $\gamma=0.6$  时,准确率达到峰值 58.66%,而在  $\gamma$  较小(0.1、0.3)或较大(0.8、1.0)时,性能均有所回落,表明中等强度的正规化因子能够在控制过拟合与保留表达能力之间取得最佳平衡。图 6(a) 则固定  $\gamma=0.6$ ,考察  $\mu$  对性能的作用;可以看到  $\mu=0.3$  时准确率最高(58.63%),当  $\mu$  过小或过大时,模型性能同样下降,说明适度的样本混合比例是提升泛化的重要因素。对于 GRAPH-SST5 数据集,图 5(b) 在  $\mu=0.5$  情况下对  $\gamma$  的扫描结果表明, $\gamma=0.8$  时 AUC 达到最大值 45.78%,而

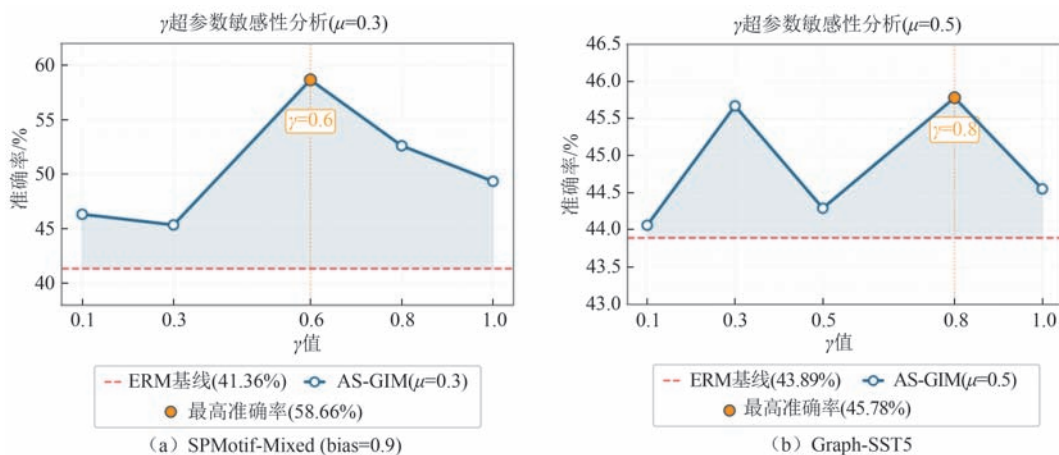


图 5  $\gamma$  超参数分析

其他位置均低于该值；图6(b)在 $\gamma=0.8$ 情形下对 $\mu$ 的敏感性分析则显示 $\mu=0.5$ 可使AUC达到约46.0%，进一步验证了在本任务中中等程度的正则化与样本混合比例为最优超参数配置。整体而言，两组实验均呈现出“过弱或过强均不利，适中取值最优”的典型超参数敏感性曲线，为后续模型调优提供了明确的取值指引。

在此基础上，图7和图8进一步展示了模型在SPMotif-Mixed (bias=0.9)与Graph-SST5数据集上对关键超参数 $k$ 和 $\alpha$ 的敏感性分析。在采样边 $k$ 上，SPMotif-Mixed数据集在 $k=0.25$ 时取得最佳性能，偏离该值会导致准确率明显下降，表明 $k$ 需在信息完整性与噪声控制之间取得平衡；过小的 $k$ 会导致不变子图信息不足，而过大的 $k$ 则可能引入

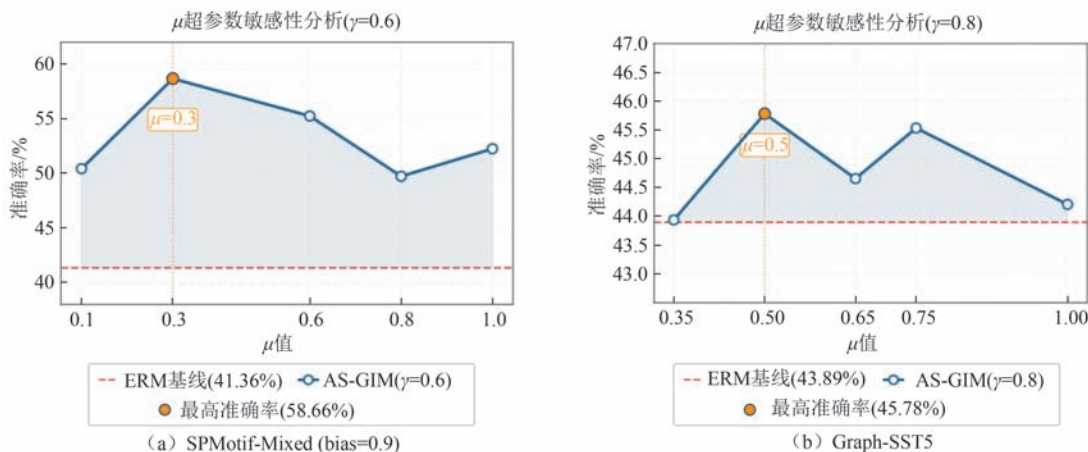


图6  $\mu$ 超参数分析

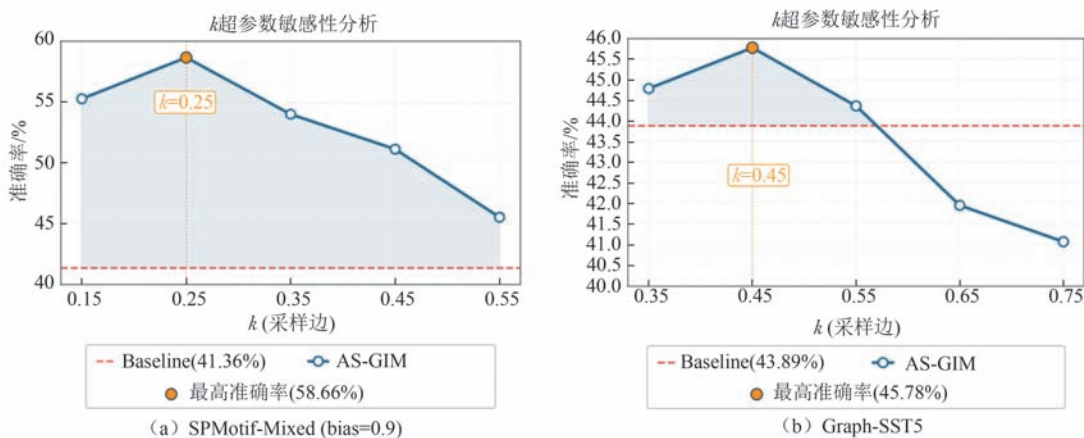


图7  $k$ 超参数分析

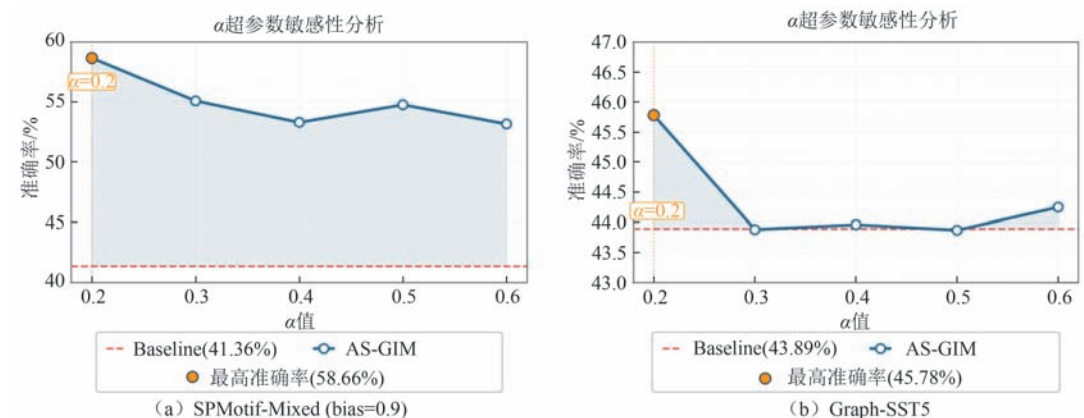


图8  $\alpha$ 超参数分析

冗余结构,削弱不变性建模效果。Graph-SST5数据集的趋势类似,最佳性能出现在 $k=0.45$ ( $ACC=45.78\%$ ),但同样表现出过小或过大取值引发的性能下降。在Beta参数 $\alpha$ 的影响下,SPMotif-Mixed数据集上模型在 $\alpha=0.2$ 时表现最佳,随着 $\alpha$ 增大,准确率略有波动(最低为 $53.17\%$ ),但在所有测试值下均优于基线模型,说明Beta参数 $\alpha$ 在不同取值下均能稳定提升模型性能。Graph-SST5上模型对 $\alpha$ 变化相对不敏感,在 $\alpha \in [0.2, 0.6]$ 范围内性能波动低于 $2\%$ ,显示出较高稳健性。实验结果表明合理的超参数选择不仅能够有效提升模型性能,还能在不同任务中增强其稳定性。

## 6 结 论

本文针对图结构数据的分布外泛化(OOD)问题,提出了一种结合特征维度加权与不变混合数据增强的方法,以缓解节点属性偏移带来的影响。通过学习维度加权向量,有效识别并抑制受属性偏移影响的特征,同时采用不变混合策略增强数据分布的多样性,从而提升模型的OOD泛化能力。实验结果表明,该方法在多个基准数据集上均优于现有不变表示学习方法,并在不同分布偏移场景下展现出较强的鲁棒性。这项研究为图数据的OOD泛化提供了一种新的视角,有助于进一步理解和优化不变子图的提取策略,未来可拓展至更广泛的图学习任务。

## 参 考 文 献

- [1] Bi W, Xu B, Sun X, et al. Predicting the silent majority on graphs: Knowledge transferable graph neural network// Proceedings of the ACM Web Conference 2023. Austin, USA, 2023: 274-285
- [2] Buterez D, Janet J P, Kiddle S J, et al. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting. Nature Communications, 2024, 15(1): 1517
- [3] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of the 5th International Conference on Learning Representation. Toulon, France, 2017: 1-14
- [4] Wu F, Souza A, Zhang T, et al. Simplifying graph convolutional networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA, 2019: 6861-6871
- [5] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs//Proceedings of the 31st International Conference on Neural Information Systems. Long Beach, USA, 2017: 1025-1035
- [6] Vapnik V. The nature of statistical learning theory. New York, USA: Springer-Verlag, 2013
- [7] Bevilacqua B, Zhou Y, Ribeiro B. Size-invariant graph representations for graph classification extrapolations//Proceedings of the 38th International Conference on Machine Learning. Virtual, 2021: 837-851
- [8] Yehudai G, Fetaya E, Meir E, et al. From local structures to size generalization in graph neural networks//Proceedings of the 38th International Conference on Machine Learning. Virtual, 2021: 11975-11986
- [9] Yoo H, Lee Y C, Shin K, et al. Disentangling degree-related biases and interest for out-of-distribution generalized directed network embedding //Proceedings of the ACM Web Conference 2023. Austin, USA, 2023: 231-239
- [10] Ji Y, Zhang L, Wu J, et al. Drugood: out-of-distribution dataset curator and benchmark for ai-aided drug discovery - a focus on affinity prediction problems with noise annotations// Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA, 2023: 8023-8031
- [11] Zhang K, Liu S, Wang S, et al. A survey of deep graph learning under distribution shifts: from graph out-of-distribution generalization to adaptation. arXiv preprint arXiv: 2410.19265, 2024
- [12] Koyama M, Yamaguchi S. Out-of-distribution generalization with maximal invariant predictor//Proceedings of the 9th International Conference on Learning Representation. Virtual, 2021: 1-20
- [13] Hu W, Fey M, Zitnik M, et al. Open graph benchmark: datasets for machine learning on graphs//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada, 2020: 22118-22133
- [14] Koh P W, Sagawa S, Marklund H, et al. Wilds: A benchmark of in-the-wild distribution shifts//Proceedings of the 38th International Conference on Machine Learning. Virtual, 2021: 5637-5664
- [15] Zhao T, Jin W, Liu Y, et al. Graph data augmentation for graph machine learning: a Survey. arXiv preprint arXiv: 2202.08871, 2022
- [16] Shi B, Wang Y, Guo F, et al. Graph domain adaptation: challenges, progress and prospects. arXiv preprint arXiv: 2402.00904, 2024
- [17] Cai R, Wu F, Li Z, et al. Graph domain adaptation: A generative view. ACM Transactions on Knowledge Discovery from Data, 2024, 18(3): 1-24
- [18] Arjovsky M, Bottou L, Gulrajani I, et al. Invariant risk minimization. arXiv preprint arXiv: 1907.02893, 2019
- [19] Krueger D, Caballero E, Jacobsen J H, et al. Out-of-distribution generalization via risk extrapolation (rex)//Proceedings of the 38th International Conference on Machine Learning. Virtual, 2021: 5815-5826

- [20] Creager E, Jacobsen J H, Zemel R. Environment inference for invariant learning//Proceedings of the 38th International Conference on Machine Learning. Virtual, 2021: 2189-2200
- [21] Jia T, Li H, Yang C, et al. Graph invariant learning with subgraph co-mixup for out-of-distribution generalization//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2024: 8562-8570
- [22] Chen Y, Zhang Y, Bian Y, et al. Learning causally invariant representations for out-of-distribution generalization on graphs//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA, 2022: 22131-22148
- [23] Miao S, Liu M, Li P. Interpretable and generalizable graph learning via stochastic attention mechanism//Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA, 2022: 15524-15543
- [24] Gui S, Liu M, Li X, et al. Joint learning of label and environment causal independence for graph out-of-distribution generalization//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA, 2023: 3945-3978
- [25] Li X, Gui S, Luo Y, et al. Graph structure and feature extrapolation for out-of-distribution generalization. arXiv preprint arXiv:2306.08076, 2023
- [26] Guo K, Wen H, Jin W, et al. Investigating out-of-distribution generalization of GNNs: an architecture perspective//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain, 2024: 932-943
- [27] Yu J, Liang J, He R. Mind the label shift of augmentation-based graph ood generalization//Proceedings of the 41st IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2023: 11620-11630
- [28] Wu Y, Wang X, Zhang A, et al. Discovering invariant rationales for graph neural networks. arXiv preprint arXiv: 2201.12872, 2022
- [29] Li H, Zhang Z, Wang X, et al. Learning invariant graph representations for out-of-distribution generalization//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA, 2022: 11828-11841
- [30] Zhu Y, Shi H, Zhang Z, et al. Mario: model agnostic recipe for improving ood generalization of graph contrastive learning//Proceedings of the ACM Web Conference 2024. Singapore, 2024: 300-311
- [31] Piao Y, Lee S, Lu Y, et al. Improving out-of-distribution generalization in graphs via hierarchical semantic environments//Proceedings of the 42nd IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2024: 27631-27640
- [32] Sun X, Wang L, Liu Q, et al. DIVE: subgraph disagreement for graph out-of-distribution generalization//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain, 2024: 2794-2805
- [33] Sui Y, Wang X, Wu J, et al. Adversarial causal augmentation for graph covariate shift//Proceedings of the 11st International Conference on Learning Representation. Kigali, Rwanda, 2023: 1-19
- [34] Sui Y, Wu Q, Wu J, et al. Unleashing the power of graph data augmentation on covariate distribution shift//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA, 2023: 18109-18131
- [35] Bengio Y, Deleu T, Rahaman N, et al. A meta-transfer objective for learning to disentangle causal mechanisms//Proceedings of the 8th International Conference on Learning Representation. Virtual, 2020: 1-26
- [36] Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2016, 78(5): 947-1012
- [37] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization//Proceedings of the 6th International Conference on Learning Representation. Vancouver, Canada, 2018: 1-13
- [38] Jang E, Gu S, Poole B. Categorical reparametrization with Gumbel-Softmax//Proceedings of the 5th International Conference on Learning Representation. Toulon, France, 2017: 1-13
- [39] Yuan H, Yu H, Gui S, et al. Explainability in graph neural networks: a taxonomic survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(5): 5782-5799
- [40] Yao T, Chen Y, Chen Z, et al. Empowering graph invariance learning with deep spurious infomax//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria, 2024: 56860-56884



**CAO Fu-Yuan**, professor, Ph. D. supervisor. His main research interests include data mining and machine learning.

**ZHAO Liang**, M. S. candidate. His main research interests include graph machine learning and out-of-distribution generalization.

**CHEN Qing-Qiang**, Ph. D., lecturer. His main research interests include data mining and machine learning.

## Background

Graph representation learning is a foundational task in graph machine learning, aiming to embed complex graph-structured data into low-dimensional vector spaces for downstream tasks such as graph classification. This approach has shown significant success in a wide range of applications, including community detection in social networks and molecular property prediction in biochemistry. In recent years, Graph Neural Networks (GNNs) have become the dominant paradigm for graph representation learning, leveraging hierarchical message-passing mechanisms to iteratively aggregate local and global neighborhood information, thus producing expressive graph-level representations in an end-to-end manner.

Despite their success, most existing GNN methods operate under the assumption of independent and identically distributed training and testing data. However, in real-world applications, this assumption is often violated due to differences in data generation processes, environmental factors, or sampling strategies. These differences result in distribution shifts in various graph aspects such as graph size, node degree, subgraph patterns, and node attribute distributions. As a consequence, models may capture spurious correlations that do not generalize to unseen distributions, leading to severe performance degradation in out-of-distribution (OOD) testing scenarios.

To enhance the OOD generalization of GNNs, recent research has explored approaches from domain adaptation and invariant representation learning. Domain-adversarial training and risk extrapolation methods, such as IRM and V-REx, aim to learn representations that remain stable across pre-defined environments. However, these methods typically rely on accurate environment labels or domain annotations, which are difficult to obtain for graph data. To address this, environment-free or self-

supervised techniques such as EIIL and IGM have been proposed, which infer latent environments or construct pseudo-environments during training. These methods have shown promising results in structural distribution shifts. Yet, they largely ignore the influence of node attribute shifts, which are prevalent and often detrimental in real-world graphs.

To tackle this issue, we focus on a relatively underexplored but crucial problem: improving the robustness of graph representation learning under node attribute distribution shifts. We propose a novel framework named Attribute-Shift-Aware Graph Invariant Modeling (AS-GIM), which explicitly models attribute perturbations while preserving structural invariance. Our method introduces adaptive feature dimension weighting and Beta-based Mixup augmentation to suppress noisy dimensions and generate cross-distribution virtual samples. By combining consistency and orthogonal constraints on original and mixed representations, AS-GIM effectively learns stable and generalizable features. Extensive experiments on synthetic and real-world benchmarks demonstrate that our approach significantly outperforms state-of-the-art methods under various OOD settings, especially when node attribute shifts are severe.

This work was supported by the Joint Fund Project of the National Natural Science Foundation of China (U24A20323), the National Natural Science Foundation of China (62376145), the Science and Technology Innovation Talent Team Project of Shanxi Province (202204051002016), the Key Technology Research Project of Taihang Laboratory in Shanxi Province (THYF-JSZX-24010700), the “Unveiling the List and Assuming Leadership” Key Core Technology Project under Taiyuan’s “Double Hundred Tackling Action” (2024TYJB0127), and the Fundamental Research Program of Shanxi Province (202503021212095).