

分布外泛化图机器学习及其在图基础模型中的应用

王鑫^{1,2}, 李昊阳¹, 钟嘉意¹, 朱文武^{1,2*}

1. 清华大学计算机科学与技术系, 北京 100084

2. 北京信息科学与技术国家研究中心, 北京 100084

* 通信作者. E-mail: wwzhu@tsinghua.edu.cn

国家重点研发计划项目(批准号: 2023YFF1205001)和北京信息科学与技术国家研究中心(批准号: BNR2026TD03005)资助

摘要 图机器学习在学术界和工业界都得到广泛关注与研究。尽管近年来涌现出大量新方法和技术, 但现有工作仍大多建立在同分布假设之上, 即测试与训练图数据服从相同分布。然而, 在许多实际图场景中, 同分布假设难以成立, 当测试和训练图数据之间存在分布偏移时, 模型性能通常会明显下降。围绕超越同分布假设的图上分布外(OOD)泛化问题, 相关研究已取得显著进展, 并引起了研究界的广泛关注。本文对图 OOD 泛化领域进行了系统论述, 回顾并总结了其最新研究进展。首先, 我们对图上的 OOD 泛化问题进行了形式化定义。其次, 我们根据现有方法在图机器学习流程中的位置, 从概念上将其划分为数据、模型和学习策略三类, 并分别加以详细讨论。此外, 本文还回顾了与图 OOD 泛化相关的理论, 介绍了用于系统评估该类方法的常用图数据集, 并总结了其在图基础模型中的应用。最后, 我们对未来研究方向进行了展望。

关键词 机器学习, 图神经网络, 分布外泛化, 表征学习, 图基础模型

1 引言

图数据在日常生活中无处不在。它已被广泛用于建模实体之间复杂的关系和依赖性, 范围涵盖从物理系统中的微观粒子相互作用、蛋白质中的分子结构, 到宏观交通网络和全球通信网络。图上的机器学习方法, 特别是图神经网络(GNNs), 在过去十年中引起了广泛关注并被深入研究。图上的机器学习方法在学术界和工业界均取得了显著进展, 展示了其在广泛现实应用中的有效性, 例如社交网络 [1]、推荐系统 [2]、知识表示 [3]、交通预测 [4] 等领域。

尽管图机器学习方法取得了显著成功, 但现有文献多依赖于测试和训练图数据来自相同分布的假设, 即分布内(in-distribution, I.D.)假设。然而, 在现实世界中, 由于不可控的底层数据产生机制 [5], 这样的假设很难得到满足。在实践中, 测试图和训练图之间将不可避免地出现分布偏移的场景 [6]。这些图机器学习方法普遍缺乏分布外(out-of-distribution, OOD)泛化的能力, 在分布偏移下往往会失效并伴随性能明显下降。因此, 开发能够在图上进行分布外泛化的方法至关重要, 特别是对于高风险的图应用, 例如分子预测 [7]、金融

引用格式: 王鑫, 李昊阳, 钟嘉意, 朱文武. 分布外泛化图机器学习及其在图基础模型中的应用. 中国科学: 信息科学, 在审文章
Xin Wang, Haoyang Li, Jiayi Zhong, Wenwu Zhu.
. Sci Sin Inform, for review

分析 [8]、刑事司法 [9]、自动驾驶 [10]、粒子物理学 [11], 以及针对 COVID-19 的大流行病预测 [12]、医疗检测 [13] 和药物再利用 [14] 等方向。

分布外 (OOD) 泛化算法 [15–17] 旨在使模型在未知分布偏移下仍能实现稳定的泛化性能。由于处理真实世界中未见数据的需求不断增加, OOD 泛化方法在研究社区中一直占据着重要地位。结合图机器学习和 OOD 泛化的优势, 图 OOD 泛化逐渐成为促进图机器学习模型在现实场景中部署的重要研究方向。然而, 由于存在以下挑战, 图的分布外泛化问题具有较高挑战性:

- **图数据的独特性:** 图结构化数据的空间具有非欧几里得特性, 这一特性催生了独特的图模型设计, 同时也为主要基于欧几里得数据 (如图像与文本) 开发的分布外泛化算法在直接迁移应用时构成了障碍。
- **图任务的多样性:** 图上的问题高度多样化, 从节点级别、链接级别到图级别的任务, 伴随着不同的设置、目标和约束, 因此需要将不同级别的图特征整合到图 OOD 泛化方法中。
- **图分布偏移类型的复杂性:** 图上的分布偏移可存在于特征级别 (例如节点特征) 和拓扑级别 (例如图大小或其他结构属性)。这种复杂类型的图分布偏移 (如图 1 所示) 给 OOD 泛化带来了更多困难。

在机遇与挑战并存的背景下, 当前有必要对图 OOD 泛化方法进行系统回顾与深入分析。本文对图上的 OOD 泛化进行了系统而全面的综述¹⁾。具体而言, 为覆盖图 OOD 泛化问题的完整研究生命周期, 本文首先给出了形式化的问题定义; 随后根据方法在图机器学习流程中所处的位置, 将现有方法划分为三个在概念上彼此区分的类别, 并对每一类别中的代表性方法进行了详细梳理与分析。进一步地, 本文系统回顾了图 OOD 泛化相关的理论基础与评测数据集, 以期促进该领域更加深入与规范化的发展。最后, 本文总结并讨论了若干值得进一步探索的研究方向。

一些相关的综述从图数据增强 [18, 19]、图自监督学习 [20, 21]、图对抗学习 [22, 23] 等角度进行回顾。然而, 这些工作与本文的关注重点存在显著差异。首先, 它们并未以图 OOD 泛化这一问题作为核心研究主题; 其次, 其中部分综述的方法仅构成图 OOD 泛化整体问题中的关键组成部分之一, 而非对该问题的系统性刻画。目前尚缺乏一篇对当前图 OOD 泛化方法最新进展进行全面系统总结的综述工作。

本文的其余部分组织如下。在第 2 节中, 我们形式化了图上 OOD 泛化的问题, 并提出了我们对现有文献的分类。我们在第 3 节至第 5 节中全面回顾了三类方法, 随后回顾了相关理论 (在第 6 节中) 和用于评测的数据集 (在第 7 节中)。我们在第 8 节介绍了分布外泛化在图基础模型中的应用。最后, 我们在第 9 节中指明了未来可能的研究方向与机遇。

2 问题定义与分类

在本节中, 我们首先给出图分布外泛化的形式化定义。在此之后我们提供了对现有图 OOD 泛化方法的分类。

2.1 问题定义

令 $G = (V, E)$ 表示一个图, 其中 V 是节点集合, $E \subseteq V \times V$ 是边集合。 $X \in \mathbb{R}^{|V| \times F}$ 表示节点特征矩阵, 其中 F 是节点特征的维度。 A 表示反映拓扑结构的邻接矩阵。因此, 图 G 可以由节点特征和拓扑结构组成, 即 $G = (X, A)$ 。主要数学符号说明在表 1 中总结。

令 $\mathbb{G}$ 为输入图空间, $\mathbb{Y}$ 为标签空间。一个带参数 θ 的图预测器 $f_\theta: \mathbb{G} \rightarrow \mathbb{Y}$ 将输入实例 $G \in \mathbb{G}$ 映射到标签 $Y \in \mathbb{Y}$ 。损失函数 ℓ 衡量预测与真实标签之间的距离。直观而言, 图 OOD 泛化关注的是: 模型只能在一

1) 本综述中回顾的图 OOD 泛化方法总结可在 <https://graph.ood-generalization.com> 找到。

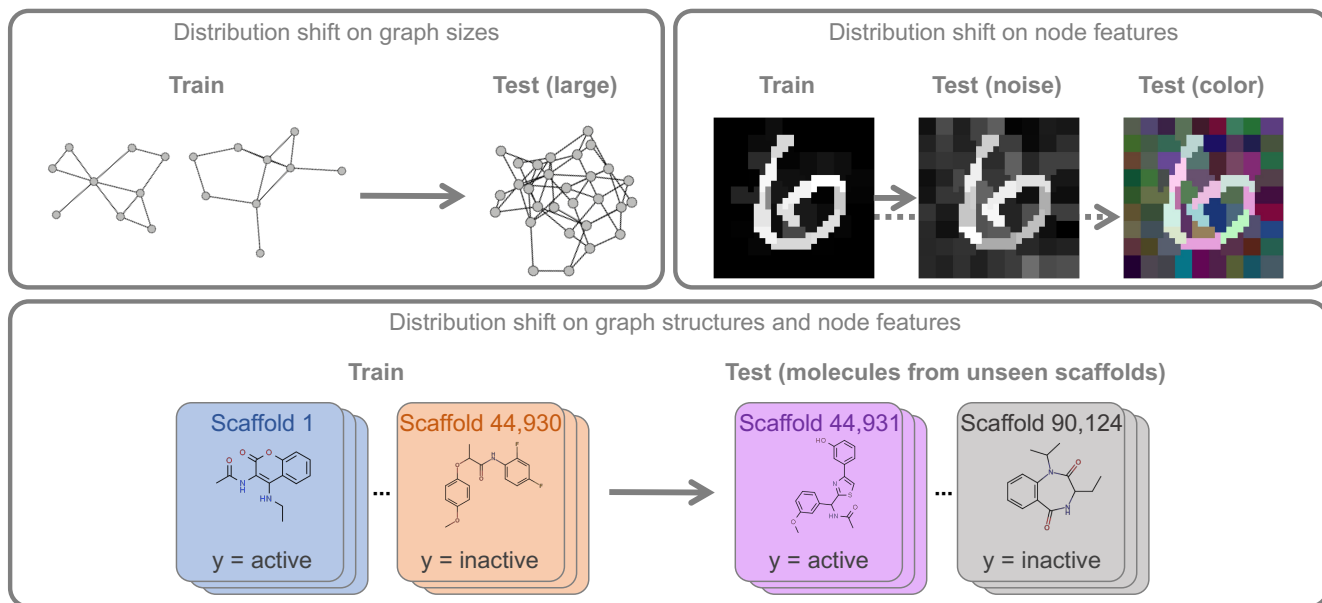


图 1 图上的复杂分布偏移类型。图数据中的分布偏移可能出现在图规模、节点特征以及图结构属性等不同方面 [6]。图分布外泛化方法旨在避免对训练数据过拟合,并在存在分布偏移的情况下仍能够在未见过的测试数据上保持良好的泛化性能。

Figure 1 Complex types of distribution shifts on graphs. The distribution shifts can exist on graph sizes, node features, and graph structural properties [6]. The OOD generalized graph approaches are expected to perform well on the unseen testing data even under distribution shifts rather than overfitting the training data.

个或若干训练图分布上学习,但在部署时需要面对与训练数据在节点特征、连接结构或图规模等方面存在差异的测试图。图 OOD 泛化问题定义如下:

定义 1 (图 OOD 泛化) 给定从训练分布 $P_{train}(G, Y)$ 中抽取了 N 个实例 (即节点、链接或图) 的训练集 $D = \{(G_i, Y_i)\}_{i=1}^N$, 其中 $G_i \in \mathbb{G}$ 且 $Y_i \in \mathbb{Y}$, 目标是学习一个最优图预测器 f_{θ}^* , 使其在从测试分布 $P_{test}(G, Y)$ 中抽取的数据上实现最佳泛化, 其中 $P_{test}(G, Y) \neq P_{train}(G, Y)$:

$$f_{\theta}^* = \arg \min_{f_{\theta}} \mathbb{E}_{G, Y \sim P_{test}} [\ell(f_{\theta}(G), Y)]. \quad (1)$$

$P_{test}(G, Y)$ 和 $P_{train}(G, Y)$ 之间的分布偏移可能导致基于分布内 (I.D.) 假设构建的图预测器失败, 因为直接最小化训练实例上的平均损失 $\mathbb{E}_{G, Y \sim P_{train}} [\ell(f_{\theta}(G), Y)]$ 无法获得一个在分布偏移下能泛化到测试实例的最优预测器。换言之, 式 (1) 强调的并不是在训练集上取得最低误差, 而是在未知测试分布上保持稳定且最低的预测误差。需要注意的是, 测试分布在训练阶段是未知的。

与传统的域泛化问题 [24] 相比, 图 OOD 泛化本质上更为复杂, 因为它需要解决潜在的多层次分布偏移, 包括特征级别 (例如, 节点特征 X) 和拓扑级别 (例如, 图大小、结构模式 A) 的偏移。这些偏移可能独立出现, 也可能同时发生, 这为学习一个能够在同一任务内乃至跨任务 (例如节点级别、边级别和图级别预测) 中有效泛化的最优图预测器 f_{θ}^* 带来了显著挑战。

2.2 方法分类

为应对未知分布偏移带来的挑战并有效解决图 OOD 泛化问题, 现有研究已进行了大量系统性的探索。总体而言, 相关方法可按照其在图机器学习流程中所作用的阶段与核心机制, 划分为以下三类:

- **数据 (Data)**: 该类方法主要通过直接操作输入图数据来提升模型的 OOD 泛化能力, 典型代表为图数据增强方法。这类工作的核心出发点在于一个普遍观察: OOD 泛化性能的退化往往源于训练数据多样性不足,

表 1 主要数学符号

Table 1 Main mathematical notation

Symbol	Meaning
$G = (V, E)$	A graph, where V denotes the set of nodes and E denotes the set of edges.
$X \in \mathbb{R}^{ V \times F}$	The node feature matrix, where each row corresponds to the feature representation of a node.
F	The dimensionality of node features.
A	The adjacency matrix, which represents the topological structure of the graph.
$\mathbb{G}$	The input graph space.
$\mathbb{Y}$	The label space.
G	A random graph instance drawn from the graph space.
Y	The random label variable corresponding to the graph instance.
$f_\theta : \mathbb{G} \rightarrow \mathbb{Y}$	A graph predictor parameterized by θ , which maps an input graph instance to the label space.
θ	The model parameters of the graph predictor.
$\ell(\cdot, \cdot)$	The loss function, which measures the discrepancy between the prediction and the ground-truth label.
$P_{train}(G, Y)$	The training distribution.
$P_{test}(G, Y)$	The test distribution.
$D = \{(G_i, Y_i)\}_{i=1}^N$	The training set drawn from the training distribution.
N	The number of training instances.
(G_i, Y_i)	The i -th training instance and its label, where $G_i \in \mathbb{G}$ and $Y_i \in \mathbb{Y}$.
f_θ^*	The graph predictor with optimal generalization performance on the test distribution.
$\mathbb{E}_{G, Y \sim P}[\cdot]$	The expectation over the random variables G and Y under distribution P .
$\mathcal{E}$	The set of environments; $ \mathcal{E} > 1$ indicates that multiple environments are used during training.

或训练分布未能充分覆盖潜在的测试分布空间。在保持模型骨干结构不变的前提下，图数据增强方法通过系统性地生成扩充样本，显著丰富训练数据的分布特征，从而在一定程度上缓解由分布偏移引发的泛化问题。

• **模型 (Model):** 该类方法侧重于设计新的图模型结构，以学习具备更强 OOD 泛化能力的图表示。其中，基于表示解耦的图模型与基于因果关系建模的图模型是两类具有代表性的方向。这些方法通常引入明确的先验知识或因果假设，将其直接编码进图神经网络的结构设计之中，以期在模型层面提升对分布偏移的鲁棒性。此类方法多在图表示学习阶段发挥作用，其核心目标是通过结构性归纳偏差，将因果相关因素与虚假相关因素加以分离。尽管在实践中它们可与图数据增强或定制化训练目标相结合，但其主要贡献集中于面向 OOD 泛化的新型模型架构设计。

• **学习策略 (Learning Strategy):** 该类方法从训练与优化的角度出发，通过设计特定的学习目标、约束或优化策略来增强模型的 OOD 泛化能力，代表性工作包括图不变性学习、图对抗训练以及图自监督学习等。与前两类方法不同，学习策略类方法通常保留原始数据形式，且不依赖于特定的新模型架构，而是作为一种通用的训练范式，与多种 GNN 骨干网络保持良好的兼容性。其核心思想在于通过引导学习过程，使模型获得对分布变化更加稳健的表示。

这些方法从三个概念上彼此区分、但又相互补充的角度，应对并缓解了图 OOD 泛化问题。我们在图 2 中给出了现有方法的整体分类框架，并将在后续章节中分别对每一类方法进行系统阐述。该分类在一定程度上

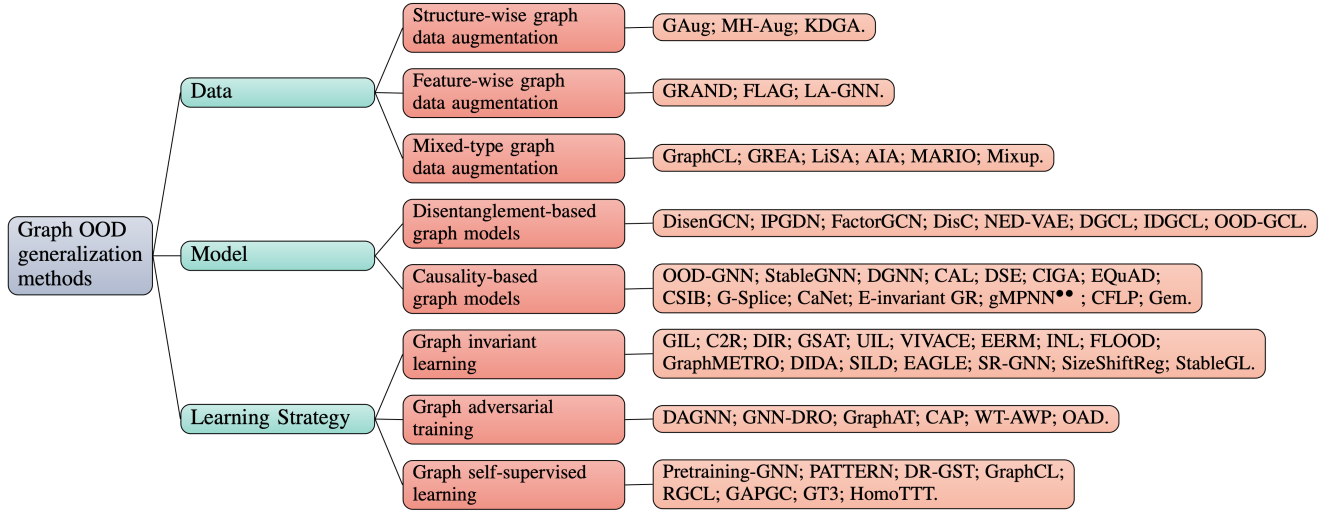


图 2 图分布外泛化方法的分类体系。根据现有方法在图机器学习流程中所处的位置，我们将图分布外泛化方法划分为三个概念上不同的类别，即基于数据的方法、基于模型的方法和基于学习策略的方法。

Figure 2 Taxonomy of graph OOD generalization methods. We categorize existing methodologies into three conceptually different branches based on their positions in the graph machine learning pipeline, i.e., data, model and learning strategy.

受到已有综述工作的启发 [15, 24–26]，同时也充分考虑了图 OOD 泛化问题自身的独特性。具体而言，该分类力图反映不同方法在原始工作中所强调的核心作用机制，即其主要研究动机或中心设计焦点，以及其在图学习流程中所作出的关键贡献环节。为进一步澄清三类方法之间的概念差异，我们在表 2 中总结并对比了它们在研究目标、主要作用组件以及潜在理论动机等方面的差异。此外，各类方法在实践层面所体现出的具体特点也进一步汇总于表 3 中。

3 数据

机器学习模型（包括图模型）的 OOD 泛化能力在很大程度上依赖于训练数据的多样性与质量 [16]。一般而言，训练数据分布越丰富、样本质量越高，图模型在未知分布下的泛化性能也越可靠。基于这一认识，图数据增强方法通过对原始图数据进行合理变换，以相对简洁而有效的方式构造出更多用于训练的图实例，从而扩展训练分布、提升模型的 OOD 泛化能力。其整体目标可形式化表示为

$$\min_{f_{\theta}} \mathbb{E}_{X', Y'} [\ell(f_{\theta}(X'), Y')], \quad (2)$$

其中 (X', Y') 表示由原始数据集 $\mathcal{D}$ 经过增强操作生成的训练样本对， $\mathcal{D}'$ 为相应的增强后训练集。

从现有研究来看，图数据增强方法大体可归纳为三类策略：结构增强 (*structure-wise augmentations*)、特征增强 (*feature-wise augmentations*) 以及混合类型增强 (*mixed-type augmentations*)。

3.1 结构级图数据增强

由于图结构（即拓扑）在图属性预测中起着关键作用，一些工作侧重于对输入图进行结构级增强，通过生成更加多样化的训练拓扑，使其能够覆盖潜在的未观测测试拓扑，从而提升模型的 OOD 泛化能力。这里我们主要回顾那些声称或已在实践中验证能够提高 OOD 泛化性能的代表性图数据增强方法。关于其他图数据增强方法的更全面讨论，可参考相关综述 [18, 19]。

表 2 图分布外泛化三类方法的概念关系与关键区别。该表从方法目标、作用对象、应用阶段、对 GNN 骨干网络的依赖程度、理论基础以及典型技术手段等方面,对基于数据、基于模型和基于学习策略的方法进行比较,从而突出本文所提出分类体系的基本依据与内在逻辑。

Table 2 Conceptual relationships and key distinctions among the three categories of methods for graph OOD generalization. The table compares data-, model-, and learning-strategy-based methods in terms of their objectives, targeted components, application stages, dependence on GNN backbones, theoretical foundations, and typical tools, highlighting the rationale behind the taxonomy proposed in this paper.

Aspect	Data	Model	Learning Strategy
Goal	Increase data diversity and quality	Encode prior knowledge or causal assumptions into model design	Enhance generalization via tailored training objectives and strategies
Targeted Component	Input graph structure or feature	Model architecture design or representations	Training procedures, loss functions or optimization schemes
Predominant Phase of Application	Primarily during data preparation stage	Typically during model design and representation learning	Mainly during optimization or training
GNN Backbone	Typically keep GNN backbone unchanged	Generally have specific GNN backbone design	Generally are compatible with different GNN backbones
Theoretical Foundation	Graph data augmentation	Representation disentanglement, causal modeling	Invariant learning, adversarial training, self-supervised learning
Typical Tools	Graph perturbation, mixup, graphon interpolation	Disentangled encoders, do-calculus, sample reweighting	Invariance loss, contrastive loss, adversarial training

GAug (Graph Augmentation) [27] 提出利用可微的边预测器生成增强图,以提升模型的泛化能力。该方法发现,边预测器能够有效编码类别同质结构,从而在给定图结构中促进类内边的形成并抑制类间边的产生。这类边层面的操作不仅有助于提升预测准确率,也有利于增强图模型的泛化性能。GAug 通过边预测模块对输入图进行结构修改,并将生成的增强图用于下游任务的训练与推理过程。同时,该方法还可以学习为输入图生成潜在的新边。即使在推理阶段不显式修改图结构,节点级分类任务的性能仍然能够得到提升。整体而言,该框架通过对图结构进行去噪并模拟结构变异性,有效增强了模型的泛化能力。

MH-Aug (Metropolis-Hastings Data Augmentation) [28] 从马尔可夫链蒙特卡洛采样 [89] 的视角提出了一种图数据增强框架,以实现增强强度与多样性的灵活控制。该方法从显式设计的增强目标分布中抽取一系列增强样本。由于从复杂分布中直接采样往往不可行,MH-Aug 采用 Metropolis-Hastings 算法高效获取增强样本。相比随机图增强策略,该方法具有更强的可控性,能够通过量化和调节增强强度,精确刻画自我中心图(或节点分类任务中样本)的结构变化。通过引入多样化的增强训练样本,模型的分布外泛化能力得到了显著提升。

3.2 特征级图数据增强

除通过增删边进行图结构增强的方法外,近期研究还提出了一系列针对节点特征的操作技术,并证明其在提升 OOD 泛化性能方面同样具有显著效果。

GRAND (Graph Random Neural Network) [30] 是一种简单而有效的特征级增强方法,旨在提升模型的泛化能力。该方法通过部分或完全随机丢弃节点特征,并将受扰动的特征在输入图上进行传播,从而减弱节点对特定邻域的过度敏感性。这种过度依赖局部邻域结构的行为往往会导致较差的 OOD 泛化性能。在同质性假设 [90] 下,GRAND 为每个节点随机生成多个增强表示,并通过最小化一致性损失,使这些增强表示在表示空间中保持接近。

FLAG (Free Large-scale Adversarial Augmentation on Graphs) [31] 是另一种简单、可扩展且通用的图数据增强方法,目标是实现更强的泛化能力。该方法在训练过程中引入基于梯度的对抗扰动,迭代地增强输入

表 3 图分布外泛化方法汇总。“Task”表示各方法所关注的任务类型,包括节点级、链接级和图级任务;“Shift Type”表示各方法能够处理的分布偏移类型,包括拓扑层面的分布偏移(如图规模和图结构变化)以及特征层面的分布偏移(如节点特征变化);“Backbone agnostic”表示该方法是否能够适用于其他 GNN 骨干网络;“ $|\mathcal{E}| > 1$ ”表示该方法在训练过程中是否依赖多个环境。

Table 3 A summary of graph OOD generalization methods. “Task” denotes the task type that each method focuses on, including node/link/graph level tasks. “Shift Type” denotes the type of distribution shifts that each method can handle, including topology-level (i.e., graph size and graph structure) and feature-level (i.e., node features) distribution shifts. “Backbone agnostic” indicates whether the method can be used for other GNN backbones. “ $|\mathcal{E}| > 1$ ” indicates whether the method relies on multiple environments during the training process.

Category	Subcategory	Method	Task				Shift Type		Backbone Agnostic	$ \mathcal{E} > 1$
			Node	Link	Graph	Size	Structure	Feature		
Data	Structure-wise graph data augmentation	GAug [27]	✓				✓		✓	
		MH-Aug [28]	✓				✓		✓	
		KDGA [29]	✓				✓		✓	
	Feature-wise graph data augmentation	GRAND [30]	✓					✓	✓	
		FLAG [31]	✓	✓	✓			✓	✓	
		LA-GNN [32]	✓					✓	✓	
	Mixed-type graph data augmentation	GraphCL [33]	✓		✓		✓	✓	✓	
		GREa [34]			✓		✓	✓	✓	
		LiSA [35]	✓		✓		✓	✓	✓	
		AIA [36]			✓		✓	✓	✓	
		MARIO [37]	✓		✓		✓	✓	✓	
		Mixup [38]	✓		✓		✓	✓	✓	
Model	Disentanglement-based graph models	DsenGCN [39]	✓				✓	✓		
		IPGDN [40]	✓				✓	✓		
		FactorGCN [41]			✓		✓	✓		
		DisC [42]			✓		✓	✓	✓	
		NED-VAE [43]			✓		✓	✓		
		DGCL [44]			✓		✓	✓	✓	
		IDGCL [45]			✓		✓	✓	✓	
		OOD-GCL [46]			✓		✓	✓	✓	
	Causality-based graph models	OOD-GNN [6]			✓	✓	✓	✓	✓	
		StableGNN [47]			✓		✓	✓	✓	
		DGNN [48]	✓				✓	✓	✓	
		CAL [49]			✓		✓	✓	✓	
		DSE [50]			✓		✓	✓	✓	
		CIGA [51]			✓	✓	✓	✓	✓	
		EQuAD [52]			✓	✓	✓	✓	✓	
		CSIB [53]			✓	✓	✓	✓	✓	
		G-Splice [54]			✓	✓	✓	✓	✓	
		CaNet [55]	✓				✓	✓	✓	
		E-invariant GR [56]			✓	✓		✓	✓	
		gMPNN** [57]		✓		✓			✓	
Learning Strategy	Graph invariant learning	CFLP [58]		✓			✓		✓	
		Gem [59]	✓		✓		✓		✓	
		GIL [60]			✓		✓	✓	✓	✓
		C2R [61]			✓		✓	✓	✓	✓
		DIR [62]			✓		✓	✓	✓	✓
		GSAT [63]			✓		✓	✓	✓	
		VIVACE [64]			✓		✓	✓	✓	
		UIL [65]			✓		✓	✓	✓	✓
		EERM [66]	✓				✓	✓	✓	✓
		INL [67]	✓				✓	✓	✓	✓
		FLOOD [68]	✓				✓	✓	✓	✓
		GraphMETRO [69]	✓		✓	✓	✓	✓	✓	
		DIDA [70]		✓			✓	✓	✓	✓
		SILD [71]		✓			✓	✓	✓	✓
		EAGLE [72]		✓			✓	✓	✓	✓
	Learning Strategy	SR-GNN [73]	✓				✓	✓	✓	✓
		SizeShiftReg [74]			✓	✓			✓	
		StableGL [75]	✓				✓	✓		✓
		DAGNN [76]			✓		✓	✓		✓
Graph adversarial training	Graph adversarial training	GNN-DRO [77]	✓				✓	✓	✓	
		GraphAT [78]	✓				✓	✓	✓	
		CAP [79]	✓				✓	✓	✓	
		WT-AWP [80]	✓		✓		✓	✓	✓	
		OAD [81]	✓				✓	✓	✓	
		Pretraining-GNN [82]			✓		✓	✓	✓	
	Graph self-supervised learning	PATTERN [83]			✓	✓			✓	
		DR-GST [84]	✓				✓	✓	✓	
		GraphCL [33]	✓		✓		✓	✓	✓	
		RGCL [85]			✓		✓	✓	✓	
		GAPGC [86]			✓		✓	✓	✓	
		GT3 [87]			✓		✓	✓	✓	
		HomoTTT [88]	✓				✓	✓	✓	

节点特征空间中的节点特征,同时保持图结构不变。FLAG 借鉴了自由对抗训练方法 [91] 来实现高效的对抗性数据增强。得益于其简洁且可扩展的设计,该方法能够在大规模数据集上高效训练,并且可以方便地集成到

常见 GNN 骨干网络的训练流程中。不同于主要针对节点级任务的 GRAND, FLAG 同时适用于节点级、链接级和图级任务。

LA-GNN (Local Augmentation for GNN) [32] 提出了一种面向 GNN 的局部特征增强策略,旨在以中心节点特征为条件的邻居节点特征分布。具体而言,该方法首先利用生成模型进行预训练,以建模中心节点特征条件下邻居节点特征的分布。随后,学习得到的条件分布被用于生成与中心节点相关的特征向量,作为每次训练迭代的附加输入。由于生成模型的预训练过程与下游 GNN 的训练相互解耦,该方法具有良好的模型无关性,能够以即插即用的方式应用于多种 GNN 骨干结构。此外,新节点的特征向量可以直接由生成模型生成,从而增强了模型对未见测试节点的泛化能力。与其他特征级增强方法相比,LA-GNN 更加关注节点邻域的局部信息,而非仅针对整体分布属性进行全局增强。

3.3 混合类型图数据增强

为结合结构级与特征级图数据增强方法的优势,一些工作采用了混合类型的增强范式,而非仅对图拓扑或节点特征进行单一类型的操作。这类方法近年来在研究社区中逐渐受到关注,并被证明在改善 OOD 泛化方面具有良好效果。

GraphCL (Graph Contrastive Learning) [33] 提出了四种通用的图结构数据增强操作,包括节点丢弃 (node dropping)、边扰动 (edge perturbation)、属性遮蔽 (attribute masking) 以及子图采样 (subgraph sampling)。其中,节点丢弃通过随机移除节点及其相关边来实现;边扰动随机添加或移除部分边;属性遮蔽将部分节点属性替换为高斯噪声;子图采样则基于随机游走策略从原图中采样包含部分节点的子图。在获得输入图的多种增强视图后,GraphCL 通过最大化不同增强下图表示的一致性进行训练,并在图分类任务中展现出良好的 OOD 泛化能力 [92]。

GREa (Graph Rationalization Enhanced by Environment-based Augmentations) [34] 提出了一种基于环境替换的数据增强策略,用以提升对图基本原理的识别能力,从而实现 OOD 泛化。该方法将图的理据定义为输入图中最能支持预测且具有 OOD 泛化能力的代表性子图。作者指出,现有增强方法(如 GraphCL)多为对输入图的启发式修改,可能难以直接支持理据子图的识别。为此,GREa 通过将输入图的环境子图替换为其他图的环境子图来构造增强样本,并约束增强样本与原始图具有相同标签。考虑到显式子图编码与解码的高复杂性,该方法转而在潜在表示空间中进行隐式的理据-环境分离与学习。实验结果表明,基于对代表性子图的准确识别,GREa 能够有效提升模型的 OOD 泛化能力。

LiSA (Label-invariant Subgraphs to Construct Augmented Environments) [35] 是一种启发式且有效的图数据增强方法,通过从源域中提取标签不变子图来构建多个增强域,以实现 OOD 泛化。该方法针对分布偏移源于域差异的情形,确保图预测器在不同域之间保持一致的预测行为。为缓解现实中难以收集足够多域的问题,LiSA 采用变分子图生成器生成多样化的子图,同时保留关键的预测信息,从而构造增强域。基于能量的正则化用于拉开不同增强域分布之间的距离以促进多样性,而信息约束则确保生成子图包含与标签相关的关键信息。这些增强域维持一致的预测关系,使模型能够在新域中的 OOD 测试图上实现有效泛化。

AIA (Adversarial Invariant Augmentation) [36] 提出了一种用于缓解协变量偏移问题的图增强技术,该问题是图 OOD 泛化中的一种典型场景。作者指出,现有图增强策略往往受到环境有限或因果特征不稳定的制约,从而限制了其在协变量偏移条件下的泛化能力。为此,AIA 提出了两条图增强原则:环境多样性与因果不变性。环境多样性原则鼓励增强策略外推到未见环境或域,而因果不变性原则则旨在缩小增强图数据与未见测试图之间的分布差距。该方法包含两个核心模块:对抗性增强器,用于在图拓扑和节点特征层面对掩码进行对抗学习以增强环境多样性;以及因果生成器,用于生成捕获因果信息的掩码。基于上述原则与设计,AIA 在协变量偏移场景下展现出更强的鲁棒性与 OOD 泛化能力。

MARIO (Model-Agnostic Recipe for Improving OOD Generalization) [37] 通过图数据增强提升不同增强视图之间表示的一致性,并结合条件互信息正则化以抑制冗余信息,同时保留与任务相关的特征。通过协同处理增强引入的变异性与表示冗余问题,MARIO 有效缓解了模型对虚假相关性的过拟合,并在节点级和图级分类任务中实现了良好的 OOD 泛化性能。

此外,随着图神经网络的发展,**Mixup** 及其变体 [38, 93] 作为一种通过对给定实例进行插值来生成新样本的通用数据增强方法,已在计算机视觉 [94] 和自然语言处理 [95] 领域从理论与实证层面被证明能够提升泛化能力。类似的思想也被引入到图数据中 [96–101]。

例如,**GraphMix** [96] 通过联合训练全连接网络 (FCN) 与 GNN,在节点分类任务中采用流形 Mixup [93]。其中,FCN 的损失通过流形 Mixup 计算,而 GNN 的损失则采用标准训练方式。两者之间通过参数共享机制进行耦合,以促进关键节点表示从 FCN 向 GNN 的迁移。**G-Mixup** [97] 在嵌入空间中对节点特征与图结构进行插值,实现对图隐式表示的混合生成。**NodeAug** [100] 将 Mixup 建模为一个双分支图卷积模块,通过混合一对节点的原始特征,并在双分支 GNN 中进一步混合其各层隐藏表示。**ifMixup** (intrusion-free Mixup) [99] 将 Mixup 直接作用于图数据本身而非潜在表示。针对图结构不规则、不同图节点间缺乏对齐关系的问题,ifMixup 采用随机分配节点索引并基于索引进行节点匹配。**G-Mixup** [101] 针对图结构不规则、难以对齐以及不同类别图拓扑差异显著的问题,提出了一种基于 Graphon 的混合策略。该方法首先利用同类别图估计对应的 Graphon,并在欧几里得空间中对不同类别的 Graphon 进行插值,得到混合 Graphon,随后通过采样生成合成图。实验结果表明,该方法在存在分布偏移的图分类数据集上表现优异,展现出良好的 OOD 泛化潜力。**OOD-GMixup** [102] 则通过可控的数据增强来应对混合结构分布偏移问题。该方法首先提取任务相关的图基本原理以消除虚假相关性,随后通过流形 Mixup 生成虚拟样本,并结合极值理论进行校准与重加权训练,从而进一步提升 OOD 泛化性能。

鉴于上述方法在思想上的相似性,本文在 Figure 2 和 Table 3 中统一使用符号“Mixup”来指代这些基于 Mixup 的图数据增强方法。

4 模型

除了通过对输入图数据进行增强以辅助实现良好的 OOD 泛化能力之外,另一类研究工作专门致力于设计新的图模型,即式 (1) 中的 f_θ 。这类方法通过在模型设计中引入一定的结构性先验,使图模型具备学习有利于 OOD 泛化的图表示的能力。沿着这一研究方向,现有工作主要形成了两条具有代表性的技术路线:基于解耦的图模型和基于因果的图模型。

基于解耦方法与基于因果方法的区别: 尽管这两类方法的共同目标都是在削弱虚假模式影响的同时提取稳定且与任务相关的信息,但它们建立在本质不同的理论基础与建模范式之上。本文的分类依据在于它们实现 OOD 泛化的核心机制,即分别基于统计假设与因果假设。

基于解耦的方法起源于表示学习领域,其核心目标是将潜在表示分解为若干统计上相互独立的组成部分,每一部分对应一个不同的潜在因子。这类方法强调表示的模块化与可解释性,通常通过多通道编码器或基于路由的机制来实现 [39]。值得注意的是,基于解耦的方法通常不依赖于关于数据生成过程的显式先验知识,也不需要引入因果假设。

基于因果的方法源于因果推断理论,其基本假设是观测到的图数据由潜在的因果机制(例如结构化因果模型)生成,并致力于学习在不同干预或环境变化下保持稳定的表示。这类方法通常涉及混杂因子平衡、后门或前门调整,以及反事实推理等技术。

尽管两类方法均以提升 OOD 泛化能力为目标,但它们关注的信息类型存在显著差异:基于解耦的方法侧重于识别数据中统计独立的潜在因素,而基于因果的方法则着眼于建模和利用支配数据生成过程的底层因

果机制。

4.1 基于解耦的图模型

本节将介绍基于解耦思想的图模型在 OOD 泛化场景下的代表性工作。

真实世界中的图结构通常由多个潜在因素相互作用所驱动,其生成过程复杂且具有高度异质性。解耦图表示学习的目标在于学习能够将图数据背后这些不同且具有信息量的潜在因素分离开的表示,并在分解后的向量表示的不同子空间中刻画这些因素 [39]。已有研究表明,这类解耦表示有助于提升模型在分布外场景下的泛化能力 [103, 104]。现有方法大致可分为三类:有监督解耦方法 [39–42]、无监督生成式解耦方法 [43],以及自监督对比式解耦方法 [44, 45]。

DisenGCN [39] 是首个用于学习解耦节点表示的图神经网络模型,其核心组件为解耦多通道卷积层 DisenConv。该层通过引入邻域路由机制,旨在识别连接中心节点与其各个邻居的潜在因果因子,并将邻居节点动态分配至负责相应因子的专用通道中。具体而言,该方法将节点及其邻居投影到多个子空间,并通过在潜在子空间聚类结果的迭代分析来推断潜在因素。作者证明,在足够多次迭代后,所提出的邻域路由机制能够收敛。因此,DisenConv 中的每一个通道都能够仅从邻居节点中提取对应某一解耦潜在因素的特征,并独立执行卷积操作。通过堆叠多个 DisenConv 层,DisenGCN 可以捕获超出局部邻域范围的信息并生成解耦的节点表示。由于节点的潜在因素被有效分离,该方法在 OOD 泛化场景下表现出更优的性能。

IPGDN (Independence Promoted Graph Disentangled Network) [40] 在 DisenGCN [39] 的基础上进行了进一步扩展。除了利用邻域路由机制对图中的潜在因素进行解耦外,IPGDN 还显式地鼓励不同潜在因素之间保持尽可能高的独立性。该方法采用核方法中的 Hilbert–Schmidt Independence Criterion (HSIC) [105] 来度量并最小化不同表示之间的统计依赖性。具体而言,在解耦目标节点时,IPGDN 的卷积层首先从邻居节点的不同方面构建解耦特征表示,随后通过最小化 HSIC 正则项来强化潜在表示之间的独立性,从而获得最终的节点表示。需要指出的是,解耦表示学习与独立性正则化在一个统一的优化框架中联合进行,使得该方法相比 DisenGCN 能够学习到更加充分解耦的表示。此外,DisenGCN [39] 与 IPGDN [40] 均主要面向图上的节点级任务。

FactorGCN (Factorizable GCN) [41] 是一种面向图级表示学习的解耦图神经网络模型。该方法基于因子分解机制,将输入图分解为多个具有明确语义的因子子图,并据此学习解耦的图级表示。具体而言,每一个因子子图都会被输入至一个专用的图卷积网络 (GCN),用于聚合与该子图对应的潜在因子相关的特征;随后,通过聚合操作将来自所有潜在因子的特征进行拼接,形成最终的图级表示。所得表示具有良好的块状可解释性,其中每一个子表示对应一个解耦且可解释的关系子空间。上述过程构成了 FactorGCN 的单层结构。通过堆叠多层,该模型能够在不同层级采用不同数量的因子子图,从而实现层次化的解耦建模,并在不同粒度上对输入图结构进行表示学习。

与仅对潜在因素进行解耦的方法不同,**DisC** (Disentangled Causal Substructure) [42] 是一种直接对输入图中的因果信息与非因果信息进行解耦的 GNN 模型。该方法通过将输入图显式分解为因果子图与偏置子图,使模型在分布偏移导致严重偏差的情况下,仍能够仅依赖因果子图结构进行稳定预测。具体而言,该方法首先利用参数化的边掩码生成器,将图中的边划分为因果边与偏置 (即非因果) 边,其参数在整个数据集范围内共享。该边掩码生成器用于刻画每条边的重要性,并据此提取因果子图与偏置子图。随后,因果子图与偏置子图分别被输入至两个独立的 GNN,并分别采用因果感知的加权交叉熵损失与偏置感知的广义交叉熵损失进行训练,从而获得解耦的表示。在此基础上,该方法进一步对来自不同图的潜在表示进行置换以生成额外的训练样本。尽管新生成的样本同时包含因果信息与偏置信息,但其因果子图与偏置子图在统计意义上是去相关的。最终,模型能够更加专注于解耦后的因果子图与标签之间的真实关联,从而实现具有 OOD 泛化能力的预测。

除了上述有监督解耦方法之外,研究者还提出了一系列无监督的解耦图表示学习方法。

NED-VAE (Node-Edge Disentangled Variational Auto-Encoder) [43] 是一种用于图解耦学习的无监督生成式方法,能够从带属性的图中自动捕捉节点与边所对应的独立潜在因素。该方法通过联合优化三个子编码器(节点编码器、边编码器以及节点-边协同编码器),实现节点与边的联合解耦建模,并学习三类潜在表示;同时,利用两个子解码器(节点解码器与边解码器)对节点与边进行联合生成,从而刻画节点与边之间的复杂依赖关系。基础版本的 NED-VAE 还可以进一步扩展至组级解耦与变量级解耦情形,从而支持更加细粒度的解耦建模。

由于无监督生成式方法通常依赖重构目标,其计算代价较高,且重构误差可能为表示学习引入不必要的偏置,**DGCL** (Disentangled Graph Contrastive Learning) [44] 首次提出通过自监督对比学习的方式来学习解耦的图表示。具体而言,该方法首先识别输入图背后的潜在因素,并通过定制的解耦图编码器获得因子化表示,其中编码器的核心组件是一个多通道消息传递层。每一个因子化表示对应于图中某一特定潜在因素所关联的解耦属性。随后, DGCL 在每一个因子对应的表示子空间内独立地进行因子维度的对比学习,而非在整体表示空间中统一进行。这种设计促使每个解耦因子仅在图的某一特定方面具有判别性,从而引导图编码器学习能够分别刻画不同潜在因素的解耦表示。与生成式模型不同,对比学习作为一种实例级的判别式方法,其目标是在表示空间中拉近相似实例、拉远不相似实例 [106, 107],从而避免了计算代价较高的图重构过程,并有助于学习信息更加充分的图表示。

为了进一步提升所学习图表示的解耦性,**IDGCL** (Independence Promoted Disentangled Graph Contrastive Learning) [45] 在 DGCL 的基础上引入了 Hilbert-Schmidt Independence Criterion (HSIC) [105],以显式地削弱反映不同潜在因素的解耦表示之间的统计依赖关系。由于解耦图表示被期望分别捕捉不同潜在因素的互补信息, IDGCL 通过独立性正则化有效增强了不同潜在表示之间的统计独立性。因子维度的对比学习与独立性约束在一个统一框架中联合优化,使得解耦图编码器能够生成更加充分解耦的图表示。相较于现有方法, IDGCL 以自监督方式为单个图学习多个解耦表示,使得对各通道语义的分析成为可能,在表示可解释性与 OOD 泛化能力方面均展现出优势。

OOD-GCL (OOD Generalized Disentangled Graph Contrastive Learning) [46] 在此基础上进一步提出了一种具有理论保证的解耦图对比学习模型,用于应对 OOD 泛化挑战。该方法结合解耦图编码器与定制的不变性自监督学习机制,能够有效捕捉在不同环境或分布下保持稳定的不变潜在因素,从而在分布偏移条件下生成具有良好泛化能力的图表示。相关理论分析表明,该方法在一定条件下能够保证学习到解耦的图表示,并基于这些解耦表示实现 OOD 泛化。

4.2 基于因果的图模型

在本节中,我们介绍用于 OOD 泛化的基于因果的图模型。

因果推断是实现 OOD 泛化的一项重要技术。图机器学习模型往往会利用训练集中存在的细微统计相关性,即便这些相关性在预测中属于虚假相关(即意外的“捷径”),以提升训练精度。当模型高度依赖这些虚假相关性时,其性能在真实的 OOD 测试环境中可能会显著下降,因为这些虚假相关性在分布外环境中往往会发生变化。相比之下,受因果推断理论支持的基于因果的图模型能够内在地捕捉输入图数据与标签之间在分布偏移下保持稳定的因果关系 [108],从而实现良好的 OOD 泛化性能。现有方法可根据其理论基础进行划分,包括混杂因子平衡方法 [6, 47, 48]、基于预定义结构化因果模型的方法 [49, 50, 56, 57],以及基于反事实推断 [58] 和 Granger 因果关系 [59] 的方法。

4.2.1 基于混杂因子平衡的方法

一些方法 [6, 47, 48] 将混杂因子平衡引入图模型中, 以缓解虚假相关性对 OOD 泛化性能的影响。

OOD-GNN [6] 基于因果推断中的混杂因子平衡理论 [109], 率先提出通过对图表示施加非线性去相关操作来解决 OOD 泛化问题。具体而言, OOD-GNN 利用随机傅里叶特征 [110] 实现非线性的图表示去相关, 以消除图编码器所学习的因果表示与非因果表示之间的统计依赖关系。该方法具有线性时间复杂度, 并能够有效削弱虚假相关性。图编码器的参数与用于图表示去相关的样本权重通过迭代方式进行联合优化, 从而学习用于预测的判别性图表示。这种去相关操作在本质上等价于混杂因子平衡, 其目标是鼓励处理变量与混杂变量之间的独立性。在加权数据集上训练得到的图编码器能够更准确地估计图表示中变量对标签的因果效应, 同时有效消除虚假相关性。通过这种方式, OOD-GNN 在多种图基准任务上 (包括图规模、节点特征以及图结构发生变化的分布偏移场景) 均取得了良好的性能, 表明其在真实环境中具备较强的 OOD 泛化能力。

StableGNN [47] 提出利用可微分的图池化层, 通过样本重加权的方式提取基于子图的去相关表示, 其基本思想与 OOD-GNN 类似。首先, 图的高层变量学习模块采用图池化层 [111, 112], 将相邻的低层节点映射到一组簇中, 其中每个簇被期望表示原始图中的一个致密连接的子图单元。随后, 通过聚合同一簇内节点的嵌入生成簇级表示, 并通过有序拼接操作在不同图之间对齐簇的语义空间, 这些簇级表示作为图的高层变量。接下来, 通过优化样本权重来消除这些高层变量之间的统计依赖关系。因此, 图编码器能够更加关注判别性子结构与标签之间的真实联系, 从而获得良好的 OOD 泛化能力。

除了上述图层级的去相关模型之外, **DGNN** (Debiased GNN) [48] 作为一种节点级去相关模型, 其方法论与 StableGNN [47] 相似, 旨在通过消除节点层面的虚假相关性, 在分布偏移下实现稳定预测。具体而言, 该方法提出了一个用于 OOD 泛化的节点表示学习统一框架, 通过联合优化去相关正则项与加权 GNN 模型来实现。其中, 去相关正则项用于学习一组样本权重, 以消除节点中因果信息与非因果信息之间的虚假相关性, 从而提升模型在分布外环境下的泛化性能。

4.2.2 基于结构化因果模型的方法

一些方法 [49–51, 56, 57] 在模型设计中显式引入结构化因果模型 (SCM), 以刻画数据生成过程背后的因果机制。在模型中融入恰当的因果结构, 有助于提升模型在分布偏移下的稳定性与泛化能力。

CAL (Causal Attention Learning) [49] 从因果视角审视 GNN 模型, 并通过刻画五个变量之间的因果关系构建了一个结构化因果模型, 这五个变量分别为图数据、因果特征、捷径特征、图表示以及预测结果。基于该结构因果模型 (SCM), 该方法重点关注因果特征 C 与预测结果之间的后门路径, 其中捷径特征 S 起到混杂因子的作用。该后门路径可能诱导模型利用捷径特征而非因果特征进行预测, 从而在分布偏移下导致较差的 OOD 泛化性能。为此, 该方法通过对因果特征实施干预演算以切断后门路径 (即进行后门调整 [113]), 从而消除混杂效应。该方法能够在不受不稳定捷径特征干扰的情况下, 学习因果特征与预测结果之间的真实关系, 从而提升图分类任务中的 OOD 泛化性能。

DSE (Deconfounded Subgraph Evaluation) [50] 提出了一种用于无偏评估解释性子图对预测结果因果效应的方法。作者指出, 在分布偏移情形下, 直接刻画混杂效应往往较为困难, 因此在预定义结构因果模型 (SCM) 的前提下, 使用后门调整来阻断因果因子图到标签之间的后门路径具有较大挑战。为此, 该方法采用前门调整, 并为因果因子图引入一个替代变量。不同于基于特征移除原则的解释子图评估方法, DSE 设计了一种生成模型, 即条件变分图自编码器, 用于生成符合数据分布的潜在替代变量。通过该机制, DSE 能够对因果因子图与标签之间的关系进行无偏估计, 从而缓解分布外效应, 并实现良好的 OOD 泛化性能。

CIGA (Causality Inspired Invariant Graph Learning) [51] 进一步将因果部分 C 与非因果部分 S 之间的潜在交互划分为完全信息不变特征 (FIIF) 和部分信息不变特征 (PIIF), 该划分取决于潜在因果部分 C 是

否对标签 Y 完全具有信息量, 即 $(S, E) \perp\!\!\!\perp Y | C$ 。在 FIIF 假设下, 非因果部分 S 直接受因果部分 C 控制; 而在 PIIF 情形下, 非因果部分 S 通过标签 Y 间接受因果部分 C 的影响。这两类结构因果模型 (SCM) 在观测到的分布偏移下表现出不同的行为。若忽略 FIIF 或 PIIF 中的任意一种情形, 图 OOD 泛化性能将会显著下降。基于上述分析, CIGA 将因果部分 C 实例化为关键子图, 该子图包含决定标签的潜在因果信息。通过识别在不同训练环境中最大程度保留类内信息的一致关键子图, 模型能够在分布偏移下实现稳定预测, 从而获得良好的 OOD 泛化性能。

EQuAD (Encoding-Quantifying Decorrelation) [52] 在 CIGA 的基础上进一步改进, 通过引入量化机制来区分虚假特征与因果特征, 并将虚假特征映射到紧凑空间中以实现高效去相关。同时, 该方法结合样本级重加权策略, 以缓解数据不平衡问题。

CSIB (Causal Subgraphs and Information Bottlenecks) [53] 利用结构化因果模型识别在不同环境中对标签具有因果影响的不变子图。该方法同时考虑 FIIF 与 PIIF 两种情形, 并引入信息瓶颈机制以抑制虚假特征, 从而在复杂分布偏移下实现图 OOD 泛化。

G-Splice (Graph Splicing for Structural Linear Extrapolation) [54] 通过显式建模因果子图与环境子图, 将结构化因果模型引入图 OOD 泛化问题。该框架通过从环境相关特征中识别因果模式, 确保外推得到的图结构在因果层面保持一致性, 并利用结构因果模型 (SCM) 在非欧几里得空间中进行线性外推以生成多样且因果一致的 OOD 样本, 从而显著提升 GNN 在复杂分布偏移下的泛化能力。

CaNet (Causal Intervention for Network Data) [55] 基于因果干预理论, 针对潜在环境诱导的节点级预测混杂偏差问题提出了解决方案。该方法引入环境估计器以推断伪环境标签, 并动态引导混合专家 GNN 预测器进行学习。该协同学习框架能够捕捉稳定且对环境不敏感的关系, 从而提升模型在多样分布偏移下的泛化能力。

E-invariant GR [56] 提出了一种双网络有向无环图 [114] 作为其结构化因果模型, 用于学习对图规模不敏感的图表示, 从而更好地在训练图与测试图之间进行外推。不同于上述方法中较为简化的结构因果模型, 该方法构建的结构因果模型 (SCM) 刻画了更加复杂且细粒度的变量关系, 包括 graphon、训练/测试环境、节点特征、边以及图规模。在该 SCM 中, 训练图由 graphon 表征, 该 graphon 同时决定图的标签以及其结构与属性特征。训练环境由一个未观测的环境变量表示, 该变量刻画了与环境相关的特定图属性, 因此在训练集与测试集之间可能发生变化。基于该结构因果模型, 作者提出了一种近似规模不变的图表示方法, 该方法能够外推到 OOD 测试数据, 并证明所学习的图表示在 OOD 测试数据上的性能不劣于在与训练数据具有相同环境分布的测试数据上的性能。此外, 该方法仅依赖单一训练环境即可实现外推 (例如所有训练图具有相同规模)。

由于 E-invariant GR [56] 仅研究了图分类任务中的 OOD 泛化问题, **gMPNN⁺⁺** [57] 在相似设定下将其进一步扩展至链路预测任务, 其中测试图的规模大于训练图。具体而言, 作者首先提出了一个用于描述数据生成过程的结构因果模型 (SCM), 以学习在图规模发生分布偏移时仍具备泛化能力的链路预测器, 并通过非渐近界 (Non-asymptotic Bounds) 证明, 随着测试图规模的增大, 基于置换等变结构节点嵌入的链路预测器将逐渐退化为随机猜测。作者进一步表明, 输出的结构化成对嵌入可以收敛到一个连续函数的嵌入形式, 从而在链路预测任务中实现 OOD 泛化。

4.2.3 基于反事实推断与 Granger 因果关系的方法

一些图 OOD 方法受到反事实学习 [113] 理论的启发。反事实推断位于因果阶梯的最高层级 [115], 用于回答“如果某些事件发生或未发生, 在另一种可能世界中会产生什么结果”这一问题。另一些方法则受到 Granger 因果关系 [116] 的启发, 该理论认为, 当利用某一特征的全部信息相比于去除该特征的信息能够更好地预测标签时, 则该特征与标签之间存在因果关系。

CFLP (Counter-Factual Link Prediction) [58] 面向 OOD 链路预测任务, 通过同时预测事实链路 with 反事实链路来学习全局图结构与链路存在性之间的因果关系。该方法试图回答如下反事实问题: “如果图结构与观测到的情况不同, 该链路是否仍然存在?” 通过回答这一问题生成的反事实链路被用于训练图编码器, 从而学习具备 OOD 泛化能力的表示。为生成反事实链路样本, 该方法构建了一个因果模型, 将节点对的信息 (即学习到的表示) 视为上下文, 将全局图结构属性作为处理变量, 并将链路存在性作为结果变量。在此基础上, 该模型能够生成反事实训练链路样本, 并从事实链路 with 反事实链路中联合学习表示, 以提升 OOD 泛化能力。

Gem [59] 方法基于 Granger 因果关系, 将原始计算图输入解释器并输出因果解释图, 从而展现出更强的泛化能力。该方法认为, 当移除某条边或某个节点导致预测性能下降时, 该边或节点与对应预测之间存在因果关系。鉴于图数据本身具有内在的相互依赖性, 节点及其边往往是相关变量, 该方法进一步引入多种图规则 (例如连通性检查), 以确保获得的解释结果为有效且符合人类理解的因果图。最终, 该方法能够为 GNN 提供具有可解释性的因果解释, 并在分布偏移场景下实现具备 OOD 泛化能力的预测。

5 学习策略

除了图数据增强和图模型之外, 还有一些工作专注于通过具有针对性的优化目标与约束的训练策略来促进 OOD 泛化能力, 包括图不变性学习、图对抗训练以及图自监督学习。

5.1 图不变性学习

首先, 我们介绍用于 OOD 泛化的图不变性学习方法。

不变性学习旨在挖掘在不同分布之间保持稳定的“特征—标签”关系, 同时忽略随分布变化的虚假相关性; 在分布偏移下, 该类方法在一定条件下可被证明能够实现令人满意的 OOD 泛化性能 [117–119]。当因果性评估具有挑战性, 或强假设在实践中可能被违反时, 不变性学习可以通过搜索在分布偏移下保持不变的特征来近似完成该任务, 从而实现 OOD 泛化 [118]。不变性学习通常假设每个实例中用于预测的信息由两部分组成: 其一是不变部分, 与标签之间的关系在不同环境中保持稳定; 其二是变化部分, 与标签之间的关系可能随环境变化。当仅基于不变信息进行预测时, 可以获得良好的 OOD 泛化性能。沿着这一思路, 图不变性学习方法主要分为两类: 不变性优化方法 [60, 62, 63, 66, 70] 和显式表示对齐方法 [73–75]。

5.1.1 不变性优化

这类方法基于不变性原理来解决图 OOD 泛化问题。不变性原理假设数据内部存在不变属性, 从而可以在多个环境中发现这种不变性以实现 OOD 泛化。该假设可以形式化为:

假设 1 (不变性假设) 在输入实例 X 中存在一部分信息 $\Phi(X)$, 使得 $\forall e, e' \in \text{supp}(\mathcal{E}), P^e(Y|\Phi(X)) = P^{e'}(Y|\Phi(X))$, 其中 $\mathcal{E}$ 表示所有可能的环境, 而 $\Phi(X)$ 通常被称为输入实例 X 的不变理据 (invariant rationales)。

遵循近期基于不变性学习的 OOD 泛化研究 [117–119], 这些不变性优化方法将测试图数据与训练图数据之间的分布偏移原因视为一个潜在的未知环境变量 e 。其优化目标可表示为:

$$\min_{f_\theta} \max_{e \in \text{supp}(\mathcal{E})} \mathcal{R}(f_\theta|e), \quad (3)$$

其中 $\mathcal{R}(f_\theta|e) = \mathbb{E}_{X, Y \sim P^e}[\ell(f_\theta(\Phi(X)), Y)]$ 表示基于不变信息 $\Phi(X)$ 进行预测的模型 f_θ 在环境 e 下的风险。因此, 如表 3 的最后一列所示, 这类方法依赖于显式的多环境划分 (由 $|\mathcal{E}| > 1$ 表示), 该划分可以事先给定, 或在训练过程中生成。

GIL (Graph Invariant Learning) [60] 被提出用于在分布偏移下捕获预测性图结构信息 (即子图或理据) 与标签之间的不变关系, 以实现图级别的 OOD 泛化。图不变性学习的一个主要挑战在于, 图的环境标签通常不可观测, 或收集成本极高, 从而使得在多个环境中学习不变性变得困难。因此, 该方法首先研究在没有显式环境划分的情况下进行不变性学习。具体而言, GIL 联合优化三个相互促进的模块, 包括不变子图识别模块、环境推断模块以及不变性学习模块。首先, 不变子图识别模块是一个基于 GNN 的子图生成器 $\Phi(\cdot)$ 。给定输入图 G , 它识别不变子图 $\Phi(G)$, 并将图中其余部分 (即不变子图的补集) 定义为变化子图, 记为 $G \setminus \Phi(G)$ 。随后, 环境推断模块对数据集中所有已识别的变化子图进行聚类, 以推断潜在环境。其直觉在于: 不变子图捕获了预测性图结构信息与标签之间的不变关系, 而变化子图刻画了不同分布下的变化相关性, 因此可作为环境判别性特征。最后, 不变性学习模块在已识别的不变子图和推断出的环境基础上, 优化所提出的最大不变子图生成准则, 以生成在分布偏移下具备 OOD 泛化能力的图表示。理论分析表明, 图上的 OOD 泛化问题等价于寻找 GIL 的最大不变子图生成器, 并进一步证明了 GIL 满足置换不变性。

C2R (Cooperative Classification and Rationalization) [61] 进一步提出了一种协同框架, 将分类模块与理据学习模块进行集成。通过在整个数据集上对非理据子图进行聚类, C2R 能够推断全局环境而非局部环境, 并将其用于增强分类性能, 从而在分布偏移下提升 OOD 泛化能力。

DIR (Discovering Invariant Rationale) [62] 用于通过在干预分布下发现不变子图 $\Phi(G)$ 来处理图级别的 OOD 泛化任务。DIR 的基本设定同样不同于传统假设环境可观测且可获取的设定, 而是与 GIL 类似, 不假设事先存在显式的环境划分。具体而言, 它使用一个基于 GNN 的子图生成器, 在分布偏移下将输入图划分为不变子图和变化子图, 并分别由编码器将其编码为表示。随后, 该方法提出的分布干预模块在变化表示上实施干预, 从而构造多个干预分布作为多个环境。最后, 分别基于不变子图和变化子图构建的两个分类器联合对输入图实例进行预测, 从而在不同环境中最小化不变风险。通过该策略, DIR 能够捕捉在不同分布下保持稳定的不变理据, 同时过滤掉虚假模式, 以实现 OOD 泛化。

GSAT (Graph Stochastic Attention) [63] 的工作聚焦于利用注意力机制来解决图级别的 OOD 泛化问题, 通过构建内在可解释的 GNN, 在分布偏移下学习不变子图 $\Phi(G)$ 。GSAT 所学习的不变子图根植于信息瓶颈理论 [120]。该方法通过向注意力机制中注入随机性, 将注意力建模为信息瓶颈, 从而约束从输入图到预测结果的信息流。在训练过程中, 与标签相关的不变子图上的随机性会被自动减小, 而与标签无关的变化子图上的随机性则会被保留。此外, GSAT 还对来自输入图数据的信息量进行惩罚。最终, 在一定假设下, GSAT 能够输出可解释且具有 OOD 泛化能力的子图, 并可证明这些子图不包含与任务虚假相关的模式。

UIL (Unified Invariant Learning) [65] 提出了一种统一的图 OOD 泛化框架, 通过联合约束结构不变性与语义不变性来实现。该方法通过节点级与边级掩码分离稳定特征与环境特征; 在此基础上, 通过估计稳定 graphon 以捕捉各类别固有的结构模式。通过最小化跨环境的 graphon 距离并约束标签一致的预测, 它能够更准确地识别最小稳定特征。

VIVACE (Variance Contrastive Estimation) [64] 强调了变化子图的重要性, 这类子图携带环境相关信息, 往往在以往工作中被忽视。该方法提出利用变化子图来估计虚假相关性, 并引导不变子图的识别。同时, 引入基于逆倾向评分的重加权机制来校正虚假效应, 从而提升 OOD 泛化性能。

除了图级别的 OOD 泛化方法外, **EERM** (Explore-to-Extrapolate Risk Minimization) [66] 被设计用于处理分布偏移下的节点级任务, 在较弱条件下即可为节点级 OOD 问题提供有效解。首先, 为了应对图中节点非独立同分布的特性, 该方法提出将图转换为以中心节点为核心的一组自我中心图 (ego-graphs), 从而借鉴图级问题来刻画节点级 OOD 泛化问题。随后, 它通过在自我中心图的 BFS 树上进行递归计算, 将不变性原理扩展至结构信息建模。最后, GNN 主干网络通过最小化环境生成器生成的多个训练环境中的风险均值与方差进行优化; 与此同时, 环境生成器则通过策略梯度方法最大化方差损失进行训练。

INL (Invariant Node Representation Learning) [67] 方法同样基于不变性学习原理来应对具有多个潜在环境的图数据分布偏移问题。通过将不变模式与变化模式定义为自我中心子图, INL 采用基于对比模块度的图聚类来推断节点环境。随后, 它对最大不变模式准则进行优化, 以生成能够有效泛化到未知分布的节点表示。理论分析为其性能提供了保障, 同时在合成数据集与真实数据集上的实验均表明, 该方法在 OOD 的节点分类任务中相比现有最优方法取得了显著提升。

FLOOD (Flexible Invariant Learning for Out-Of-Distribution Generalization) [68] 进一步引入了一种双重策略: 不变表示学习与自举表示学习。通过构造训练环境并在测试阶段进一步优化共享编码器, FLOOD 实现了更强的 OOD 泛化能力。该框架能够有效应对传导式与归纳式设定下的分布偏移问题。

GraphMETRO (Graph Mixture-of-Experts for OOD generalization) [69] 通过利用专家混合 (MoE) 框架来应对分布偏移。每个专家针对一种特定的偏移成分 (如图规模、节点度或特征噪声) 进行训练, 并通过随机图变换进行缓解。门控模型能够动态识别每个输入所对应的偏移成分, 并引导专家输出的聚合以形成不变表示。该架构支持对异质分布偏移的灵活建模。

DIDA (Disentangled Intervention-based Dynamic Graph Attention Network) [70] 是首个用于处理更复杂时空分布偏移下图 OOD 泛化问题的方法。现有方法通常仅关注节点特征或图结构中的空间分布偏移, 难以直接应用于同时存在空间与时间分布偏移的复杂场景。具体而言, 该方法首先设计了一种解耦的时空注意力网络, 用于发现动态图中潜在的不变模式与变化模式, 使得每个节点能够通过解耦的注意力消息传递机制关注其所有历史邻居。随后, 引入时空干预机制, 通过在邻域与时间维度上对变化模式进行采样与重组, 构造多个干预分布, 从而消除变化模式与标签之间的虚假相关性。需要注意的是, 变化模式在节点之间高度纠缠, 若直接对结构与特征的子集进行生成与混合干预, 将带来较高的计算开销。因此, 该方法使用由解耦时空注意力网络获得的汇总模式来近似干预过程, 而非直接操作原始结构与特征。最后, 该方法利用不变性正则化来最小化模型在多个干预分布下的预测方差, 以此学习具有不变性的模式。

除了上述方法之外, **SILD** (Spectral Invariant Learning for Dynamic Graphs) [71] 将 DIDA 扩展到谱域, 通过发现不变与变化的谱模式来处理动态图中的分布偏移。**EAGLE** (Environment-Aware Dynamic Graph Learning) [72] 通过对时空潜在环境进行建模, 并结合环境感知卷积与解耦机制, 解决动态图上的 OOD 泛化问题。EAGLE 方法能够在缓解虚假相关性的同时建模不变模式, 并在分布偏移下取得了良好的泛化性能。

5.1.2 显式表示对齐

显式表示对齐方向工作的核心思想是通过在多个环境 (或域) 之间显式对齐图表示, 学习环境不变的图表示, 从而实现 OOD 泛化。图表示对齐通常通过引入正则化策略来最小化不同环境之间的差异或增强其相似性, 其形式化表述为:

$$\min_{f_\theta} \mathbb{E}_{X,Y} [\ell(f_\theta(X), Y)] + \ell_{reg}(\mathcal{E}), \quad (4)$$

其中 $\ell_{reg}(\mathcal{E})$ 表示所采用的正则化损失项。在大多数图学习场景中, 用于计算正则项的多个环境 $\mathcal{E}$ 通常无法事先获得, 而是需要在训练过程中构造或生成。

SR-GNN (Shift-Robust GNN) [73] 通过显式最小化有偏训练数据与图真实推断分布之间的分布差异, 解决 GNN 中的节点级 OOD 泛化问题。该方法通过引导带偏训练节点样本在分布特性上更好地对齐图中独立同分布 (i.i.d.) 采样节点的统计特征, 从而缓解由分布偏移引起的性能退化。SR-GNN 能够同时处理在深层 GNN 模型以及近年来提出的线性化 (浅层) GNN 模型中普遍存在的两类偏置问题。具体而言, 对于依赖图结构进行迭代信息传播的标准 GNN 模型 (如 GCN [121]), SR-GNN 在网络的隐藏层中引入分布对齐正则项以缓解分布偏移, 其中用于度量不同分布间差异的正则化指标可以采用最大均值差异 (Maximum Mean Discrepancy, MMD) [122] 或中心矩差异 (Central Moment Discrepancy, CMD) [123]。而对于将 GNN 解耦

为非线性特征编码与线性消息传递两部分的线性化模型 (如 SimpleGCN [124]), 由于图结构会在所有可学习层之后对特征分布引入偏置, SR-GNN 进一步引入实例重加权策略, 以鼓励训练样本在整体图数据分布上具有更好的代表性。该方法通过核均值匹配 (Kernel Mean Matching, KMM) [125] 学习一组最优的实例权重, 从而实现训练分布与目标分布之间的有效对齐。

SizeShiftReg [74] 旨在训练能够从小规模图泛化到大规模图的 GNN, 其基本思想与 SR-GNN [73] 相似。该方法不依赖于基于特定先验知识或假设的人为模型设计, 而是研究一种通用的正则化策略, 使任意 GNN 均能够在图规模分布发生偏移时实现 OOD 泛化。具体而言, SizeShiftReg 通过引入图粗化 (graph coarsening) 策略来模拟训练阶段图规模的分布变化, 并在此基础上设计相应的正则化项, 以鼓励 GNN 学习到具有更强 OOD 泛化能力的表示。对于给定的训练图, 该方法最小化 GNN 在原始训练图与其粗化版本上所学习到的节点表示分布之间的差异, 该差异通过中心矩差异 (CMD) [123] 进行度量。在上述训练范式下, 所学习到的 GNN 不仅能够在不同粗化版本之间实现 OOD 泛化, 还能够推广至规模未知的新图。

StableGL [75] 关注稳定图学习 (Stable Graph Learning), 旨在捕获环境不变的节点属性, 并通过显式平衡多个环境来实现分布偏移下的稳健泛化性能。给定一个输入图作为训练环境, 该方法的目标是训练一个在多个未知测试环境中具有较高平均预测性能、且预测性能方差较低的 GNN。具体而言, 该方法首先对输入训练图进行有偏采样, 以构造多个训练环境。从局部视角来看, 由于图中的节点可以由其邻居节点部分表征, 该方法通过对邻域聚合过程进行重加权来捕获稳定的节点属性。从全局视角来看, 在有偏训练过程中, 不同环境下的预测误差可能逐渐发散, 从而导致跨环境的不稳定性能。因此, 该方法通过显式减少不同训练环境之间的训练差距, 对训练过程进行对齐, 从而约束所学习的 GNN 在未知测试环境中保持良好的泛化能力。不同于 SR-GNN [73] 和 SizeShiftReg [74] 采用 MMD 或 CMD 等分布差异度量方式, StableGL 的正则化目标是直接最小化多个环境下训练损失的方差。

5.2 图对抗训练

本节讨论用于 OOD 泛化的图对抗学习方法。已有研究表明, 对抗训练不仅能够提升模型对对抗攻击的鲁棒性, 同时也有助于增强其 OOD 泛化能力。本文主要关注以提升泛化性能为目标的图对抗训练方法, 而针对 GNN 抗攻击性的相关研究可参考已有综述 [22]。

DAGNN (Domain Adversarial GNN) [76] 是一种受 DANN [126] 启发的方法, 通过在域分类器与编码器之间引入域对抗学习, 来学习域 (或环境) 不变的图表示, 从而实现 OOD 泛化。具体而言, 该方法包含两个目标: 一是最小化编码器在源域数据上的分类损失; 二是促进域分类器区分源域与目标域。通过这种图对抗训练策略, 模型能够充分利用域信息, 从而学习具备 OOD 泛化能力的分类器。需要注意的是, 该方法最初应用于文本分类任务, 其中图由文档构建, 因此数据集中显式提供了域 (或环境) 划分。

GNN-DRO [77] 采用分布鲁棒优化 (Distributionally Robust Optimization, DRO) [127] 框架, 这是一类用于应对分布偏移的经典方法, 主要针对节点级任务。该方法假设真实数据分布位于以经验分布为中心的 Wasserstein 球内, 并通过最小化该 Wasserstein 球中最坏情况下的期望损失来训练 GNN, 从而提升模型在分布变化下的鲁棒性。

除上述直接将通用 OOD 方法扩展至图数据的研究外, 还有一些工作更加充分地考虑了图数据自身的结构特性。

GraphAT (Graph Adversarial Training) [78] 通过在图上引入对抗训练来提升模型的泛化能力。在生成目标样本的对抗扰动时, GraphAT 通过最大化目标样本与其邻居样本预测结果之间的差异, 使对抗扰动尽可能破坏图的平滑性; 随后, 在参数更新阶段最小化图对抗正则项, 从而减小扰动后目标样本与其邻居节点之间的预测差异。此外, 该方法基于反向传播推导了一种线性近似策略, 以高效计算对抗扰动。通过抵御最坏情

况下的扰动, GraphAT 能够显著增强模型的鲁棒性与泛化性能。

CAP (Co-Adversarial Perturbation) [79] 从训练过程中损失景观 (loss landscape) 的角度出发, 指出 GNN 容易在模型权重和特征空间中陷入局部极小值。为此, 该方法提出协同对抗扰动优化策略, 通过交替平坦化权重和特征的损失景观, 避免模型收敛至泛化能力较差的局部最优解。具体而言, CAP 将训练目标表述为在模型权重与节点特征扰动邻域内最小化最大训练损失。为提高训练效率, 该方法将目标函数进行解耦, 并采用交替对抗扰动策略: 一步对模型权重施加对抗扰动并更新 GNN, 另一步计算节点特征的对抗扰动以进一步更新模型参数。

WT-AWP (Weighted Truncated Adversarial Weight Perturbation) [80] 延续了通过平坦化局部极小值以提升 OOD 图数据泛化能力的研究思路。由于直接将现有对抗权重扰动技术应用于 GNN 训练时容易受到梯度消失问题的影响, WT-AWP 将对抗权重扰动对应的损失作为额外正则项, 与原始损失函数 (如标准交叉熵) 共同用于训练 GNN。此外, 该方法在 GNN 的最后一层中移除了扰动, 以实现训练动态的更精细控制。除训练策略设计外, 作者还进一步推导了针对 OOD 图分类任务的泛化误差上界。

OAD (Online Adversarial Distillation) [81] 是一种用于 GNN 的在线对抗知识蒸馏方法。不同于前述直接将对抗训练引入 GNN 训练过程的方法, OAD 通过对抗机制缓解知识蒸馏过程中引入的问题。受知识蒸馏能够提升 OOD 泛化性能的启发, OAD 以在线方式训练一组学生 GNN, 并利用全局与局部知识进行指导。通过迁移教师网络中的有效知识, 学生模型的 OOD 泛化能力得以提升。为更好地学习复杂的局部结构信息, 该方法进一步引入对抗循环学习机制, 实现学生模型之间更精确的嵌入对齐。相比传统知识蒸馏方法, OAD 在参数规模更小的同时具有更高的训练效率, 并且在应对图分布偏移方面表现更为有效。

5.3 图自监督学习

最后, 我们介绍用于 OOD 泛化的图自监督学习方法。

作为一种新兴技术, 自监督学习已被广泛应用于图像领域, 用于训练具有更强泛化能力的神经网络 [128–130]。已有研究同样表明, 自监督学习也能够帮助 GNN 获得更强的泛化能力 [131], 其动机主要体现在以下几个方面。一方面, 自监督学习任务鼓励 GNN 捕获输入图中的关键信息, 同时避免监督学习中对“捷径”特征的平凡过拟合, 从而有助于提升 OOD 泛化能力。另一方面, Xu *et al.* [132] 指出, 自监督学习能够将语义相近的数据映射至相似的表示空间, 因此部分 OOD 测试样本在映射后可能落入训练分布内部。

本节主要回顾那些明确声称能够提升图 OOD 泛化能力的代表性图自监督学习方法。关于图自监督学习的更多细节, 读者可参考相关综述 [20, 21]。

Pretraining-GNN [82] 探索了多种图预训练技术, 涵盖节点级和图级任务, 以提升 GNN 的 OOD 泛化能力。该方法鼓励 GNN 在学习图级知识的同时, 捕获关于节点和边的领域特定信息, 从而使学习到的表示具有更强的 OOD 泛化能力。在节点级预训练方面, 作者提出了两种自监督任务, 即上下文预测和属性掩码。在图级预训练方面, 同样设计了两类任务, 包括预测整图的领域特定属性 (如监督标签) 以及预测图结构 (即建模两张图之间的结构相似性)。总体而言, 该预训练策略首先进行节点级自监督预训练, 随后进行图级多任务监督预训练。

PATTERN [83] 通过提出一种用于学习 d -pattern 表示的自监督前置任务, 研究了 GNN 从小规模图泛化到大规模图的能力。尽管 GNN 在形式上可以直接应用于不同规模的图, 但其在图规模 OOD 泛化方面的内在机制仍不清晰。为此, 作者首先形式化了一种称为 d -pattern 的局部结构表示, 用于刻画模型对新图规模的泛化能力。 d -pattern 将节点度的概念推广到以中心节点为核心的 d 跳邻域, 用于建模 GNN 所感知的节点及其 d 跳邻居的取值分布。理论分析表明, 即使测试分布与训练分布在 d -pattern 分布上仅存在较小差异, 也可能导致模型权重分配无法良好泛化, 从而揭示了泛化性能较差的全局最小值的存在。基于此, 作者提出自监督

预文本任务,从小规模图与大规模图中联合学习 d -pattern 表示,并在图规模 OOD 泛化任务上取得了显著性能提升。

DR-GST (Distribution Recovered Graph Self-Training) [84] 是一种在存在分布偏移条件下恢复原始标注数据分布的图自训练框架。该方法首先证明:在自训练框架中,若为每个伪标注节点分配合适的权重系数,则在分布偏移下的损失函数等价于总体分布下的损失函数。由于该权重系数难以直接计算,作者发现其与信息增益具有相同的变化趋势,因此采用信息增益进行近似。信息增益分别通过 dropout 变分推断和 dropedge 变分推断进行估计。随后,通过引入信息增益加权的损失函数,可以恢复发生偏移的分布,并促使 GNN 关注具有较高信息增益的节点。总体而言,DR-GST 从信息增益的角度处理分布偏移问题,并提出了一种损失修正策略以提升伪标签质量,从而为更多未标注节点分配与标注节点分布一致的伪标签,进而提升 OOD 泛化能力。

此外,图对比学习 (Graph Contrastive Learning) 也可用于提升模型的 OOD 泛化能力。

GraphCL (Graph Contrastive Learning) [33] 是图神经网络中具有代表性的自监督学习方法之一,并已在实践中展示出良好的泛化能力。作者指出,基于人工设计预任务的自监督学习方法依赖启发式规则构造,这种方式可能限制所学习图表示的通用性。为此,他们提出了对比学习方法 GraphCL,其核心思想是在对输入图施加四种不同变换的情况下,使得到的图表示保持一致。GraphCL 的泛化能力在化学领域的分子性质预测任务以及生物领域的蛋白质功能预测任务中得到了验证。

RGCL (Rationale-aware Graph Contrastive Learning) [85] 通过在对图学习框架中自动发现用于图增强的关键子结构,进一步提升模型在存在分布偏移的未知域上的泛化性能。作者认为,尽管 GraphCL 等方法取得了较好效果,但其内在随机性可能导致语义信息的潜在丢失,从而削弱泛化能力。为解决这一问题,RGCL 引入关键子结构生成器和对比学习器两个模块。其中,关键子结构生成器决定图中需要保留和隐藏的部分,并生成能够封装实例判别信息的关键子结构;对比学习器则利用这些感知关键子结构的视图进行图实例判别。通过这种方式,RGCL 能够避免随机增强导致的判别语义丢失,从而在增强视图中保留更多具有泛化能力的关键信息。

GAPGC (Graph Adversarial Pseudo Group Contrast) [86] 是一种面向图神经网络的测试时训练方法,在测试阶段采用对比损失的变体作为自监督学习目标。近年来,测试时训练在提升模型 OOD 测试性能方面的有效性已得到验证,其中通常引入自监督辅助任务。作者指出,传统自监督训练中常用的简单图增强策略(如随机删除节点或边)可能破坏与标签相关的关键信息,从而影响图表示质量。为此,GAPGC 通过生成相对可靠的伪标签,避免由错误正样本引入的严重分布偏移。此外,该方法提出的对抗式可学习增强器以及分组伪正样本机制增强了自监督任务与主任务之间的相关性,从而提升了主任务性能。作者还从信息论角度给出了理论分析,表明 GAPGC 能够为主任务捕获最小充分信息,有助于提升模型在 OOD 测试数据上的预测表现。

GT3 (Graph Test-Time Training with Constraint) [87] 是另一种面向图数据的测试时训练方法,提出了一种分层自监督学习框架。具体而言,该方法首先引入全局对比学习策略,通过最大化局部节点表示与全局图表示之间的互信息,促使节点表示捕获全图的全局信息。随后,通过局部对比学习区分同一图在不同增强视图下的节点,使节点表示进一步捕获局部结构细节。此外,该方法还引入额外约束项,旨在缩小测试样本与训练样本表示之间的距离。依托这种结合自监督学习的测试时训练策略,模型在图分类任务中的 OOD 泛化能力得到了显著提升。

HomoTTT (Homophily-guided Fully Test-Time Training) [88] 是一个针对 OOD 环境下节点分类任务的模型无关框架。该方法利用基于同质性的无参数自监督对比学习目标及自适应图增强,实现完全测试时训练 (Fully Test-Time Training)。为防止性能退化,该框架进一步引入基于同质性的模型选择机制,对每个节点选择性地应用适配后的模型。

6 理论分析

本节回顾从理论角度分析图神经网络泛化能力的相关研究工作。

首先,基于不同的统计学习理论,已有多种理论方法被提出,主要用于推导图神经网络的泛化界(generalization bound)。**Scarselli et al.** [133] 基于 VC 维理论 [134] 给出了图神经网络的泛化界。作者发现, GNN 的 VC 维上界与循环神经网络的上界具有可比性,这表明随着连通节点数量的增加, GNN 的泛化能力也会随之增强。**Verma & Zhang** [135] 基于算法稳定性理论 [136], 对 GCN [121] 进行了进一步的理论分析, 并给出了单层 GCN 的泛化界。其结论表明, 当图卷积滤波器满足稳定性条件时, 单层 GCN 可以满足强一致稳定性, 从而具备良好的泛化性能。

Garg et al. [137] 基于 Rademacher 复杂度分析了 GNN 在图分类任务上的泛化性质。该工作显式考虑了 GNN 聚合函数的局部置换不变性。相较于 [133] 中基于 VC 维得到的泛化界, 该研究所推导的 Rademacher 泛化界在理论上更加紧致。在此基础上, **Lv** [138] 进一步推导了单隐层 GCN 的 Rademacher 复杂度界。二者的主要区别在于, [138] 专门针对节点级任务进行分析, 并假设邻接矩阵是固定不变的。

Liao et al. [139] 建立了图神经网络在图分类任务上的 PAC-Bayesian 泛化界。该结果在 [137] 基于 Rademacher 复杂度的分析基础上进行了改进, 使所得泛化界在对图中最大节点度数和模型最大隐层维度的依赖上更加紧致。此外, **Ma et al.** [140] 针对非 I.D. 的节点级任务, 在节点子群体层面给出了 GNN 泛化性能的 PAC-Bayesian 分析, 这是其相较于 [139] 的关键区别。

Du et al. [141] 提出了图神经切线核 (Graph Neural Tangent Kernel, GNTK), 用于刻画 GNN 在图分类任务上的泛化性质。GNTK 由无限宽的 GNN 所诱导, 其预测仅依赖于图之间的成对核值, 并且可以通过解析形式高效计算。该方法在保留 GNN 表达能力的同时, 继承了图核方法易于训练且具有理论保证的优点。在此基础上, **Xu et al.** [132] 给出了单层 GNN 泛化能力的理论分析, 并研究了网络结构与目标算法任务之间的对齐程度对 OOD 泛化性能的影响。沿着这一研究方向, **Zhang et al.** [142] 证明, 在合适的张量初始化和加速梯度下降条件下, 其算法能够学习到一个单隐层 GNN, 并在回归任务中实现零泛化误差, 从而在假设真实模型存在的条件下逼近真实模型。

鉴于上述多数方法通常假设图数据可以以任意方式生成和标注, 而这一假设在实际应用中往往难以满足, 部分研究进一步建立了依赖于特定图数据生成机制的泛化界。**Baranwal et al.** [143] 在上下文随机块模型这一特定生成机制下研究了 GNN 的 OOD 泛化问题, 分析了线性可分性与图 OOD 泛化之间的关系, 并推导了单层 GCN 在二分类节点任务上的泛化保证。此外, **Maskey et al.** [144] 采用 graphon 作为图生成模型, 该模型在理论上具有更强的表达能力和普适性, 并能够得到更紧的泛化界。

除了直接推导泛化界之外, 还有一些理论框架从因果性、不变学习以及信息瓶颈等角度对 OOD 泛化能力进行分析。

因果推断通过关注输入特征与标签之间稳定的因果关系, 为提升 OOD 泛化能力提供了坚实的理论基础。与对分布偏移高度敏感的伪相关不同, 因果特征在不同环境中保持不变, 从而为 OOD 场景下的预测提供了可靠依据。结构化因果模型 (SCMs) [49, 51, 54] 以及因果干预方法 [55] 为识别和利用这些因果特征提供了理论工具, 使模型能够捕获决定标签的真实因素。此外, 反事实推理 [58] 通过考虑假设性情形, 进一步增强了模型在未知分布下的泛化能力。这些方法共同强调了因果性在应对分布偏移以及建立泛化理论基础中的重要作用。

不变学习通过识别并利用在不同环境中与标签保持稳定关系的特征, 为 OOD 泛化提供了一种系统性的理论框架 [117–119]。该框架通常假设输入数据可分解为不变成分和变化成分, 其中不变成分是目标的稳定预测因子, 而变化成分则对应于与环境相关的伪相关因素。其核心思想是在优化预测性能的同时, 约束模型在不同训练环境下保持不变性, 从而使预测结果与数据背后的稳定因果机制对齐。从理论上讲, 不变学习依赖

于不变性原理, 即在不同环境中, 给定不变特征时标签的条件分布应保持一致。这一原理通常通过跨环境风险最小化目标或显式约束表示空间对齐的正则化方式来实现 [60, 62, 67]。通过关注不变模式并抑制变化模式, 不变学习不仅提升了 OOD 泛化能力, 还在某些假设条件下 (如环境多样性充分 [72] 或数据中存在潜在不变性 [70, 71]) 提供了相应的理论保证, 为图 OOD 泛化奠定了理论基础。

信息瓶颈理论同样被用于广义图学习中的 OOD 泛化分析。其核心思想是在压缩与任务无关的图结构信息的同时, 最大化任务相关子图表示与标签之间的互信息。例如, GSAT [63] 通过联合训练预测器和子图提取器, 并利用随机注意力机制实现信息控制。InfoIGL [145] 引入冗余过滤器并结合多层次对比学习, 从而提取图中的不变特征, 在增强同类图之间互信息的同时降低任务无关噪声的影响。基于信息瓶颈的优化目标在理论上能够保证去除伪相关因素, 从而提升模型在 OOD 场景下的泛化能力。

7 用于评估的数据集

为了促进图的 OOD 泛化的进一步研究, 我们在表 4 中总结了现有流行的图数据集用于评估。

表 4 用于图分布外泛化研究的常用合成数据集和真实世界图数据集。“Task”表示各数据集可用于图级、节点级或链接级任务; “Type”表示各数据集所包含的图数据类型; “Cause of Shifts”表示训练数据与测试数据之间产生分布偏移的原因; “Metric”表示各数据集采用的性能评价指标; “References”表示提出或构建相应数据集的相关工作。

Table 4 Commonly used synthetic and real-world graph datasets for OOD generalization. “Task” denotes each dataset can be used in graph-level, node-level task or link-level task. “Type” indicates what kind of graph data that each dataset includes. “Cause of Shifts” indicates the reason for inducing distribution shifts between training and testing data. “Metric” is the evaluation metric adopted by each dataset. And “References” denotes the work developing each dataset.

Dataset	Type	Task	Cause of Shifts	Metric	References
Spurious-Motif	Synthetic graph	Graph	Correlations	Accuracy	[62]
MNIST-75sp	Superpixel graph	Graph	Feature noises	Accuracy	[146]
CMNIST-75sp	Superpixel graph	Graph	Feature colors	Accuracy	[92, 147]
D&D ₂₀₀	Molecular graph	Graph	Graph size	Accuracy	[146]
Graph-SST2	Text sentiment	Graph	Node degree	Accuracy	[148]
OGBG-Molhiv	Molecular graph	Graph	Scaffold	ROC-AUC	[7]
OGBG-Molpcba	Molecular graph	Graph	Scaffold	Average precision	[7]
OGBG-PPA	Protein network	Graph	Species	Accuracy	[7]
DrugOOD	Molecular graph	Graph	Assay/Scaffold/Size	Accuracy/AUC	[149]
CBA-Shapes	Synthetic graph	Node	Feature colors	Accuracy	[147]
Facebook-100	Social network	Node	Structure	Accuracy	[66]
WebKB	Webpage network	Node	Structure	Accuracy	[147]
Twitch-Explicit	Social network	Node	Structure	ROC-AUC	[150]
Elliptic	Bitcoin transactions	Node	Time	F1 Score	[151]
OGBN-Arxiv	Citation network	Node	Time	Accuracy	[7]
OGBN-Proteins	Protein network	Node	Species	ROC-AUC	[7]
OGBN-Products	Co-purchasing	Node	Popularity	Accuracy	[7]
COLLAB	Collaboration network	Link	Field	ROC-AUC	[70]
Yelp	Social network	Link	Food category	ROC-AUC	[70]
ACT	Social network	Link	Attribute	ROC-AUC	[72]
OGBL-PPA	Protein network	Link	Biological throughput	Hits@100	[7]
OGBL-DDI	Drug interaction network	Link	Protein-target	Hits@20	[7]

这些数据集分为三类任务: 图级、节点级和边级任务的数据集。这些数据集涵盖了多种图来源 (例如社交网络、引用网络、分子图等), 其分布偏移的原因也复杂多样 (例如时间、物种、骨架等)。

7.1 图级任务数据集

首先,我们回顾一些用于评估图分类任务模型性能的代表性数据集。

Spurious-Motif [62]: 这是一个合成数据集,按照 [152] 的工作创建,用于研究图结构上的分布偏移。每个图由一个模体 (motif) 和一个基图组成。基图包括 Tree、Ladder 和 Wheel 三类 (分别记为 $S = 0, 1, 2$), 而模体包括 Cycle、House 和 Crane 三类 (分别记为 $I = 0, 1, 2$)。真实标签 Y 仅依赖于模体 I , 且 I 服从均匀分布。通过控制基础子图的分布方式向数据中注入 S 与 Y 之间的虚假相关性: 当 $S = I$ 时, $P(S) = b$; 当 $S \neq I$ 时, $P(S) = (1 - b)/2$ 。直观地, 参数 b 控制虚假相关性的强度。通过在训练集与测试集中设置不同的 b 值, 可以模拟分布偏移场景。

MNIST-75sp [146]: 这是一个半人工构造的数据集, 其中每个图由 MNIST [153] 图像通过超像素方法 [154] 转换而来。节点对应超像素, 边根据节点之间的空间距离计算得到。节点特征包括超像素的坐标与像素强度。任务目标是将每个图分类为对应的手写数字 (0-9)。为模拟图特征层面的分布偏移, 该数据集通过对测试图像进行颜色化处理生成测试样本, 即在原始图像基础上增加两个通道, 并向每个通道添加独立的高斯噪声来实现。

CMNIST-75sp [92, 147]: 该数据集同样是一个半人工构造数据集, 由 MNIST 图像通过超像素方法转换为图结构。不同于 MNIST-75sp 通过噪声模拟分布偏移, CMNIST-75sp 借鉴工作 [117], 根据数字标签或数据集划分对数字进行颜色化处理。需要注意的是, CMNIST-75sp 提供了两种设置, 分别用于模拟协变量偏移和概念偏移。在前一种设置中, 测试数据使用训练数据中未出现过的颜色; 在后一种设置中, 训练数据中颜色与数字标签相关, 而测试数据中颜色与标签之间的相关性发生变化。

D&D₂₀₀ [146]: 这是一个真实世界的图分类数据集, 包含 1178 个蛋白质网络结构, 每个节点具有 82 个离散标签。任务目标是判断每个图是否对应酶 (enzyme) 或非酶 (non-enzyme)。为引入图规模层面的分布偏移, 训练集与测试集按照图的大小进行划分, 即模型在小规模图上训练, 而在更大规模图上测试。具体而言, 训练集包含节点数在 30 至 200 之间的图, 而测试集包含节点数在 201 至 5748 之间的图。

Graph-SST2 [148]: 这是一个来源于自然语言情感分析任务的真实世界图数据集。每个图由一段文本序列转换而来, 其中节点表示单词, 边表示单词之间的关系, 图标签对应句子的情感极性。通过按照平均节点度对图进行划分来构造分布偏移。节点特征由预训练的 BERT 词向量 [155] 初始化。得益于其明确的语义结构, 该数据集在可视化或中间结果分析方面具有较强的可解释性。

OGBG [7]: Open Graph Benchmark (OGB) 是一个面向图机器学习的基准测试平台, 包含一系列真实、大规模且多样化的数据集, 其中 OGBG 是其用于图级任务 OOD 泛化评估的一个子集, 代表性数据集包括 OGBG-Molhiv、OGBG-Molpcba 和 OGBG-PPA 等。具体而言, **OGBG-Molhiv** 和 **OGBG-Molpcba** 是两个具有分布偏移的图属性预测数据集, 任务目标是预测分子的目标属性。数据集提供了默认的骨架划分策略, 即根据分子的二维结构骨架对图进行划分。该策略旨在将结构差异显著的分子划分至不同子集, 从而构造更加真实且具有挑战性的图 OOD 泛化评估场景。**OGBG-PPA** 由来自 1581 个不同物种的蛋白质-蛋白质关联网络中提取的无向蛋白质关联邻域图构成, 任务目标是预测给定蛋白质关联邻域图所属的分类学类别。该数据集采用按物种划分的方式, 将不同物种来源的图分配至不同子集。

DrugOOD [149]: 这是一个面向 AI 辅助药物发现的基准数据集, 包含多个真实的分子图数据集。该基准基于大规模生物测定数据集 ChEMBL [156], 提供了一套自动化流水线用于构建 OOD 数据集。相比 OGB 仅使用分子骨架作为数据划分指标, DrugOOD 引入了更加多样化的数据划分标准, 以生成符合生物化学领域知识的特定域。除 scaffold 外, 该基准还可基于实验测定 (assay) 和分子规模等因素对图进行划分, 从而构造更加丰富的分布偏移场景。

7.2 节点级任务数据集

接着,我们回顾一些用于评估节点分类任务模型性能的代表性数据集。

CBA-Shapes [147]: 这是一个合成的数据集,它沿用 BA-Shapes 数据集的生成方式 [152]。输入图由一个基图和一组模体 (motifs) 构成,其中基图为一个包含 300 个节点的 Barabási-Albert (BA) 图,模体集合包含 80 个房屋结构模体。任务目标是预测每个节点的结构角色,包括房屋结构模体中的顶部节点、中部节点或底部节点,或属于基图的节点,即一个四分类问题。为了构造分布偏移,节点特征通过颜色进行赋值,并设计了两种设置以分别模拟协变量偏移和概念偏移。在协变量偏移设置中,测试节点被赋予在训练阶段未出现过的颜色;而在概念偏移设置中,颜色与训练节点标签之间存在相关性,但在测试阶段,颜色与标签之间的相关关系发生改变。

Facebook-100 [66]: 这是一个真实节点分类数据集,由 2005 年的 100 个 Facebook 社交网络快照组成。每个网络包含来自美国某一所特定大学的 Facebook 用户节点。通过按照用户所属的不同大学来划分训练集与测试集,可以自然地引入分布偏移,因为不同大学对应的社交网络在规模、网络密度以及度分布等方面均存在显著差异。例如,**Facebook-100** [66] 中采用的默认数据划分方式是:将来自 14 所大学中的 3 所(如 John Hopkins、Cornell 等)的网络作为训练集,并将来自另外 3 所大学(即 Penn、Brown 和 Texas)的网络作为测试集。当然,也可以采用其他大学组合来评估模型在节点层面的 OOD 泛化能力。

WebKB [147]: 这是一个真实大学网页网络数据集,用于节点分类。节点为网页,边为超链接,节点特征为网页文本词向量。任务是预测网页类别,包括学生、项目、课程、职员或教师。分布偏移通过按所属大学对数据集进行划分而引入。在仅使用网页的文本内容和超链接结构、而不依赖大学相关特征的情况下,模型需要实现对分布外数据的泛化预测。

Twitch-Explicit [150]: 这是一个真实世界的社交网络数据集,其中节点表示 Twitch 用户,边表示用户之间的好友关系。节点特征包括用户偏好的游戏、地理位置以及直播行为习惯。每个网络对应于一个特定地区,包括 DE、ENGB、ES、FR、PTBR、RU 和 TW。不同地区的网络在结构属性上存在显著差异,例如网络密度和最大节点度等 [66]。训练集与测试集之间的分布偏移通过按照网络所属地区进行数据划分而产生。

Elliptic [151]: 这是一个真实场景下的比特币交易网络数据集,由多个时间快照构成,其中节点表示交易,边表示交易之间的支付流。任务目标是在未来数据中区分合法交易与非法交易。通过将时间较早的快照作为训练集、时间较新的快照作为测试集,可以由于市场中新出现的事件而自然观察到分布偏移。

OGBN [7]: 该数据集包含多个节点属性预测任务数据集,如 **OGBN-Arxiv**、**OGBN-Proteins** 和 **OGBN-Products**,它是完整 OGB [7] 数据集的一个子集。具体而言,**OGBN-Arxiv** 是一个真实世界的引文网络数据集,其中节点表示 arXiv 论文,边表示论文之间的引用关系。其 40 类分类任务旨在预测论文的研究领域。节点层面的分布偏移通过按照不同时间区间划分论文作为训练集和测试集而引入。**OGBN-Proteins** 是一个蛋白质图数据集,其中节点表示蛋白质,边表示蛋白质之间不同类型的、具有生物学意义的关联关系。任务目标是预测蛋白质功能的存在性。分布偏移通过按照蛋白质所属物种对节点进行划分而产生。**OGBN-Products** 是一个 Amazon 商品共购网络,其中节点表示 Amazon 平台上的商品,边表示两种商品被共同购买的关系。任务目标是预测商品类别。分布偏移通过一种更具挑战性且更贴近真实场景的数据划分方式构造,即使用高流行度商品作为训练集,而将相对低流行度商品作为测试集。

7.3 边级任务数据集

此外,我们回顾一些用于评估边预测任务模型性能的代表性数据集。

COLLAB [70]: 这是一个源自学术合作网络的边预测数据集。节点表示作者,边表示作者在 1990 至 2006 年间共同发表论文所形成的合著关系。该数据集进一步引入了领域相关的信息,根据合著论文所属的研究领域

对边进行分类,例如“数据挖掘(Data Mining)”、“数据库(Database)”、“医学信息学(Medical Informatics)”、“理论(Theory)”和“可视化(Visualization)”,数据集覆盖16个按年份划分的时间切片,刻画了学术合作关系随时间的演化过程。分布偏移通过按合著论文的研究领域进行划分而引入,其中数据挖掘(Data Mining)被作为训练阶段不可见的领域,从而构造了一个具有挑战性的分布外(OOD)链路预测测试场景。

Yelp [70]: 这是一个真实世界的边预测数据集,来源于用户与商家之间的交互记录。节点对应于用户和商家,边表示随时间发生的评论交互行为。数据集涵盖2019年1月至2020年12月期间的数据。交互关系按照商家类别进行标注,例如披萨类(Pizza)、美国食物类(American Food)、咖啡与茶类(Coffee & Tea)、寿司类(Sushi Bars)和快餐类(Fast Food)。分布偏移通过将涉及披萨(Pizza)类别的交互保留为测试域而引入,该工作为评估模型在真实分布偏移场景下的泛化能力提供了一个现实案例。

ACT [72]: 这是一个记录了大规模在线课程(MOOC)平台中学生行为的动态变化的数据集。节点表示学生,边表示学生的行为活动,例如课程参与或学习材料的交互。数据集中跟踪了多种行为类别,包括“课程观看(Lecture Viewing)”、“作业提交(Assignment Submissions)”和“论坛参与(Forum Participation)”,从而刻画不同类型的交互模式。分布偏移通过在训练阶段排除特定类别的行为而构造,使模型在测试阶段需要泛化至未见过的学生行为模式,从而对其泛化能力提出挑战。

OGBL-PPA [7]: 这是一个来源于蛋白质-蛋白质关联网络的真实世界图数据集。节点表示来自58个不同物种的蛋白质,边表示具有生物学意义的蛋白质关联关系,包括物理相互作用、共表达、同源性以及基因组邻域等。每个节点均关联一个58维的one-hot特征向量,用以指示其物种来源。该数据集聚焦于边预测任务,其目标是将真实存在的蛋白质-蛋白质关联在排序中置于随机采样的负边之前。评估指标采用Hits@100,用于衡量真实正样本边在排序结果前100名中的比例。分布偏移通过一种基于生物实验通量的划分策略引入:训练边来源于成本较低的高通量实验方法或计算技术,而验证集与测试集中的边则由低通量、资源密集型的实验室实验所确认,从而模拟现实生物实验条件下的分布差异。

OGBL-DDI [7]: 这是一个源自药物-药物相互作用网络的真实世界图数据集。节点表示已获FDA批准或处于实验阶段的药物,边表示两种药物联合使用时,其总体效应相较于各自独立作用显著偏离的相互作用关系。任务目标是通过排序的方式预测新的药物-药物相互作用,即将已知真实相互作用的边排在约100,000条随机采样的负边之前。评估指标采用Hits@20,用于衡量真实相互作用在排序结果前20名中的比例,从而构成一个具有较高难度的评测基准。该数据集采用蛋白靶点划分策略引入分布偏移:训练集与验证集中的药物靶向同一组蛋白,而测试集中的药物靶向完全不同的蛋白。这种划分方式确保模型在评测时需要泛化至具有不同生物作用机制的药物,符合药物发现中真实存在的OOD场景。

7.4 其他基准

此外,还有一些工作将上述常用或多个数据集整合为标准化的开源评测基准,并在其提出的评测协议下,对若干经典的通用OOD泛化算法以及图OOD方法进行了系统性的实验评估。鉴于大多数数据集的具体细节已在前文中进行了详细介绍,这里对这些评测基准进行简要回顾。具体而言,**GDS** [92] 汇集了八个用于图级任务的数据集,涵盖了多种类型的图分布偏移情形,用以比较主流OOD泛化算法与不同GNN骨干模型在分布外场景下的性能表现。**GOOD** [147] 总结并整理了十余个同时涵盖图级任务与节点级任务的数据集,这些数据集通过结合不同的域划分策略与分布偏移类型,引入了多样化的分布偏移设置。该基准还系统性地报告了模型在同分布与分布外场景下的显著性能差距,并对通用机器学习领域与图学习领域中的多种OOD方法进行了全面对比分析。

8 图基础模型中的应用

近年来,随着基础模型在自然语言处理和计算机视觉等领域取得显著进展 [157, 158],图基础模型(Graph Foundation Models, GFMs)也逐渐成为图机器学习中的重要研究方向 [159, 160]。与传统图神经网络通常针对特定数据集或特定任务进行训练不同,图基础模型旨在通过大规模、多来源、多任务的图数据预训练,学习可迁移、可复用且具有更强泛化能力的图表示。由于真实应用中的图数据常常存在结构、领域、模态和任务层面的分布偏移,图基础模型为提升图 OOD 泛化能力提供了新的研究范式。

从 OOD 泛化的角度看,图基础模型主要面临四类分布偏移。首先是结构层面的分布偏移,即不同图数据之间可能在节点特征、边连接模式、图规模、度分布、motif 结构以及关系语义等方面存在显著差异。其次是领域层面的分布偏移,即不同图数据集可能来自不同应用领域或采集机制,从而导致数据偏差、标注偏差以及虚假相关性的差异。第三是模态层面的分布偏移,尤其是在文本属性图、多模态图或分子图等场景中,辅助模态的可用性、质量和对齐关系可能在训练和测试阶段发生变化。最后是任务层面的分布偏移,即模型在部署时可能需要处理与预训练阶段不同的监督形式、输出空间或任务粒度,例如从节点分类迁移到链接预测或图级预测。

现有图基础模型在处理上述 OOD 泛化问题时,大体可以分为两类。第一类是面向同构任务设定的图基础模型 [161–163],即预训练和下游推理通常围绕固定任务形式展开,例如节点分类、链接预测或图分类。在这一设定下,任务形式和输出空间基本保持一致,OOD 泛化主要体现为模型在结构偏移、领域偏移或模态偏移下仍能保持稳定表现。相关方法通常通过多图预训练、跨域表示对齐、结构不变表示学习、对比学习、专家混合机制或轻量化适配策略,提升模型在未见图或未见领域中的迁移能力。第二类是面向异构任务设定的图基础模型 [164–166],即模型不仅需要适应不同图数据分布,还需要支持不同任务形式之间的迁移。此类方法通常试图构建统一的图表示空间或统一的任务接口,使节点级、边级和图级任务能够在同一框架下表达和求解。代表性的技术思路包括基于提示学习或指令学习的任务统一、基于文本语义的图任务描述、图结构与语言模型的对齐、通用图 tokenization,以及面向多任务的预训练目标设计。与同构任务设定相比,异构任务图基础模型需要同时应对结构、领域、模态和任务层面的复合分布偏移,因此对模型的表示能力、任务适配能力和跨域稳定性提出了更高要求。

图 OOD 泛化方法与图基础模型之间具有天然联系。一方面,OOD 泛化研究中强调的不变表示学习、因果建模、分布鲁棒优化和自监督学习等思想,可以为图基础模型提供更稳健的预训练目标和适配机制,减少模型对训练数据中特定结构模式或领域偏差的依赖。另一方面,图基础模型通过大规模预训练和多任务学习,能够扩大训练分布的覆盖范围,为缓解未知分布偏移提供更丰富的先验知识和可迁移表示。因此,将图 OOD 泛化方法融入图基础模型的预训练、对齐、提示和微调过程中,是提升图基础模型可靠性和实用性的重要方向。

尽管图基础模型为图 OOD 泛化提供了新的机遇,但其在真实部署中仍面临若干挑战。首先,如何构建能够覆盖结构、领域、模态和任务多样性的预训练语料仍不明确。其次,现有图基础模型的预训练目标并不一定直接对应 OOD 泛化目标,甚至可能在某些场景下学习到领域特定的虚假相关性。再次,针对图基础模型的 OOD 评测仍缺乏统一标准,未来需要更系统地设计覆盖结构偏移、领域偏移、模态缺失和任务迁移的评测协议。最后,图基础模型在分布偏移下的理论理解仍然有限,如何刻画大规模预训练表示在未见图、未见领域和未见任务中的泛化边界,将是后续研究的重要问题。

9 讨论

在本节中,我们对本文的综述内容进行总结,并围绕未来值得深入探索的若干挑战与机遇展开讨论。

9.1 总结

图任务训练数据的多样性与质量对于图机器学习方法的分布外泛化性能至关重要。当前,研究者已提出多种图数据增强技术,涵盖结构级、特征级以及混合范式等不同类别。这些方法通常具有实现简洁、开销可控等优势,并在提升模型泛化能力方面取得了显著效果。

不同于上述基于图数据增强的思路,另一类研究聚焦于通过设计新型图模型来提升 OOD 泛化能力。相较于数据增强方法,这些模型往往具有更明确的理论动机与更具针对性的图结构建模机制。其中,基于解耦的图模型提供了较为清晰的动机引导,而基于因果的图模型则依托丰富的因果推断理论工具。这些面向特定偏移机理的模型设计在实践中也展现出有希望的 OOD 泛化性能。

近期,围绕训练过程本身的图学习策略亦发展迅速,包括图不变性学习、图对抗训练与图自监督学习等。与直接改造模型结构的方法相比,这类方法更强调学习过程的设计,因此在提升 OOD 泛化能力时通常能够更灵活地兼容不同的 GNN 骨干架构。

在理论层面,已有多项工作尝试构建图泛化的理论框架,以帮助更深入理解图 OOD 泛化方法的有效性边界与适用条件。在评测层面,该领域已采用覆盖图级与节点级任务的多样化数据集,并通过模拟复杂且贴近真实场景的分布偏移,为验证各类 OOD 泛化方法的有效性提供了相对全面且公正的基准。

9.2 方法选择的实践建议

在系统梳理图 OOD 泛化方法之后,一个自然的问题是:在实际应用中应如何根据具体场景选择合适的方法。总体而言,方法选择不仅取决于模型在基准数据集上的平均性能,还应综合考虑分布偏移的来源、训练数据规模、环境标签是否可获得、是否允许修改模型结构、部署阶段的计算成本,以及应用场景对可解释性和稳定性的要求。结合本文提出的数据、模型和学习策略三类方法,可以得到如下实践性结论。

首先,当主要问题来自训练数据覆盖不足,且希望在不显著改变模型结构的前提下提升泛化性能时,数据层面的方法通常是较为直接和易于部署的选择。图数据增强方法可以通过扰动节点特征、修改局部结构或构造混合样本来扩展训练分布,因而适用于训练样本有限、测试偏移形式相对可预期、且已有 GNN 骨干网络较为成熟的场景。其优势在于实现简单、模型无关性较强,并且可以与多种下游模型结合;但其局限在于增强策略本身需要与真实分布偏移相匹配。如果增强操作不能合理反映测试环境中的变化,甚至破坏了与标签相关的关键结构,则可能引入额外噪声并削弱模型性能。

其次,当分布偏移主要源于虚假相关、潜在混杂因素或图生成机制变化时,模型层面的方法更具针对性。基于解耦表示或因果建模的图模型能够显式区分稳定因素与不稳定因素,因此更适合用于医疗、金融、分子性质预测等对可靠性和可解释性要求较高的风险敏感场景。这类方法的优势在于具有较明确的建模动机,并有助于解释模型为何能够在分布偏移下保持稳定;但其通常依赖更强的先验假设或更复杂的模型设计,在缺乏领域知识或因果假设难以验证的场景中需要谨慎使用。

第三,当已有模型架构不便大幅修改,或者研究者希望通过训练目标和优化策略来提升泛化能力时,学习策略层面的方法通常更具灵活性。图不变性学习、图对抗训练和图自监督学习等方法可以作为通用训练范式,与不同 GNN 骨干网络结合。若能够获得多个训练环境,基于不变性学习的方法更适合挖掘跨环境稳定关系;若标注数据有限或希望提升表示的可迁移性,自监督学习和预训练策略通常更具优势;若部署场景存在扰动、攻击或局部噪声,对抗训练可用于提升模型鲁棒性。此类方法的主要挑战在于训练目标的设计往往会影响到稳定性与判别性之间的平衡,过强的不变性约束也可能抑制对任务有用的环境特异信息。

最后,在复杂真实应用中,单一类别的方法往往难以充分应对多源分布偏移。更可行的策略是根据偏移类型和部署需求组合使用不同方法。例如,在结构和特征偏移同时存在的场景中,可以先通过数据增强扩大训练分布覆盖范围,再通过不变性学习或自监督学习约束表示稳定性;在风险敏感任务中,可以进一步引入因果

建模或解耦结构,以提升模型的可信性和可解释性;在跨领域、跨任务或多模态图学习场景中,图基础模型可作为新的选择,但仍需结合具体任务评估其是否真正缓解了结构、领域、模态或任务层面的分布偏移。总体而言,图 OOD 泛化方法的选择应从“偏移来源—可用信息—模型约束—部署需求”这一链条出发,而不应仅依据单一性能指标进行判断。

9.3 未来研究方向

尽管近年来图 OOD 泛化研究取得了显著进展,但在方法、理论与应用层面仍存在大量尚未解决的问题,亟需进一步探索。

9.3.1 更多的理论保证

尽管部分图 OOD 泛化方法在实验评测中已取得显著进展,但其与第 6 节所讨论的理论基础之间仍存在明显脱节,而弥合这一差距有赖于对可学习的图 OOD 泛化问题进行更严格且更系统的理论刻画。此外,深入理解不同类型分布偏移的作用机制同样至关重要,例如协变量偏移、概念偏移与标签偏移等;这些偏移在图结构数据中往往以复杂方式相互交织。现有研究已在处理特定偏移类型方面取得初步成果,但仍缺乏统一的理论框架。未来研究有必要探索能够同时刻画多种分布偏移情形的图 OOD 泛化理论,例如同时建模节点特征偏移、图结构偏移、图规模偏移与标签分布偏移之间的相互作用,进一步回答哪些图属性在多重偏移下仍然可学习、可辨识且可泛化,并结合泛化界、因果推断、不变性学习或信息瓶颈等理论工具,以进一步提升方法的可解释性与可靠性。

9.3.2 GNN 架构

近期工作 [132, 146, 167–169] 指出, GNN 的架构设计(例如读出函数或聚合机制)在实现 OOD 泛化中发挥着关键作用。这些研究为理解 GNN 架构与分布偏移之间的交互提供了基础性见解。为了更系统地提升 GNN 的 OOD 泛化能力,未来可以针对每个图实例自动定制适合的 GNN 架构 [170],以增强分布偏移下的预测表现,这将是一个有前景的方向。例如,一个具体问题是如何根据图规模、同质性程度、局部结构模式或潜在偏移类型,自适应地选择消息传递深度、邻域聚合机制和读出函数。此外,仍需更多研究探索如何自动学习能够适应多样化环境的 OOD 泛化 GNN 架构,以减少人工设计带来的局限性。

9.3.3 环境划分

大多数通用 OOD 泛化算法依赖于访问多个训练环境 [15]。然而,为真实图数据获取准确的环境标签通常成本高昂,从而限制了这类方法的适用性。未来研究可进一步探索单环境 OOD 泛化方法,利用图结构与特征的异质性动态学习环境划分。其中一个关键挑战是如何区分由真实环境变化导致的分布差异与由噪声、采样偏差或图稀疏性导致的局部扰动,并避免自动构造的伪环境进一步强化虚假相关性。此外,现实图数据往往随时间演化,需要模型适应动态或连续环境。已有关于终身学习与持续图学习的工作 [171, 172] 为开发能够高效更新图模型与学习策略、以应对时间分布偏移的方法提供了基础。将这些方法扩展至在未知分布偏移下动态演化的图学习问题,仍是一个具有潜力且尚未充分探索的研究方向。

9.3.4 测试时训练以提升泛化能力

图测试时训练(Test-Time Training)为提升未知分布偏移下的 OOD 泛化能力提供了一种灵活的范式,其通过在推理阶段利用未标注测试数据并引入测试时任务,实现对模型的自适应调整。除第 5.3 节介绍的基

于对比学习的测试时训练方法 [86, 87] 外, 近期工作 **GTrans** [173] 提出了在测试阶段自适应调整图数据的策略。此外, **LEBED** [174] 采用无参数重训练机制, 通过度量节点预测与结构重建之间的差异, 估计训练良好的 GNN 在分布偏移条件下对未标注测试图的泛化误差。未来仍需进一步研究如何设计不会破坏已学习不变表示的测试时更新目标, 如何避免测试时自适应过程中的错误累积, 以及如何在大规模图或动态图场景中实现高效更新。总体而言, 设计更丰富的测试时训练任务或探索新的测试阶段自适应策略, 是进一步提升图 OOD 泛化能力的一个重要研究方向。

9.3.5 更广泛的应用场景

分布外 (OOD) 图数据在日常生活中广泛存在。尽管经典的图机器学习方法已在多种应用场景中取得成功, 但在存在分布偏移的真实环境中部署具备 OOD 泛化能力的图学习方法, 仍然是一个关键且尚未得到充分研究的挑战。诸如推荐系统、社交网络、交通预测、材料科学, 以及医疗和金融等风险敏感领域, 不仅对预测精度提出要求, 还对决策过程的可靠性与可信性具有更高标准 [175–180]。因此, 未来研究需要进一步围绕具体应用场景提出更有针对性的技术问题, 例如推荐系统中的用户兴趣漂移、交通网络中的时空结构变化、分子图中的支架分布偏移, 以及医疗和金融图中的跨机构、跨人群分布差异。因此, 将领域知识有效融入图 OOD 泛化方法, 被认为是提升模型在实际应用中泛化能力与决策可信度的一条潜在而有前景的途径。

参考文献

- 1 Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. Deepinf: Social influence prediction with deep learning. In *KDD*, 2018.
- 2 Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 2022.
- 3 Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE TKDE*, 2017.
- 4 Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *IJCAI*, 2018.
- 5 Yoshua Bengio et al. A meta-transfer objective for learning to disentangle causal mechanisms. *ICLR*, 2019.
- 6 Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Ood-gnn: Out-of-distribution generalized graph neural network. *IEEE TKDE*, 2022.
- 7 Weihua Hu et al. Open graph benchmark: Datasets for machine learning on graphs. *NeurIPS*, 2020.
- 8 Shuo Yang et al. Financial risk analysis for smes with graph-based supply chain mining. In *IJCAI*, 2020.
- 9 Chirag Agarwal, Himabindu Lakkaraju, and Marinka Zitnik. Towards a unified framework for fair and stable graph representation learning. In *UAI*, 2021.
- 10 Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *ECCV*, 2020.
- 11 Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2020.
- 12 George Panagopoulos et al. Transfer graph neural networks for pandemic forecasting. *AAAI*, 2021.

- 13 Michael J Horry et al. Covid-19 detection through transfer learning using multimodal imaging data. *IEEE Access*, 2020.
- 14 Kanglin Hsieh et al. Drug repurposing for covid-19 using graph neural network and harmonizing multiple evidence. *Scientific reports*, 2021.
- 15 Zheyang Shen, Jiashuo Liu, Yue He, Xingxuan Zhang, Renzhe Xu, Han Yu, and Peng Cui. Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*, 2021.
- 16 Jindong Wang et al. Generalizing to unseen domains: A survey on domain generalization. *IJCAI*, 2021.
- 17 Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE TPAMI*, 2022.
- 18 Kaize Ding, Zhe Xu, Hanghang Tong, and Huan Liu. Data augmentation for deep graph learning: A survey. *arXiv preprint arXiv:2202.08235*, 2022.
- 19 Tong Zhao, Gang Liu, Stephan Günnemann, and Meng Jiang. Graph data augmentation for graph machine learning: A survey. *arXiv preprint arXiv:2202.08871*, 2022.
- 20 Yixin Liu et al. Graph self-supervised learning: A survey. *IEEE TKDE*, 2022.
- 21 Yaochen Xie, Zhao Xu, Jingtun Zhang, Zhengyang Wang, and Shuiwang Ji. Self-supervised learning of graph neural networks: A unified review. *IEEE TPAMI*, 2022.
- 22 Lichao Sun, Yingdong Dou, Carl Yang, Kai Zhang, Ji Wang, S Yu Philip, Lifang He, and Bo Li. Adversarial attack and defense on graph data: A survey. *IEEE TKDE*, 2022.
- 23 Liang Chen, Jintang Li, Jiaying Peng, Tao Xie, Zengxu Cao, Kun Xu, Xiangnan He, and Zibin Zheng. A survey of adversarial learning on graphs. *arXiv preprint arXiv:2003.05730*, 2020.
- 24 Jindong Wang et al. Generalizing to unseen domains: A survey on domain generalization. *IEEE TKDE*, 2022.
- 25 Jingjing Li, Zhiqi Yu, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE TPAMI*, 2024.
- 26 Xin Wu et al. Out-of-distribution generalization in time series: A survey. *arXiv preprint arXiv:2503.13868*, 2025.
- 27 Tong Zhao, Yozen Liu, Leonardo Neves, Oliver Woodford, Meng Jiang, and Neil Shah. Data augmentation for graph neural networks. *AAAI*, 2021.
- 28 Hyeonjin Park et al. Metropolis-hastings data augmentation for graph neural networks. *NeurIPS*, 34, 2021.
- 29 Lirong Wu, Haitao Lin, Yufei Huang, and Stan Z Li. Knowledge distillation improves graph structure augmentation for graph neural networks. In *NeurIPS*, 2022.
- 30 Wenzheng Feng et al. Graph random neural networks for semi-supervised learning on graphs. *NeurIPS*, 2020.
- 31 Kezhi Kong, Guohao Li, Mucong Ding, Zuxuan Wu, Chen Zhu, Bernard Ghanem, Gavin Taylor, and Tom Goldstein. Robust optimization as data augmentation for large-scale graphs. In *CVPR*, 2022.
- 32 Songtao Liu et al. Local augmentation for graph neural networks. In *ICML*, 2022.
- 33 Yuning You et al. Graph contrastive learning with augmentations. *NeurIPS*, 2020.
- 34 Gang Liu et al. Graph rationalization with environment-based augmentations. In *KDD*, 2022.

- 35 Junchi Yu, Jian Liang, and Ran He. Mind the label shift of augmentation-based graph ood generalization. In *CVPR*, 2023.
- 36 Yongduo Sui et al. Unleashing the power of graph data augmentation on covariate distribution shift. In *NeurIPS*, 2023.
- 37 Yun Zhu, Haizhou Shi, Zhenshuo Zhang, and Siliang Tang. Mario: Model agnostic recipe for improving ood generalization of graph contrastive learning. In *WWW*, 2024.
- 38 Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *ICLR*, 2018.
- 39 Jianxin Ma, Peng Cui, Kun Kuang, Xin Wang, and Wenwu Zhu. Disentangled graph convolutional networks. In *ICML*, 2019.
- 40 Yanbei Liu, Xiao Wang, Shu Wu, and Zhitao Xiao. Independence promoted graph disentangled networks. In *AAAI*, 2020.
- 41 Yiding Yang, Zunlei Feng, Mingli Song, and Xinchao Wang. Factorizable graph convolutional networks. *NeurIPS*, 2020.
- 42 Shaohua Fan, Xiao Wang, Yanhu Mo, Chuan Shi, and Jian Tang. Debiasing graph neural networks via learning disentangled causal substructure. In *NeurIPS*, 2022.
- 43 Xiaojie Guo, Liang Zhao, Zhao Qin, Lingfei Wu, Amarda Shehu, and Yanfang Ye. Interpretable deep graph generation with node-edge co-disentanglement. In *KDD*, 2020.
- 44 Haoyang Li, Xin Wang, Ziwei Zhang, Zehuan Yuan, Hang Li, and Wenwu Zhu. Disentangled contrastive learning on graphs. *NeurIPS*, 2021.
- 45 Haoyang Li, Ziwei Zhang, Xin Wang, and Wenwu Zhu. Disentangled graph contrastive learning with independence promotion. *IEEE TKDE*, 2022.
- 46 Haoyang Li, Xin Wang, Zeyang Zhang, Haibo Chen, Ziwei Zhang, and Wenwu Zhu. Disentangled graph self-supervised learning for out-of-distribution generalization. In *ICML*, 2024.
- 47 Shaohua Fan, Xiao Wang, Chuan Shi, Peng Cui, and Bai Wang. Generalizing graph neural networks on out-of-distribution graphs. *arXiv preprint arXiv:2111.10657*, 2021.
- 48 Shaohua Fan, Xiao Wang, Chuan Shi, Kun Kuang, Nian Liu, and Bai Wang. Debaised graph neural networks with agnostic label selection bias. *TNNLS*, 2022.
- 49 Yongduo Sui, Xiang Wang, Jiancan Wu, Min Lin, Xiangnan He, and Tat-Seng Chua. Causal attention for interpretable and generalizable graph classification. *KDD*, 2022.
- 50 Yingxin Wu et al. Deconfounding to explanation evaluation in graph neural networks. *arXiv preprint arXiv:2201.08802*, 2022.
- 51 Yongqiang Chen et al. Learning causally invariant representations for out-of-distribution generalization on graphs. In *NeurIPS*, 2022.
- 52 Tianjun Yao, Yongqiang Chen, Zhenhao Chen, Kai Hu, Zhiqiang Shen, and Kun Zhang. Empowering graph invariance learning with deep spurious infomax. *ICML*, 2024.
- 53 Weizhi An et al. Causal subgraphs and information bottlenecks: Redefining ood robustness in graph neural networks. In *ECCV*, 2024.

- 54 Xiner Li, Shurui Gui, Youzhi Luo, and Shuiwang Ji. Graph structure extrapolation for out-of-distribution generalization. In *ICML*, 2024.
- 55 Qitian Wu, Fan Nie, Chenxiao Yang, Tianyi Bao, and Junchi Yan. Graph out-of-distribution generalization via causal intervention. In *WWW*, 2024.
- 56 Beatrice Bevilacqua, Yangze Zhou, and Bruno Ribeiro. Size-invariant graph representations for graph classification extrapolations. In *ICML*, 2021.
- 57 Yangze Zhou, Gitta Kutyniok, and Bruno Ribeiro. Ood link prediction generalization capabilities of message-passing gnns in larger test graphs. *NeurIPS*, 2022.
- 58 Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, and Meng Jiang. Learning from counterfactual links for link prediction. In *ICML*, pages 26911–26926, 2022.
- 59 Wanyu Lin, Hao Lan, and Baochun Li. Generative causal explanations for graph neural networks. In *ICML*, 2021.
- 60 Haoyang Li, Ziwei Zhang, Xin Wang, and Wenwu Zhu. Learning invariant graph representations under distribution shifts. In *NeurIPS*, 2022.
- 61 Linan Yue, Qi Liu, Ye Liu, Weibo Gao, Fangzhou Yao, and Wenfeng Li. Cooperative classification and rationalization for graph generalization. In *WWW*, 2024.
- 62 Ying-Xin Wu, Xiang Wang, An Zhang, Xiangnan He, and Tat seng Chua. Discovering invariant rationales for graph neural networks. In *ICLR*, 2022.
- 63 Siqi Miao, Mia Liu, and Pan Li. Interpretable and generalizable graph learning via stochastic attention mechanism. In *ICML*, 2022.
- 64 Haoyang Li, Xin Wang, Xueling Zhu, Weigao Wen, and Wenwu Zhu. Disentangling invariant subgraph via variance contrastive estimation under distribution shifts. In *ICML*, 2025.
- 65 Yongduo Sui, Jie Sun, Shuyao Wang, Zemin Liu, Qing Cui, Longfei Li, and Xiang Wang. A unified invariant learning framework for graph classification. *KDD*, 2025.
- 66 Qitian Wu, Hengrui Zhang, Junchi Yan, and David Wipf. Handling distribution shifts on graphs: An invariance perspective. In *ICLR*, 2022.
- 67 Haoyang Li, Ziwei Zhang, Xin Wang, and Wenwu Zhu. Invariant node representation learning under distribution shifts with multiple latent environments. *ACM TOIS*, 2023.
- 68 Yang Liu et al. Flood: A flexible invariant learning framework for out-of-distribution generalization on graphs. In *KDD*, 2023.
- 69 Shirley Wu, Kaidi Cao, Bruno Ribeiro, James Y Zou, and Jure Leskovec. Graphmetro: Mitigating complex graph distribution shifts via mixture of aligned experts. *NeurIPS*, 2024.
- 70 Zeyang Zhang, Xin Wang, Ziwei Zhang, Haoyang Li, Zhou Qin, and Wenwu Zhu. Dynamic graph neural networks under spatio-temporal distribution shift. In *NeurIPS*, 2022.
- 71 Zeyang Zhang et al. Spectral invariant learning for dynamic graphs under distribution shifts. *NeurIPS*, 2023.
- 72 Haonan Yuan, Qingyun Sun, Xingcheng Fu, Ziwei Zhang, Cheng Ji, Hao Peng, and Jianxin Li. Environment-aware dynamic graph learning for out-of-distribution generalization. *NeurIPS*, 2023.

- 73 Qi Zhu, Natalia Ponomareva, Jiawei Han, and Bryan Perozzi. Shift-robust gnns: Overcoming the limitations of localized graph training data. *NeurIPS*, 2021.
- 74 Davide Buffelli, Pietro Liò, and Fabio Vandin. Sizeshiftreg: a regularization method for improving size-generalization in graph neural networks. *NeurIPS*, 2022.
- 75 Shengyu Zhang et al. Stable prediction on graphs with agnostic distribution shift. *arXiv preprint arXiv:2110.03865*, 2021.
- 76 Man Wu, Shirui Pan, Xingquan Zhu, Chuan Zhou, and Lei Pan. Domain-adversarial graph neural networks for text classification. In *ICDM*, 2019.
- 77 Alireza Sadeghi, Meng Ma, Bingcong Li, and Georgios B Giannakis. Distributionally robust semi-supervised learning over graphs. *arXiv preprint arXiv:2110.10582*, 2021.
- 78 Fuli Feng, Xiangnan He, Jie Tang, and Tat-Seng Chua. Graph adversarial training: Dynamically regularizing based on graph structure. *IEEE TKDE*, 2019.
- 79 Haotian Xue, Kaixiong Zhou, Tianlong Chen, Kai Guo, Xia Hu, Yi Chang, and Xin Wang. Cap: Co-adversarial perturbation on weights and features for improving generalization of graph neural networks. *arXiv preprint arXiv:2110.14855*, 2021.
- 80 Yihan Wu, Aleksandar Bojchevski, and Heng Huang. Adversarial weight perturbation improves generalization in graph neural networks. *AAAI*, 2023.
- 81 Can Wang, Zhe Wang, Defang Chen, Sheng Zhou, Yan Feng, and Chun Chen. Online adversarial distillation for graph neural networks. *arXiv preprint arXiv:2112.13966*, 2021.
- 82 Weihua Hu et al. Strategies for pre-training graph neural networks. *ICLR*, 2020.
- 83 Gilad Yehudai, Ethan Fetaya, Eli Meir, Gal Chechik, and Hagai Maron. From local structures to size generalization in graph neural networks. In *ICML*, 2021.
- 84 Hongrui Liu, Binbin Hu, Xiao Wang, Chuan Shi, Zhiqiang Zhang, and Jun Zhou. Confidence may cheat: Self-training on graph neural networks under distribution shift. *WWW*, 2022.
- 85 Sihang Li, Xiang Wang, An Zhang, Yingxin Wu, Xiangnan He, and Tat-Seng Chua. Let invariant rationale discovery inspire graph contrastive learning. In *ICML*, 2022.
- 86 Guanzi Chen, Jiyang Zhang, Xi Xiao, and Yang Li. Graphtta: Test time adaptation on graph neural networks. *arXiv preprint arXiv:2208.09126*, 2022.
- 87 Yiqi Wang, Chaozhuo Li, Wei Jin, Rui Li, Jianan Zhao, Jiliang Tang, and Xing Xie. Test-time training for graph neural networks. *arXiv preprint arXiv:2210.08813*, 2022.
- 88 Jiaxin Zhang, Yiqi Wang, Xihong Yang, and En Zhu. A fully test-time training framework for semi-supervised node classification on out-of-distribution graphs. *TKDD*, 2024.
- 89 W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.
- 90 Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 2001.
- 91 Ali Shafahi, Mahyar Najibi, Mohammad Amin Ghiasi, Zheng Xu, John Dickerson, Christoph Studer, Larry S Davis, Gavin Taylor, and Tom Goldstein. Adversarial training for free! *NeurIPS*, 32, 2019.

- 92 Mucong Ding et al. A closer look at distribution shifts and out-of-distribution generalization on graphs. *NeurIPS Workshop*, 2021.
- 93 Vikas Verma et al. Manifold mixup: Better representations by interpolating hidden states. In *ICML*, 2019.
- 94 Linjun Zhang, Zhun Deng, Kenji Kawaguchi, Amirata Ghorbani, and James Zou. How does mixup help with robustness and generalization? In *ICLR*, 2021.
- 95 Hongyu Guo. Nonlinear mixup: Out-of-manifold data augmentation for text classification. In *AAAI*, 2020.
- 96 Vikas Verma et al. Graphmix: Improved training of gnn for semi-supervised learning. In *AAAI*, 2021.
- 97 Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, and Bryan Hooi. Mixup for node and graph classification. In *WWW*, 2021.
- 98 Lirong Wu, Jun Xia, Zhangyang Gao, Haitao Lin, Cheng Tan, and Stan Z Li. Graphmixup: Improving class-imbalanced node classification by reinforcement mixup and self-supervised context prediction. In *ECML-PKDD*, 2022.
- 99 Hongyu Guo and Yongyi Mao. Intrusion-free graph mixup. *arXiv preprint arXiv:2110.09344*, 2021.
- 100 Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Juncheng Liu, and Bryan Hooi. Nodeaug: Semi-supervised node classification with data augmentation. In *KDD*, 2020.
- 101 Xiaotian Han, Zhimeng Jiang, Ninghao Liu, and Xia Hu. G-mixup: Graph data augmentation for graph classification. *ICML*, 2022.
- 102 Bin Lu, Ze Zhao, Xiaoying Gan, Shiyu Liang, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. Graph out-of-distribution generalization with controllable data augmentation. *TKDE*, 2024.
- 103 Milton Llera Montero, Casimir JH Ludwig, Rui Ponte Costa, Gaurav Malhotra, and Jeffrey Bowers. The role of disentanglement in generalisation. In *ICLR*, 2020.
- 104 Andrea Dittadi et al. On the transfer of disentangled representations in realistic settings. *arXiv preprint arXiv:2010.14407*, 2020.
- 105 Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. *NeurIPS*, 2007.
- 106 Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A survey on contrastive self-supervised learning. *Technologies*, 2020.
- 107 Phuc H Le-Khac, Graham Healy, and Alan F Smeaton. Contrastive representation learning: A framework and review. *IEEE Access*, 2020.
- 108 Bernhard Schölkopf et al. Toward causal representation learning. *Proceedings of the IEEE*, 2021.
- 109 Kun Kuang et al. Stable prediction across unknown environments. In *KDD*, 2018.
- 110 Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *NeurIPS*, 2007.
- 111 Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2019.
- 112 Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. *NeurIPS*, 2018.
- 113 Madelyn Glymour, Judea Pearl, and Nicholas P Jewell. *Causal inference in statistics: A primer*. 2016.

- 114 Alexander Balke and Judea Pearl. Probabilistic evaluation of counterfactual queries. *AAAI*, 1994.
- 115 Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. 2018.
- 116 Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 1969.
- 117 Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.
- 118 Shiyu Chang et al. Invariant rationalization. In *ICML*, 2020.
- 119 Kartik Ahuja et al. Invariance principle meets information bottleneck for out-of-distribution generalization. *NeurIPS*, 2021.
- 120 Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *IEEE information theory workshop*, 2015.
- 121 Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- 122 Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *ICML*, 2015.
- 123 Werner Zellinger et al. Robust unsupervised domain adaptation for neural networks via moment alignment. *Information Sciences*, 2019.
- 124 Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *ICML*, 2019.
- 125 Arthur Gretton et al. Covariate shift by kernel mean matching. *Dataset shift in machine learning*, 2009.
- 126 Yaroslav Ganin et al. Domain-adversarial training of neural networks. *JMLR*, 2016.
- 127 Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.
- 128 Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. *NeurIPS*, 32, 2019.
- 129 Divyat Mahajan, Shruti Tople, and Amit Sharma. Domain generalization using causal matching. In *ICML*, 2021.
- 130 Michael Zhang et al. Correct-n-contrast: A contrastive approach for improving robustness to spurious correlations. *arXiv preprint arXiv:2203.01517*, 2022.
- 131 Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. When does self-supervision help graph convolutional networks? In *ICML*, 2020.
- 132 Keyulu Xu et al. How neural networks extrapolate: From feedforward to graph neural networks. *ICLR*, 2021.
- 133 Franco Scarselli, Ah Chung Tsoi, and Markus Hagenbuchner. The vapnik–chervonenkis dimension of graph and recursive neural networks. *Neural Networks*, 2018.
- 134 Vladimir N Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of complexity*. 2015.
- 135 Saurabh Verma and Zhi-Li Zhang. Stability and generalization of graph convolutional neural networks. In *KDD*, 2019.

- 136 Olivier Bousquet and André Elisseeff. Stability and generalization. *JMLR*, 2002.
- 137 Vikas Garg, Stefanie Jegelka, and Tommi Jaakkola. Generalization and representational limits of graph neural networks. In *ICML*, 2020.
- 138 Shaogao Lv. Generalization bounds for graph convolutional neural networks via rademacher complexity. *arXiv preprint arXiv:2102.10234*, 2021.
- 139 Renjie Liao, Raquel Urtasun, and Richard Zemel. A pac-bayesian approach to generalization bounds for graph neural networks. *ICLR*, 2021.
- 140 Jiaqi Ma, Junwei Deng, and Qiaozhu Mei. Subgroup generalization and fairness of graph neural networks. *NeurIPS*, 34, 2021.
- 141 Simon S Du et al. Graph neural tangent kernel: Fusing graph neural networks with graph kernels. *NeurIPS*, 2019.
- 142 Shuai Zhang, Meng Wang, Sijia Liu, Pin-Yu Chen, and Jinjun Xiong. Fast learning of graph neural networks with guaranteed generalizability: one-hidden-layer case. In *ICML*, 2020.
- 143 Aseem Baranwal et al. Graph convolution for semi-supervised classification: Improved linear separability and out-of-distribution generalization. *ICML*, 2021.
- 144 Sohir Maskey, Ron Levie, Yunseok Lee, and Gitta Kutyniok. Generalization analysis of message passing neural networks on large random graphs. In *NeurIPS*, 2022.
- 145 Wenyu Mao et al. Invariant graph learning meets information bottleneck for out-of-distribution generalization. *FCS*, 2025.
- 146 Boris Knyazev, Graham W Taylor, and Mohamed Amer. Understanding attention and generalization in graph neural networks. *NeurIPS*, 32, 2019.
- 147 Shurui Gui, Xiner Li, Limei Wang, and Shuiwang Ji. GOOD: A graph out-of-distribution benchmark. In *NeurIPS Datasets and Benchmarks Track*, 2022.
- 148 Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE TPAMI*, 2022.
- 149 Yuanfeng Ji et al. Drugood: Out-of-distribution (ood) dataset curator and benchmark for ai-aided drug discovery. *arXiv preprint arXiv:2201.09637*, 2022.
- 150 Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 2021.
- 151 Aldo Pareja et al. Evolvegc: Evolving graph convolutional networks for dynamic graphs. In *AAAI*, 2020.
- 152 Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. *NeurIPS*, 2019.
- 153 Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998.
- 154 Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI*, 2012.
- 155 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, 2019.

- 156 David Mendez et al. Chembl: towards direct deposition of bioassay data. *Nucleic acids research*, 2019.
- 157 Rishi Bommasani. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.
- 158 Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, et al. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *International Journal of Machine Learning and Cybernetics*, 2025.
- 159 Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S Yu, et al. Graph foundation models: Concepts, opportunities and challenges. *TPAMI*, 2025.
- 160 Zehong Wang, Zheyuan Liu, Tianyi Ma, Jiazheng Li, Zheyuan Zhang, Xingbo Fu, Yiyang Li, Zhengqing Yuan, Wei Song, Yijun Ma, et al. Graph foundation models: A comprehensive survey. *arXiv preprint arXiv:2505.15116*, 2025.
- 161 Yun Zhu, Haizhou Shi, Xiaotang Wang, Yongchao Liu, Yaoke Wang, Boci Peng, Chuntao Hong, and Siliang Tang. Graphclip: Enhancing transferability in graph foundation models for text-attributed graphs. In *WWW*, 2025.
- 162 Dmitry Eremeev, Oleg Platonov, Gleb Bazhenov, Artem Babenko, and Liudmila Prokhorenkova. Graphpfn: A prior-data fitted graph foundation model. *arXiv preprint arXiv:2509.21489*, 2025.
- 163 Tianyin Liao, Ziwei Zhang, Yufei Sun, Chunyu Hu, and Jianxin Li. Invariant graph transformer for out-of-distribution generalization. In *KDD*, 2026.
- 164 Hao Liu, Jiarui Feng, Lecheng Kong, Ningyue Liang, Dacheng Tao, Yixin Chen, and Muhan Zhang. One for all: Towards training one graph model for all classification tasks. *ICLR*, 2024.
- 165 Yufei He, Yuan Sui, Xiaoxin He, and Bryan Hooi. Unigraph: Learning a unified cross-domain foundation model for text-attributed graphs. In *KDD*, 2025.
- 166 Haibo Chen, Xin Wang, Zeyang Zhang, Haoyang Li, Ling Feng, and Wenwu Zhu. Autogfm: Automated graph foundation model with adaptive architecture customization. In *ICML*, 2025.
- 167 Petar Veličković, Rex Ying, Matilde Padovano, Raia Hadsell, and Charles Blundell. Neural execution of graph algorithms. *ICLR*, 2020.
- 168 Jihoon Ko, Taehyung Kwon, Kijung Shin, and Juho Lee. Learning to pool in graph neural networks for extrapolation. *arXiv preprint arXiv:2106.06210*, 2021.
- 169 Yiqi Wang, Yao Ma, Wei Jin, Chaozhuo Li, Charu C. Aggarwal, and Jiliang Tang. Customized graph neural networks. *arXiv preprint arXiv:2005.12386*, 2020.
- 170 Yijian Qin, Xin Wang, Ziwei Zhang, Pengtao Xie, and Wenwu Zhu. Graph neural architecture search under distribution shifts. In *ICML*, 2022.
- 171 Lukas Galke, Benedikt Franke, Tobias Zielke, and Ansgar Scherp. Lifelong learning of graph neural networks for open-world node classification. In *IJCNN*, 2021.
- 172 Zeyang Zhang et al. Disentangled continual graph neural architecture search with invariant modular supernet. In *ICML*, 2024.
- 173 Wei Jin, Tong Zhao, Jiayuan Ding, Yozen Liu, Jiliang Tang, and Neil Shah. Empowering graph representation learning with test-time graph transformation. *ICLR*, 2023.
- 174 Xin Zheng, Dongjin Song, Qingsong Wen, Bo Du, and Shirui Pan. Online gnn evaluation under test-time graph distribution shifts. *ICLR*, 2024.

- 175 Kehang Han, Balaji Lakshminarayanan, and Jeremiah Liu. Reliable graph neural networks for drug discovery under distributional shift. *NeurIPS Workshop*, 2021.
- 176 Nianzu Yang, Kaipeng Zeng, Qitian Wu, Xiaosong Jia, and Junchi Yan. Learning substructure invariance for out-of-distribution molecular representations. In *NeurIPS*, 2022.
- 177 Koustuv Sinha, Shagun Sodhani, Joelle Pineau, and William L Hamilton. Evaluating logical generalization in graph neural networks. *ICML Workshop*, 2020.
- 178 Ren Li et al. How does knowledge graph embedding extrapolate to unseen data: a semantic evidence view. *AAAI*, 2022.
- 179 Kangming Li et al. A critical examination of robustness and generalizability of machine learning prediction of materials properties. *npj Computational Materials*, 2023.
- 180 Saba Heidari Gheshlaghi, Milan Aryal, et al. Explainability-based graph augmentation for out-of-distribution robustness in digital pathology. *KBS*, 2025.

Out-of-distribution generalized graph machine learning and its applications in graph foundation models

Xin Wang^{1,2}, Haoyang Li¹, Jiayi Zhong¹ & Wenwu Zhu^{1,2*}

1. *Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*

2. *BNRist, Tsinghua University, Beijing 100084, China*

* Corresponding author. E-mail: wwzhu@tsinghua.edu.cn

Abstract Graph machine learning has been widely explored in both academic research and industrial applications. Despite the rapid emergence of numerous novel methods and techniques, most existing studies are grounded in the in-distribution (I.D.) assumption—namely, that training and testing graph data are drawn from identical distributions. However, this assumption rarely holds in real-world graph scenarios, where model performance often deteriorates significantly in the presence of distribution shifts between training and test data. To address this critical challenge, out-of-distribution (OOD) generalization on graphs, which moves beyond the in-distribution hypothesis, has made substantial progress and garnered growing attention from the research community. In this paper, we provide a comprehensive overview of OOD generalization on graphs, offering a detailed review of recent advancements in the field. We begin by formally defining the problem of OOD generalization on graphs. Next, we categorize existing approaches into three conceptually distinct classes—data, model, and learning strategy—based on their roles in the graph machine learning pipeline, followed by in-depth discussions of each category. We also review theoretical foundations related to OOD generalization on graphs, introduce commonly used graph datasets for rigorous evaluation, and discuss its applications in graph foundation models. Finally, we present our insights into promising directions for future research.

Keywords machine learning, graph neural network, out-of-distribution generalization, representation learning, graph foundation models