

图基础模型研究进展与挑战：图神经网络的视角

吴涛^{1,2}, 聂发志¹, 先兴平², 王超³, 袁霖², 乔少杰⁴, 牛伟纳⁵

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 重庆邮电大学网络空间安全与信息法学院, 重庆 400065;
3. 重庆师范大学计算机与信息科学学院, 重庆 401331; 4. 成都信息工程大学软件工程学院, 四川 成都 610225;
5. 电子科技大学计算机科学与工程学院, 四川 成都 611731)

摘要: 图基础模型 (GFM) 是基础模型思想在图学习领域中的延伸, 是在广泛图数据上预训练并微调适配多种下游任务的图模型。与借助大语言模型 (LLM) 实现 GFM 的技术路线不同, 主要关注从图神经网络 (GNN) 的角度构建 GFM。首先, 分析了 GFM 的研究现状并定义了关键概念。其次, 总结了 GFM 骨干架构和基础表示单元的研究成果。再次, 根据代理任务和微调策略的不同, 分别总结了图模型的预训练技术与微调方法。然后, 介绍了与 GFM 相关的评价指标。最后, 分析了面临的挑战并展望了未来的研究方向。

关键词: 图基础模型; 图神经网络; 预训练; 模型微调; 提示调优

中图分类号: TP183

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025114

Research advances and challenges on graph foundation model: perspective from graph neural network

WU Tao^{1,2}, NIE Fazhi¹, XIAN Xingping², WANG Chao³, YUAN Lin², QIAO Shaojie⁴, NIU Weina⁵

1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
3. School of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China
4. School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China
5. School of Computer Science and Engineering, University of Electronics Science and Technology of China, Chengdu 611731, China

Abstract: Graph foundation model (GFM) represents the extension of foundation model concepts in graph learning. These models were pre-trained on extensive graph data and fine-tuned for various downstream tasks. Unlike current approaches that utilized large language model (LLM) for GFM, the construction of GFM was emphasized from the perspective of graph neural network (GNN). Firstly, the current research of GFM was analyzed, and the key concepts were defined. Secondly, the research in the backbone architectures and fundamental representation units of GFM were summarized. Then, based on the differences in pretext tasks and fine-tuning strategies, the pre-training techniques and fine-tuning methods of graph models were summarized. Additionally, the evaluation metrics related to GFM were introduced. Finally, the unresolved issues and future research directions were discussed.

Keywords: graph foundation model, graph neural network, pre-training, fine-tuning, prompt-tuning

收稿日期: 2025-03-10; 修回日期: 2025-06-09

通信作者: 先兴平, xianxp@cqupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62376047, No.62106030); 重庆市自然科学基金创新发展联合基金重点资助项目 (No.CSTB2023NSCQ-LZX0003); 重庆市教委科学技术研究计划基金重点资助项目 (No.KJZD-K202300603); 重庆市教委科学技术研究基金资助项目 (No.KJQN202300606)

Foundation Items: The National Natural Science Foundation of China (No.62376047, No.62106030), The Key Projects of the Joint Fund for Innovation and Development of Chongqing Natural Science Foundation (No.CSTB2023NSCQ-LZX0003), The Key Project of Science and Technology Research Program of Chongqing Education Commission (No.KJZD-K202300603), Science and Technology Research Program of Chongqing Municipal Education Commission (No.KJQN202300606)

*第二十七届中国科协年会“AI时代网络技术创新”专题

0 引言

为了满足海量、泛在图数据的处理需求,深度学习自提出以来就与图挖掘、社交网络分析以及复杂网络等领域交叉融合^[1-3],形成了以图神经网络(GNN, graph neural network)为代表的图深度学习模型,并被广泛应用于社交网络、分子结构、知识图谱以及交通网络等领域,逐渐发展为图数据建模学习领域的代表性方法^[4]。然而,大多数图深度学习模型都是针对特定任务在特定图数据上训练的,已训练好的图模型无法有效地用于其他新任务的求解。特别是,在当前图数据集数量不断增加、图学习任务范围不断扩展的趋势下,面向特定图数据和特定任务的图模型训练方式表现出明显的局限性,影响模型在其他图数据和任务中的泛化性和通用性,极大地阻碍了GNN的实际应用,也限制了海量图数据内在价值的发挥。

随着ChatGPT的横空出世,大语言模型(LLM, large language model)开启了新一轮的科技浪潮。LLM通常基于Transformer架构,使用自监督学习范式、利用海量无标记数据进行训练,能够学习到数据隐含的世界知识,在解决复杂任务时表现出与小型预训练模型不同的行为,并可以通过统一形式解决各种下游任务,在自然语言处理和计算机视觉等领域取得了巨大的成功。然而,LLM在图数据建模方面未能表现出同样优秀的性能水平。为了利用LLM的思想处理泛在图数据,支撑节点分类、链路预测以及图分类等图任务,探索构建能够利用跨领域图数据、具有良好泛化性的图基础模型(GFM, graph foundation model)具有重要的科学意义。

关于GFM的构建,当前研究领域主要有3类技术路线,即基于LLM的方法、基于GNN+LLM的方法以及基于GNN的方法。其中,基于LLM的GFM主要是进行图数据和自然语言的对齐,从而将图学习任务转化为自然语言处理任务,相关工作包括NLGraph^[5]、Graph-LLM^[6]、LLMtoGraph^[7]等;基于GNN+LLM的GFM综合利用了GNN和LLM在图结构和自然语言处理方面的优势,包括以GNN为主的方法TAPE^[8]、GALM^[9]、TOUCHUP-G^[10]等和以LLM为主的方法GraphGPT^[11]、THLM^[12]、ChatGraph^[13]等;基于GNN的GFM直接利用各种图深度学习模型进行图数据的建模学

习。在以上方法中,基于LLM的GFM的不足之处在于图数据到自然语言的转化会导致图结构信息的丢失,从而影响图的推理计算。以GNN为主的GNN+LLM方法首先利用LLM提取节点的特征,然后利用GNN进行推理计算,其本质上还是GNN方法。相对地,以LLM为主的GNN+LLM方法的关键在于利用GNN生成保留图结构上下文的提示,从而基于LLM的语言理解能力实现对图数据的处理。此外,图数据具有复杂的特征模式,要保证图数据分类、预测方法的性能,图模型对图数据必须具有强大的表达刻画能力,这一直是GNN领域的主要研究方向。由于GFM的根本是进行大规模图数据的建模学习,大量研究已经证明了GNN对图数据建模学习的优良性能,并且LLM至今仍无法证明其可以保证图数据建模学习过程中置换不变性的基本要求。因此本文重点关注以GNN为主的GFM研究工作。

GFM对推动图深度学习从理论研究迈向大规模实际应用具有重要作用,研究领域关于GFM构建已进行了初步的研究探索。Wang等^[14]利用来自13个领域的超过5 000个图数据提出了基于概率分布的大图生成模型(LGGM),此模型通过学习在不同领域之间具有迁移性的基础结构模式获得了良好的泛化能力。为了实现跨领域图数据的建模与泛化,He等^[15]提出了图基础模型(UniGraph),该模型利用文本属性图(TAG, text-attributed graph)进行图数据的统一表示,采用编码器-解码器架构,其中以级联的预训练语言模型(LM, language model)和图注意力网络(GAT, graph attention network)作为核心。Liu等^[16]提出图基础模型OFA(one for all),此模型利用文本属性图(TAG)对来自不同领域的图数据进行统一表示,并基于文本对图中的节点和边进行描述,然后采用LLM对节点特征和下游任务进行编码,提出图提示范式来操作输入图,从而使图模型能够从输入数据中获取与任务相关的信息。Lachi等^[17]将领域特定的特征转化到共同的隐空间进行表示,并提出基于Graph Transformer架构的通用模型GraphFM。从模型零样本泛化的角度,为了使模型能够在没有见过的数据上进行准确预测,Xia等^[18]认为GNN在泛化到与训练数据有差异的下游数据时经常面临困难,需要构建能理解不同图数据的模式规律并适应下游图

数据中零样本学习任务的模型,为此,其通过将图转化为统一标记序列,提出了基于 Graph Transformer 架构的 OpenGraph 方法。

Li 等^[19]研究了跨 TAG 的模型迁移泛化问题,提出了零样本迁移学习方法 ZeroG,该方法利用预训练的语言模型编码节点和类别的特征属性以确保各数据集特征维度的一致性,从而支撑 GFM 对不同数据的适用性。从图数据表征学习角度, Hu 等^[20]研究了上下游任务不一致性对图模型性能的影响,并以子图作为基本单位,结合图自编码器 (GAE, graph autoencoder) 和生成对抗网络思想提出了适用于多种学习任务的图数据表征方法 GA²E。从 GFM 训练的角度, Liu 等^[21]研究了尺度定律中参数规模、数据量与图模型性能的关系。不同于直推式学习范式, Zhao 等^[22]基于简单图卷积网络 (SGCN, simple graph convolutional network)^[23]提出了图基础模型 GraphAny 以实现归纳式的节点分类,从而基于预训练图数据学习到的模型能够在新图上进行推理预测。Frasca 等^[24]通过实证分析评估了预训练的 GNN 在数据集之间的可迁移程度,发现可迁移性与下游数据的数据量以及预训练数据的质量紧密相关。Klaser 等^[25]在图卷积网络 (GCN, graph convolutional network)、图同构网络 (GIN, graph isomorphism network) 和消息传递图神经网络的基础上提出了分子指纹基础模型 MiniMol。Sypetkowski 等^[26]从模型深度、样本规模、数据多样性等角度研究了不同 GNN 架构的扩展行为,并提出了图基础模型 MolGPS。面向原子性质的建模预测需求, Shoghi 等^[27]基于图神经网络 GemNet-OC 提出了基础模型 JMP。

总体而言,目前面向 GFM 的相关研究处于起步但快速发展的阶段。尽管有不少研究人员利用 LLM 进行 GFM 的实现,但越来越多的研究工作^[18,20,22,24-26]证明了以 GNN 作为骨干架构构建 GFM 的可行性。考虑 KDD、WWW、NeurIPS、TKDE 等会议期刊以及预印版论文,图1给出了近5年 GFM 相关学术论文数量。

为了促进 GFM 的发展,已有研究人员进行了相关研究的梳理综述(如表1所示)。Zhang 等^[28]首先讨论了 GFM 应具备的基本特性,然后从统一表征、图数据、图模型3个方面梳理了研究现状,分析了存在的问题与挑战,阐述了关于交互指令、模型架

构、预训练等方面的观点,讨论了 GFM 在推荐系统、知识图谱、化学分子等领域的应用前景。Mao 等^[29]认为构建 GFM 的关键在于迁移性,通过梳理图数据模式、模型表达能力以及模型稳定性与迁移性的关系分析了如何构建有效的“图词汇”,从而更好地编码图数据之间的不变性,然后讨论了数据规模、模型规模和 LLM 在 GFM 构建中的作用和影响。Liu 等^[30]讨论了 GFM 的概念、特征和影响因素,对 GFM 的相关方法给出了类别框架,在基于 GNN 的模型、基于 LLM 的模型以及基于 GNN+LLM 的模型方面分别综述了与之相关的模型架构、预训练、适应机制等研究工作,并分析了面临的挑战和未来研究方向。Li 等^[31]聚焦基于 GNN+LLM 的 GFM 构建方法,根据 LLM 在图学习任务中所扮演的角色,即增强器、预测器和对齐组件,将现有方法组织成3类。Fan 等^[32]分别从利用 LLM 强化图机器学习和利用知识图谱优化 LLM 这2个方面对相关研究进行了总结。Ren 等^[33]关注基于 LLM 的图学习问题,并将相关研究分为以 GNN 作为 LLM 的前缀、以 LLM 作为 GNN 前缀、GNN 与 LLM 融合以及 LLM 的4类方法。Jin 等^[34]从纯图、文本属性图和文本配对图的角度对在图模型中用作编码器、预测器和对齐器的 LLM 的相关研究进行了梳理论。

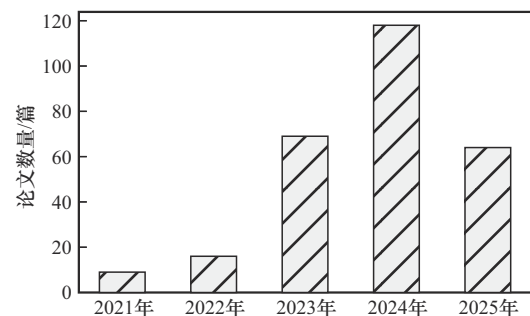


图1 近5年 GFM 相关学术论文数量(截至2025年6月)

以上综述主要聚焦于对 GFM 概念、特性和挑战的讨论以及结合 LLM 进行 GFM 构建工作的总结。然而,大量研究已证明 GNN 对图数据建模的有效性,当前 LLM 的主要作用在于更好地进行节点特征表示。无论是基于 GNN 的还是基于 GNN+LLM 的 GFM 构建,其核心都离不开对图的建模学习,图模型构建仍是 GFM 的核心。因此,有必要结合图学习任务和图数据特性,从 GNN 的角度对与 GFM 相关的模型架构、预训练、适应机制等进

表1 GFM相关的研究综述与本文的对比

文献	主要内容	框架类型	与本文的差异
文献[28]	阐述 GFM 应具备的特征以及在统一表征、数据资源、模型构建等方面需要开展的工作	GNN 或 LLM	文献重点给出了 GFM 的宏观概念性定义, 本文梳理了研究领域内与 GFM 构建相关的模型架构、预训练、微调方面的具体研究工作
文献[29]	讨论图的基础可迁移单元和规模定律的影响	GNN 或 LLM	文献重点关注基础可迁移单元和规模定律, 本文更侧重于从统一的 GFM 概念模型出发, 更加全面地归纳了相关的具体研究工作
文献[30]	从 GNN、LLM 以及 GNN 和 LLM 结合这 3 个角度阐述 GFM 构建的研究工作	GNN、LLM 或 GNN 结合 LLM	文献关注基于不同技术路径的 GFM 构建方法, 本文聚焦基于 GNN 构建 GFM 的理论体系
文献[31]	综述 LLM 和 GNN 不同的结合方式	GNN 结合 LLM	文献结合 LLM 和 GNN 以进行图数据建模学习, 本文聚焦面向 GFM 的 GNN 预训练和微调
文献[32]	基于 LLM 强化 GNN 和基于知识图谱强化 LLM	GNN 结合 LLM	文献关注 LLM 和图数据之间相互强化作用, 本文聚焦基于 GNN 的 GFM 构建过程
文献[33]	综述 LLM 和 GNN 不同的结合方式	GNN 结合 LLM 或 LLM	文献关注利用 LLM 处理图数据的架构流程, 本文关注 GFM 构建工作中模型的架构设计、模型训练与评价问题
文献[34]	针对不同的图数据特征, 综述如何利用 LLM 增强对图数据的建模学习能力	GNN 结合 LLM 或 LLM	文献关注针对不同类型的图数据 LLM 和 GNN 的不同结合方式, 本文关注 GFM 构建工作中 GNN 模型自身的问题

行综述, 其整体框架如图 2 所示。本文的创新之处在于系统性论证和介绍了基于 GNN 进行 GFM 构建的研究框架和具体进展, 包括分析了基于 GNN 的 GFM 研究对 GFM 发展的重要意义, 给出了 GFM 的形式化定义, 阐述了从 GNN 的角度构建 GFM 的整体框架以及 GFM 的具体实例, 并对相关技术的研究进展进行了全面的梳理分析, 这将有益于后续基于 GNN 的 GFM 的发展。

1 图基础模型:概念框架

1.1 图基础模型概念

1.1.1 图基础模型定义

GFM 本质上是一种具有大量参数的图机器学习

模型, 旨在利用来自不同领域、具有不同特征的图数据, 通过强大的学习能力来更好地理解、分析以及处理图数据中蕴含的模式规律, 形成统一的特征单元和数据表征, 能灵活适用于多种下游任务, 并在不同应用领域具有良好的跨数据集迁移泛化能力。GFM 的目标是学习不同领域中的多个图数据的结构和特征信息, 进行特征空间对齐, 形成能够很容易地用于各种下游任务的统一表征。GFM 的实质是增强模型泛化性, 使模型面对下游任务时具有零样本迁移能力, 能有效处理与训练图数据集中有显著差异的、未见过的图数据并做出准确预测。

令 $S = \{G_S\}$ 表示训练图数据集, 其中每个图

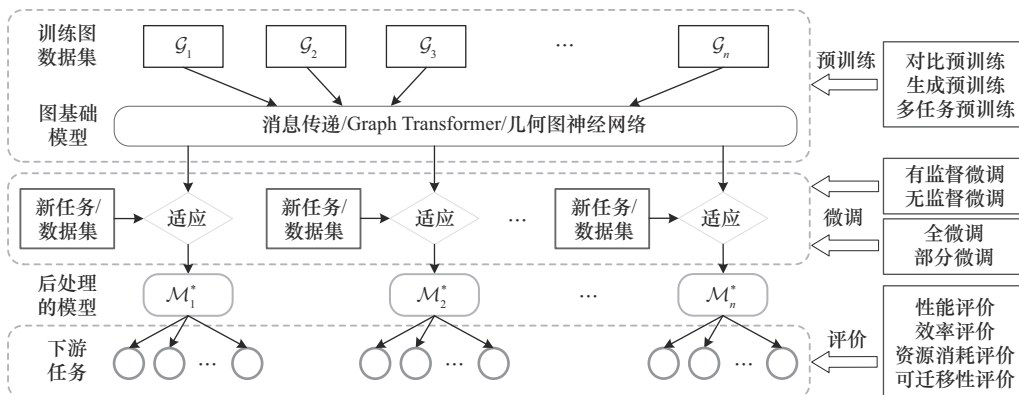


图2 从 GNN 的角度构建 GFM 的整体框架

数据 \mathcal{G}_S 对应标签集合 \mathcal{Y}_S 。同时，令目标图数据为 \mathcal{G}_T 对应的标签集合为 \mathcal{Y}_T 且 $\mathcal{G}_S \cap \mathcal{G}_T = \emptyset$, $\mathcal{Y}_S \cap \mathcal{Y}_T = \emptyset$ 。GFM 可以形式化定义为

$$\Theta = \arg \min_{\theta} \sum_{\mathcal{G}_S \in \mathcal{S}} \mathcal{L}(f_{\theta}(\mathcal{G}_S), \mathcal{Y}_S) \quad (1)$$

$$f = \arg \max_{f, \mathcal{L}} \sum_{\mathcal{G}_T} \mathcal{C}(f_{\theta}(\mathcal{G}_T), \mathcal{Y}_T) \quad (2)$$

其中， Θ 为可训练的模型参数， \mathcal{L} 为损失函数， \mathcal{C} 用于衡量模型在下游任务预测精度的评价指标。GFM 不要求训练图数据和目标图数据具有一致的标签空间，它能够适用不同于训练图数据 $\mathcal{S} = \{\mathcal{G}_S\}$ 的全新图数据 \mathcal{G}_T 。GFM 构建的关键在于设计能够编码图数据中不同特征空间和结构模式的模型架构 f ，从而实现不同训练图数据集之间以及训练集与目标图数据之间结构、特征和标签的对齐。

1.1.2 模型预训练

为了减轻对标签数据的依赖，研究人员主要以自监督的方式进行 GNN 预训练，在丰富的未标记图数据上训练 GNN 以捕获可转移的知识，通过梯度下降更新参数以优化模型，并将优化后的模型的权重和参数迁移到下游任务，通过少量有标签的下游数据进行微调，形式化定义为

$$\Theta^*, \Phi^* = \arg \min_{\theta, \phi} \mathcal{L}(f_{\theta}(\cdot), g_{\phi}(\cdot)) \quad (3)$$

$$\Theta^{**}, \Psi^* = \arg \min_{\theta^*, \psi} \mathcal{L}_{\text{sup}}(f_{\theta^*}(\cdot), g_{\psi}(\cdot)) \quad (4)$$

其中， $f_{\theta}(\cdot)$ 和 $g_{\phi}(\cdot)$ 分别表示预训练模型中的编码器和解码器， $f_{\theta^*}(\cdot)$ 和 $g_{\psi}(\cdot)$ 对应下游模型的编码器和解码器。值得注意的是，此处的解码器泛指能够执行类似解码作用的组件，即能够将编码器生成的特征表示转换为任务特定输出的模块，其中也包括对比学习中的投影头以及下游任务的预测层，而不仅限于图自编码结构中的狭义解码器。模型预训练的目的是通过大量无标签数据先训练一个编码器 $f_{\theta}(\cdot)$

形成参数的初始化取值，然后在下游少量有标签数据中将初始化的编码器 $f_{\theta^*}(\cdot)$ 和下游解码器 $g_{\psi}(\cdot)$ 同时进行训练，以微调模型参数。其基本原理架构如图 3 所示。

1.1.3 模型适应

在模型预训练阶段，模型对下游任务是完全不可知的，这不利于预训练模型的迁移泛化，对其在实际场景中的应用造成了影响。因此，在通过预训练获得大量知识后，模型仍然需要后处理来增强其对下游任务的适应性。其中，微调是一种常用的基于模型参数细微调整的适应方法，可以利用预训练模型获得通用知识，然后基于少量标记样本对模型的所有层或特定层参数进行细微调整，将预训练模型推广到下游任务或特定数据集上，在适应特定任务的同时保留模型的一般功能。另外，提示调优可以避免更新模型参数，其通过构造针对特定任务的指令指导模型如何处理下游数据并产生期望的输出，从而使预训练模型更好地适应下游任务。

1.2 与大语言模型的联系与区别

作为代表性的基础模型，LLM 在自然语言处理等领域取得了显著成效，但在图数据建模方面没有达到类似的水平。受 LLM 等基础模型的启发，近年来研究人员开始探索面向图数据建模的 GFM。虽然有研究人员尝试直接采用 LLM 作为 GFM，但基于 GNN 的 GFM 构建方法仍然是主要研究思路（本文主要关注的方向）。GFM 具有与 LLM 类似的特征，同样追求涌现与同质化，希望模型能随着规模的扩大自发形成新的能力，能学习数据中蕴含的常识知识并广泛部署于各种应用场景。它们都遵循“预训练-适应”的训练范式，都希望通过在广泛数据集上的充分训练增强模型的表达能力以及在各种任务上的泛化能力。

GFM 与 LLM 的不同之处在于，相对于文本、图像数据，图数据更加难以收集且天然缺乏语义

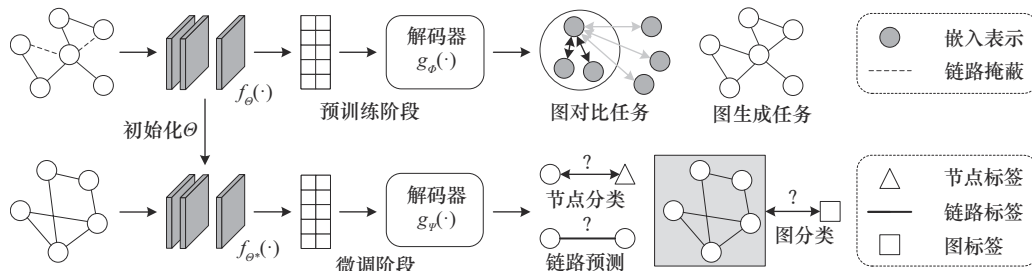


图 3 面向 GFM 的预训练与微调基本原理架构

性,不同领域的图数据在结构模式、节点特征以及标记空间等方面往往具有明显差异。而且,文本与图像处理领域发展较为成熟,而图挖掘、图神经网络领域近年来才得到研究人员的关注,缺乏研究积累。因此,相比于LLM以词元作为基础表示单元并且直接采用Transformer作为骨干架构,设计有效且合理的图数据基础表示单元和GFM模型骨干架构更具挑战性,更加难以实现良好的泛化性。另外,相对于LLM的下游任务可以用统一的形式进行处理,图学习任务类型多样且难以统一表示,实现GFM的同质化更具挑战性。

2 图基础模型架构

2.1 模型骨干架构

虽然图学习领域面向图数据建模学习提出了多种类型的模型算法,但当前还没有能够完全适用于GFM的骨干架构。本节对各类GNN模型架构进行概述,从而为GFM的设计提供支撑。

2.1.1 消息传递架构

GNN中最具代表性的模型架构是基于消息传递的方法,包括GCN、GAT以及GIN等,此类模型利用每层消息传递,从邻域节点收集特征信息,并利用聚合函数将其与中心节点自身特征融合,以生成更新后的节点嵌入。通过堆叠多层消息传递机制,模型可在图结构中实现信息的递归传播,使节点嵌入在保留原始特征的同时,还能深度捕捉到图数据的拓扑特性,从而构建出兼具语义属性和结构感知的节点表征。在此基础上,研究人员针对有向、时空、异质等各类图数据的建模学习需求提出各种优化改进。

至今为止,研究人员提出的基于消息传递架构的代表性GFM方法包括GA²E^[20]、GraphAny^[22]、MiniMol^[24]、MolGPS^[26]以及JMP^[27]。此类方法的优势在于具有内在的置换不变性,能产生与输入节点顺序无关的表征,对图结构特征具有良好的表达能力。同时,消息传递架构基于图结构进行消息传递和计算更新,具有较低的复杂度。另外,MolGPS等方法表明,基于消息传递架构的模型性能能够随着模型大小、数据集规模和训练计算量的增加而持续优化,这证明其符合尺度定律、具有良好的扩展能力。近年来,研究人员基于消息传递架构从聚合机制、模型深度等角度不断提升模型的表达能

力^[35-36],相关成果可直接用于提升GFM的性能。

2.1.2 Graph Transformer 架构

由于Transformer在自然语言处理和计算机视觉领域的良好性能,Graph Transformer的研究引起了学术界和工业界的高度关注^[37]。Graph Transformer是多头注意力机制和全连接前馈网络的交替堆叠。为了刻画图数据中的结构信息,Graph Transformer依赖结构编码和位置编码,其中结构编码使模型感知到图的结构信息,位置编码使节点知道在图数据中自己与其他节点的相对位置。近年来Graph Transformer在解决图学习的挑战方面显示出优势,在各个应用领域证明了其有效性和通用性。

至今为止,研究人员提出的基于Graph Transformer架构的代表性GFM方法包括GraphFM^[17]、OpenGraph^[18]以及MolE^[38]。基于Graph Transformer的方法将图视为全连接的,通过注意分数考虑并度量图中每对节点之间的相似性,其优点在于能捕获图数据中的长距离依赖并减轻过平滑现象,但不足之处在于其具有二次时间复杂度,从而会影响其在大规模图数据中的应用^[39]。为了解决以上问题,研究人员提出了线性注意力机制和局部注意力机制^[40-41],然而以上机制当前还未被成功应用于GFM的构建。另外,GFM的预训练数据集的规模差异较大,这对通过填充形成标准化输入大小的模型训练十分具有挑战性。同时,不同领域的图数据往往具有各自的特征,从而跨领域的多样性要求Graph Transformer能够理解和处理不同的图拓扑以及领域特定属性。为了解决以上问题,当前主要方法^[17-18]是先构建跨数据集的统一表征,再进行Graph Transformer的训练。

2.1.3 几何图神经网络架构

为了建模学习空间几何结构,几何图神经网络被提出并被用于蛋白质、机器人控制等众多科学领域中物理系统的建模。与传统GNN在二维空间中表征拓扑结构不同,几何图神经网络通过几何向量的形式为每个节点分配一种特殊类型的节点特征建模三维空间结构。值得注意的是,几何图具有平移、旋转和/或反射的对称性,通过将几何性质嵌入模型的学习流程中,此类模型能够更好地刻画数据的内在结构和变化规律,展现出良好的解释性、泛化性和通用性。近年来研究人员提出了不变图神经网络、等变图神经网络、几何Graph Transformer

等多种类型的模型算法,详细情况可参考文献[42]。

至今为止,研究人员提出的与几何图神经网络相关的GFM方法包括EPT^[43]和RiemannGFM^[44]。EPT对几何图基础模型进行了初步研究探索。RiemannGFM将图结构映射至非欧几何空间,并构建基于黎曼几何的模型架构,捕捉跨领域共享的拓扑模式,从而显著提升了模型的迁移能力。与消息传递架构、Graph Transformer架构相比,几何图神经网络架构的优势在于其具有更精细的刻画能力,能够建模节点之间的几何关系或空间结构,适用于对分子和蛋白质等的几何结构的描述。除了具有置换不变性之外,几何图神经网络还支持平移、旋转以及翻转等操作。

总体而言,以上3种GFM的架构各有特点。消息传递架构依赖于固定的图结构进行消息传递,专门设计用于处理图结构数据,其数据项之间的关系是显式定义的,复杂度较低。Graph Transformer架构基于内容相似度来决定信息的流动,可被视为在一个完全有向图上定义的特殊形式的GNN。其考虑全部节点对之间的相似性,具有捕获长距离依赖的优势;具有更强的表征能力,能够避免过平滑问题;具有并行处理能力,但复杂度较高。与消息传递架构、Graph Transformer架构相比,几何图神经网络架构主要关注几何关系或空间结构的刻画,描述能力更强,适用于科学领域图数据的处理。

2.2 基础表示单元

实现基础模型的关键挑战在于找到跨领域数据集背后可迁移的基本单元,从而编码不同数据之间的可共享知识。例如,LLM成功的关键在于各种自然语言处理任务背后存在以词元为代表的基本数据单元,语义大不相同的文本也能用同一字典进行编码。然而,对于图学习任务,图数据中的节点和边在不同应用领域可能具有不同的含义,从而无法作为不同领域的图数据之间通用的基础表示。相对地,传统GNN主要面向特定图数据和任务,缺乏将所学知识迁移到不同领域的的能力。

为了构建具有良好泛化性的GFM,He等^[15]和Liu等^[16]提出构建TAG,将来自不同领域图数据的节点/边特征转化为自然语言文本,实现以词元为基本表示单元的跨领域知识迁移,并借助LLM将文本表示映射到共享嵌入空间实现特征对齐。类比语言中的词元,Hu等^[20]提出图模型GA²E,将原始

图重新表述成多个子图来执行预训练,确保不同图任务共享相同的单元结构,并有效消除了预训练和下游任务之间的差异。Sun等^[44]进一步扩展该思路,提出“结构词元”的概念,将树、环等基础结构视为图的构造单元,但该方法在处理复杂且无规律的图数据时效果较为有限。

与依赖图结构进行显式对齐的方法不同,现有方法通过引入额外标记,能够有效促进跨领域图数据在结构上的统一。Zhao等^[45]利用虚拟节点连接同一域和跨领域图数据,在保持域内局部结构完整性的同时,建立跨领域全局交互通道,将孤立的域图整合成一个全局图,并构建联合邻接矩阵以统一结构。Cheng等^[46]融合锚节点和所有类标签构建虚拟超级节点,舍弃原始图中的边,并在其与锚节点的邻居之间建立虚拟边,同时利用注意力机制动态调整边权重,实现跨领域结构信息的自适应聚合,从而弱化不同图数据在拓扑连接上的差异。针对不同源域的图数据,Yu等^[47]在编码器的各层中嵌入可学习的结构标记,逐层聚合并更新节点表示,从而捕获不同粒度的结构信息,并利用最终堆叠的多源域节点嵌入矩阵,形成统一的结构对齐表示。

文献[18,48]基于快速奇异值分解(SVD, singular value decomposition)将邻接矩阵或特征矩阵投影到统一、固定维度的表示空间中,从而实现结构或特征的统一。Xia等^[18]提出图分词器,对邻接矩阵进行平滑处理,并结合不同阶数的邻接矩阵以捕捉图中的高阶连接和稀疏节点关系,继而利用SVD将邻接矩阵转化为包含拓扑信息的标记序列,使模型能够通过学习理解图结构。Wang等^[48]将不同节点的语义特征(文本嵌入和本体特征)作为投影对象,以增强模型跨知识图谱的推理能力。

总体而言,研究人员已经提出了多种面向GFM的基础表示单元设计方法,表2对这些方法的设计方式、适用图类型以及优缺点进行了系统性总结。其中,基于TAG的方法简单易实现,可利用成熟的LLM进行处理,但高度依赖图数据的文本特征。为此,研究人员提出利用基础图结构、额外标记以及SVD等方法,但这些方法需要额外定制模块且计算复杂度较高。另外,尽管GFM致力于实现全领域普适性,但部分方法仍仅适用于同质图或TAG。因此,当前领域仍然还有待提出效率更高、适用性更好的基础表示单元设计方法。

表2 代表性的基础表示单元设计方法

类型	文献	设计方式	适用图类型	优点	缺点
TAG	文献[16]	构建 TAG, 利用文本特征作为不同领域图数据的统一表示方式, 实现特征对齐	TAG	方法实现相对简单且技术较为成熟	需要借助 LLM 处理文本特征
基础图结构	文献[20]	利用随机游走将不同图数据统一为子图的形式, 实现结构统一	所有类型的图	实现图数据与下游任务的双重统一	对不同领域的语义信息需进行调整和优化
	文献[44]	将树和环等子结构作为构建任意图数据的基本表示单元, 实现结构统一	所有类型的图	利用黎曼几何空间能更好地捕捉复杂结构	对复杂且无规律的图数据的适用性受限
额外标记	文献[45]	引入全连接虚拟节点并建立跨图连接, 构建联合邻接矩阵实现结构统一	所有类型的图	能够实现多个图数据的联合建模与协同分析	构建联合邻接矩阵, 导致时间开销较大
	文献[46]	融合锚节点与类标签构建虚拟超级节点, 以虚拟边连接邻域节点, 并基于超级节点表示实现结构统一	TAG	有效避免了对不同领域原始结构的依赖	依赖图数据的文本特征, 通用性不足
	文献[47]	在编码器中逐层引入结构标记形成统一的结构对齐表示, 实现结构统一	所有类型的图	降低了直接合并多域图时所产生的结构干扰	对极端结构差异的泛化能力有限
SVD	文献[18]	利用 SVD 投影将不同规模的邻接矩阵压缩为固定维度的向量, 实现结构统一	同质图	保留图数据的高阶连接性且计算效率相对较高	无法适用于异构图数据, 通用性不足
	文献[48]	利用 SVD 将不同领域的节点语义特征转化到固定维度的特征空间中, 实现特征对齐	知识图谱	捕捉到不同领域图数据中的语义关联	在大规模图上的可扩展性仍需进一步验证

3 图模型预训练技术

GFM 通过学习图结构与特征信息, 从而构建统一表征以适配跨领域多任务场景。尽管当前从 GNN 的角度构建 GFM 的研究尚处于初步研究阶段, 但利用预训练策略赋予图模型“基础认知”, 进而实现下游任务的高效适配, 被认为是构建 GFM 的有效途径^[17,20,44-45,47]。在此背景下, 系统梳理现有预训练方法具有重要的理论价值和实践意义。此外, 在训练过程中, 对于标注数据易于获取的特定图任务, 可采用有监督的方式进行训练。例如, Hu 等^[49]联合使用单个分子图的多种监督标签, 针对每个属性设计对应的二元分类任务。Hou 等^[50]利用获取到的粗粒度标签执行有监督预训练任务。然而, 现实中标签的收集通常成本高昂且困难, 海量无标签数据则更易获取。因此, 本节将重点探讨图模型的自监督预训练方法。根据学习目标和借口任务的不同, 自监督预训练可分为对比预训练、生成预训练以及多任务预训练 3 类方法。

3.1 对比预训练

对比预训练基于对比学习框架, 通过节点丢弃、边扰动、属性掩码等多样化的数据增强技术, 对大规模无标记图数据进行变换, 从而生成不同视

角的增强图样本, 常利用 InfoNCE^[51]、JSD^[52]等损失优化预训练模型, 迫使相似图样本在低维嵌入表示空间相互靠近, 同时推远不相似图样本。根据图对比学习中目标对象的不同, 即节点与节点、图与图以及节点与图, 对比预训练方法可分为节点级对比、图级对比以及跨级对比 3 种类型。对比预训练基本原理框架如图 4 所示。

3.1.1 节点级对比

节点级对比学习以节点为基本单元, 通过对比节点自身的不同特征表示, 使模型学习节点在图结构中的局部信息。不同方法的核心差异在于正样本对的选择范围, 主要是将不同增强视图下的同一节点作为正样本对, 促使模型学习节点在不同视图下的不变特征。基于这一思想, Qiu 等^[53]提出图对比编码 (GCC) 的预训练框架。该方法针对每个目标顶点, 通过对其多跳自我网络进行子图采样以构建实例样本, 并将两次随机采样结果作为正样本对, 从而达到区分不同顶点采样子图的目的。

文献[54]基于同质性假设, 将相邻节点视为正样本对, 以减轻对数据增强的依赖。然而, 异质图的节点和边具有多样性, 导致该假设失效, 并且常用的数据增强侧重于局部结构, 会破坏异质图的语义完整性。为此, Xiao 等^[55]在原有的节点

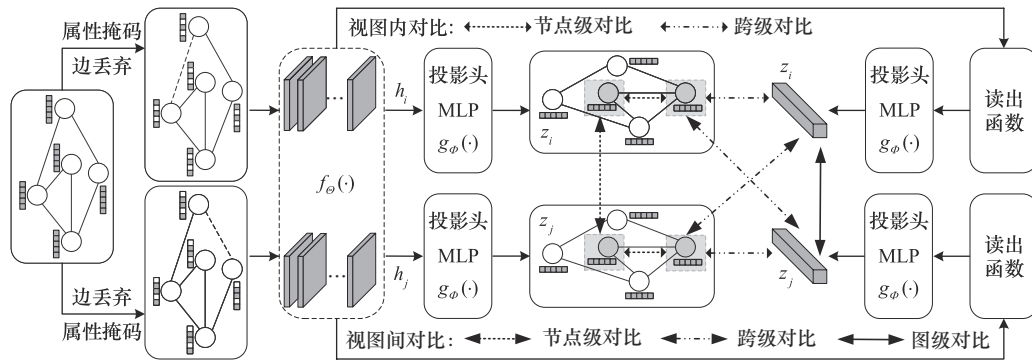


图4 对比预训练基本原理框架

级对比学习框架中引入非对称预测器，利用中心节点重构邻域表示以捕捉上下文语义，进而筛选出具有相似上下文的节点作为正样本对。同时基于不同中心节点重构相同节点的邻域上下文，实现两跳邻居节点表示的隐式对齐。该方法不依赖同质性假设和数据增强，仅以低阶（一跳或两跳）邻域信息选择正样本，但对高阶多跳交互场景的处理能力相对有限。

针对同质图而言，也并非完全遵循同质性假设，尤其在决策边界附近，不同类别且相邻的节点直接作为正样本可能导致误报，且传统语义相似性度量在邻域外节点的适用性较弱。针对上述问题，Zhuo 等^[56]提出构建具有块对角属性且幂等的亲和矩阵来确定最优正样本集。该方法本质是从全局视角筛选出具有代表性的正样本，而并不仅限于同一视图或不同视图，但在生成亲和矩阵时需调用批次内所有节点特征，导致计算的开销大、效率低。

3.1.2 图级对比

图级对比学习以整个图为研究对象，通过对原始图实施不同的数据增强生成多个增强图样本，并将同一原始图的不同增强图作为正样本对，其本质与节点级对比学习类似，但图级对比更侧重于捕捉全局信息，利用读出函数将获取到的节点表示池化为全局表示，并且其研究重点不在于正样本的选取，而在于如何筛选出合适的数据增强策略，从而减轻对原始图结构和特征的破坏，保留图数据的关键信息，并提升模型性能。

You 等^[57]提出 GraphCL 方法。该方法基于不同先验设计多种图数据增强策略，对于不同的增强图采用 GNN 进行编码表示，然后将编码表示投影到隐空间进行对比损失的计算，并采用 InfoNCE 损失的变体 NT-Xent^[58]优化模型参数，从而提高模型鲁

棒性，但其效果取决于特定的数据增强方法，必须通过经验法或试错法手动地进行选择，这极大地限制了 GraphCL 的适用性。为此，You 等^[59]提出双层优化框架 JOAO，它在对特定图数据执行数据增强时能够自适应和动态地选择，但其可选空间都是预先定义的，而不是通过模型自动学习得出的，因此只能称为半自动选择。

Yao 等^[60]在数据增强的基础上引入模型增强，提出 3 种基于模型剪枝的增强策略，分别针对模型的不同组件（权重、层以及注意力头）进行扰动，以消除冗余参数，降低模型复杂度并生成更为稳健的样本。此方法利用双视角增强机制显著提升样本多样性，但在剪枝增强时会修改模型架构。Ji 等^[61]基于实验证明了保留特定维度表示对下游任务存在积极影响。因此利用可学习的维度权重对编码器的输出进行加权，并设计基于维度感知的 InfoNCE 损失优化模型，促使网络学习到更具判别力的包含维度的表示。该方法显式建模了图表示的维度重要性，有效突破了传统图对比学习方法高度依赖于固定数据增强操作的局限性。

3.1.3 跨级对比

跨级对比学习通过对比节点与图之间的表示，使模型学习到不同层次特征之间的内在联系，在挖掘整体图特征的同时，也涉及局部节点特征的学习。

对于正样本的选取，与节点级对比类似，跨级对比也存在多种选择方式，通常是将同一图中局部节点表示和全局图表示作为正样本对。基于该思想，Veličković 等^[62]提出 DGI 方法，首次引入节点级表示与图级表示的对比学习机制，旨在最大化同一图中节点表示与图表示之间的互信息，同时最小化不同图中两者的互信息。针对不同增强图的特点，Hassani 等^[63]提出 MVGRL 方法，其本质上是

在不同增强图中分别选择节点和图作为正样本对,从而实现跨视图对比学习,但该方法仅限于结构空间的增强,对特征空间的增强适用性相对较差。

针对单图生成负样本时, DGI 仅改变样本的节点位置而未引入语义扰动,导致编码器仅能区分节点来源于原始图或噪声图,难以辨别不同语义图之间的特征差异,且正样本表示难以向最优解收敛。为了解决该问题, Zhao 等^[64]融合生成对抗思想提出 GA-GGD 方法,利用语义攻击生成器生成具有错误语义的节点表示,并主动干扰图组判别器,从而辅助其强化语义区分能力。同时,图组判别器采用 DGI 框架,用于区分节点的不同来源(原始图、噪声图或生成图),从而实现了节点层与图层级之间的信息交互与对比。

总体而言,研究人员针对不同对比类型提出了多种预训练方法,表 3 从数据增强方式、下游任务、目标函数、评价指标以及优缺点等方面对其进

行了系统性总结。从方法设计的角度出发,节点级对比学习侧重于优化正样本选择机制,以应对多样化图数据的复杂特性;图级对比学习致力于设计高效的数据增强方法,最大限度地减少增强过程中的信息损失;跨级对比学习虽能实现多层次表征,但其瓶颈在于节点表示与图表示之间会存在特征异质性,导致语义对齐困难,并且图池化函数的设计缺陷会引发图表示的退化,进而直接影响 DGI 等经典跨级对比学习方法的性能表现^[65]。从任务适配性的角度出发,图级对比学习主要用于图分类任务;节点级与跨级对比学习因其多粒度表征能力,在节点分类和链接预测等多任务场景中展现出更强的泛化能力。

3.2 生成预训练

对比预训练主要强调特征表示的判别性,容易忽视图数据的局部细节以及分布信息。相比之下,生成预训练利用图结构生成和图属性预测的

表 3 代表性的对比预训练方法

类型	文献	增强方式	下游任务	目标函数	评价指标	优点	缺点
节点级对比	文献[53]	子图采样	节点分类、图分类、相似性搜索	InfoNCE	ACC	能够扩展到大规模的图学习任务	未能有效地利用到节点属性
	文献[55]	边丢弃、属性掩码	节点分类	InfoNCE	ACC	不依赖同质性假设和数据增强构架正样本	高阶多跳交互场景的适用相对受限
	文献[56]	边丢弃、属性掩码	节点分类	InfoNCE	ACC	从全局视角中筛选出代表性正样本	计算的开销较大且效率较低
图级对比	文献[57]	节点丢弃、边扰动、属性掩码、子图采样	图分类	InfoNCE	ROC-AUC	模型具有良好泛化性和鲁棒性	效果取决于特定的数据增强方法
	文献[59]	节点丢弃、边扰动、属性掩码、子图采样	图分类	JSD	AUC-ROC	自适应和动态地选择数据增强策略	增强方法的可选空间是预定义的
	文献[60]	位置掩码、节点丢弃、生成增强、模型剪枝	图分类	InfoNCE	ROC-AUC	双视角增强机制显著提升正样本的多样性	模型剪枝增强时会修改相应架构
	文献[61]	节点丢弃、边扰动、属性掩码、子图采样	图分类	InfoNCE	ROC-AUC	不依赖于特定的数据增强方法构建正样本	训练的复杂性和时间成本相对较高
跨级对比	文献[62]	随机增强	节点分类	JSD	ACC、Micro F1	首次提出跨级图对比学习框架	语义特征区分不足且正样本收敛困难
	文献[63]	图扩散、子图采样	节点分类、图分类、节点聚类	JSD	ACC、NMI、ARI	采用多视图对比学习,计算效率较高	对特征空间的增强适用性较差
	文献[64]	随机增强	节点分类、节点聚类	JSD	ACC、NMI、ARI	提升语义区分性并促进正样本表示的收敛	生成器的训练难以平衡且模型较复杂

方式，直接建模图数据中的内在规律，通常采用交叉熵 (CE, cross entropy)、均方误差 (MSE, mean squared error) 以及缩放余弦误差 (SCE, scaled cosine error) 等损失函数优化模型参数。依据图重建任务的目标是一次性重建或序列生成，可将其分为图自编码生成和图自回归生成两类。此外，非重建任务侧重于预测图数据本身的固有属性，可将其归类为图属性预测。生成预训练基本原理框架如图 5 所示。

3.2.1 图自编码生成

图自编码生成通过掩蔽部分图信息 (结构或特征)，使模型在不完整信息下学习相应表征，并重建缺失内容，但结构掩蔽会改变图的拓扑结构，使重建图与原始图产生结构性偏差^[66]。因此，现有方法多关注特征掩蔽，即将图中部分节点特征值随机替换为特定的掩蔽标记来破坏原始特征，并依赖剩余节点特征与图结构重建被掩蔽特征，从而学习到节点特征间的关系及其与图结构的关联。

传统特征重建方法^[49,67-70]常用 MSE 作为重建损失函数，但该损失对特征向量范数和维度敏感，容易引发训练崩溃，且在不平衡样本情形下缺乏选择性。同时，上述方法还多以架构简单的多层感知机 (MLP, multi-layer perception) 作为解码器，但由于图数据中目标特征向量的信息较少，MLP 的表达能力相对受限，难以有效弥合编码器表示和目标特征之间的差距。针对上述问题，Hou 等^[71]提出图掩蔽自编码器 GraphMAE。该框架将 SCE 作为误差度量函数，以消除维度和向量范数的影响，增强训练稳定性，并通过调整缩放因子降低简单样本的

贡献度，从而有效应对样本的不平衡问题。在解码阶段，GraphMAE 提出利用更具表达能力的单层 GNN 作为解码器，基于多个节点恢复目标节点的原始特征，并在解码前再次掩蔽压缩表示提升重建难度，促使编码器学习更紧凑的表示。此方法提出了一种新的图自编码器架构，并在后续研究中广泛应用，但其性能表现依赖于输入特征的可判别性，容易受到特征扰动的影响。

由于现有特征重建方法在模型预训练时多采用简单的随机掩蔽，其特征恢复仅依赖于局部邻域信息，难以有效利用高级子结构知识。因此，Inac 等^[72]将分子图拆解为多个互不相交的基元，使掩蔽层级从节点级提升至子图级，并对图中部分基元的所有节点特征予以掩蔽，通过借助不同基元之间的特征和内部结构重建被掩蔽的节点特征，但此方法会过度依赖领域辅助知识，且需手动预定义基元结构。Liu 等^[73]提出了一种更为通用的特征掩蔽策略，即按照节点重要性程度依次掩蔽，此方法能够有效避免关键节点的过度掩蔽，并维持节点对图结构的感知，同时降低特征的恢复难度。

3.2.2 图自回归生成

自回归生成是指在生成序列或结构时，模型依次生成每个元素，并且当前步骤都依赖于前一步生成的内容。研究人员将自回归的思想扩展到图域中，设计预训练方法使模型迭代生成图的节点特征和边。其中，在特征生成阶段，模型根据已有的节点特征和相连的边生成新的特征。在边生成阶段，基于已观察到的特征和边，再生成与目标节点相关联的其他边，以此类推，交替重建图结构及其特

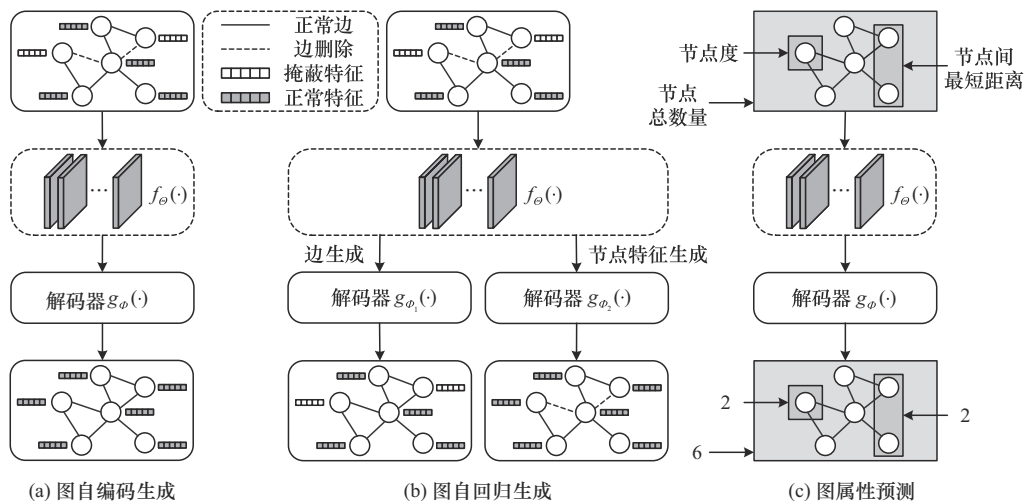


图 5 生成预训练基本原理框架

征,从而使模型学习到图数据的生成过程。

Hu等^[74]首次将自回归生成引入图模型的借口任务中,提出图生成预训练框架GPT-GNN。为了提高预训练GNN的效率,该方法在图生成过程中将节点分为边生成节点和特征生成节点。其中,边生成节点保留原始特征作为模型输入,特征生成节点则利用虚拟标记和可学习的共享向量来处理其特征,从而有效避免边生成和特征生成之间的信息干扰,并且仅需执行一次模型就可同时计算每个节点的边生成损失和特征生成损失。此外,在不同的数据集中,通过实验发现特征生成任务和边生成任务对模型的贡献存在差异,表明模型在不同数据分布下对2种任务的依赖程度不同,导致该方法需要针对特定数据集优化相应节点和边的掩蔽概率。

Zhang等^[70]将图自回归生成思想应用于分子基元树的生成过程,与GPT-GNN的随机生成顺序不同,该方法采用2种经典遍历策略,即深度优先搜索(DFS, depth-first search)和广度优先搜索(BFS, breadth-first search)。基于结构化生成路径,在有效学习分子结构和特征的同时,显著增强了自回归生成过程的有序性和可控性。Chen等^[75]将自回归生成过程扩展至动态图中,提出利用时间感知掩蔽替代GPT-GNN的随机边掩蔽,并依据先前时间点的边状态和节点特征预测未来时刻的图结构或节点状态,但该方法会面临时间依赖性问题。此外,当前基于自回归机制的其他前沿方法^[76-77]还将图结构数据转换为序列化表示,并采用Transformer模型实现图序列的生成。

3.2.3 图属性预测

图属性预测综合利用图数据的结构和语义特征,从而实现对节点、边以及子图等多层次属性的精准推断。Jin等^[67]证明了该策略在局部属性预测(如节点度数)上的有效性,并成功应用于介数中心性、接近中心性等节点重要性指标的预测^[78]。此外,该方法还将代理任务目标扩展至关系属性预测,包括节点对的连接性预测和多节点的聚类关系分析等任务。其核心思想是利用最短路径长度衡量节点间的距离,并在每个训练轮次中随机抽取部分节点构建最短路径长度预测任务,同时引入由节点聚类定义的锚节点集,通过计算锚节点与其他节点的最短距离来降低整体计算开销。Hu等^[78]进一步提出聚类保持任务,采用已知的聚类信息训练GNN模

型以维持节点聚类关系的表征,从而估计出目标节点的所属类别。目前该思路已被广泛应用到基于节点聚类和图划分任务的预训练工作中^[69,79]。

为了在有限标记数据上构建具有良好泛化性的GNN Transformer来进行分子图的表示学习,Rong等^[80]提出GROVER方法,通过提取目标节点的上下文,统计原子类型、链路数量以及连接模式等全局属性,构建了节点嵌入与上下文属性的预测关系,从而突破传统图生成任务对孤立特征的依赖,有效提升了表示清晰度。此外,该方法还结合化学领域知识预测高频基元子图,强化了模型对语义信息的捕获能力。Zang等^[81]提出的HiMol框架在图水平的表示上预测原子数量和键数量,利用这些统计信息从全局视角理解分子的结构和组成。

总体而言,研究人员对生成预训练已经进行了较为充分的研究,表4对代表性方法的生成目标、下游任务、目标函数、评价指标以及优缺点进行了系统性总结。从方法设计的角度出发,图自编码生成的研究重点在于掩蔽策略的优化,从随机掩蔽转向为有针对性的掩蔽节点特征,从而保留关键信息并降低恢复难度;图自回归生成将自回归思想应用于多种图数据,以实现更具针对性和可控性的生成过程;图属性预测呈现出目标层级的递进特性,从节点级预测拓展至关系级推理,最终实现子图层面的全局建模。从下游任务适配性的角度出发,图自编码生成因其在节点特征重建的优势,主要应用于图分类任务;图自回归生成通过学习现有连接模式和属性来推断潜在边的存在,多用于链路预测任务;图属性预测随着预测层级的提升,其适配的下游任务逐步从节点级向图级转变。

3.3 多任务预训练

在图模型的预训练中,对比任务主要关注图间判别,生成任务侧重于图内生成。以上2种单一类型的借口任务难以充分建模图数据的多维度特性,对模型的性能提升较为有限。因此,当前研究趋势正从特定任务转向多任务学习,构建节点、边、子图等多维度任务协同挖掘图数据知识。例如,Wang等^[82]针对推荐任务,结合节点、路径以及图对比学习,设计由易到难的借口任务。Yan等^[83]则关注分子图表示的学习,在原子水平重建图结构和特征,并在子图层面预测基元结构。

Luo等^[84]融合对比学习与生成学习范式,设计

表 4 代表性的生成预训练方法

类型	文献	生成目标	下游任务	目标函数	评价指标	优点	缺点
图自编码生成	文献[71]	节点特征	图分类	SCE	ROC-AUC	提出了一种典型的图自编码架构	性能表现容易受到特征扰动的影响
	文献[72]	节点特征	图分类	CE、MSE、SCE	ROC-AUC	利用基元结构，能学习到子图级别的信息	需要手动预定义基元结构
	文献[73]	节点特征	图分类	SCE	ROC-AUC	可避免关键节点的过度掩蔽，降低恢复难度	对动态图或稀疏图的处理能力有限
图自回归生成	文献[74]	节点特征和边	节点分类、链路预测	InfoNCE、MSE	MRR、Micro F1	首次将自回归思想引入借口任务中	不同数据集需调整节点和边的掩蔽概率
	文献[70] (基元级任务)	节点特征和边	图分类	MSE	ROC-AUC	显著增强自回归生成的有序性和可控性	仅适用于分子图，通用性不足
	文献[75]	节点特征和边	链路预测	InfoNCE、MSE	AUC、AP、F1-score	使用时间感知掩蔽策略，鲁棒性较强	会面临时间依赖性的问题
图属性预测	文献[67] (节点性质)	节点度	节点分类	MSE	ACC	首次引入局部属性预测任务	任务较为简单，性能提升有限
	文献[78] (聚类保持)	节点特征相似性	节点分类、图分类、链路分类	CE	Micro F1	有效提升模型的全局结构感知能力	模型效果过度依赖于聚类质量
	文献[80]	子图属性和基元子图	图分类、图回归	MSE	ROC-AUC、RMSE、MAE	模型架构灵活，且具备强大的泛化能力	模型参数规模较大且计算复杂度较高
	文献[81] (图级任务)	图属性	图分类、图回归	L1	ROC-AUC、RMSE、MAE	实现不同层级之间的信息交互	仅适用于分子图，可扩展性不足

了共享同一编码器的双分支架构。其中，重建分支分别掩蔽图的节点特征、节点和边，构建相应的视图，并驱动模型从局部视角恢复特征视图中被掩蔽的节点特征；对分支则将节点视图和边视图作为正样本对，同时将其他图的特征视图和节点视图作为负样本对，从而实现多视图对比以捕获图数据的全局信息。Wang 等^[85]在 Luo 等^[84]基础上改进了原有的多任务学习框架，在对比分支上直接将特征图和数据增强图作为正样本对，简化了正样本的选择并强化了局部特征与全局表示的协同学习；在重建分支上则通过提升节点嵌入间的方差，有效降低了 GraphMAE^[71]等图自编码器对输入特征可判别性的依赖，同时重建节点特征与整个邻接矩阵，帮助模型捕获全局图结构，但由于双分支架构和邻接矩阵重建的复杂性，该方法的训练时间相对较长。

尽管多任务学习能够组合不同借口任务提升模型性能，但传统方法通常依赖启发式策略来选择任务组合，容易忽视不同任务之间的兼容性，并且需要手动分配任务权重。为了解决以上问题，Fan 等^[86]利用多教师知识蒸馏策略提出 WAS 框架。该框架针对不同借口任务，训练多个教师模型从不同

角度提取知识，随后通过动态试错机制探索最优任务组合，并逐渐淘汰冲突任务。同时计算实例特征与不同任务潜在因子的相关性，实现任务组合权重的自适应分配。Yu 等^[87]则基于轻量级的提示机制来解决多任务协同困难的问题。与 WAS 相比，该框架理论上能够适配任何类型的借口任务，可自适应调整任务对不同层级特征的关注，从而有效避免多任务对同一层级特征参数争夺问题，但此方法仍依赖于人工设定最优的任务权重系数。

针对代表性的多任务预训练方法，表 5 从借口任务类型、下游任务、目标函数、评价指标以及优缺点等方面进行了系统总结。总体而言，多任务预训练通过联合优化框架融合图结构与语义的多维度信息，有效突破了单一任务的学习局限。当前研究重心已从人工设计任务组合与静态权重分配，转向基于任务相关性和动态优化的自适应策略。此外，得益于对图数据信息的全面捕捉，此类方法可灵活支撑节点分类、图分类等多样化的下游任务，未来可成为构建 GFM 的主流技术路径。

3.4 讨论分析

虽然研究人员关于图模型预训练已提出众多方

表5 代表性的多任务预训练方法

文献	借口任务类型	下游任务	目标函数	评价指标	优点	缺点
文献[82]	掩蔽节点预测、掩蔽链路预测、图属性预测、图级对比学习	推荐任务	NLL、CE、InfoNCE	HR、NDCG	从易到难逐步提取和学习有用信息	模型参数规模相对较大
文献[83]	节点特征重建、图属性预测	图分类	SCE、L2、CE	ROC-AUC	能自动学习到分子图的关键结构模式	仅适用于分子图,通用性不足
文献[84]	节点特征重建、图级对比学习	图分类	SCE、InfoNCE	ROC-AUC	同时学习到局部与全局信息	高度依赖数据增强策略
文献[85]	节点特征重建、掩蔽链路预测、图级对比学习	节点分类、链路预测、节点聚类、图分类	SCE、CE、MSE、DIST、VAR、InfoNCE	ACC、AUC、AP、NMI、ARI	实现局部特征与全局表示的协同学习	训练时间相对较长
文献[86]	节点特征重建、掩蔽链路预测、图自回归生成、图级对比学习、跨级对比学习、图属性预测、节点级对比学习	节点分类、图分类	MAE、CE、JSD、InfoNCE、MSE	ACC、ROC-AUC	自适应选择最优的任务组合以及任务权重	计算复杂度较高且依赖于任务池的质量
文献[87]	图级对比学习、跨级对比学习、掩蔽链路预测	节点分类、图分类	CE、InfoNCE	ACC	适配任何类型的借口任务,通用性强	仍需手动选择最优任务权重

法,但GFM具有自身的特征和需求,直接利用现有方法进行GFM的预训练仍然面临很多挑战。

1) 图模型预训练方法往往设计某种借口任务,然后将学习到的图表示应用于下游任务中,但预训练借口任务和下游任务之间的差异性使此过程可能存在“负迁移”现象^[88]。因此,GFM预训练需要统一预训练任务与下游任务的目标,从而平滑地将预训练模型的能力推广到不同的下游任务。

2) 传统的图模型预训练方法主要是在相似领域中的不同图数据之间进行模型的预训练和微调应用。然而,GFM预训练需要训练具有良好泛化性的模型,要求其能够适用于不同领域中未曾见过的图数据,因此需要进行数据特征维度、数据标签等的对齐表示操作。

3) 预训练GFM涉及大量的图数据,它们的大小差异显著,因此传统的依赖标准化输入大小的预训练模型可能效率低下。与此同时,由于领域背景存在差异,图数据的概率分布和属性特征也各不相同。鉴于此,预训练的GFM需具备识别并提取图数据中潜在各类不同模式特征的能力。

4 图模型微调策略

图模型微调是指在预训练模型的基础上,使用特定任务数据进行二次训练的技术。尽管预训练模型早已通过海量无标签图数据学习到通用的图知识,但由于预训练阶段与下游任务在数据分布上存在潜

在差异,即便同属相似领域的图数据,在节点特征、图结构、标签分布等方面都可能会存在不同。通过具有针对性的参数微调,可使模型更好地适应目标任务的特性。根据不同的技术维度,模型微调存在多种实现范式:从样本标注维度,可分为利用标注数据的有监督微调与依赖无标签数据的无监督微调;从参数更新范围,可划分为全参数优化的全微调和仅调整顶层参数或少数几层参数的部分微调。此外,从任务适用性来看,当新任务与借口任务目标存在显著差异时,由于全微调会更新全局参数,能够突破预训练模型约束重新学习新知识,从而更充分适配新任务场景需求,但需消耗大量图数据与计算资源。相对而言,部分微调限制参数更新规模,有效降低了计算成本并抑制过拟合,在资源有限或小样本场景中仍能保持模型的鲁棒性和泛化能力。

4.1 有监督微调

在有监督的图模型全微调方面,从任务差异性的角度出发,Han等^[89]认为预训练任务与下游任务在优化目标和训练数据方面存在天然差异,模型迁移性不足,因此提出了基于任务相似度的自适应辅助损失加权方法AUX-TS,通过量化任务集合中各个借口任务和下游任务的梯度之间的相似性以设计借口任务自适应加权模型,从而在微调阶段选择和组合不同的预训练和下游任务。最终,通过监督数据、基于借口任务和下游任务的联合损失对模型进行微调,提高模型的迁移性。Huang等^[90]提出了量

化借口任务和下游任务相似性的“任务一致性度量”，从而评估下游任务可以从特定的借口任务中获益的程度。在此基础上，作者提出了微调方法 Bridge-Tune，通过引入中间任务减少上下游任务不一致性的影响。

从数据差异性的角度出发，Zhang 等^[91]认为已有的微调策略主要来自文本和图像处理领域，无法充分考虑图数据的结构信息，从而提出了基于最优传输的微调方法 GTOT-Tuning。此方法将预训练图数据和下游任务图数据之间节点嵌入距离作为微调的正则化项，保持微调和预训练模型之间的局部特征不变性，实现预训练知识的高效迁移。Sun 等^[92]认为模型预训练微调过程中预训练和下游数据的结构模式差异性会明显影响迁移效果，出现“负迁移”问题，从而提出了图模型微调方法 G-TUNING，通过设计结构重构模块使预训练模型学习下游图数据的生成模式，从而适应下游任务的数据特征，缩小预训练和下游数据之间的差异。另外，为了设计更加高效、普适的图模型微调策略，Wang 等^[93]提出为给定的预训练 GNN 和下游图数据搜索合适微调策略的方法 S2PGNN，包括系统性微调框架搜索空间和高效的搜索算法。

针对全微调方法存在的高计算资源和内存占用需求，以及在下游任务数据稀缺场景中易产生灾难性遗忘与过拟合的问题，近年来研究人员开始研究图模型的部分微调方法。Gui 等^[94]分析了 NLP 领域的参数高效微调 (PEFT, parameter-efficient fine-tuning) 方法在图模型微调中的迁移可行性，并在此基础上面向 Graph Transformer 提出结构感知的图模型微调方法 G-Adapter，该方法以图结构信息作为归纳偏置指导微调过程，在保持模型轻量化与任务普适性的同时，还可无缝集成到大部分 Graph Transformer 模型中。Li 等^[95]打破现有 PEFT 方法对 Transformer 架构的依赖，面向消息传递机制的 GNN 提出双适配器架构 AdapterGNN。此模型在编码器前后并行部署可训练的适配器模块，协同捕获节点特征在消息传递前后的互补信息，并且仅需优化 5% 的模型参数即可超越传统 PEFT 与全微调方法的性能表现。考虑到上述方法在跨领域迁移场景中无法有效应对不同图的结构差异，Yang 等^[96]在预训练 GNN 的基础上引入小型可训练 GNN 模块，在微调阶段冻结预训练模型参数，并采用低秩分解

优化新增模块，显著减少了可调参数的数量，并有效整合目标图的特定结构信息。

4.2 无监督微调

为了降低有监督微调的标记开销，研究人员开始探索基于无标签数据的模型微调方法。在实际应用中，训练数据与测试数据往往存在分布偏移，导致依赖独立同分布假设的模型性能显著下降^[97]。因此，测试时自适应 (TTA, test-time adaptation) 利用测试样本自身特性动态调整模型参数，以提升其在分布外 (OOD, out of distribution) 数据的表现。根据能否访问源数据 (训练数据) 和修改源模型 (预训练模型) 训练过程，TTA 可细分为测试时训练 (TTT, test-time training) 与完全测试时自适应 (FTTA, fully test-time adaptation)。

图 6 对不同测试时适应方法的实现原理和阶段差异进行了对比，主要体现在参数更新策略的不同。针对包含 \mathcal{L} 层结构的图编码器，将其参数表示为 $\theta_f = \{\theta_e, \theta_m\}$ 。其中， $\theta_e = \{\theta_1, \theta_2, \dots, \theta_{\mathcal{L}-1}\}$ 表示共享特征编码层参数， $\theta_m = \{\theta_{\mathcal{K}}, \theta_{\mathcal{K}+1}, \dots, \theta_{\mathcal{L}}\}$ 为主任务头参数。TTT 在训练阶段联合优化主任务损失 $\mathcal{L}(\mathcal{X}_s)$ 和自监督任务损失 $\mathcal{L}(\mathcal{Y}_s, \hat{\mathcal{Y}}_s)$ ，同时更新共享特征提取器参数 θ_e 、主任务头参数 θ_m 以及自监督头参数 θ_s 。在微调阶段针对单个样本，利用自监督任务损失仅更新 θ_e 和 θ_s 并固定 θ_m ，结合优化后的 θ_e' 与原始 θ_m 进行推理。相比之下，FTTA 在训练过程中仅利用主任务损失 $\mathcal{L}(\mathcal{X}_s)$ 优化 θ_e 和 θ_m ，不需要引入额外的 θ_s 。与 FTTA 类似，TTT 也可进一步扩展为完全测试时训练 (FTTT, fully test-time training)，两者均不依赖训练数据且不改变模型训练过程，但 FTTT 测试时仅更新 θ_e ，属于部分微调。

Sun 等^[98]首次提出 TTT 的“Y”形架构，模型前 \mathcal{K} 层为共享特征提取器 (底部)，后续层则为独立的主任务头 (分支 1) 与自监督任务头 (分支 2)。虽然最初针对计算机视觉领域而设计，但 Wang 等^[99]将其扩展至 GNN 中，提出基于分层对比学习任务的 GT3 框架。该方法在训练阶段利用图分类与多层对比学习的联合损失预训练模型，并在测试时仅微调共享特征层与自监督任务头，同时冻结主任务头，实现高效的无监督部分微调。现有面向 GNN 的 TTA 方法普遍依赖图对比学习构建自监督任务，容易忽视简单的数据增强方法会破坏标签的一致性。因此，Chen 等^[100]提出基于伪标签构建正负样本组的

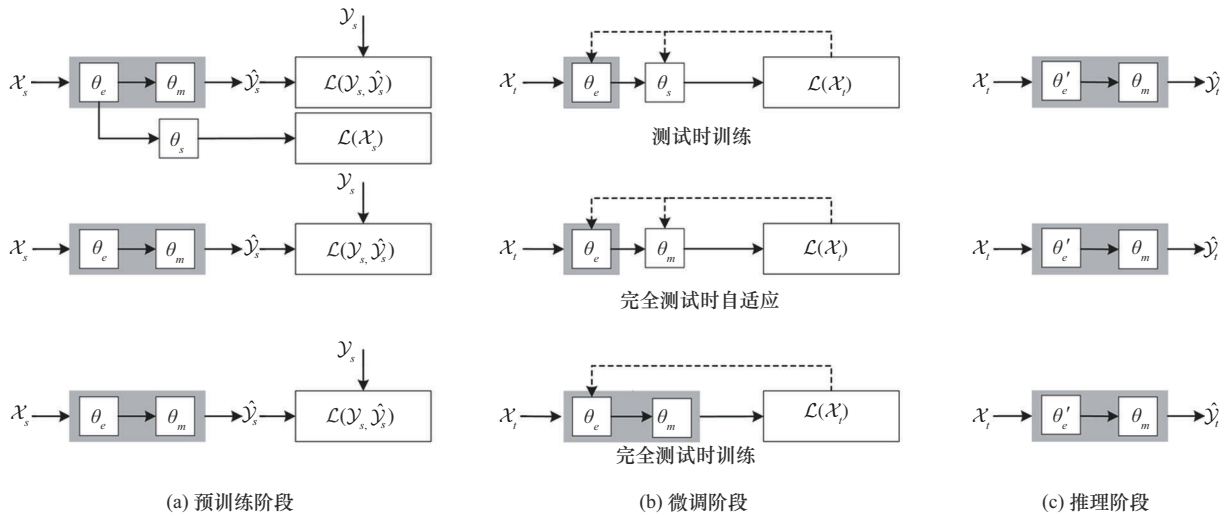


图 6 不同测试时适应方法的基本原理框架

FTTA 方法 GAPGC。该方法利用模型预测结果作为伪标签，动态聚合同类伪标签数据构建正样本组，异类则作为负样本组，并进一步通过正样本组在特征空间的固有相似性削弱部分错误伪标签的干扰。Zhang 等^[101]将 FTTT 应用到 HomoTTT 框架中，摒弃了传统方法中需要参数更新的投影头，采用无参数图对比学习，有效保持了架构的轻量化，但其性能会受限于共享特征提取器的特征区分能力。

4.3 提示调优

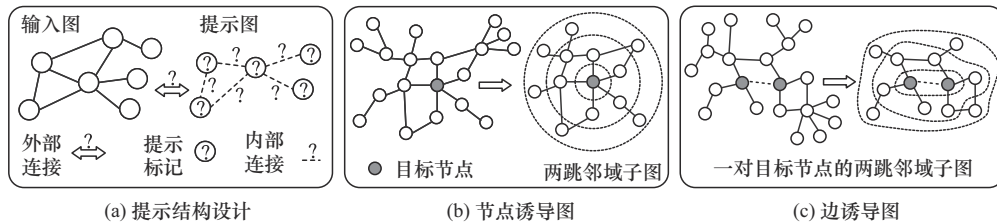
为了构建具有广泛适用性且对应用场景数据资源依赖性较低的 GFM，研究人员逐渐关注提示调优方法。作为一种高效且低成本的模型适应方法，提示调优能在不需要重新训练模型和更新模型参数的前提下，根据借口任务的特征对下游任务的数据进行转化，使其适合被预训练模型处理。与微调不同，提示调优本质上会冻结预训练模型且不改变预训练模型的参数，而是通过在数据空间中转换下游任务的输入来实现，并在小样本场景中发挥出优异表现。其形式化定义为

$$\max_{\phi} P_{f, \phi}(y | g_{\phi}(\mathcal{G})) \quad (5)$$

其目标是选择最优参数 ϕ ，在不调整预训练模型 f 的情况下最大化预测正确标签 y 的似然性。在提示调优过程中，下游任务的输入图数据 \mathcal{G} 被提示器 $g_{\phi}(\cdot)$ 转化，随后将转化后的图数据 $g_{\phi}(\mathcal{G})$ 注入已冻结的预训练模型 f 中进行处理。值得注意的是，提示调优的目的是统一多种下游任务并缩小上下游任务之间的差异，并在参数更新时固定预训练模型参数，通过优化提示标记实现对预训练模型所含知

识的高效利用。

Sun 等^[102]首次提出提示调优方法 GPPT。该方法通过掩蔽链路预测进行预训练，并将图提示设计为由任务标记和结构标记组成的标记对。其中，任务标记是一个可训练的连续向量，代表节点的原型，结构标记表示目标节点及其邻域信息，通过计算标记对的链接概率，将概率最高的任务标记类别作为目标节点的标签，从而将下游任务转化为链路预测任务。在此过程中，编码器的参数被固定，仅优化 MLP 的参数和提示标记对的嵌入表示，显著减少了参数的更新次数，但此方法仅适用于节点分类任务。为了统一多样化的下游任务，Sun 等^[103]基于邻域结构的诱导图生成机制，将节点、链路等各级的图学习任务统一表示为图水平任务，并将对应的节点/图标签作为诱导图标签，其原理如图 7 所示。同时，该方法基于语言提示设计结构提示，其中包含多个节点标记以及标记之间的链路，并借助点积计算标记之间以及标记与下游图节点之间的相似度，从而将提示图插入下游图数据以提供额外的知识引导。此外，现有方法在图提示设计时主要依赖于相似性度量，未能有效借鉴 NLP 中的掩蔽语言建模 (MLM, masked language modeling) 思想。尽管 Zhu 等^[104]引入连接全图的超级节点作为提示，并采用掩蔽特征重建策略将图分类转换为节点分类任务，但其本质上仍未实现真正的 MLM 范式，其关键缺陷在于该方法仅将重建损失作为正则化项，从而辅助提示嵌入与类原型表示的对齐，而节点类别却并未在重建信息中体现。

图7 Sun等^[103]提出的提示调优框架

上述方法主要为同质图而设计,未能有效应对异质图的相关特性。针对异质图节点类型多样化问题, Ma等^[105]为每种节点类型引入一组独立基向量,并与节点表示加权融合,同时综合考虑基于类型与元路径的邻域信息,生成更具代表性的目标节点嵌入,随后将其与类原型表示进行相似性度量。尽管该方法在思想上与GPPT类似,但由于多视图邻域聚合机制,提示调优方法能够有效扩展至异质图场景中。Yu等^[106]从图结构分解的角度出发,基于节点类型将异质图分解为多个简单的同质子图,并利用子图相似性构建统一任务模板。在预训练阶段,对比候选节点上下文子图表示的相似性。针对下游任务,将节点分类转化为目标节点与类原型节点的子图相似性度量,而图分类则直接度量目标图与类原型图的嵌入相似性。该方法采用双提示机制,在池化层中引入可学习的特征提示,对子图中的节点表示进行动态加权,以辅助子图特征的有效聚合。此外,在聚合多个同质子图的池化结果时,利用异质性提示自适应调整不同子图的重要性,从而获得具有判别性的图表示。

现有提示调优方法主要针对特定的预训练模型,缺乏广泛适用性,而图学习领域多样化的预训练策略对设计通用的GNN提示调优方法提出了挑战。Fang等^[107]提出一种通用提示微调方法GPF,将可学习向量添加到所有的节点特征中,理论上可实现与任何提示函数相同的效果。进一步地,作者以多个基础向量加权组合而成的独立可学习向量,取代所有节点共享的单一可学习向量,显著提升了模型的表现力与灵活性。类似地, Jiang等^[108]和 Lee等^[109]分别将提示向量的作用对象扩展至不同重要性的节点、边及子图,从不同尺度捕捉图数据的复杂信息。此外,不同于特征提示机制, Duan等^[110]基于图元理论为不同类别的下游数据生成图级子提示,并通过集成同类子提示构建精确的图级提示,将其与类原型表示进行度量,从而确定最优

类别。总体上,虽然上述方法不再局限于特定的预训练模型,但其本质上并没有统一多样化的下游任务以及消除上下游任务间的目标差异。

4.4 讨论分析

虽然研究人员对图模型微调进行了探索,但现有方法无法直接适用于GFM的微调工作,同时相关研究还不够充分深入, GFM微调仍面临诸多挑战。

1) 当前图模型微调方法主要针对特定下游任务而设计,但不同任务目标具有差异性,导致现有方法难以满足GFM对不同下游任务的同质性适配需求。虽然提示调优方法已实现部分下游任务的统一,但其应用范围仍主要局限于分类场景。为了实现GFM的跨任务通用性,未来需构建更加统一的任务框架,揭示任务之间的本质关联,进而开发出具备普适性的图模型微调方法。

2) 现有图模型微调方法假设上下游图数据具有相同的结构模式和节点特征。然而,在实际应用中,不同领域的图数据往往存在拓扑结构差异、边语义异构以及节点/边不匹配等问题,导致预训练知识难以有效地迁移至下游任务。未来应利用基础表示单元实现统一表征,以适应不同领域的图数据,进而设计出更具泛化性的模型微调方法。

3) 现有预训练方法主要采用节点特征重建和图对比学习等自监督任务训练模型,而下游任务多为节点分类或图分类任务,导致上下游目标存在差异,使模型迁移效果受限。此外,现有部分提示调优方法局限于链路预测或子图相似性度量等特定预训练任务,缺乏通用性。因此,亟须设计通用性强且目标任务差异相对较小的提示调优策略。

4) 当前提示调优策略多依赖于复杂而精巧的提示设计,因而需要探索更加简单且轻量化的图提示构建方法。此外,类似于自然语言中的提示机制,图领域中的提示设计也应进一步研究基于MLM思想的提示调优策略,将分类任务转换为生

成任务,并在信息重建过程中直接或间接地反映出类别信息,从而达到分类的目的。

5 评价指标

5.1 性能评价指标

GFM利用预训练提取图数据的通用拓扑与语义特征,并基于微调机制适配于各种下游任务,同时采用差异化的性能评价指标衡量模型的性能表现。其中,节点分类任务^[54-56,62,67]通常以准确率衡量分类精度,图分类任务^[57-61,71-73]通过特征曲线下面积(ROC)-AUC评估类别区分度的鲁棒性。链路预测^[66,74-75]作为二元分类问题,主要选取F1-score及其细粒度变体Micro-F1指标,以综合平衡精确率与召回率之间的关系。对于图回归任务^[80-81],多选用均方根误差(RMSE, root mean square error)与平均绝对误差(MAE, mean absolute error)等连续误差指标衡量预测值与真实值间的偏差。此外,归一化互信息(NMI, normalized mutual information)与调整兰德指数(ARI, adjusted rand index)被广泛用于评估节点聚类结果与真实标签的一致性^[63-64]。

5.2 效率评价指标

效率评价指标主要利用训练时间和推理速度量化GFM的运行效率。其中,训练时间反映模型在训练过程中的耗时,其长短直接影响GFM的开发周期和部署效率,并与数据集的规模和参数更新量呈正相关。此外,优化模型训练算法,如学习率调度、提前停止等策略,有助于提升训练效率,同时保证模型性能。推理速度衡量的是模型针对测试样本的预测速度,即模型在接收输入数据后做出准确预测的速度,通常以“单位时间处理的样本数”或“单位样本的处理时间”来度量。该指标直接影响模型响应的及时性,即模型推理速度越快,响应越迅速,决策效率越高,从而能够在实时性要求较高的场景中提供良好的用户体验。

5.3 资源消耗评价指标

在实际应用中,GFM的训练过程需要消耗大量的计算资源,其中计算资源消耗和内存占用是评价模型表现的关键指标,直接决定其在有限资源条件下的可行性与高效性。具体而言,计算资源消耗反映模型运行时对硬件资源(如CPU、GPU等)的需求程度,随着现有模型复杂度和数

据集规模的增加,计算资源需求往往呈指数级增长,从而显著提高了模型的运行成本。此外,内存占用则是度量模型在训练过程中处理数据时所消耗的内存量,涉及数据加载、中间结果存储以及参数更新等操作。过高的内存占用会导致模型响应迟缓甚至崩溃,严重影响训练效率并制约模型的实际性能表现。

5.4 可迁移性评价指标

可迁移性评价指标主要用于衡量模型从源任务学习到的知识向目标任务迁移的难易程度。对于追求通用表现的GFM而言,该指标可筛选出具有高迁移性的借口任务,并将其作为优先选项,同时针对性地优化特定下游任务的迁移瓶颈以及揭示影响迁移效果的关键因素。Nguyen等^[111]提出的LEEP指标利用伪标签与下游任务标签的联合概率,以经验预测值的对数期望评价迁移性能,但其评价结果易受分类头参数的影响,并且仅适用于有监督模型到下游分类任务的迁移。针对这些问题,You等^[112]设计了LogME指标,通过建模预训练模型提取的目标数据特征与标签的平均最大对数证据,直接分析特征-标签相关性而非依赖伪标签,显著提升了评价指标的通用性。随后的研究依然围绕特征与标签的关联性展开,但采用了多样化的技术路线。例如,TransRate^[113]基于互信息衡量特征与标签的兼容性,但其受限于隐藏层表示的熵估计,导致计算效率较低。TMI^[114]通过计算类内特征方差的变化程度,实现最优预训练模型的选择与最佳迁移层的判定。KITE^[115]针对特征在不同任务场景下的应用,动态调整迁移评价策略。该方法在粗粒度分类任务中计算特征和理想特征(真实标签)的关联性,而在细粒度分类任务中对比特征与随机特征的差异性。此外,BeST^[116]将源模型的Softmax输出量化,并建立量化级别与验证准确率的映射关系,以最优量化级对应的验证准确率作为度量指标,间接揭示特征与标签之间的潜在关系。至今为止,可迁移性评价指标的实验验证多集中于语言模型和视觉模型中,对图模型的实证研究仍较为有限,除TransRate外,其他方法^[111-112,114-116]在GFM迁移场景的实际有效性尚待进一步验证。

6 面临的挑战与展望

GFM对于推动图智能系统实际应用有着重要

意义,受到越来越多的关注。近年来随着大模型技术的进步,该领域取得了快速发展,但仍面临一些挑战和值得研究的方向。

1) 与 LLM 领域直接采用 Transformer 作为基础架构进行模型构建不同,图学习领域虽然提出了消息传递、Graph Transformer、几何图神经网络等多种模型架构,但在模型表达能力、计算复杂度、扩展能力、适用范围等方面难以兼顾,至今仍然缺少具有强大表达能力且能适用于 GFM 特性要求的图模型架构,需要不断探索和创新。

2) GFM 的基础模型特性要求提示调优以统一的方式处理多样化的下游任务,同时实现与预训练阶段借口任务的适配。然而,节点级、链路级和图级的图学习任务非常多样化。虽然研究人员对此问题有了初步的探索,但往往假设已知下游任务或者面向特定借口任务,具有针对性,仍缺乏适用于各种任务的统一表达机制,难以满足 GFM 的实际应用需求。因此,研究灵活且通用的图学习任务的统一表达形式,对于推动 GFM 的发展和具有重要应用意义,也是当前亟须解决的关键问题。

3) 图数据源自多个不同领域,其特征、标签及语义各异,当前缺乏跨数据集、跨领域图数据的基础结构单元和统一的数据表征形式,这阻碍了 GFM 的有效训练,因此设计跨领域适用的基础结构单元和统一数据表征,对于促进 GFM 的跨数据集训练和泛化至关重要。根据解耦表征思想,通过建模图数据整体模式规律学习其背后的基本单元,揭示“多层次、多尺度的隐因子”,提取并融合不同领域图数据的主要结构模式,可能是实现 GFM 基础结构单元和统一数据表征的一种有效途径。

4) 现有图模型微调相关工作研究较少,且大部分都关注有监督全微调和提示调优,由于基础模型的规模性,进行有监督全微调需要大量标记数据和计算资源,成本高昂。因此,对于资源依赖性较小的图模型,有监督部分微调、无监督微调以及提示调优等方法的研究对 GFM 的发展具有重要的现实意义。另外,为了降低 GFM 的资源消耗,可以考虑采用模型蒸馏的方式,以其他模型指导构建小规模 GFM,从而提升训练效率,并使其在保持高性能的同时降低计算资源的需求。

5) 研究人员对 GFM 可解释性的研究关注较少。GFM 的可解释性对于阐明其行为、局限性和社会

影响至关重要。通过阐明 GFM 预测背后的推理机制,理解其决策过程,可以避免算法偏见,提升社会对模型的信任水平,也可以发现模型的局限性和潜在缺陷,支撑研究人员进行模型的优化改进。虽然研究人员关于 GNN 提出了基于梯度的、基于扰动的、基于分解的等各类可解释性方法,然而以上方法对于 GFM 的有效性还缺乏实践证明。同时,由于大量的模型参数、大规模的训练数据以及模型微调等训练策略,仍需要设计与 GFM 的特征相适配的新型可解释性方法。

6) GFM 的构建高度依赖于丰富图数据。尽管当前图学习领域已有一些开放数据集,如 Open Graph Benchmark^[117],但与构建 LLM 所使用的庞大数据量相比,这些数据在多样性和规模上仍存在明显不足。因此,为推动 GFM 的构建,收集整理高质量、多样化的图数据显得尤为重要。同时,为了保障 GFM 训练所需的数据资源,面向合成图数据构造的图生成模型、图对抗生成网络等应该受到更多关注。

7 结束语

本文首先总结了 GFM 的基本概念,阐述了研究现状,给出了形式化定义和问题框架。然后,对与 GFM 构建紧密相关的图模型骨干架构和基础表示单元进行了介绍。在此基础上,分类梳理了图模型预训练和微调的相关研究工作,讨论了对 GFM 的支撑作用和不足之处。此外,将 GFM 相关的评价指标进行了总结。最后,对 GFM 构建面临的挑战和未来的工作方向进行了阐述。通过持续的研究和创新,GFM 有望为泛在的图结构化数据的建模学习和图智能系统提供通用的解决方案。

参考文献:

- [1] ZANIN M, PAPO D, SOUSA P A, et al. Combining complex networks and data mining: why and how[J]. *Physics Reports*, 2016, 635: 1-44.
- [2] MUSCOLONI A, THOMAS J M, CIUCCI S, et al. Machine learning meets complex networks via coalescent embedding in the hyperbolic space[J]. *Nature Communications*, 2017, 8(1): 1615.
- [3] TANG Y, KURTHS J, LIN W, et al. Introduction to focus issue: when machine learning meets complex systems: networks, chaos, and nonlinear dynamics[J]. *Chaos: an Interdisciplinary Journal of Nonlinear Science*, 2020, 30(6): 063151.
- [4] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.

- [5] WANG H, FENG S B, HE T X, et al. Can language models solve graph problems in natural language[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2023: 30840-30861.
- [6] CHEN Z K, MAO H T, LI H, et al. Exploring the potential of large language models (LLMs) in learning on graphs[J]. ACM SIGKDD Explorations Newsletter, 2024, 25(2): 42-61.
- [7] LIU C, WU B. Evaluating large language models on graphs: performance insights and comparative analysis[J]. arXiv Preprint, arXiv: 2308.11224, 2023.
- [8] HE X X, BRESSON X, LAURENT T, et al. Harnessing explanations: LLM-to-LM interpreter for enhanced text-attributed graph representation learning[J]. arXiv Preprint, arXiv: 2305.19523, 2023.
- [9] XIE H, ZHENG D, MA J, et al. Graph-aware language model pre-training on a large graph corpus can help multiple graph applications[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2023: 5270-5281.
- [10] ZHU J, SONG X, IOANNIDIS V, et al. TouchUp-G: improving feature representation through graph-centric finetuning[C]//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2024: 2662-2666.
- [11] TANG J B, YANG Y H, WEI W, et al. GraphGPT: graph instruction tuning for large language models[C]//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2024: 491-500.
- [12] ZOU T, YU L, HUANG Y F, et al. Pretraining language models with text-attributed heterogeneous graphs[J]. arXiv Preprint, arXiv: 2310.12580, 2023.
- [13] PENG Y, LIN S, CHEN Q, et al. ChatGraph: chat with your graphs[C]//Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2024: 5445-5448.
- [14] WANG Y, ROSSIRA, PARK N, et al. Large generative graph models[J]. arXiv Preprint, arXiv: 2406.05109, 2024.
- [15] HE Y F, SUI Y, HE X X, et al. UniGraph: learning a cross-domain graph foundation model from natural language[J]. arXiv Preprint, arXiv: 2402.13630, 2024.
- [16] LIU H, FENG J R, KONG L C, et al. One for all: towards training one graph model for all classification tasks[J]. arXiv Preprint, arXiv: 2310.00149, 2023.
- [17] LACHI D, AZABOU M, ARORA V, et al. GraphFM: a scalable framework for multi-graph pretraining[J]. arXiv Preprint, arXiv: 2407.11907, 2024.
- [18] XIA L H, KAO B, HUANG C. OpenGraph: towards open graph foundation models[J]. arXiv Preprint, arXiv: 2403.01121, 2024.
- [19] LI Y H, WANG P S, LI Z X, et al. ZeroG: investigating cross-dataset zero-shot transferability in graphs[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2024: 1725-1735.
- [20] HU Y L, OUYANG S, YANG Z R, et al. Exploring task unification in graph representation learning via generative approach[J]. arXiv Preprint, arXiv: 2403.14340, 2024.
- [21] LIU J Z, MAO H T, CHEN Z K, et al. Neural scaling laws on graphs[J]. arXiv Preprint, arXiv: 2402.02054, 2024.
- [22] ZHAO J A, ZHU Z C, GALKIN M, et al. Fully-inductive node classification on arbitrary graphs[J]. arXiv Preprint, arXiv: 2405.20445, 2024.
- [23] WU F, ZHANG T Y, SOUZA A, et al. Simplifying graph convolutional networks[C]//Proceedings of the 36th International Conference on Machine Learning (ICML). New York: ACM Press, 2019: 6861-6871.
- [24] FRASCA F, JOGL F, ELIASOF M, et al. Towards foundation models on graphs: an analysis on cross-dataset transfer of pretrained GNN[J]. arXiv Preprint, arXiv: 2412.17609, 2024.
- [25] KLASER K, BANASZEWSKI B, MADDRELL-MANDER S, et al. Minimol: a parameter-efficient foundation model for molecular learning[J]. arXiv Preprint, arXiv: 2404.14986, 2024.
- [26] SYPETKOWSKI M, WENKEL F, POURSARFAEI F, et al. On the scalability of gnns for molecular graphs[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2024: 19870-19906.
- [27] SHOGHI N, KOLLURU A, KITCHIN J R, et al. From molecules to materials: pre-training large generalizable models for atomic property prediction[J]. arXiv Preprint, arXiv: 2310.16802, 2023.
- [28] ZHANG Z W, LI H Y, ZHANG Z Y, et al. Graph meets llms: towards large graph models[J]. arXiv Preprint, arXiv: 2308.14522, 2023.
- [29] MAO H T, CHEN Z K, TANG W Z, et al. Position: graph foundation models are already here[J]. arXiv Preprint, arXiv: 2402.02216, 2024.
- [30] LIU J W, YANG C, LU Z Y, et al. Towards graph foundation models: a survey and beyond[J]. arXiv Preprint, arXiv: 2310.11829, 2023.
- [31] LI Y H, LI Z X, WANG P S, et al. A survey of graph meets large language model: progress and future directions[J]. arXiv Preprint, arXiv: 2311.12399, 2023.
- [32] FAN W Q, WANG S J, HUANG J N, et al. Graph machine learning in the era of large language models (LLMs)[J]. arXiv Preprint, arXiv: 2404.14928, 2024.
- [33] REN X B, TANG J B, YIN D W, et al. A survey of large language models for graphs[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2024: 6616-6626.
- [34] JIN B W, LIU G, HAN C, et al. Large language models on graphs: a comprehensive survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(12): 8622-8642.
- [35] ZHANG B X, FAN C J, LIU S X, et al. The expressive power of graph neural networks: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2025, 37(3): 1455-1474.
- [36] LIU M, GAO H Y, JI S W. Towards deeper graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 338-348.
- [37] SHEHZAD A, XIA F, ABID S, et al. Graph transformers: a survey[J]. arXiv Preprint, arXiv: 2407.09777, 2024.
- [38] MÉNDEZ-LUCIO O, NICOLAOU C A, EARNSHAW B. MolE: a foundation model for molecular graphs using disentangled attention[J]. Nature Communications, 2024, 15(1): 9431.
- [39] MÜLLER L, GALKIN M, MORRIS C, et al. Attending to graph transformers[J]. arXiv Preprint, arXiv: 2302.04181, 2023.
- [40] TAY Y, DEGHANI M, BAHRI D, et al. Efficient transformers: a survey[J]. ACM Computing Surveys, 2023, 55(6): 1-28.
- [41] SOYDANER D. Attention mechanism in neural networks: where it comes and where it goes[J]. Neural Computing and Applications, 2022, 34(16): 13371-13385.
- [42] HAN J Q, CEN J C, WU L M, et al. A survey of geometric graph neural networks: data structures, models and applications[J]. arXiv Preprint, arXiv: 2403.00485, 2024.
- [43] JIAO R, KONG X Z, YU Z Y, et al. Equivariant pretrained transformer for unified geometric learning on multi-domain 3D molecules[J]. arXiv Preprint, arXiv: 2402.12714, 2024.

- [44] SUN L, HUANG Z H, ZHOU S Y, et al. RiemannGFM: learning a graph foundation model from riemannian geometry[J]. arXiv Preprint, arXiv: 2502.03251, 2025.
- [45] ZHAO H H, CHEN A C, SUN X G, et al. All in one and one for all: a simple yet effective method towards cross-domain graph pretraining[C]// Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2024: 4443-4454.
- [46] CHENG Y, ZHAO Y G, YU J X, et al. Boosting graph foundation model from structural perspective[J]. arXiv Preprint, arXiv: 2407.19941, 2024.
- [47] YU X T, GONG Z C, ZHOU C, et al. SAMGPT: text-free graph foundation model for multi-domain pre-training and cross-domain adaptation[J]. arXiv Preprint, arXiv: 2502.05424, 2025.
- [48] WANG K, LUO S Q. Towards graph foundation models: the perspective of zero-shot reasoning on knowledge graphs[J]. arXiv Preprint, arXiv: 2410.12609, 2024.
- [49] HU W H, LIU B W, GOMES J, et al. Strategies for pre-training graph neural networks[J]. arXiv Preprint, arXiv: 1905.12265, 2019.
- [50] HOU Y P, HU B B, ZHAO W X, et al. Neural graph matching for pre-training graph neural networks[C]// Proceedings of the 2022 SIAM International Conference on Data Mining (SDM). Philadelphia: Society for Industrial and Applied Mathematics, 2022: 172-180.
- [51] OORD A V D, LI Y Z, VINYALS O. Representation learning with contrastive predictive coding[J]. arXiv Preprint, arXiv: 1807.03748, 2018.
- [52] HJELM R D, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning deep representations by mutual information estimation and maximization[J]. arXiv Preprint, arXiv: 1808.06670, 2018.
- [53] QIU J Z, CHEN Q B, DONG Y X, et al. GCC: graph contrastive coding for graph neural network pre-training[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 1150-1160.
- [54] ZHANG H R, WU Q T, WANG Y, et al. Localized contrastive learning on graphs[J]. arXiv Preprint, arXiv: 2212.04604, 2022.
- [55] XIAO T, ZHU H S, CHEN Z Y, et al. Simple and asymmetric graph contrastive learning without augmentations[C]// Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2023: 16129-16152.
- [56] ZHUO J M, QIN F Y, CUI C, et al. Improving graph contrastive learning via adaptive positive sampling[C]// Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 23179-23187.
- [57] YOU Y N, CHEN T L, SUI Y D, et al. Graph contrastive learning with augmentations[C]// Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2020: 5812-5823.
- [58] SOHN K. Improved deep metric learning with multi-class n-pair loss objective[C]// Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2016, 1857-1865.
- [59] YOU Y N, CHEN T L, SHEN Y, et al. Graph contrastive learning automated[C]// Proceedings of International Conference on Machine Learning (ICML). New York: ACM Press, 2021: 12121-12132.
- [60] YAO Z L, LIU C, MA X Q, et al. Dual-perspective cross contrastive learning in graph transformers[J]. arXiv Preprint, arXiv: 2406.00403, 2024.
- [61] JI Q R, LI J M, HU J, et al. Rethinking dimensional rationale in graph contrastive learning from causal perspective[C]// Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2024: 12810-12820.
- [62] VELIČKOVIĆ P, FEDUS W, HAMILTON W L, et al. Deep graph infomax[J]. arXiv Preprint, arXiv: 1809.10341, 2018.
- [63] HASSANI K, KHASAHMADI A H. Contrastive multi-view representation learning on graphs[C]// Proceedings of the 37th International Conference on Machine Learning (ICML). New York: ACM Press, 2020: 4116-4126.
- [64] ZHAO J T, HE D X, GE M, et al. GA-GGD: improving semantic discriminability in graph contrastive learning via generative adversarial network[J]. Information Fusion, 2024, 110: 102465.
- [65] ZHENG Y Z, PAN S R, LEE V C, et al. Rethinking and scaling up graph contrastive learning: an extremely efficient approach with group discrimination[C]// Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2022: 10809-10820.
- [66] ZUO S M, JIANG H M, YIN Q Y, et al. Dip-GNN: discriminative pre-training of graph neural networks[J]. arXiv Preprint, arXiv: 2209.07499, 2022.
- [67] JIN W, DERR T, LIU H C, et al. Self-supervised learning on graphs: deep insights and new direction[J]. arXiv Preprint, arXiv: 2006.10141, 2020.
- [68] LIU Y, YANG S S, LEI C Y, et al. Pre-training graph transformer with multimodal side information for recommendation[C]// Proceedings of the 29th ACM International Conference on Multimedia (ACM MM). New York: ACM Press, 2021: 2853-2861.
- [69] YOU Y N, CHEN T L, WANG Z Y, et al. When does self-supervision help graph convolutional networks [C]// Proceedings of the 37th International Conference on Machine Learning (ICML). New York: ACM Press, 2020: 10871-10880.
- [70] ZHANG Z X, LIU Q, WANG H, et al. Motif-based graph self-supervised learning for molecular property prediction[C]// Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2021: 15870-15882.
- [71] HOU Z Y, LIU X, CEN Y K, et al. Graphmae: self-supervised masked graph autoencoders[C]// Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2022: 594-604.
- [72] INAE E, LIU G, JIANG M. Motif-aware attribute masking for molecular graph pre-training[J]. arXiv Preprint, arXiv: 2309.04589, 2023.
- [73] LIU C, WANG Y Y, ZHAN Y B, et al. Where to mask: structure-guided masking for graph masked autoencoders[J]. arXiv Preprint, arXiv: 2404.15806, 2024.
- [74] HU Z N, DONG Y X, WANG K S, et al. GPT-GNN: generative pre-training of graph neural networks[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2020: 1857-1867.
- [75] CHEN K J, ZHANG J J, JIANG L P, et al. Pre-training on dynamic graph neural networks[J]. Neurocomputing, 2022, 500: 679-687.
- [76] ZHAO Q F, REN W D, LI T Y, et al. Graphgpt: graph learning with generative pre-trained transformers[J]. arXiv Preprint, arXiv: 2401.00529, 2023.
- [77] CHEN X H, WANG Y K, HE J X, et al. Graph generative pre-trained transformer[J]. arXiv Preprint, arXiv: 2501.01073, 2025.
- [78] HU Z N, FAN C J, CHEN T, et al. Pre-training graph neural networks for generic structural feature extraction[J]. arXiv Preprint, arXiv: 1905.13728, 2019.
- [79] YANG Y M, GUAN Z Y, WANG Z, et al. Self-supervised heterogeneous graph pre-training based on structural clustering[C]// Proceed-

- ings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2022: 16962-16974.
- [80] RONG Y, BIAN Y T, XU T Y, et al. Self-supervised graph transformer on large-scale molecular data[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2020: 12559-12571.
- [81] ZANG X, ZHAO X B, TANG B Z. Hierarchical molecular graph self-supervised learning for property prediction[J]. Communications Chemistry, 2023, 6(1): 34.
- [82] WANG H, ZHOU K, ZHAO X, et al. Curriculum pre-training heterogeneous subgraph transformer for top-n recommendation[J]. ACM Transactions on Information Systems, 2023, 41(1): 1-28.
- [83] YAN P W, SONG K S, JIANG Z R, et al. Empowering dual-level graph self-supervised pre-training with motif discovery[C]//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2024: 9223-9231.
- [84] LUO Y C, LI S H, SUI Y D, et al. Masked graph modeling with multi-view contrast[C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2024: 2584-2597.
- [85] WANG Y X, YAN X, HU C, et al. Generative and contrastive paradigms are complementary for graph self-supervised learning[C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2024: 3364-3378.
- [86] FAN T Y, WU L R, HUANG Y F, et al. Decoupling weighing and selecting for integrating multiple graph pre-training tasks[J]. arXiv Preprint, arXiv: 2403.01400, 2024.
- [87] YU X T, ZHOU C, FANG Y, et al. Multigprompt for multi-task pre-training and prompting on graphs[C]//Proceedings of the ACM Web Conference 2024 (WWW). New York: ACM Press, 2024: 515-526.
- [88] CAO Y X, XU J R, YANG C, et al. When to pre-train graph neural networks From data generation perspective![C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2023: 142-153.
- [89] HAN X T, HUANG Z H, AN B, et al. Adaptive transfer learning on graph neural networks[C]//Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2021: 565-574.
- [90] HUANG R H, XU J R, JIANG X, et al. Measuring task similarity and its implication in fine-tuning graph neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2024: 12617-12625.
- [91] ZHANG J Y, XIAO X, HUANG L K, et al. Fine-tuning graph neural networks via graph topology induced optimal transport[J]. arXiv Preprint, arXiv: 2203.10453, 2022.
- [92] SUN Y F, ZHU Q, YANG Y, et al. Fine-tuning graph neural networks by preserving graph generative patterns[C]//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2024: 9053-9061.
- [93] WANG Z L, DI S M, CHEN L, et al. Search to fine-tune pre-trained graph neural networks for graph-level tasks[C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2024: 2805-2819.
- [94] GUI A C, YE J Q, XIAO H. G-adapter: towards structure-aware parameter-efficient transfer learning for graph transformer networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2024: 12226-12234.
- [95] LI S R, HAN X T, BAI J. Adaptergnn: parameter-efficient fine-tuning improves generalization in gnns[C]//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2024: 13600-13608.
- [96] YANG Z R, HAN J D, WANG C D, et al. Graphlora: structure-aware contrastive low-rank adaptation for cross-graph transfer learning[J]. arXiv Preprint, arXiv: 2409.16670, 2024.
- [97] LI H Y, WANG X, ZHANG Z W, et al. Out-of-distribution generalization on graphs: a survey[J]. arXiv Preprint, arXiv: 2202.07987, 2022.
- [98] SUN Y, WANG X L, LIU Z, et al. Test-time training with self-supervision for generalization under distribution shifts[C]//Proceedings of International conference on machine learning (ICML). New York: ACM Press, 2020: 9229-9248.
- [99] WANG Y Q, LI C Z, JIN W, et al. Test-time training for graph neural networks[J]. arXiv Preprint, arXiv: 2210.08813, 2022.
- [100] CHEN G Z, ZHANG J Y, XIAO X, et al. Graphtta: test time adaptation on graph neural networks[J]. arXiv Preprint, arXiv: 2208.09126, 2022.
- [101] ZHANG J X, WANG Y Q, YANG X H, 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.
- [102] SUN M C, ZHOU K X, HE X, et al. Gpnt: graph pre-training and prompt tuning to generalize graph neural networks[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2022: 1717-1727.
- [103] SUN X G, CHENG H, LI J, et al. All in one: multi-task prompting for graph neural networks[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2023: 2120-2131.
- [104] ZHU Y, GUO J H, TANG S L. Sgl-pt: a strong graph learner with graph prompt tuning[J]. arXiv Preprint, arXiv: 2302.12449, 2023.
- [105] MA Y H, YAN N, LI J Y, et al. Hetgpt: harnessing the power of prompt tuning in pre-trained heterogeneous graph neural networks[C]//Proceedings of the ACM Web Conference 2024 (WWW). New York: ACM Press, 2024: 1015-1023.
- [106] YU X T, FANG Y, LIU Z M, et al. Hgprompt: bridging homogeneous and heterogeneous graphs for few-shot prompt learning[C]//Proceedings of the AAAI conference on artificial intelligence (AAAI). Palo Alto: AAAI Press, 2024: 16578-16586.
- [107] FANG T R, ZHANG Y C, YANG Y, et al. Universal prompt tuning for graph neural networks[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2023: 52464-52489.
- [108] JIANG B, WU H, ZHANG Z Y, et al. A unified graph selective prompt learning for graph neural networks[J]. arXiv Preprint, arXiv: 2406.10498, 2024.
- [109] LEE J, YANG W, KANG J. Subgraph-level universal prompt tuning[J]. arXiv Preprint, arXiv: 2402.10380, 2024.
- [110] DUAN Y T, LIU J, CHEN S W, et al. G-Prompt: graphon-based prompt tuning for graph classification[J]. Information Processing & Management, 2024, 61(3): 103639.
- [111] NGUYEN C, HASSNER T, SEEGER M, et al. Leep: a new measure

to evaluate transferability of learned representations[C]//Proceedings of International Conference on Machine Learning (ICML). New York: ACM Press, 2020: 7294-7305.

- [112] YOU K C, LIU Y, WANG J M, et al. Logme: practical assessment of pre-trained models for transfer learning[C]//Proceedings of International Conference on Machine Learning (ICML). New York: ACM Press, 2021: 12133-12143.
- [113] HUANG L K, HUANG J Z, RONG Y, et al. Frustratingly easy transferability estimation[C]//Proceedings of International conference on machine learning (ICML). New York: ACM Press, 2022: 9201-9225.
- [114] XU H W, KANG U. Fast and accurate transferability measurement by evaluating intra-class feature variance[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 11474-11482.
- [115] GUO Y H. KITE: a kernel-based improved transferability estimation method[J]. arXiv Preprint, arXiv: 2405.01603, 2024.
- [116] SONI A, JU P Z, ERYILMAZ A, et al. BeST--a novel source selection metric for transfer learning[J]. arXiv Preprint, arXiv: 2501.10933, 2025.
- [117] HU W H, FEY M, ZITNIK M, et al. Open graph benchmark: datasets for machine learning on graphs[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Massachusetts: MIT Press, 2020: 22118-22133.



先兴平 (1984-), 女, 四川泸州人, 博士, 重庆邮电大学副教授、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。



王超 (1986-), 男, 博士, 重庆师范大学副教授、硕士生导师, 主要研究方向为时序数据和图数据表征、自然语言处理及其应用等。

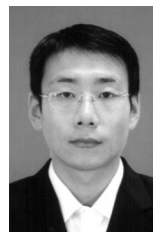


袁霖 (1989-), 男, 河南新蔡人, 博士, 重庆邮电大学副教授、硕士生导师, 主要研究方向为计算机视觉、深度图像伪造、数据隐私保护等。

[作者简介]



吴涛 (1987-), 男, 甘肃庆阳人, 博士, 重庆邮电大学教授、博士生导师, 主要研究方向为图神经网络、知识图谱、人工智能安全、图数据挖掘等。



乔少杰 (1981-), 男, 博士, 成都信息工程大学教授、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、空间人工智能等。



聂发志 (2001-), 男, 陕西安康人, 重庆邮电大学硕士生, 主要研究方向为图神经网络、人工智能安全、图数据挖掘等。



牛伟纳 (1991-), 女, 博士, 电子科技大学特聘研究员、硕士生导师, 主要研究方向为网络与软件安全、人工智能安全等。