

分布偏移场景下图神经网络测试时自适应方法

吴 涛¹, 周圆庆¹, 先兴平^{1*}, 崔灿一星¹, 刘逸琛¹, 乔少杰², 牛伟纳³

(1. 重庆邮电大学网络空间安全与信息法学院, 重庆 400065; 2. 成都信息工程大学软件工程学院, 四川成都 610225;
3. 电子科技大学网络空间安全学院, 四川成都 611731)

摘 要: 图神经网络(Graph Neural Networks, GNNs)凭借其强大的表征能力,在多种图学习任务中表现出色。然而,当训练集与测试集之间的数据分布发生偏移时,GNNs的预测性能会显著下降。现有研究普遍将分布偏移视为单一的整体扰动,缺乏对协变量偏移与概念偏移的区分建模,这种粗粒度的处理难以刻画两类偏移对模型性能的差异化影响,进而制约了传统的图神经网络泛化方法在分布偏移场景下的针对性与有效性。针对此问题,本工作从协变量偏移和概念偏移的特性出发,提出一种基于原型引导的测试时自适应方法ProGTA(Prototype-Guided Test-time Adaptation)。该方法突破了传统静态域适应的局限,主要由原型图对比学习模块和多层次对齐模块构成,能够在模型推理之前进行自适应更新。具体而言,原型图对比学习模块在编码器后引入可学习的类别原型层,利用源域先验指导目标域节点的软分配,以建模图数据的类别语义结构,进而构造出具备更高语义可靠性的同质性增强视图。同时,该模块结合原型对比损失和原型聚类损失联合优化节点表征与类别原型,将局部视图一致性提升为类别语义一致性,为后续类别原型对齐提供可靠的语义锚点。多层次对齐模块则在测试时训练阶段通过对齐源域与目标域嵌入表征的一阶和二阶统计量来缓解协变量偏移造成的特征分布变化,同时通过对齐源域与目标域的类别原型分布来缓解概念偏移造成的类别语义变化。为验证所提方法的有效性,本工作在涵盖人工偏移、时间演化及跨域偏移的多个图分布偏移基准数据集上进行了详尽的实验评估。实验结果表明,ProGTA在分布外泛化性能上显著优于Tent、EERM和GTrans等代表性方法。进一步地,本研究从域适应视角分析了对抗攻击和概念偏移在破坏模型决策边界方面的一致性,将对抗攻击问题转化为分布偏移的一种特例,证实了该方法可同时提升模型的分布外泛化性与对抗鲁棒性。

关键词: 图神经网络;分布偏移;测试时自适应;图对比学习;模型泛化性;对抗鲁棒性

基金项目: 国家自然科学基金(No.62376047);重庆市自然科学基金创新发展联合基金重点项目(No.CSTB2023NSCQ-LZX0003);重庆市教委科学技术研究计划重点项目(No.KJZD-K202300603, No.KJZD-K202500604)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2026)04-1732-16

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20251206

Test-Time Adaptation of Graph Neural Networks under Distribution Shifts

WU Tao¹, ZHOU Yuanqing¹, XIAN Xingping^{1*}, CUI Canyixing¹,
LIU Yichen¹, QIAO Shaojie², NIU Weina³

(1. School of Cyberspace Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. School of Software Engineering, Chengdu University of Information Technology, Chengdu, Sichuan 610225, China;

3. School of Cyber Security, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China)

Abstract: Graph neural networks (GNNs) have achieved outstanding performance in various graph learning tasks owing to their powerful representation capabilities. However, the predictive performance of GNNs degrades significantly when data distribution shifts occur between the training and test sets. Existing studies generally treat distribution shift as a single holistic perturbation and lack differentiated modeling of covariate shift and concept shift. Such coarse-grained processing fails to characterize the distinct impacts of these two types of shifts on model performance, thereby restricting the pertinence and effectiveness of traditional GNN generalization methods in distribution shift scenarios. To address this issue, starting from the characteristics of covariate shift and concept shift, this work proposes a prototype-guided test-time adaptation method named ProGTA (Prototype-Guided Test-time Adaptation). This method breaks through the limitations of traditional static domain adaptation, mainly consisting of a prototype graph contrastive learning module and a multi-level alignment module, which enables adaptive updating prior to model inference. Specifically, the prototype graph contrastive learning module introduces a learnable class prototype layer after the encoder, leverages source domain priors to guide the soft

assignment of target domain nodes, models the class semantic structure of graph data, and further constructs homogeneity-enhanced views with higher semantic reliability. Meanwhile, this module jointly optimizes node representations and class prototypes by combining prototype contrastive loss and prototype clustering loss, elevating local view consistency to class semantic consistency and providing reliable semantic anchors for subsequent class prototype alignment. The multi-level alignment module mitigates feature distribution variations caused by covariate shift by aligning the first-order and second-order statistics of embedding representations between the source and target domains during the test-time training phase, and addresses class semantic variations induced by concept shift by aligning the class prototype distributions of the source and target domains. To verify the effectiveness of the proposed method, this work conducts comprehensive experimental evaluations on multiple benchmark datasets for graph distribution shift, covering synthetic shift, temporal evolution, and cross-domain shift scenarios. Experimental results demonstrate that ProGTA significantly outperforms representative methods such as Tent, EERM, and GTrans in terms of out-of-distribution generalization performance. Furthermore, this study analyzes the consistency of adversarial attacks and concept shift in disrupting the model's decision boundaries from the domain adaptation perspective, formulates the adversarial attack problem as a special case of out-of-distribution shift, and confirms that the proposed method can simultaneously improve the model's out-of-distribution generalization and adversarial robustness.

Keywords: graph neural networks; out-of-distribution shift; test-time adaptation; graph contrastive learning; model generalization; adversarial robustness

Foundation Item(s): National Natural Science Foundation of China (No.62376047); The Key Project of Innovation and Development Joint Fund of Chongqing Natural Science Foundation (No.CSTB2023NSCQ-LZX0003); The Key Project of Science and Technology Research Program of Chongqing Education Commission (No.KJZD-K202300603, No.KJZD-K202500604)

0 引言

图数据广泛存在于社交网络分析^[1]、分子建模^[2]、药物设计^[3]、知识图谱构建^[4]等领域。图神经网络通过联合建模图拓扑结构与节点属性,在节点分类^[5]、链路预测^[6]等任务中表现优异,已成为图学习领域的研究热点。然而,现有GNNs(Graph Neural Networks)通常基于训练与测试数据分布一致的假设^[7]。在开放环境中,分布偏移广泛存在,即测试数据分布相较于训练数据发生系统性变化,导致模型在测试阶段性能显著下降,因此亟需提升其分布外(Out-Of-Distribution, OOD)泛化能力^[8]。从统计学习视角看,分布偏移可细分为两类具有不同机理的典型形式。其一为协变量偏移(covariate shift)^[9],指输入特征的边缘分布发生变化,而特征与标签之间的条件关系保持不变。在图数据中,这类偏移通常表现为节点特征分布或图结构统计特性随时间、场景等发生系统性变化。其二是概念偏移(concept shift)^[10],指特征与标签之间的映射关系发生变化,而输入特征本身的边缘分布相对稳定。在图学习场景中,这种偏移往往表现为相同或相似的结构模式在不同时间或环境中对应不同的类别标签^[11-12]。在实际学习任务中,这两类偏移常常交织出现,共同削弱了GNNs在动态复杂环境中的泛化能力。

针对图数据的分布偏移问题,现有研究大体可归纳为三类解决范式。第一类为数据层面的方法,通过在训练阶段进行数据增强构造多样化样本以模拟潜在分布变化,提升模型对偏移的适应能力。典型策略

包括特征扰动^[13]、结构扰动^[14]以及多视图生成^[15]等,旨在通过构造不同的“伪环境”使模型学习更具不变性的表征。第二类为模型层面的方法,通过在GNNs架构中融入先验知识或因果假设,在架构层面增强模型的表达与泛化能力,代表方法包括Disen-GCN^[16]、OOD-GNN^[17]。第三类为训练策略层面的方法,通过设计新的学习目标或优化机制,引导模型学习跨环境稳定的特征与决策规则。其中,域适应方法较为典型,通过借助对抗式域判别器或最大均值差异(Maximum Mean Discrepancy, MMD)等统计对齐手段,使源域和目标域分布在隐空间中趋于一致,代表工作有UDA-GCN^[18]、DAGCN^[19]等。近年来,测试时自适应(Test-Time Adaptation, TTA)^[20]作为一种新兴范式,在测试阶段利用未标注测试样本对模型进行在线自适应更新,使其在部署过程中动态调整表征,以应对分布偏移带来的性能退化。

尽管上述方法在一定程度上缓解了分布偏移带来的性能退化问题,但仍面临诸多挑战。首先,基于数据增强的方法依赖人为设计的扰动策略,其覆盖的偏移空间有限,难以刻画真实复杂环境中出现的多样性结构变化。其次,基于模型的方法虽能从表示层面提升泛化能力,但往往会引入较高的模型复杂度,增加训练与部署成本。再次,基于训练策略的方法大多隐含环境稳定或条件分布一致的假设,当测试数据随时间或场景动态演化时,难以保持类别语义边界的持续稳定。此外,这三类方法普遍将分布偏移视为单一

整体进行处理,缺乏对协变量偏移与概念偏移的区分建模,导致难以针对性地消除两种偏移的影响。

针对上述问题,本工作从协变量偏移与概念偏移的特性出发,提出一种原型引导的图神经网络测试时自适应方法 ProGTA (Prototype-guided Test-time Adaptation)。针对协变量偏移所表现出的整体特征分布变化,ProGTA 在测试时自适应阶段对源域与目标域的全局特征嵌入进行对齐,从而缩小两者在特征空间中的表示差异。针对概念偏移所导致的类别条件结构及决策边界变化,ProGTA 在图对比学习框架中引入可学习的类别原型层,对类别条件关系进行显式建模,在测试时自适应过程中持续更新原型,并进一步对齐源域与目标域的类别原型分布,从而间接校正类别条件关系,实现对决策边界的自适应调整。通过上述机制,ProGTA 在动态开放环境中提升了 GNNs 的分布外泛化能力与鲁棒性。此外,本工作分析了对抗攻击与概念偏移在扰动决策边界方面的内在一致性,将对抗攻击视为分布偏移的一种特例。实验结果表明,原型分布的对齐本质上是对决策边界的动态优化。

1 相关工作

1.1 图分布外泛化方法

1.1.1 基于数据增强的图分布外泛化方法

该类方法从数据层面出发,在训练阶段通过数据增强构造多样化的样本,扩展训练分布的覆盖范围,从而提升模型对潜在测试分布偏移的适应能力。根据增强对象的不同,主流增强方式大致可分为特征增强、结构增强和混合增强三类。其中,特征增强聚焦于节点特征维度,旨在降低模型对特定邻域或特征模式的过度依赖。例如,GRAND^[21]通过随机丢弃节点特征并在图上传播扰动,使模型学习更加鲁棒的节点表示;FLAG^[13]利用基于梯度的对抗扰动在特征空间进行迭代增强,使模型在对抗性特征扰动下仍能保持稳定预测;LA-GNN^[22]通过预训练生成模型刻画邻域特征的条件分布,并生成补充特征以增强模型对未见节点的泛化能力。结构增强侧重于对图拓扑结构进行编辑或重构,以生成更加多样的图实例。例如,GAug^[23]通过可微分的边预测器对输入图的边进行添加或删除,增强类内连接,抑制类间连接,从而提升节点分类任务的泛化能力;MH-Aug^[24]采用马尔科夫蒙特卡罗 (Markov Chain Monte Carlo, MCMC) 采样机制精细控制增强强度,避免过度扰动带来的性能退化;KDGA^[14]结合知识蒸馏,在增强图与原始测试图之间进行知识迁移,缓解增强分布与真实分布之间的偏移。混合增强综合特征与结构两种策略,以获得更全面的增强效果。例如,GraphCL^[15]提出了节点丢

弃、边扰动、属性掩码和子图采样等多种增强方式,并通过对比学习最大化不同视图表示的一致性,在图分类任务中展现出较好的 OOD 泛化能力;GRE^[25]通过环境子图替换生成增强样本,以支持对因果图图的精确识别;MARIO^[26]结合表示一致性约束与条件互信息正则化,抑制与任务无关的噪声特征。

1.1.2 基于模型的图分布外泛化方法

该类方法从模型架构设计出发,将先验知识或因果假设融入 GNNs 架构中,直接学习具有 OOD 泛化能力的图表示。按照建模思路的不同可分为解耦模型和因果模型两类。其中,解耦模型旨在分离因果相关因素与虚假相关因素,将图表示解耦为统计上尽可能独立的潜在因素,每个因素对应数据生成过程中的一个独立变化维度。例如,DisenGCN^[16]首次在图上提出节点表示解耦思想,通过多通道卷积层实现类似邻域路由的机制,迭代识别潜在因素并将邻居分配到不同通道;NED-VAE^[27]在变分自编码器框架下联合优化节点与边的编码器和解码器,捕捉节点和边联合的独立潜在因素;DGCL^[28]引入多通道消息传递机制,在每个因素子空间内分别进行因素级对比学习,保证每个解耦因素只关注特定语义维度。因果图模型利用因果推理思想刻画图数据中的稳定因果关系,显式消除混杂因素造成的虚假相关,从而实现跨环境的鲁棒预测。例如,OOD-GNN^[17]基于混杂平衡理论,引入非线性去相关操作,并利用随机傅里叶特征实现线性缩放,削弱因果与非因果表示之间的统计依赖,在图大小、节点特征和结构偏移等多种 OOD 基准上表现出色;StableGNN^[29]通过可微分的图池化层提取子图级高阶变量,对簇级嵌入进行对齐,以提升图级任务下的 OOD 泛化;DGNN^[30]面向节点级任务,联合优化去相关正则化和加权 GNNs,通过学习样本权重方式抑制与标签无关的虚假相关样本的学习。

1.1.3 基于训练策略的图分布外泛化方法

该类方法通过定制化的优化目标与约束来引导模型的训练过程,以提升模型在未知分布下的泛化能力,而不依赖特定模型架构。典型的不变性方法如 EERM^[31],通过构造虚拟环境并最大化不同环境间的风险差异,引导模型在优化过程中主动排除与特定环境相关的噪声特征,从而提取对分布变化具有稳健性的因果不变表示。然而,后续研究指出,仅依赖环境增强学习不变特征,在假设空间受限或环境数量有限的条件下,泛化效果依然受限^[32]。因此,GIL^[33]等工作尝试在子图级别识别稳定模式,即不变性表征,并通过多模块联合优化获得可迁移的图表示,以强化模型在多域环境中的结构一致性与语义鲁棒性。此外,域适应 (Domain Adaptation, DA) 理论通过缩小源域与

目标域之间的分布差异可提升跨域泛化能力,相关方法通常在训练阶段通过统计匹配或对抗对齐实现跨域表征一致性。例如, CORAL^[34]通过对齐源域和目标域嵌入空间中的一阶和二阶统计特征以逼近整体分布,对抗式方法则通过借助域判别器与特征提取器的对抗博弈学习“不可分域”的隐空间表示,并结合结构保持与图读出模块以兼顾拓扑信息与节点属性。然而,上述方法普遍依赖目标域先验与静态统计假设,难以适应开放环境下不断演化的分布。这一局限推动了测试时自适应(Test-Time Adaptation, TTA)策略的发展。TTA 通过在测试阶段利用测试数据对模型进行在线更新,使其在不访问源域数据的情况下实现部分自适应。典型方法如 Tent^[35]以“预测熵最小化”为适应目标,通过降低模型对当前测试样本的预测熵提高判别置信度,从而减弱分布偏移带来的性能下降; GTrans^[36]提出在测试阶段针对图数据的节点特征与结构采用无参数对比代理损失进行适配,而无需修改模型参数,显著提升了模型在分布偏移场景下的泛化能力; GT3^[37]面向图分类任务设计了融合局部与全局图信息的测试时训练(Test-Time Training, TTT)框架,通过构建全局对比学习与局部对比学习的分层自监督学习任务来对测试数据进行适应; HomeTTT^[38]通过动态调整图结构与节点特征,进一步提升了模型在分布外环境下的泛化能力。

2 问题定义

定义 1 协变量偏移。在测试数据中,输入特征的边缘分布发生变化,但特征与标签之间的条件关系保持不变,可表示为

$$P_T(\mathbf{x}) \neq P_S(\mathbf{x}), P_T(y|\mathbf{x}) = P_S(y|\mathbf{x}) \quad (1)$$

其中, $P_T(\mathbf{x})$ 和 $P_S(\mathbf{x})$ 分别表示目标域和源域的输入分布, $P_T(y|\mathbf{x})$ 和 $P_S(y|\mathbf{x})$ 表示给定输入特征 $\mathbf{x}$ 的标签条件分布。协变量偏移的挑战在于模型在测试时遇到了训练时未见过的新特征分布。

定义 2 概念偏移。在测试数据中,特征与标签之间的决策关系发生变化,导致模型在目标域上难以泛化。可表示为

$$P_T(\mathbf{x}) = P_S(\mathbf{x}), P_T(y|\mathbf{x}) \neq P_S(y|\mathbf{x}) \quad (2)$$

概念偏移下的挑战是模型在训练时基于源域特征学到的条件关系可能在目标域不再适用。

为了解决以上两类偏移问题,本工作基于测试时自适应框架,旨在通过在线更新模型参数来适应目标域的分布变化。具体而言,给定一个带有节点特征与标签信息的图 $G=(V, E, \mathbf{X}, Y)$, 其中 $V=\{v_1, v_2, \dots, v_N\}$ 为节点集, $E=\{e_1, e_2, \dots, e_M\}$ 为边集, $\mathbf{X} \in \mathbb{R}^{N \times d}$ 为节点特征矩阵, $Y \in \{1, 2, \dots, C\}^N$ 为节点标签集合。在训练阶

段,仅可获得源域图 $G_s=(V_s, E_s, \mathbf{X}_s, Y_s)$ 用于监督训练,以训练得到参数为 θ_s 的模型 $f_{\theta_s}(\cdot)$ 。在测试阶段,模型需在无标签的目标域图 $G_t=(V_t, E_t, \mathbf{X}_t)$ 上进行推理,且测试数据分布与训练分布存在显著差异。该问题可形式化为在无标签条件下最小化目标域期望风险:

$$\mathcal{R}_t(f_{\theta_t}) = \mathbb{E}_{(\mathbf{x}, y) \sim P_t} [\ell(f_{\theta_t}(\mathbf{x}), y)] \quad (3)$$

其中, $\ell(\cdot)$ 为分类损失函数, θ_t 为更新后的模型参数。

3 方法描述

3.1 框架描述

本节详细介绍所提出的分布偏移场景下 GNNs 测试时自适应方法 ProGTA。ProGTA 将整体流程划分为三个阶段:训练阶段、测试时训练阶段以及测试阶段。①训练阶段。在源域图上联合优化主任务损失与基于原型聚类的图对比学习损失(详见 3.2),以学习共享编码器和初始类别原型。通过 GradNorm 自适应平衡两任务梯度,得到既具判别性又具语义聚合性的节点表示,并将各类别原型作为后续测试时自适应的语义先验。②测试时训练阶段。冻结主任务分类头和自监督头参数,仅更新共享编码器。利用源域原型对目标域节点进行软分配,构造原型引导的伪标签视图,并在此基础上执行自监督对比学习优化。同时,引入多层次对齐机制,一方面对齐源域和目标域嵌入的均值与协方差,缓解协变量偏移;另一方面对齐两域类别原型分布,减弱概念偏移对决策边界的影响。③测试阶段。使用更新后的编码器和主任务的分类头,对连续到来的测试节点进行预测。ProGTA 方法的整体框架如图 1 所示。

3.1.1 训练阶段

在训练阶段,模型在源域图 $G_s=(V_s, E_s, \mathbf{X}_s, Y_s)$ 上同时执行主任务和原型图对比学习任务(详见 3.2),以联合优化模型的参数 $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 。采用模型的前 l 层作为共享编码器 $\theta_e = \{\theta_1, \theta_2, \dots, \theta_l\}$, 主任务的模型参数表示为 (θ_e, θ_m) , 其中 $\theta_m = (\theta_{l+1}, \theta_{l+2}, \dots, \theta_n)$ 为主任务头。原型图对比学习任务的模型参数表示为 (θ_e, θ_s) , 其中 $\theta_s = (\theta'_{l+1}, \theta'_{l+2}, \dots, \theta'_n)$ 为对比学习头。两任务输出经反向传播同步优化参数 θ 和 θ_s , 实现主任务监督损失 $\mathcal{L}_m(\cdot)$ 和对比学习的自监督损失 $\mathcal{L}_s(\cdot)$ 的协同训练。为避免某一任务主导优化过程,本方法采用 GradNorm 策略对两种任务损失的权重进行动态调整,保持多任务训练的梯度平衡,该阶段的总损失为

$$\mathcal{L}_{\text{Train}} = w_1 \mathcal{L}_m + w_2 \mathcal{L}_s \quad (4)$$

其中, w_1 和 w_2 分别为动态更新的权重系数,由 GradNorm 基于两任务梯度范数自适应调节。

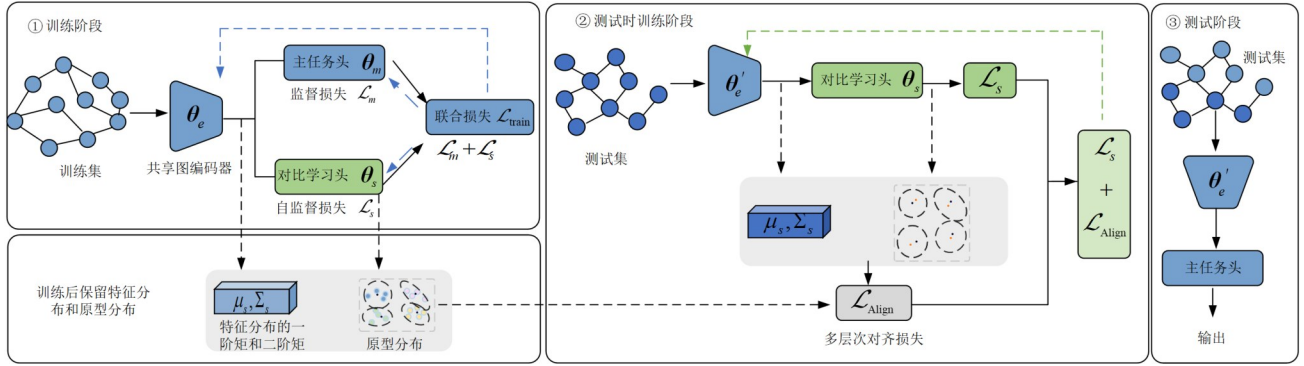


图1 基于原型引导的图神经网络测试时自适应方法框架

Figure 1 Framework of prototype-guided graph neural networks for test-time adaptation

3.1.2 测试时训练阶段

针对目标域图 G_t , ProGTA 在测试阶段采用测试时训练范式对模型进行部分更新。具体做法是在保持主任务头参数 θ_m 和对比学习头 θ_s 不变的前提下, 仅对共享编码器参数 θ_e 进行优化, 即冻结任务特定参数, 专注于调整表征空间以适配目标域分布。该阶段以自监督损失 $\mathcal{L}_s(\cdot)$ 为核心优化目标, 同时引入基于域适应的多层次对齐机制(详见 3.3), 在全局统计层面对齐源域与目标域的嵌入分布, 在类别条件层面对齐两域的原型分布, 从而同时缓解协变量偏移和概念偏移。综合上述两部分, 测试时训练阶段的总优化目标如下:

$$\mathcal{L}_{TTA} = \mathcal{L}_s + \lambda \mathcal{L}_{Align} \quad (5)$$

其中, $\mathcal{L}_{Align}$ 为多层次对齐损失, 见式(17), λ 为控制对齐强度的权重超参数。

3.1.3 测试阶段

完成测试时训练后, 可得到更新后的共享编码器

θ'_e 。此时不再进行额外的参数更新, 模型 (θ'_e, θ_m) 直接用于对后续到来的目标域节点进行预测。

3.2 基于原型聚类的图对比学习

在测试时自适应框架中, 自监督任务的设计质量直接决定了模型能否在目标域上有效缓解分布偏移, 同时为了解决概念偏移造成的语义变化, 本工作提出基于原型聚类的图对比学习模块(Prototypical Graph Contrastive Learning, PGCL), 通过显式引入一组可学习的类别原型, 将传统图对比学习中的局部视图一致性提升为原型语义一致性。PGCL 的核心思想有两点: 一是利用源域训练得到的类别原型对目标域节点进行软聚类, 为目标域节点构造更稳定的伪标签, 从而进行更具质量的同质性视图增强。二是通过原型对比损失与原型聚类损失联合优化, 使节点嵌入围绕各自类别原型形成类内紧致、类间分离的结构, 挖掘目标域的潜在类别语义结构, 为后续多层次对齐提供可靠的语义锚点, 具体方法如图 2 所示。

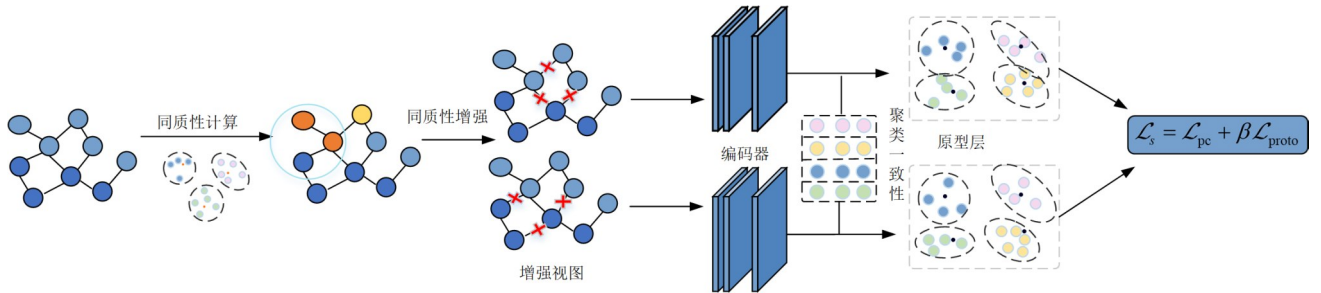


图2 基于原型聚类的图对比学习框架图

Figure 2 Framework of graph contrastive learning based on prototype clustering

3.2.1 基于原型伪标签的同质性增强

现有图对比学习方法在构造增强视图时, 多采用对节点特征或图结构进行均匀的随机扰动, 例如随机删除添加边、随机掩蔽节点特征等。这类与语义无关的增强策略往往忽略图的同质性与类别结构信息, 可能破坏关键拓扑模式, 导致生成的视图与真实语义偏

离, 从而削弱对比学习对任务相关特征的刻画能力。为此, PGCL 引入类别原型先验, 在测试阶段利用源域学习到的类别原型对目标域节点进行软聚类, 从而在无标签条件下估计边的语义同质性, 进而指导结构增强过程。具体地, 假设在训练阶段得到 K 个原型向量(对应 K 个类别), 组成原型矩阵 $\mathbf{P} \in \mathbb{R}^{K \times d}$, 其中 d 为嵌

入维度,对目标域中每个节点 v_i , 记其编码器输出的嵌入为 $\mathbf{h}_i \in \mathbb{R}^d$, 则定义其相对于各类别原型的软分配概率为

$$q_{i,k} = \frac{\exp(\text{sim}(\mathbf{h}_i, \mathbf{p}_k)/\tau_1)}{\sum_{k'=1}^K \exp(\text{sim}(\mathbf{h}_i, \mathbf{p}_{k'})/\tau_1)}, \quad k=1, 2, \dots, K \quad (6)$$

其中, $\text{sim}(\cdot, \cdot)$ 表示余弦相似度, τ_1 为温度系数。由此, 可将节点 v_i 的伪标签定义为

$$\hat{y}_i = \arg\max_k q_{i,k} \quad (7)$$

即每个节点更倾向于其所属软概率最大的原型簇。通过引入源域原型先验, 将源域类别级知识先验以伪标签的形式显式迁移至目标域, 因此可以在存在分布偏移时提供更稳定的语义锚定。

在得到伪标签后, 为每条边 (i, j) 构造其同质性分数以度量其语义可靠程度。记节点 v_i 的邻居集合为 $\mathcal{N}(i)$, 定义其同质性分数 η_i 如下:

$$\eta_i = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \mathbb{I}(\hat{y}_j = \hat{y}_i) \quad (8)$$

其中, $\mathbb{I}(\cdot)$ 为指示函数。边 (i, j) 的同质性分数表示为语义是否一致, 并结合其局部邻域内的一致性程度进行加权。在具体增强策略中, PGCL 根据同质性分数 η_i 为每条边分配保留或扰动的概率, 对高同质性边的保留概率高, 以保护类别内的关键结构, 对低同质性边的扰动概率更大, 用以削弱跨类噪声连接。

3.2.2 基于原型聚类的表示对比

传统图对比学习多采用实例级对比范式, 将同一节点在不同增强视图下的表示视为正样本, 将其他节点一律视为负样本。该对比学习范式本身就高度依赖大规模负采样和复杂采样策略, 而在仅有小批量无标签样本的测试时自适应场景中, 这种依赖被进一步放大, 难以稳定构造可靠的对比目标。更重要的是, 由于缺乏类别级语义约束, 在概念偏移下, 语义相近的同类节点很容易被误当作负样本, 不仅引入噪声梯度, 还会在测试时自适应过程中扰动决策边界, 从而削弱测试时自适应的稳定性。为缓解上述问题, 本工作提出一种基于原型聚类的表征对比机制, 将类别原型作为语义锚点建模全局类别结构, 并通过跨视图原型分配的一致性约束, 提升图对比学习的语义稳定性与鲁棒性。

具体而言, 本工作首先引入一组可训练的类别原型向量 $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K\} \in \mathbb{R}^{K \times d}$ 。其中 $\mathbf{p}_k$ 表示第 k 类的语义中心, 在训练与测试时训练阶段作为模型参数共同更新。为提供合理的语义起点, $\mathbf{P}$ 初始化为源域中各类别节点嵌入的类均值。同时将原型引入对比目标, 对节点嵌入 $\mathbf{h}_i$ 与其增强视图嵌入 $\mathbf{h}'_i$ 组成正样本对, 除所属原型之外的其他原型作为负样本, 定义

原型对比损失:

$$\mathcal{L}_{\text{pc}} = -\log \frac{\exp(\text{sim}(\mathbf{h}_i, \mathbf{h}'_i)/\tau_2)}{\exp(\text{sim}(\mathbf{h}_i, \mathbf{h}'_i)/\tau_2) + \sum_{k \neq \hat{c}(i)} \exp(\text{sim}(\mathbf{h}_i, \mathbf{p}_k)/\tau_2)} \quad (9)$$

其中, $\text{sim}(\cdot, \cdot)$ 表示余弦相似度, $\hat{c}(i)$ 是节点 i 的分配原型。该损失一方面鼓励节点在不同增强视图下保持一致性的原型归属, 另一方面通过显式远离异类原型, 为后续原型对齐提供更稳定的类别语义锚点。

此外, 为进一步强化原型聚类自身的类内紧致性和类间分离性, 提升原型自身的判别能力, 进而引入原型聚类一致性损失:

$$\mathcal{L}_{\text{proto}} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{h}_i - \mathbf{p}_{\hat{c}(i)}\|^2 + \lambda_c \sum_{k \neq l} \max(0, m - \|\mathbf{p}_k - \mathbf{p}_l\|) \quad (10)$$

其中: 第一项通过最小化节点嵌入与其分配原型之间的欧氏距离, 强化类内紧致; 第二项通过调整边距 m 约束不同原型之间的最小间隔, 最大化类间距离, 防止原型坍塌。综合对比项与聚类项, 测试时训练阶段原型聚类图对比学习总损失为

$$\mathcal{L}_s = \mathcal{L}_{\text{pc}} + \beta \mathcal{L}_{\text{proto}} \quad (11)$$

其中, β 为平衡两项损失的权重超参数。

3.3 基于域适应的多层次对齐

尽管基于原型聚类的自监督对比学习能够在一定程度上提升模型对目标域的适应能力, 但当源域与目标域存在显著的分布偏移时, 仅依赖自监督信号往往不足以刻画复杂的域间偏移关系, 仍可能导致泛化性能退化^[39]。为此, 本工作基于域适应理论^[34]提出了一种多层次的域对齐机制, 在测试时训练阶段通过全局特征和类别原型对齐, 缓解了协变量偏移与概念偏移对模型性能的影响。

经典的域适应误差上界^[34]指出, 对于任意假设 h , 目标域误差满足:

$$\varepsilon_t(h) \leq \varepsilon_s(h) + \frac{1}{\gamma} d_{\mathcal{H}\Delta\mathcal{H}}(D_s, D_t) + \lambda \quad (12)$$

其中: $\varepsilon_s(h)$ 和 $\varepsilon_t(h)$ 分别为源域和目标域的期望误差; $d_{\mathcal{H}\Delta\mathcal{H}}(D_s, D_t)$ 为两域分布差异的距离度量, 反映了源域和目标域的协变量偏移带来的影响; λ 为联合最优假设误差, 反映了源域和目标域的概念偏移所带来的不可约误差。该界限表明降低目标域泛化误差的关键在于缩小源域和目标域的协变量偏移和概念偏移。

(1) 为缓解协变量偏移, 本工作在测试时训练阶段对源域与目标域的特征嵌入分布进行对齐。具体而言, 计算源域与目标域节点嵌入分别为 $\{\mathbf{h}'_s\}_{i=1}^{N_s}$ 和 $\{\mathbf{h}'_t\}_{i=1}^{N_t}$, 然后计算节点嵌入的均值向量与协方差矩阵为

$$\mathbf{u}_d = \frac{1}{N_d} \sum_{i=1}^{N_d} \mathbf{h}_i^d, \quad (13)$$

$$\sum_d = \frac{1}{N_d - 1} \sum_{i=1}^{N_d} (\mathbf{h}_i^d - \mathbf{u}_d)(\mathbf{h}_i^d - \mathbf{u}_d)^T, \quad d \in \{s, t\}$$

其中,源域统计量在训练阶段时预先计算并进行缓存,因此测试时无需再访问源域样本。目标域统计量 $(\mathbf{u}_t, \Sigma_t)$ 则是在测试时自适应过程中进行在线更新。借鉴 CORAL^[34]等域适应方法的思想,以对两域的一阶与二阶统计量施加对齐约束,以最小化两域统计量之间的差异,得到全局统计对齐损失:

$$\mathcal{L}_{\text{global}} = \|\mathbf{u}_t - \mathbf{u}_s\|_F^2 + \|\Sigma_t - \Sigma_s\|_F^2 \quad (14)$$

其中, $\|\cdot\|_F$ 为 Frobenius 范数, $(\mathbf{u}_s, \Sigma_s)$ 和 $(\mathbf{u}_t, \Sigma_t)$ 分别为源域与目标域的均值和协方差统计量。该损失鼓励目标域嵌入的整体分布贴近源域,使模型在隐空间中学习到分布一致的表示。该对齐策略能够有效减少由特征分布偏移带来的偏差,从而降低协变量偏移对模型性能的影响。

(2) 为缓解概念偏移,本工作将源域中学习得到的类别原型视为稳定的语义锚点,并在测试时训练阶段约束目标域样本在隐空间中的聚类结构与源域保持一致。通过对齐源域与目标域的原型分布,引导目标域样本向对应的语义中心聚集,从而在无标签条件下隐式稳定决策边界,降低概念偏移的影响。具体而言,记源域与目标域在第 k 个类别原型上的聚类比例分别为 $\pi_s(k)$ 与 $\pi_t(k)$ 。其中, $\pi_s(k)$ 由源域训练后缓存的原型层参数确定,作为稳定的类别语义先验; $\pi_t(k)$ 则根据测试时训练阶段节点对原型的软分配结果计算确定,其定义如下:

$$\pi_t(k) = \frac{1}{N_t} \sum_{i=1}^{N_t} q_{ik}, \quad k = 1, 2, \dots, K \quad (15)$$

其中, q_{ik} 表示目标域节点 i 分配到第 k 个原型的概率。为了对齐源域与目标域的类别原型分布,选择通过对称 KL (Kullback-Leibler) 散度来最小化两域原型的分布差异,由此定义原型分布对齐损失为

$$\mathcal{L}_{\text{pa}} = \text{KL}(\pi_t \| \pi_s) + \text{KL}(\pi_s \| \pi_t) \quad (16)$$

其中,

$$\pi_s = [\pi_s(1), \pi_s(2), \dots, \pi_s(K)],$$

$$\pi_t = [\pi_t(1), \pi_t(2), \dots, \pi_t(K)].$$

该对齐约束一方面抑制目标域原型分布的剧烈变化,避免类别语义信息在适应过程中丢失,另一方面在类别条件分布层面将目标域与源域对齐到同一分布空间中,避免类别语义对模型的影响,从而有效缓解概念偏移对决策边界的负面影响。具体流程如图3所示,在测试时自适应前目标域概念偏移影响,导致模型对类别不可分,随着测试时自适应进行,目标域的类别原型

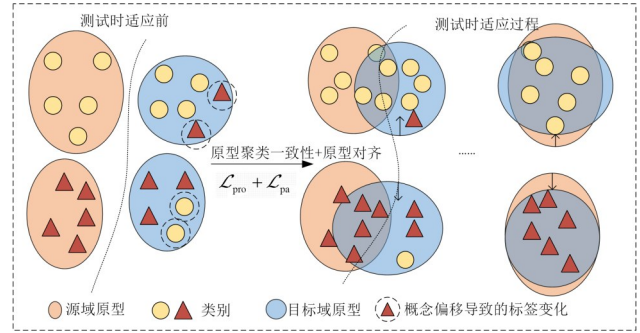


图3 概念偏移下原型分布对齐过程示意图

Figure 3 Schematic diagram of prototype distribution alignment under concept shift

逐渐可分且与源域原型在语义空间中逐渐接近。

综合全局嵌入分布对齐与类别原型分布对齐,本工作在测试时训练阶段构建多层次对齐损失函数:

$$\mathcal{L}_{\text{Align}} = \mathcal{L}_{\text{global}} + \lambda_c \mathcal{L}_{\text{pa}} \quad (17)$$

其中,超参数 λ_c 用于平衡两种对齐的占比。该联合优化策略使模型在自适应过程中既能够修正由特征分布变化引起的协变量偏移,又能够维持稳定的类别判别结构,从而在无标签条件下有效缓解协变量偏移和概念偏移的影响。

基于以上的原型图对比学习与多层次对齐策略,本工作提出基于原型引导的测试时自适应算法,详细算法流程如算法1所示。

4 实验及结果分析

4.1 面向分布偏移的泛化性

为了验证所提方法的分布外泛化效果,本节进行了全面的实证评估,首先概述实验设置,包括所使用的数据集和对比方法,以及参数设置和评价指标。然后呈现两个不同基准数据集上的实验结果并进行讨论分析。

4.1.1 实验设置

实验采用 EERM^[31]方法中使用的图 OOD 评测数据集以及 GOOD^[40]官方提出的图 OOD 评测数据集对所提出方法的性能进行系统评估。其中, EERM 评测数据集涵盖了多种典型的分布偏移场景,包括:(1)人工构造偏移的 Cora^[11]和 Amazon-Photo^[41]数据集,用于模拟节点特征和结构扰动等偏移;(2)时序演化的 OGB-Arxiv^[42]和 Elliptic^[43]数据集,用于刻画真实场景中的时间偏移;(3)跨域偏移的 Twitch-E^[44]数据集,用于评估模型在不同域之间的泛化能力。此外,为进一步系统性地分析模型在不同类型分布偏移下的行为差异, GOOD 评测数据集将偏移显式划分为协变量偏移和概念偏移,包括 GOOD-Cora、GOOD-Arxiv 和 GOOD-Twitch 数据集。上述数据集的统计信息见表1和表2。

算法 1 基于原型引导的测试时自适应算法

输入: 预训练编码器参数 θ_e , 原型矩阵 $\mathbf{P}_s$, 目标域图 $G_T=(V_t, E_t, \mathbf{X}_t)$,

TTA 步数 S , 源域统计量 $(\mathbf{u}_s, \sum_t)$

输出: 更新后的编码器参数 θ'_e

```

1.   $s=0$  /*初始化*/
2.  WHILE  $s \leq S$  /*判断测试时自适应步数*/
3.     $\mathbf{H} = f_{\theta_e}(G_T)$  /*计算节点嵌入*/
4.    For  $v_i \in V_t$  /*计算同质性分数*/
5.      根据式(6), 计算原型软分配概率
      
$$q_{i,k} = \frac{\exp(\text{sim}(\mathbf{h}_i, \mathbf{p}_k)/\tau_1)}{\sum_{k=1}^K \exp(\text{sim}(\mathbf{h}_i, \mathbf{p}_k)/\tau_1)}, k=1, 2, \dots, K$$

6.      根据式(7)和式(8), 计算节点伪标签和同质性分数
      
$$\hat{y}_i = \text{argmax}_k q_{i,k}$$

      
$$\eta_i = \text{HomophilyScore}(\hat{y}_i, \hat{y}_j)$$

7.      根据同质性分数进行数据增强
8.    End FOR
9.     $\mathbf{H}' = f_{\theta_e}(G'_T)$  /*增强视图嵌入*/
10.   根据式(11), 计算原型对比学习损失
11.    $\mathcal{L}_s = \mathcal{L}_{\text{pc}} + \lambda_c \mathcal{L}_{\text{proto}}$ 
12.   根据式(13), 计算目标域嵌入分布和原型分布
      
$$\mathbf{u}_t, \sum_t = \text{MenaCov}(\mathbf{H}'), \mathbf{P}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{P}_s s$$

13.   根据式(17), 计算多层次对齐损失
      
$$\mathcal{L}_{\text{Align}} = \mathcal{L}_{\text{globe}} + \lambda_c \mathcal{L}_{\text{pa}}$$

14.   根据式(5), 计算 TTA 总损失
      
$$\mathcal{L}_{\text{TTA}} = \mathcal{L}_s + \lambda \mathcal{L}_{\text{Align}}$$

15.    $s = s + 1$ 
16. End WHILE
17. RETURN  $\theta'_e$  /*输出更新后的编码器参数  $\theta'_e$ */
```

为了验证所提方法的有效性, 本工作使用 5 种基线方法与所提 ProGTA 方法进行对比。

经验风险最小化 (Empirical Risk Minimization,

ERM): 通过最小化模型在训练数据上的平均损失 (经验风险) 来训练模型的参数。

DropEdge^[45]: 通过在训练时随机删除或增加输入图中一定比例的边来实现数据增强, 从而增加输入数据的随机性和多样性, 旨在缓解训练时产生的过平滑问题, 并提升 GNN 模型对噪声和分布变化的鲁棒性。

Tent^[35]: 一种完全测试时自适应方法, 通过最小化模型对测试样本预测的香浓熵 (即置信度不确定性) 来在线优化模型的参数, 核心在于使用无标签测试数据逐步更新批归一化统计量和少量参数, 从而减少分布偏移带来的泛化能力下降问题。

EERM^[31]: 从不变学习视角处理图数据上的分布偏移, 基于输入图数据生成多个不同的图, 通过优化不同环境对应风险损失的均值和方差使模型学习跨域的不变表示。

GTrans^[36]: 一种测试时图变换框架, 通过在测试阶段动态调整和优化图数据来提升模型性能。该方法无需修改模型参数, 仅通过数据级变换缓解分布偏移。

本工作中骨干网络采用主流的图卷积网络 (Graph Convolutional Networks, GCN)^[46] 和图注意力网络 (Graph Attention Networks, GAT)^[47] 架构, 所有模型在 PyTorch 框架下实现, 训练使用 Adam 优化器 (学习率 0.01, 权重衰减 5×10^{-4})。实验使用准确率 (Accuracy, ACC) 和 F1 分数作为衡量模型分布外泛化能力的指标。所有实验重复 5 次, 取平均值作为最终结果。

4.1.2 实验结果与分析

表 3 汇总了在 EERM 评测数据集集合上的实验结果。实验结果表明, 所提出的 ProGTA 方法在不同数据集和不同骨干网络设置下取得了最优或次优性能, 且在多种分布偏移场景中表现出稳定的泛化性。

表 1 EERM 基准数据集的统计信息

Table 1 Statistics of the EERM benchmark datasets

数据集	分布偏移	节点数	连边数	特征数	分类数	划分方式	指标
Cora	人工偏移	2 703	5 278	1 443	10	Domain-Level	Accuracy
Amazon-Photo		7 650	119 081	745	10	Domain-Level	Accuracy
OGB-Arxiv	时间演化	169 343	1 166 243	128	40	Time-Aware	Accuracy
Elliptic		203 769	234 355	166	2	Time-Aware	F1 Score
Twitch-E	跨域偏移	1 912~9 498	31 299~153 138	3 170	2	Domain-Level	ROC-AUC

表 2 GOOD 基准数据集的统计信息

Table 2 Statistics of the GOOD benchmark datasets

数据集	偏移方式	节点数	连边数	特征数	分类数	划分方式	指标
GOOD-Cora	人工偏移	19 793	892 346	8 710	70	Word/Degree	Accuracy
GOOD-Arxiv	时间演化	169 343	1 166 243	128	40	Time/Degree	Accuracy
GOOD-Twitch	跨域偏移	34 120	892 346	128	2	Language	ROC-AUC

表3 EERM评测数据集下的实验结果

单位: %

Table 3 Experimental results on the EERM evaluation datasets

unit: %

方法	Cora		Amazon-Photo		Elliptic		OGB-Arxiv		Twitch-E	
	GCN	GAT	GCN	GAT	GCN	GAT	GCN	GAT	GCN	GAT
ERM	90.35	96.30	93.26	96.27	50.90	65.36	38.59	40.63	59.38	58.53
DropEdge	81.32	90.70	91.97	90.35	53.96	63.78	41.22	42.48	59.95	58.89
Tent	91.52	95.99	94.03	95.66	51.71	66.07	39.62	40.06	59.46	58.33
EERM	87.52	95.57	94.05	95.37	53.96	58.14	OOM	OOM	59.85	59.84
GTrans	94.13	96.67	94.13	96.42	55.88	66.43	41.32	43.76	60.42	58.59
ProGTA	97.12	98.32	96.27	98.13	57.20	68.36	43.32	45.86	64.29	62.83

注:加粗数据为最优结果。

具体地,在人工构造偏移与跨域偏移场景(Cora、Amazon-Photo和Twitch-E数据集)中,ProGTA相较于DropEdge和Tent等方法取得了显著提升,表明仅依赖随机结构扰动或预测熵最小化难以充分应对复杂的分布偏移。相比之下,ProGTA引入原型引导的同质性增强与测试时自适应机制,使模型更有针对性地适

应目标域特征和局部结构的变化,从而在人工偏移和跨域偏移场景下获得更优的泛化性能。此外,图4显示ProGTA在Cora数据集的多个OOD测试图上整体保持更高的准确率,并呈现更一致的变化趋势,表明其性能提升具有跨越测试图的稳定性,而非仅在个别图上出现偶然收益。

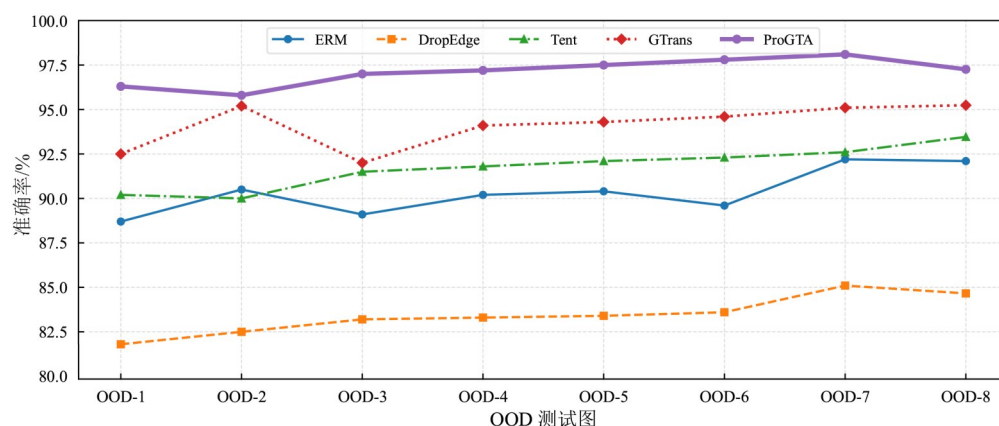


图4 Cora数据集的多个OOD测试图的实验结果

Figure 4 Experimental results of multiple OOD test graphs on the Cora dataset

在GOOD图OOD评测数据集上的实验结果如表4所示,分别以GCN和GAT作为骨干模型进行评测来验证所提方法的通用性。整体来看,ProGTA在不同数据集和不同偏移类型下均取得最优或次优性能,表现出较为稳定的泛化能力。为进一步直观展示不同方法在两类偏移设置下的性能差异,图5给出了在GOOD-Cora评测数据集上,各方法在协变量偏移与概念偏移场景下的实验结果。

(1)在协变量偏移下的性能表现,ProGTA相较于ERM、DropEdge和Tent等方法取得稳定提升,实验结果显示,协变量偏移主要导致特征分布和结构统计发生变化,使得模型特征不匹配导致性能下降。相比之下,ProGTA在测试时自适应过程中能够对目标域特征进行持续调整,从而有效缓解由特征分布变化引起的性能下降。

(2)在概念偏移下的性能表现,不同方法之间的性能差异更加显著。从表4可以看出,EERM和GTrans等方法在该类场景下的提升幅度有限,部分设置下甚至出现性能下降,说明仅依赖特征层面的适配难以有效应对类别语义变化。相比之下,ProGTA在概念偏移场景中表现出明显优势。实验结果表明,ProGTA在该类任务中能够保持更高且更稳定的预测性能,使模型在类别语义发生变化时仍能维持较好的判别效果。这一现象说明,在概念偏移下,引入类别层面的约束对模型泛化性能具有积极作用。

综合GOOD评测数据集下不同偏移类型的实验结果可以看出,ProGTA在协变量和概念偏移场景中均表现出较好的适应能力。其中,其在概念偏移场景下的优势尤为明显,进一步验证了该方法在复杂分布偏移设置下的稳定泛化性与有效性。

表 4 GOOD 评测数据集下的实验结果

单位: %

Table 4 Experimental results on the GOOD evaluation datasets

unit: %

	方法	GOOD-Cora				GOOD-Arxiv		GOOD-Twitch	
		协变量偏移		概念偏移		协变量偏移	概念偏移	协变量偏移	概念偏移
		word	degree	word	degree	time		language	
GCN	ERM	64.86	56.30	63.60	59.54	71.08	67.32	48.95	57.32
	DropEdge	60.32	50.74	63.80	61.23	71.12	66.25	47.21	55.48
	Tent	65.39	56.59	64.83	60.88	73.25	69.75	50.52	55.53
	EERM	61.98	56.88	63.09	58.38	OOM	OOM	52.43	57.37
	GTrans	66.58	58.82	65.32	60.67	OOM	OOM	53.62	58.95
	ProGTA	69.83	58.32	68.11	64.41	76.37	72.83	55.38	61.42
GAT	ERM	66.04	58.18	65.22	62.03	73.38	69.17	50.97	58.51
	DropEdge	60.21	51.43	63.63	60.62	72.18	68.63	48.16	60.72
	Tent	67.73	57.82	66.51	60.25	75.57	71.07	52.25	58.47
	EERM	62.96	56.87	65.87	63.33	OOM	OOM	53.57	58.81
	GTrans	66.33	60.35	65.16	64.31	OOM	OOM	54.78	59.69
	ProGTA	69.82	60.58	67.38	66.48	77.26	75.25	55.16	63.44

注:加粗数据为最优结果。

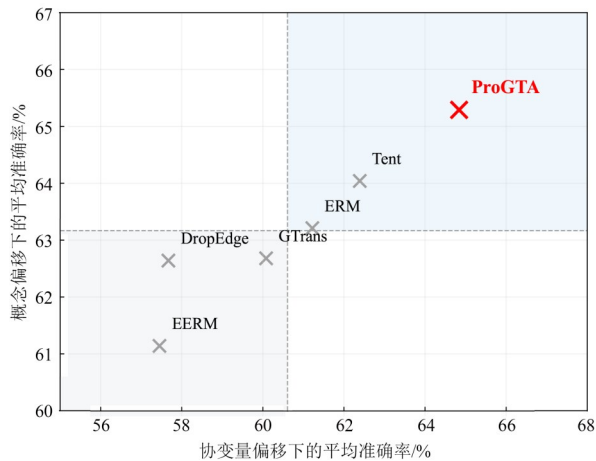


图 5 GOOD-Cora 数据集上的两种偏移下的实验结果

Figure 5 Experimental results under two types of shifts on the GOOD-Cora dataset

4.2 面向对抗攻击的鲁棒性

除常规的分布外泛化评估外,本节进一步考察所提方法在对抗攻击场景下的鲁棒性。已有研究表明,对抗攻击通过对节点特征或图结构施加精心设计的扰动,改变样本的分布特性,从而导致模型的预测性能显著下降。而从分布学习的视角来看,模型遭受的对抗扰动本质上是一种具有特定方向性的分布偏移。因此,对抗鲁棒性与分布外泛化能力在理论层面存在一定的内在关联,进而对抗场景可被视为检验模型分布外适应能力的一种特殊且更具挑战性的情形。

基于上述认识,本节在对抗攻击设置下对 ProGTA 的对抗鲁棒性进行实验评估。具体而言,采用 PR-BCD^[48] 攻击方法对 Cora、Citeseer 和 OGB-Arxiv

数据集的测试图进行扰动(即逃逸攻击),扰动比例从 5% 逐步增加至 25%,以模拟不同强度的对抗攻击。实验采用 GCN 作为骨干网络,并与 RobustGCN^[49]、SimPGCN^[50]、GCNJaccard^[51] 及 GTrans^[36] 等方法进行对比。

实验结果如图 6 所示。在不同扰动强度下,ProGTA 均能够保持相对更稳定的预测性能,尤其在高扰动率设置下,其性能下降幅度明显小于对比方法。该结果表明,ProGTA 在对抗攻击场景中具备较强的抗扰动能力,且随着攻击强度的增加,其对抗鲁棒性优势进一步凸显。实验现象表明,模型的鲁棒性提升可能源于其在测试时自适应过程中对决策边界的动态修正机制。对抗攻击通常通过施加定向扰动,使样本跨越原有判别边界并产生误分类,若模型能够在测试阶段根据扰动后的数据分布对判别结构进行自适应调整,则有助于缓解性能退化。在 ProGTA 中,原型对齐策略通过在特征空间中引入稳定的类别语义中心,对样本的判别区域形成显式约束,在一定程度上压缩类别间的模糊区域,促使决策边界向更具区分性的方向调整。这种对判别结构的稳定与校正作用,有助于削弱对抗扰动对预测结果的影响,从而提升模型的对抗鲁棒性。

4.3 消融实验

为进一步验证 ProGTA 各组成模块的必要性与相对贡献,本节在 Cora、OGB-Arxiv、Elliptic、Amazon-Photo 和 GOOD-Cora 这五个 OOD 数据集上进行了系统的消融实验,其余设置与有效性实验保持一致,消融设置如下。

(1) TTT: 移除原型聚类图对比学习模块和多层次对齐模块,退化为标准测试时训练方法。

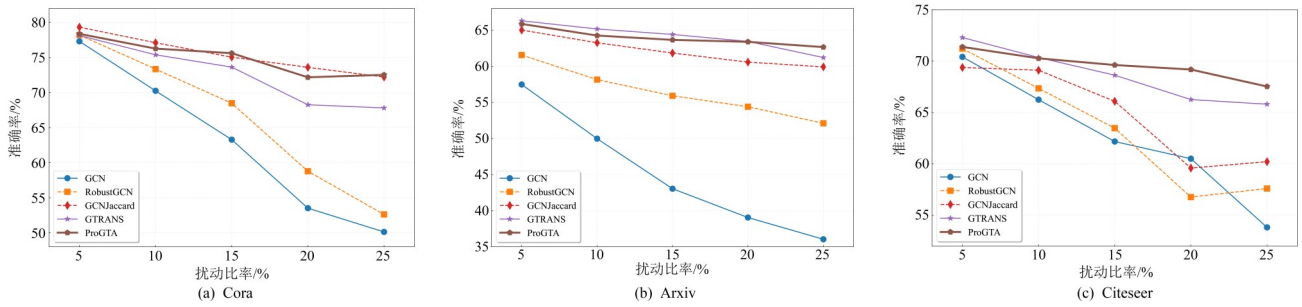


图6 不同扰动比率下各方法的性能表现

Figure 6 Performance of various methods under different perturbation ratios

(2) ProGTA-w/o Align: 移除多层次对齐模块, 保留原型聚类图对比学习模块。

(3) ProGTA-w/o Pro-Align: 移除原型对齐模块, 保留原型聚类图对比学习和特征分布对齐模块。

图7中的实验结果表明, 当移除原型聚类对比学习模块和多层次对齐模块后, 退化为标准 TTT 方法, 此时模型在所有 OOD 数据集上的性能均出现明显下降趋势, 该现象表明仅依赖实例级自监督信号难以稳

定类别语义结构, 容易在分布偏移条件下引发语义偏移。进一步地, 移除多层次对齐模块后模型退化为 ProGTA-w/o Align, 此时模型性能同样出现下降。其中移除原型分布对齐后退化为 ProGTA-w/o Pro-Align, 此时模型在概念偏移场景下的性能显著下降, 表明仅进行特征层面的适应难以应对类别条件分布变化, 这一结果验证了通过对齐源域和目标域的原型分布, 可以有效缓解概念偏移对模型判别能力的影响。

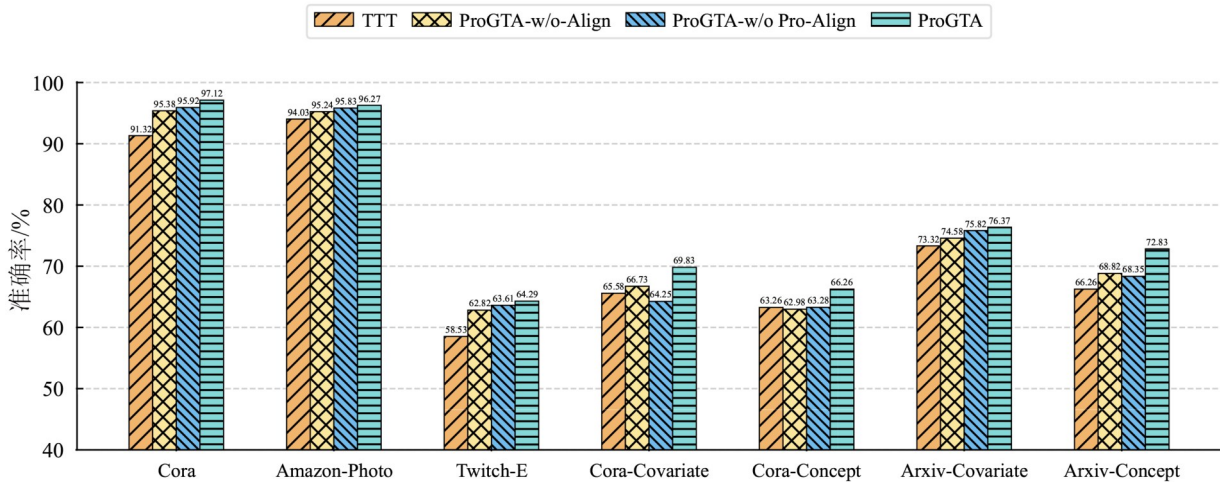


图7 消融各模块后的实验结果

Figure 7 Experimental results after ablating different modules

4.4 超参数分析

为评估 ProGTA 在不同偏移场景下的泛化性, 本节针对两个关键超参数进行敏感性分析。

(1) 测试时自适应时多层次对齐损失的权重系数 λ 。

(2) 测试时自适应更新的步数 S 。

实验在 Cora、OGB-Arxiv、GOOD-Cora (Covariate) 和 GOOD-Cora (Concept) 数据集上进行, 除被分析参数外, 其余参数均与主实验一致。

多层次对齐损失权重 λ 的实验结果如图8所示。具体地, 随着 λ 从0增大, 模型性能在各个数据集上都

呈现先提升后下降的趋势, 表明测试时自适应过程中需要原型对比自监督信号与多层次对齐约束之间保持平衡。进一步比较不同数据集可以发现, 在 OGB-Arxiv 上性能最优对应的 λ 略高于 Cora, 说明在分布偏移更复杂、变化更连续的场景中, 适当增强多层次对齐约束有助于提升模型的泛化能力。同时, 从 GOOD-Cora 数据集的协变量偏移和概念偏移结果来看, 概念偏移任务对 λ 更为敏感, 表明适度的对齐约束在稳定类别语义方面尤为重要。总体而言, 0.4~0.6 的取值在不同场景下表现出较好的稳定性。

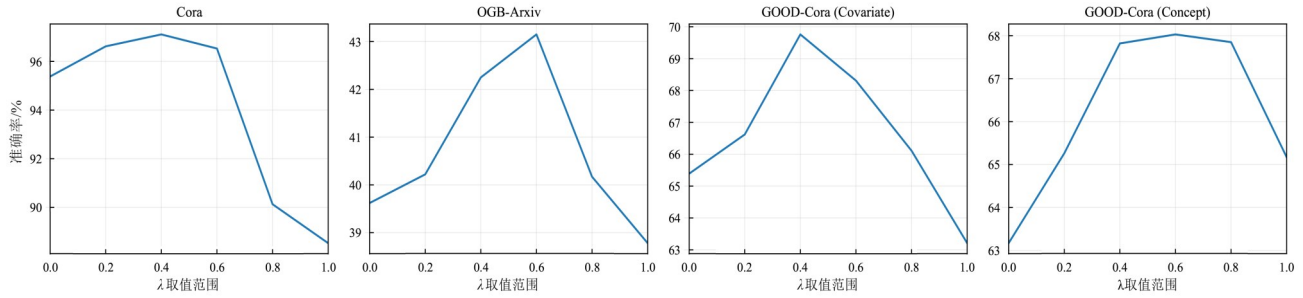
图8 多层次对齐权重 λ 的不同取值对测试时自适应的影响

Figure 8 Impact of different values of multi-level alignment weight values on test-time adaptation

测试时自适应步数 S 的实验结果如图9所示。具体地,随着TTA步数 S 增加,模型性能在各个数据集上均表现为先快速提升后趋于饱和甚至略有下降,说明少量测试时自适应更新即可显著提升模型

对目标域分布的适应能力,而过多更新可能导致对测试样本的过拟合。该结果表明,ProGTA在较小的测试时自适应步数即可获得稳定收益,无需进行大量迭代。

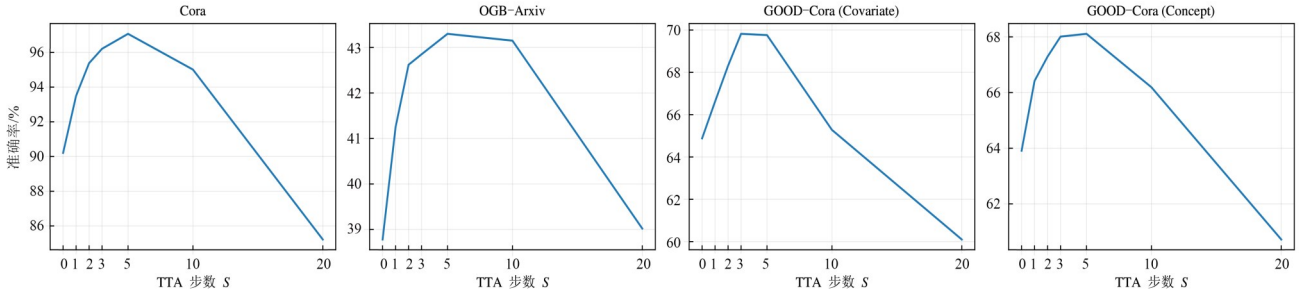
图9 测试时自适应步数 S 的取值对测试时自适应的影响

Figure 9 Impact of test-time adaptation step size on test-time adaptation

4.5 复杂度分析

为评估ProGTA的计算代价,本节对其主要模块的时间复杂度进行分析。设目标域图为 $G=(V,E)$,其中 $N=|V|$, $M=|E|$, d 为表示维度, L 为GNNs层数, C 为类别数, S 为测试时自适应步数。在每一步TTA更新中,模型首先进行一次GNNs前向传播,其复杂度为 $O(L \cdot M \cdot d)$ 。随后,原型聚类对比学习计算节点表示与原型之间的相似度构建损失,其复杂度为 $O(N \cdot C \cdot d)$ 。多层次对齐包含全局统计对齐与原型分布对齐,复杂度分别为 $O(N \cdot d^2)$ 与 $O(N \cdot C)$,因此,进行 S 步更新的复杂度为

$$O(L \cdot M \cdot d + N \cdot C \cdot d + N \cdot d^2 + N \cdot C) \quad (18)$$

为评估所提ProGTA方法的微调效率,本工作将其与EERM方法以及GTrans方法在时间开销和显存占用两方面进行了对比分析。所有实验均在相同硬件环境下完成(NVIDIA RTX 3090 GPU, 24 GB显存, Intel(R) Core(TM) i9-10940X CPU @ 3.30 GHz),并记录各模型完成一次完整微调所需的平均时间。表5给出了详细的实验结果。

实验结果表明,ProGTA在保证适应性能的同时,

表5 复杂度分析

Table 5 Complexity analysis

	微调时间/s			GPU显存占用/GB		
	Cora	Elliptic	Arxiv	Cora	Elliptic	Arxiv
EERM	32.2	617.9	—	2.5	12.8	>24
GTrans	3.4	6.2	10.7	1.4	1.5	1.3
ProGTA	7.5	21.2	35.3	1.2	2.3	5.6

显著提升了计算效率。在小规模图数据集(如Cora)上,其时间开销略高于GTrans;而在大规模图数据集(如Elliptic和OGB-Arxiv)上,由于原型引导机制能够有效减少冗余对比与迭代次数,时间开销远低于需对全模型进行重训练的EERM方法。在显存占用方面,ProGTA因仅需要进行少量自适应更新,显存消耗远低于EERM;而与仅对数据进行微调的GTrans相比,其显存占用略高。

5 结束语

本研究针对GNNs在分布外泛化能力方面的不足,提出了一种原型引导的测试时自适应框架ProGTA,以应对协变量偏移与概念偏移并存的复杂

场景。通过引入原型聚类对比学习与多层次分布对齐机制,通过充分的实验验证了所提方法有效提升了模型的分布外泛化性与对抗鲁棒性。未来工作将进一步面向开放环境下的实际应用需求,重点研究目标域存在未知类别时的测试时自适应问题,探索开放集条件下的原型建模与动态自适应机制,从而深化对GNNs在复杂分布外场景中泛化行为的理解,并为其在安全、开放环境中的部署提供更加系统的理论与方法支撑。同时,将相关思想拓展至图基础模型,探索其在大规模预训练与跨域迁移场景下的适应与鲁棒性提升机制。

参考文献

- [1] Fakhræi S, Foulds J, Shashanka M, et al. Collective spammer detection in evolving multi-relational social networks[C]//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1769-1778.
- [2] 徐敏, 胡春玲, 胡婷, 等. 基于序列与跨模态对齐的蛋白质功能预测模型[J]. 电子学报, 2025, 53(11): 4022-4034.
Xu Min, Hu Chunling, Hu Ting, et al. Sequence-based and cross-modal alignment model for protein function prediction[J]. Acta Electronica Sinica, 2025, 53(11): 4022-4034. (in Chinese)
- [3] Zhang Zaixi, Liu Qi. Learning subpocket prototypes for generalizable structure-based drug design[C]//Proceedings of the 40th International Conference on Machine Learning. Cambridge, MA: JMLR.org, 2023: 1735.
- [4] Wang Quan, Mao Zhendong, Wang Bin, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [5] Zhao Tianxiang, Zhang Xiang, Wang Suhang. GraphSMOTE: Imbalanced node classification on graphs with graph neural networks[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 833-841.
- [6] Zhang Muhan, Chen Yixin. Link prediction based on graph neural networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2018: 5171-5181.
- [7] Wu Man, Zheng Xin, Zhang Qin, et al. Graph learning under distribution shifts: A comprehensive survey on domain adaptation, out-of-distribution, and continual learning[PP/OL]. V2. arXiv (2024-03-07) [2025-11-12]. <https://arxiv.org/abs/2402.16374>.
- [8] Yu Junchi, Liang Jian, He Ran. Mind the label shift of augmentation-based graph OOD generalization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 11620-11630.
- [9] Sugiyama M, Nakajima S, Kashima H, et al. Direct importance estimation with model selection and its application to covariate shift adaptation[C]//Proceedings of the 21st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2007: 1433-1440.
- [10] Long Mingsheng, Cao Zhangjie, Wang Jianmin, et al. Conditional adversarial domain adaptation[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2018: 1647-1657.
- [11] Yang Zhilin, Cohen W W, Salakhutdinov R. Revisiting semi-supervised learning with graph embeddings[C]//Proceedings of the 33rd International Conference on Machine Learning. Cambridge, MA: JMLR.org, 2016: 40-48.
- [12] Yoshikai Y, Mizuno T, Nemoto S, et al. A novel molecule generative model of VAE combined with Transformer for unseen structure generation[PP/OL]. V2. arXiv (2024-04-05) [2025-12-18]. <https://arxiv.org/abs/2402.11950>.
- [13] Kong Kezhi, Li Guohao, Ding Mucong, et al. Robust optimization as data augmentation for large-scale graphs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 60-69.
- [14] Huang Yufei, Li S Z, LIN Haitao, et al. Knowledge distillation improves graph structure augmentation for graph neural networks[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2022: 068431-0858.
- [15] You Yuning, Chen Tianlong, Sui Yongduo, et al. Graph contrastive learning with augmentations[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 488.
- [16] Ma Jianxin, Cui Peng, Kuang Kun, et al. Disentangled graph convolutional networks[C]//Proceedings of the 36th International Conference on Machine Learning. Cambridge, MA: JMLR.org, 2019: 4212-4221.
- [17] Li Haoyang, Wang Xin, Zhang Ziwei, et al. OOD-GNN: Out-of-distribution generalized graph neural network[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7328-7340.
- [18] Wu Man, Pan Shirui, Zhou Chuan, et al. Unsupervised do-

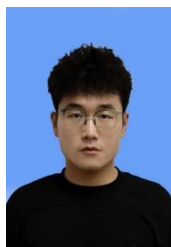
- main adaptive graph convolutional networks[C]//Proceedings of the Web Conference. New York: ACM, 2020: 1457-1467.
- [19] Li Tianfu, Zhao Zhibin, Sun Chuang, et al. Domain adversarial graph convolutional network for fault diagnosis under variable working conditions[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3515010.
- [20] Chen Guanxi, Zhang Jiying, Xiao Xi, et al. GraphTTA: Test time adaptation on graph neural networks[PP/OL]. V1. arXiv (2022-08-19)[2025-11-12]. <https://arxiv.org/abs/2208.09126>.
- [21] Feng Wenzheng, Zhang Jie, Dong Yuxiao, et al. Graph random neural networks for semi-supervised learning on graphs[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 1853.
- [22] Liu Songtao, Ying R, Dong Hanze, et al. Local augmentation for graph neural networks[C]//Proceedings of the 39th International Conference on Machine Learning. Baltimore, Cambridge, MA: JMLR.org, 2022: 14054-14072.
- [23] Zhao Tong, Liu Yozen, Neves L, et al. Data augmentation for graph neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11015-11023.
- [24] Park H, Lee S, Kim S, et al. Metropolis-hastings data augmentation for graph neural networks[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2021: 1453.
- [25] Liu Gang, Zhao Tong, Xu Jiabin, et al. Graph rationalization with environment-based augmentations[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM, 2022: 1069-1078.
- [26] Zhu Yun, Shi Haizhou, Zhang Zhenshuo, et al. MARIO: Model agnostic recipe for improving OOD generalization of graph contrastive learning[C]//Proceedings of the ACM Web Conference 2024. New York: ACM, 2024: 300-311.
- [27] Guo Xiaojie, Zhao Liang, Qin Zhao, et al. Interpretable deep graph generation with node-edge co-disentanglement[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 1697-1707.
- [28] Li Haoyang, Wang Xin, Zhang Ziwei, et al. Disentangled contrastive learning on graphs[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2021: 1674.
- [29] Fan Shaohua, Wang Xiao, Shi Chuan, et al. Generalizing graph neural networks on out-of-distribution graphs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(1): 322-337.
- [30] Fan Shaohua, Wang Xiao, Shi Chuan, et al. Debiased graph neural networks with agnostic label selection bias[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(4): 4411-4422.
- [31] Wu Qitian, Zhang Hengrui, Yan Junchi, et al. Handling distribution shifts on graphs: An invariance perspective[C/OL]//Proceedings of the 10th International Conference on Learning Representations, 2022. <https://openreview.net/forum?id=FQOC5u-1egI>.
- [32] Bian Yatao, Chen Yongqiang, Cheng J, et al. Does invariant graph learning via environment augmentation learn invariance [C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2023: 075280-3130.
- [33] Li Haoyang, Zhang Ziwei, Wang Xin, et al. Learning invariant graph representations for out-of-distribution generalization[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2022: 068431-0859.
- [34] Sun Baochen, Saenko K. Deep CORAL: Correlation alignment for deep domain adaptation[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2016: 443-450.
- [35] Wang Dequan, Shelhamer E, Liu Shaoteng, et al. Tent: Fully test-time adaptation by entropy minimization[C/OL]//Proceedings of the 9th International Conference on Learning Representations, 2021: 2928-2943. <https://openreview.net/forum?id=uXl3bZLkr3c>.
- [36] Jin Wei, Zhao Tong, Ding Jiayuan, et al. Empowering graph representation learning with test-time graph transformation[C/OL]//Proceedings of the 11th International Conference on Learning Representations, 2023: 32327-32354. <https://openreview.net/forum?id=Lnx15pr018>.
- [37] Wang Yiqi, Li Chaozhuo, Zhao Jianan, et al. Test-time training for graph neural networks[M]//Database systems for advanced applications. Singapore: Springer Nature Singapore, 2026: 349-364.
- [38] Zhang Jiabin, Wang Yiqi, Yang Xihong, et al. A fully test-time training framework for semi-supervised node classification on out-of-distribution graphs[J]. ACM Transactions on Knowledge Discovery from Data, 2024,

- 18(7): 1-19.
- [39] Liu Yuejiang, Kothari P, Van Delft B, et al. TTT++: When does self-supervised test-time training fail or thrive[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2021: 1669.
- [40] Gui Shurui, Ji Shuiwang, Li X, et al. GOOD: A graph out-of-distribution benchmark[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2022: 068431-0150.
- [41] Shchur O, Mumme M, Bojchevski A, et al. Pitfalls of graph neural network evaluation[PP/OL]. V2.arXiv (2019-06-18)[2025-11-22]. <https://arxiv.org/abs/1811.05868>.
- [42] Hu Weihua, Fey M, Zitnik M, et al. Open graph benchmark: Datasets for machine learning on graphs[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 1855.
- [43] Pareja A, Domeniconi G, Chen Jie, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5363-5370.
- [44] Rozemberczki B, Allen C, Sarkar R. Multi-scale attributed node embedding[J]. Journal of Complex Networks, 2021, 9(2): cnab014.
- [45] Rong Yu, Huang Wenbing, Xu Tingyang, et al. DropEdge: Towards deep graph convolutional networks on node classification[C/OL]//Proceedings of the 8th International Conference on Learning Representations, 2020: 2502-2519. <https://openreview.net/forum?id=Hkx1qkrKPr>.
- [46] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C/OL]//Proceedings of the 5th International Conference on Learning Representations, 2017: 2713-2727. <https://openreview.net/forum?id=SJU4ayYgl>.
- [47] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[C/OL]//Proceedings of the 6th International Conference on Learning Representations, 2018: 2920-2932. <https://openreview.net/forum?id=rJXMpikCZ>.
- [48] Geisler S, Schmidt T, Şirin H, et al. Robustness of graph neural networks at scale[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2021: 585.
- [49] Zhu Dingyuan, Zhang Ziwei, Cui Peng, et al. Robust graph convolutional networks against adversarial attacks[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 1399-1407.
- [50] Jin Wei, Derr T, Wang Yiqi, et al. Node similarity preserving graph convolutional networks[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 148-156.
- [51] Xu Kaidi, Chen Hongge, Liu Sijia, et al. Topology attack and defense for graph neural networks: An optimization perspective[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. ijcai.org, 2019: 3961-3967.

作者简介



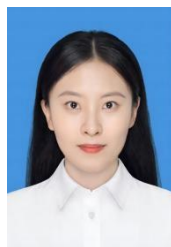
吴 涛 男,1987年出生于甘肃省庆阳市。现为重庆邮电大学网络空间安全与信息法学院教授、博士生导师。主要研究方向为人工智能安全、图数据挖掘、图机器学习、图模型安全等。
E-mail: wutao@cqupt.edu.cn



周圆庆 男,1998年出生于河南省商丘市。现为重庆邮电大学网络空间安全与信息法学院硕士研究生。主要研究方向为图神经网络、模型分布外泛化性等。
E-mail: 18837018370@163.com



先兴平 女,1984年出生于四川省泸州市。现为重庆邮电大学网络空间安全与信息法学院副教授、硕士生导师。主要研究方向为图数据挖掘、数据隐私保护、智能算法安全等。
E-mail: xianxp@cqupt.edu.cn



崔灿一星 女,1999年出生于山东省淄博市。现为重庆邮电大学计算机科学与技术学院博士研究生。主要研究方向为图神经网络、人工智能安全等。
E-mail: 1137051533@qq.com



刘逸琛 男,1995年出生于甘肃省兰州市。现为重庆邮电大学网络空间安全与信息法学院博士研究生。主要研究方向为图神经网络、人工智能安全等。

E-mail: 1013218192@qq.com



牛伟纳 女,1990年出生于河南省驻马店市。现为电子科技大学网络空间安全学院研究员。主要研究方向为恶意软件分析、网络攻击检测、数据安全等。中国电子学会会员编号:E190029983S。

E-mail: vinusniu@uestc.edu.cn



乔少杰 男,1981年出生于黑龙江省牡丹江市。现为成都信息工程大学软件工程学院教授、博士生导师。主要研究方向为大数据技术与应用、领域大数据分析、智能算法安全等。中国电子学会会员编号:E190157673M。

E-mail: sjqiao@cuit.edu.cn