



大模型驱动的文本文属性图学习研究进展

郑艳萍 季嘉蕊 胡雨韦 魏哲巍

Recent Advances in Text-attributed Graph Learning via Large Models

ZHENG Yan-Ping, JI Jia-Rui, HU Yu-Wei, WEI Zhe-Wei

在线阅读 View online: <https://doi.org/10.16383/j.aas.c250499>

您可能感兴趣的其他文章

大语言模型的工具使用综述

Survey of Tool Use in Large Language Models

自动化学报. 2025, 51(11): 2371–2386 <https://doi.org/10.16383/j.aas.c240793>

基于模体增强对比学习的图神经网络后门防御方法

Motif-augmented Contrastive Learning-based Defense Against Backdoor Attack on Graph Neural Networks

自动化学报. 2026, 52(4): 765–779 <https://doi.org/10.16383/j.aas.c240767>

基于大语言模型的复杂任务自主规划处理框架

Autonomous Planning and Processing Framework for Complex Tasks Based on Large Language Models

自动化学报. 2024, 50(4): 862–872 <https://doi.org/10.16383/j.aas.c240088>

基于大语言模型的多智能体自动渗透测试框架构建与评估

Construction and Evaluation of Multi-agent Automated Penetration Testing Framework Based on Large Language Models

自动化学报. 2026, 52(4): 821–832 <https://doi.org/10.16383/j.aas.c250293>

基于大语言模型的中文实体链接实证研究

An Empirical Study of Chinese Entity Linking Based on Large Language Model

自动化学报. 2025, 51(2): 327–342 <https://doi.org/10.16383/j.aas.c240069>

基于大模型的具身智能系统综述

Embodied Intelligence Systems Based on Large Models: A Survey

自动化学报. 2025, 51(1): 1–19 <https://doi.org/10.16383/j.aas.c240542>

大模型驱动文本属性图学习研究进展

郑艳萍^{1,2,3} 季嘉蕊^{1,2,3} 胡雨韦^{1,2,3} 魏哲巍^{1,2,3}

摘要 近年来,随着大模型在多模态领域的快速发展,文本属性图学习也迎来了新的范式. 文本属性图中包含图结构与丰富的节点/边文本属性信息,广泛存在于社交网络、论文引用网络与知识图谱等场景. 然而,传统图神经网络通常依赖浅层文本嵌入作为初始节点表示,难以充分捕捉上下文依赖与复杂语义信息,并在图结构与文本语义的融合上存在天然局限. 为克服这一瓶颈,近年来研究者开始探索利用大模型强大的语言理解、知识记忆与推理能力来提升文本属性图学习的效果,推动了该领域的新一轮发展. 本文系统梳理大模型驱动文本属性图学习研究进展,围绕图生成、图预测、图推理三大核心任务,回顾代表性方法与主要范式,并总结现有研究中常用的数据集与评测基准. 在文本属性图研究中,当前在图生成、图预测与图推理三大任务中分别面临结构与属性表达效率不足、生成方法匮乏,评估体系与可扩展性不完善,以及复杂关系建模与多跳推理受限等挑战. 结合上述问题提出未来发展方向,期望为后续研究提供系统化参考与整体性视角,进一步促进大模型与图学习的深度耦合.

关键词 大模型; 文本属性图; 图神经网络; 大语言模型

引用格式 郑艳萍, 季嘉蕊, 胡雨韦, 魏哲巍. 大模型驱动文本属性图学习研究进展. 自动化学报, 2026, 52(8): 1557–1592

DOI 10.16383/j.aas.c250499 **CSTR** 32138.14.j.aas.c250499

Recent Advances in Text-attributed Graph Learning via Large Models

ZHENG Yan-Ping^{1,2,3} JI Jia-Rui^{1,2,3} HU Yu-Wei^{1,2,3} WEI Zhe-Wei^{1,2,3}

Abstract In recent years, the rapid advancement of large models in multimodal learning has introduced new paradigms for text-attributed graph learning. Text-attributed graphs, which combine graph structures with rich textual attribute information on nodes and edges, are widely found in scenarios such as social networks, paper citation networks, and knowledge graphs. However, traditional graph neural networks typically rely on shallow text embeddings as initial node representations, making it difficult to capture contextual dependencies and complex semantic information, and they face inherent limitations in integrating graph structures with textual semantics. To address this bottleneck, recent studies leverage the powerful language understanding, knowledge retention, and reasoning capabilities of large models to enhance text-attributed graph learning, promoting further development in this field. This paper provides a systematic review of recent advances in large model-driven text-attributed graph learning, covering representative methods and major paradigms across three core tasks: Graph generation, graph prediction, and graph reasoning. We also summarize commonly used datasets and evaluation benchmarks. In text-attributed graph research, current work in the three core tasks faces persistent challenges: Inefficient structure-attribute representation and limited graph generation methods, inadequate evaluation frameworks and scalability in graph prediction, and constrained capacities for complex relational modeling and multi-hop reasoning in graph reasoning. Building on these challenges, we outline future research directions to provide a structured reference and holistic perspective, fostering deeper integration of large models and graph learning.

Keywords large models; text-attributed graphs; graph neural networks; large language models

Citation Zheng Yan-Ping, Ji Jia-Rui, Hu Yu-Wei, Wei Zhe-Wei. Recent advances in text-attributed graph learning via large models. *Acta Automatica Sinica*, 2026, 52(8): 1557–1592

收稿日期 2025-09-28 录用日期 2026-01-12

Manuscript received September 28, 2025; accepted January 12, 2026

国家自然科学基金(92470128, U2241212), 中国人民大学校级智能计算云平台, 中国人民大学“中央高校建设世界一流大学(学科)和特色发展引导专项资金”资助

Supported by National Natural Science Foundation of China (92470128, U2241212), Public Computing Cloud, Renmin University of China, and Fund for Building World-class Universities (Disciplines) of Renmin University of China

本文责任编辑 金连文

Recommended by Associate Editor JIN Lian-Wen

1. 中国人民大学高瓴人工智能学院 北京 100872 2. 大模型与

图 (graph) 作为计算机科学中一类重要的数据结构,具备强大的关系建模与结构表达能力,广泛应用于社交网络^[1]、论文引用网络^[2]、知识图谱^[3]、生

智慧治理北京市重点实验室 北京 100872 3. “新一代智能搜索与推荐”教育部工程研究中心 北京 100872

1. Gaoling School of Artificial Intelligence, Renmin University of China, Beijing 100872 2. Beijing Key Laboratory of Research on Large Models and Intelligent Governance, Beijing 100872 3. Editorial Engineering Research Center of Next-generation Intelligent Search and Recommendation, MOE, Beijing 100872

物分子网络^[4]等领域的结构建模与性质预测任务. 现有图学习研究大多依赖图神经网络 (graph neural networks, GNNs) 及其变体^[5-7], 通过在固定图结构上聚合邻居特征实现节点或图级表示学习, 取得大量成功. 然而, GNN 系列方法在捕捉远程依赖、应对稀疏结构以及适配跨域迁移方面仍存在瓶颈. 另一方面, 现实世界中的图数据往往不仅包含结构信息, 还伴随着丰富的非结构化文本描述, 例如学术引文网络中节点对应的论文摘要, 社交网络中节点对应的用户简介, 电商知识图谱中节点对应的商品文本描述等. 这类既包含结构关系又包含文本属性的图数据, 被称为文本属性图 (text-attributed graph, TAG). 与仅包含结构信息的纯图相比, 文本属性图中的节点或边往往附带丰富的文本描述, 这些文本蕴含了关于实体语义、上下文关系及潜在知识的关键信息. 如果忽视这些文本属性, 仅依赖图结构进行学习, 模型将难以捕捉节点间的深层语义关联, 导致表示能力受限^[8]. 因此, 如何有效地融合图结构信息与文本语义信息, 充分挖掘文本所蕴含的知识, 已成为提升图学习模型性能的重要方向.

传统图神经网络在处理文本属性图时, 通常将节点或边的文本信息通过非上下文化的浅层嵌入方法 (如 Bag-of-Words、Word2Vec 等) 编码为初始节点表示, 再交由 GNN 进行后续建模, 这类做法也在常见的基准数据集^[9-10]中被广泛采用. 然而, 这种基于浅层嵌入的做法在通用知识建模与深层语义理解方面存在显著局限: 一方面, 它们无法有效捕捉词语的多义性和上下文依赖, 导致语义信息匮乏^[11-12]; 另一方面, 图结构与文本语义在模态、表示空间和噪声分布等方面存在天然差异, 使得信息融合过程面临挑战. 这些不足使得基于浅层嵌入的 GNN 模型在面对长文本、复杂语义以及需要知识迁移的任务时表现往往不尽如人意.

在近年来的大模型浪潮中, 文本、图像、语音等模态的基础模型架构与训练范式已逐渐成熟, 推动了通用人工智能的快速发展. 这类大模型 (如 LLaMA^[13]、PaLM^[14]、CLIP^[15] 等) 基于海量数据和参数训练, 展现出强大的表征学习、知识记忆、推理和生成能力, 能够捕捉复杂的语义关系, 并在零样本或少样本设置下实现跨任务迁移. 随着这一趋势的发展, 研究者开始探索将大模型引入文本属性图学习中, 以克服传统方法在语义建模和知识推理上的瓶颈. 大模型与图学习的结合为文本属性图学习带来新的范式: 一方面, 大模型可作为强大的文本表示器, 提升节点文本属性的语义质量; 另一方面, 大模型具备推理与知识迁移能力, 可辅助进行图结构的生成

与推断, 从而突破传统基于局部邻域聚合的限制.

现有综述 [8, 16-18] 已对文本属性图学习进行了较为系统的总结, 但大多集中于图预测任务 (如节点分类、链接预测等), 而对图生成与图推理两类任务的涉及较少. 同时, 这些综述主要基于传统图神经网络范式展开, 对近年来快速发展的大语言模型驱动的新兴方法缺乏全面梳理. 为此, 本文从大模型驱动的视角出发, 针对图生成、图预测、图推理三大核心任务进行系统回顾. 在图生成方面, 现有方法主要面临图结构与属性表达效率不足以及文本属性图生成方法匮乏等挑战; 对于图预测任务, 尽管已有显著进展, 但在评估体系和可扩展性方面仍存在不足; 而在图推理领域, 现有方法受限于复杂关系建模和多跳推理的能力. 如图 1 所示, 本文通过详细分析上述三个方向的现状与挑战, 旨在补充现有综述中的不足, 并为后续研究提供参考.

本文的整体结构如下: 第 1 节介绍基本概念与研究现状; 第 2 节介绍大模型赋能的图生成方法, 涵盖图生成模型的分类与典型应用; 第 3 节回顾大模型在图预测任务中的研究进展, 围绕其在任务中扮演的预测器 (predictor)、增强器 (enhancer) 与对齐器 (aligner) 三类角色展开讨论, 分析不同角色下的大模型如何参与图预测过程并发挥作用; 第 4 节梳理大模型在图推理任务中的研究进展, 涵盖经典图推理与知识图谱推理两类任务, 重点探讨大模型在复杂关系建模与多跳推理等场景中的能力与应用; 第 5 节汇总当前在图生成、图预测与图推理三类任务中常用的数据集与评测基准; 第 6 节对现有方法存在的问题进行总结, 并结合当前研究现状讨论未来发展方向, 旨在为构建可扩展、可信赖、可解释且任务通用的图智能系统提供方向指引.

1 基本概念

1.1 文本属性图的定义与特征

在计算机科学领域, 图 $G = (\mathcal{V}, \mathcal{E})$ 是一种非线性的抽象数据结构, 由节点集 $\mathcal{V}$ 和边集 $\mathcal{E}$ 共同构成. 节点通常用于表示实体, 而边则描述这些实体之间的关系. 每一条边 $e \in \mathcal{E}$ 都和一对节点 (u, v) 相关联, 其中 $u, v \in \mathcal{V}$ 是该边的两个端点. 边可分为有向边和无向边: 有向边表示从 u 指向 v , 而无向边则不区分起点与终点, 可视为同时存在从 u 到 v 与从 v 到 u 的连接. 据此, 图可对应地划分为有向图与无向图. 文本属性图是一类在传统图结构基础上, 进一步将节点和/或边关联文本信息作为属性的图结构, 通常表示为 $G = (\mathcal{V}, \mathcal{E}, \{s_i\}, \{t_{ij}\})$. 其中, 对于

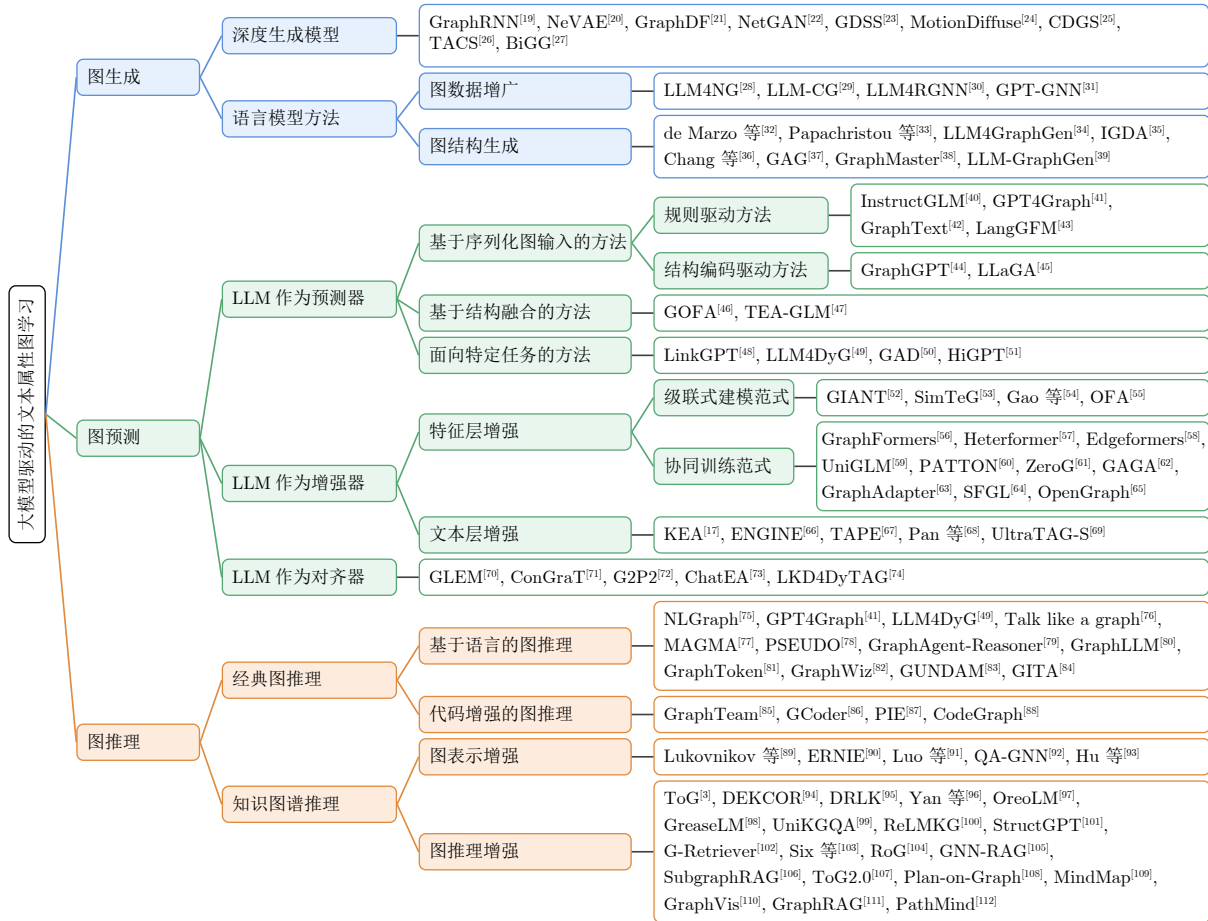


图1 大模型驱动的文本文属性图学习研究框架

Fig.1 Research framework of large-model-driven text-attributed graph learning

每个节点 $v_i \in \mathcal{V}$, 对应一个文本属性 s_i ; 对于每条边 $(v_i, v_j) = e_{ij} \in \mathcal{E}$, 则关联一个边文本属性 t_{ij} . 与仅建模结构关系的图相比, 文本文属性图同时建模节点与边的语义内容, 具备更强的表示能力. 例如, 在社交网络中, 用户 (节点) 可附带状态更新、帖子或评论 (节点文本), 而好友关系 (边) 则可能关联具体的交流消息 (边文本). 在实际应用中, 文本文属性图广泛应用于社交网络分析、知识图谱构建与自然语言处理等领域, 通过学习图中节点和边的文本属性, 能够挖掘出更丰富的语义信息与潜在结构信息.

1.2 基础模型与大模型范式

近年来, 大语言模型 (large language model, LLM)^[13, 113-114] 在自然语言处理领域取得巨大进展, 彻底改变了语言模型的发展方向, 推动从文本生成、机器翻译到语义理解等多个方面的研究和应用. 这些模型通过在大规模文本数据集上进行预训练, 学习到丰富的语义表示, 从而能够进行复杂的语言处理 (如复杂的零样本推断). 尺度定律 (scaling

laws)^[115] 揭示了模型损失与模型参数量、训练数据量、计算量之间的幂律关系, 表明在给定计算预算下, 模型性能可以通过扩大参数规模或增加训练数据量实现可预测的提升. 涌现能力 (emergent abilities) 是大模型研究中的重要现象, Wei 等^[116] 将其定义为“在小规模模型中不存在或仅表现为随机水平, 但在模型规模跨越某一临界阈值后突然显著提升的能力”. 这类能力有突变性, 即性能从近乎为零到显著提升是突然发生而非渐进的; 此外还具有不可预测性, 即能力涌现的规模阈值难以推测. 常见的涌现能力包括多步骤推理、复杂算术运算、指令遵循以及上下文学习等. 此外, 提示工程 (prompt engineering) 在挖掘和提升大语言模型推理能力方面展现出巨大潜力, 在经典输入输出提示的基础上, Kojima 等^[117] 提出思维链 (chain of thought, CoT) 提示, 其能够引导大语言模型将复杂问题拆解成若干步骤, 提升大语言模型的复杂推理能力; Wang 等^[118] 在 CoT 的基础上提出自洽思维链 (self-consistency with CoT, CoT-SC), 从多条思维链中选择生成数

目最多的回答;为使大语言模型在推理方面更贴近人类的思考方式,树型思维链 (tree of thoughts, ToT)^[119-121] 和图型思维链 (graph of thoughts, GoT)^[122] 被提出,大语言模型可以对生成的想法进行分支延伸,同时也能比较和聚合多个思维路径以及回溯修正想法.虽然大语言模型主要处理文本序列,但由于语言的通用性,其可以作为各个领域知识的载体从而连接各个领域,越来越多的研究致力于探索大语言模型处理多模态信息的能力^[123].Liu等^[124]提出用一个可学习的映射器将图像编码投影到预训练语言模型的词向量空间中,让大语言模型理解图像内容,并通过图像指令微调增强大语言模型对于图像的理解;Huang等^[125]提出 AudioGPT,使得大语言模型能够理解和生成包括语音、音乐、自然声音及多轮对话在内的多种音频;OpenAI 新推出的视频大模型 Sora^[126] 使用 Diffusion Transformer,并用类似大语言模型的训练方式在视频数据上进行训练,能够生成任意尺寸且符合物理规律的视频,引发了人们的极大关注和广泛讨论.综上,大语言模型在多种数据类型的处理上展现出巨大的潜力,极大地推进了 AI 社区实现通用人工智能的边界.然而大语言模型对于图这种非欧氏数据的处理还存在很大缺陷,亟须人们去改良和探索,为最终实现通用人工智能作出贡献.

- 判别式模型. 判别式模型的核心思想是直接建模标签 y 在给定输入样本 z 条件下的条件分布 $p(y|z)$,从而学习输入与输出之间的映射关系.其广泛应用于图分类、文本分类、语音识别、情感分析、人脸识别等任务.其优势在于在有充分标注数据的情况下,判别式模型通常能达到较高的预测准确率;同时学习输入与输出的条件分布函数较为容易,计算资源需求较低.

- 生成式模型. 生成式模型旨在学习数据的联合分布 $p(z, y)$ 或输入样本的边缘分布 $p(z)$,从而刻画训练数据的潜在生成机制,并生成与训练数据分布相似的新样本.该类模型广泛应用于文本生成、图像生成、音频合成、视频生成等任务场景.其优势在于具备生成新内容、建模复杂数据分布以及支持多样化下游任务的能力;但同时也面临计算资源需求高、生成内容可能不符合事实或预期、生成过程可控性较差等挑战.

1.3 文本属性图学习的三大核心任务

根据研究目标与任务层次的不同,文本属性图学习可划分为图生成、图预测与图推理三类核心任务.图生成关注如何从无到有构建或重构具有合理结构与属性分布的图,是后续任务的基础;在此基

础上,图预测以已知的部分结构或属性为条件,对缺失节点、边或标签进行估计与补全,进一步提升图的完整性与可用性;而图推理则面向更高层次的语义目标,利用已构建和补全的图结构进行逻辑推断、关系演绎与知识发现.三类任务在数据表示、模型设计与能力需求上逐级递进、相互支撑,共同构成文本属性图学习的重要研究框架.

- 图生成. 图生成任务的核心在于给定一组观测图 $\mathcal{G} = \{G_i\}_{i=1}^M$,其中 M 表示观测数量,目标是让模型学习这些图背后的真实分布 $p(\mathcal{G})$,使得能够从该分布中采样生成新的合理的图 $G_{\text{model}} \sim p(\mathcal{G})$,同时保持原始图的拓扑结构特征(如节点连接性、社区结构、子图模式)和属性相关性(如节点属性与边属性之间的依赖关系).其中每个图 $G = (\mathcal{V}, \mathcal{E}, \{s_i\}, \{t_{ij}\})$ 包含节点集 $\mathcal{V}$ (大小为 $|\mathcal{V}|$)、边集 $\mathcal{E}$ (大小为 $|\mathcal{E}|$) 以及每个节点 $v_i \in \mathcal{V}$ 对应的文本属性 s_i ,这些文本属性经过处理后形成节点特征矩阵 $X \in \mathbf{R}^{|\mathcal{V}| \times a}$ (每行对应一个节点的 a 维特征),边属性集合 $\{t_{ij}\}$ 转化为边特征张量 $E \in \mathbf{R}^{|\mathcal{E}| \times b}$, $|\mathcal{E}|$ 表示边集数量,边特征张量维度为 b 维.此外,图的连接结构通过邻接矩阵 $A \in \mathbf{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ 来表示,其中 $A_{ij} = 1$ 表示节点 v_i 与节点 v_j 之间存在边, $A_{ij} = 0$ 表示两者之间没有直接连接.在建模过程中,图的表示主要分为两类:序列化表示 (sequential representation) 将图的生成过程分解为一系列逐步添加节点或边的操作序列,适合使用循环神经网络 (recurrent neural networks, RNN) 或变换器 (Transformer) 等序列模型进行生成;而矩阵表示 (matrix representation) 则直接对邻接矩阵和属性矩阵进行整体建模,常用于基于变分自编码器 (variational auto-encoder, VAE)、生成对抗网络 (generative adversarial network, GAN) 或扩散模型 (diffusion) 的图生成方法.两类表示方式各有优劣,核心均在于如何高效且准确地捕捉图的结构规律与属性关联,从而生成在统计特性上与真实图一致且具有多样性的新图.其中矩阵表示更适合传统深度生成模型,如图自编码器、生成对抗网络和扩散模型,这类方法直接在图的结构矩阵上建模,能更精确地保留全局拓扑信息,适用于对结构精度要求高的场景(如分子生成、蛋白质设计);而序列化表示更适用于基于 LLM 的图生成方法,因为 LLM 天然擅长处理和生成文本序列,通过将图结构编码为线性序列(如节点遍历顺序、边添加序列或图到文本的提示模板),可以利用预训练语言模型的强大生成能力进行自回归生成,尤其适合开放域、语义丰富的图生成任务(如知识图谱补全、社交网络模拟).

- 图预测. 在图表示学习中,图预测是一类核

心任务, 致力于基于已有图结构与属性信息, 预测图中尚未观测的语义标签或结构关系. 图预测任务根据其预测对象的不同, 可分为节点级、边级与图级三类, 分别对应对节点标签、边关系以及整个图属性的预测.

定义 1 (节点分类). 给定一个部分节点带有标签的文本文属性图 G , 其中已标注节点集合为 $\mathcal{L} \subseteq \mathcal{V}$, 每个节点 $v_i \in \mathcal{L}$ 关联一个标签 $y_i \in \mathcal{Y}$, 标签空间 $\mathcal{Y}$ 为节点的类别集合. 节点分类任务是学习一个映射函数 $f_\theta: \mathcal{V} \rightarrow \mathcal{Y}$, 使得对于任意未标注节点 $v_i \in \mathcal{V} \setminus \mathcal{L}$, 能够根据其文本文属性 s_i 以及邻接关系 A 等信息, 预测其标签 $\hat{y}_i = f_\theta(v_i)$.

定义 2 (链接预测). 链接预测任务旨在判断图中任意两个节点之间是否存在潜在的连接关系. 给定一个文本文属性图 $G = (\mathcal{V}, \mathcal{E}, \{s_i\}, \{t_{ij}\})$, 其中每个节点 $v_i \in \mathcal{V}$ 对应一个文本文属性 s_i , 每条边 $(v_i, v_j) \in \mathcal{E}$ 可能关联一个边文本文属性 t_{ij} , 目标是预测节点对 (v_i, v_j) 是否存在边连接. 具体地, 对于任意两个节点 $v_i, v_j \in \mathcal{V}$, 定义二分类标签:

$$y_{ij} = \begin{cases} 1, & (v_i, v_j) \in \mathcal{E} \\ 0, & \text{其他} \end{cases}$$

链接预测任务学习一个评分函数:

$$g_\theta: \mathcal{V} \times \mathcal{V} \rightarrow [0, 1], \quad \hat{y}_{ij} = g_\theta(v_i, v_j)$$

其中, $\hat{y}_{ij}$ 表示节点对 (v_i, v_j) 存在边连接的概率. 在实际建模中, 通常将图中已有的边作为正样本, 从图中未连接的节点对中采样负样本.

定义 3 (图分类). 图分类任务旨在对多个图样本进行语义类别判定. 给定一个包含多个文本文属性图的集合:

$$\mathcal{G} = \{G_1, G_2, \dots, G_M\}$$

其中, 每个图 $G_m = (\mathcal{V}_m, \mathcal{E}_m, \{s_n\}_{n \in \mathcal{V}_m})$, 并且部分图关联有类别标签 $y_m \in \mathcal{Y}$, 模型的目标是预测未标注图的类别标签. 具体地, 图分类任务学习一个映射函数:

$$h_\theta: \mathcal{G} \rightarrow \mathcal{Y}$$

使得对于任意图 $G_m \in \mathcal{G}$, 模型能够基于其节点特征 $\{X_n\}_{n \in \mathcal{V}_m}$ 、图结构 A_m 、文本文属性 $\{s_n\}_{n \in \mathcal{V}_m}$ 等信息, 输出预测标签:

$$\hat{y}_m = h_\theta(G_m)$$

图分类任务的关键在于构建图的全局表示 (graph-level representation), 通常通过图级池化操作 (graph pooling) 对节点表示进行聚合, 以形成图的向量表示, 进而完成分类.

随着 LLM 在理解与生成任务中的广泛应用, 越来越多的研究开始探索将 LLM 应用于图预测任务. 利用其强大的文本理解与关系建模能力, LLM 可在多个层面辅助图预测建模: 将节点或边的属性信息转化为自然语言提示, 以增强节点表示的语义表达^[42, 52]; 借助语言模型的上下文聚合能力, 对邻域结构与全图语义进行联合建模^[46, 59–60]. 此外, LLM 在零样本与小样本学习方面的优势, 也使其在泛化性要求较高的图预测任务中具有应用潜力.

● **图推理.** 图推理任务的核心在于, 给定一个文本文属性图, 模型需要利用图结构、属性及它们之间的复杂关联, 通过逻辑推断和关系演算, 推导出图中隐含的、未直接陈述的信息或结论. 图推理旨在发掘深层的语义关联和逻辑规律, 其挑战性在于如何处理关系路径的长程依赖、应对观测数据的不完备性以及保证推理过程的合理性与可解释性.

图推理任务一般可分为经典图推理和知识图谱推理两类. 1) 经典图推理适用于一般的图结构数据, 主要考察模型对图结构的理解能力, 例如推断两个未直接相连的节点是否在语义或功能上相似, 查找图中一点到另一点的有效路径, 或在图中寻找符合特定模式的子结构. 2) 知识图谱 (knowledge graph, KG) 是一种特殊的文本文属性图, 其节点表示实体 (entity), 边表示实体间的关系 (relation), 节点和边的文本文属性通常直接建模为实体名和关系名. 知识图谱推理 (又称知识图谱补全) 的核心任务是预测图谱中缺失的事实或关系, 其关键在于预测缺失的三元组 (头实体、关系、尾实体), 例如根据已知的头实体和关系推断出尾实体, 或学习并运用“出生地所在城市 $\rightarrow$ 国籍”之类的逻辑规则来推导新知识. 此外, 它还包含推断实体的语义类型以及回答基于图谱的复杂自然语言问题, 这要求模型具备多跳推理能力, 能联系并在多个事实三元组上推理.

随着 LLM 的发展, 图推理任务也获得了新的赋能. LLM 强大的语义理解和上下文学习能力使其能够将结构化的图查询转化为自然语言提示, 极大地增强了对节点、关系文本描述的语义感知. 通过思维链等提示技术, LLM 可以模拟出一步步的逻辑推理过程, 有效解决多跳推理的挑战, 并在零样本/少样本场景下展现出优异的泛化性能, 为神经符号推理 (neuro-symbolic reasoning) 提供了新的实现路径.

2 图生成

2.1 图生成模型分类

早期网络建模框架通常假设现实世界的复杂网

络遵循某些预定义的结构规则(如特定的度分布、聚类模式等),并通过概率抽样方法生成具有目标统计特性的网络.后来,随着深度学习的发展,在文本属性图领域的生成模型主要分为两类:1) 基于深度学习的方法,采用神经架构(例如图神经网络和变分自编码器)以数据驱动的方式学习和生成复杂的图分布^[127],其中文本通常作为采样器的条件进行条件生成;2) 基于语言模型的方法,该类方法通常面向包含丰富文本属性的图数据,利用节点或边上的文本信息进行图数据增强或图数据生成.

2.1.1 基于深度生成模型的图生成

基于深度学习的图生成模型,作为图生成模型的主流方式,主要存在三个组成部分:编码器、解码器以及采样器.一般流程如图2所示.

● 编码器.在编码器部分,基于深度学习的生成模型定义生成数据的概率建模方式,大致可分为两种范式:显式和隐式概率模型^[128].显式模型旨在直接参数化图数据的概率分布,从而实现似然评估和样本生成.这类模型涵盖多种方法,包括自回归模型、变分自编码器和基于流的生成模型.相比之下,隐式模型不直接指定数据分布.相反,它们通过神经网络生成器学习从简单的先验分布(例如高斯噪声)到目标图分布的映射,从而以隐式方式有效地近似数据分布.此类经典方法包括生成对抗网络(generative adversarial network, GAN)、扩散模型.下文按此分类梳理不同深度生成模型中的代表性图生成方法,所选模型(如 GraphRNN^[19]、MoFlow^[129]、DiGress^[130]等)均在其对应范式或应用场景中具备典型性与影响力.

1) 自回归生成模型.自回归模型通常用于序列图表示^[19, 131],其中输入序列由节点、边或模体组成,模型根据先前生成的序列预测下一个节点或边.具体而言,该模型将生成过程分解为多个序列步骤,每个步骤根据当前子图确定下一步操作.这种方法通常通过对节点、边或模体进行序列采样,灵活地生成具有不同结构和大小的图. GraphRNN^[19] 和

NetGAN^[22] 等代表性工作采用节点列表顺序表示法,即先逐个添加新节点,再将这些节点通过边连接到先前生成的组件上.相比之下,像 BiGG^[27] 这样的模型则采用边列表顺序表示法,并采用树状自回归方法生成与每个节点相关的边.

2) 变分自编码器.作为开创性的深度生成模型,变分自编码器已在图生成任务中展现出显著的成功.通过学习图域与潜在空间之间的概率映射,图编码器将输入图映射到低维连续潜在表示,而图解码器则根据采样的潜在变量重建图^[132].这类工作使用图神经网络(例如 GCN^[5] 或 GAT^[6]) 来参数化编码器和解码器,以捕获关系归纳偏差.在该框架基础上,一系列重要变体被提出以提升生成质量与领域适应性: NeVAE^[20] 在解码过程中引入掩蔽机制,以满足生成分子的化学有效性约束; Graphite^[133] 将谱图卷积集成到编码器和解码器中,利用了节点表示中的置换不变性和局部性; MiCaM^[134] 使用基于频率的迭代子图合并算法来识别频繁出现的分子片段(模体)及其相互连接.

3) 流模型.正则化流通过变量变换定理^[135-136] 在潜在变量与图结构之间建立可逆且确定性的映射,从而采用一系列可逆函数将简单的先验分布(例如标准高斯分布)转换为复杂的图数据分布,从而生成新样本. MoFlow^[129] 是一项早期研究,采用两步生成流程:首先使用基于 Glow 的流生成分子键(边),然后通过基于键结构的图条件流预测原子类型(节点);随后的后处理校正通过解决无效价态等结构性问题来确保化学有效性.在此基础上, GraphDF^[21] 利用基于可逆模移位变换的离散正则化流将潜在变量映射到图结构,从而实现精确可逆性,并缓解了连续松弛方法固有的局限性.

4) 生成对抗网络.典型的 GAN 框架通常包含生成器和鉴别器两个模块,生成器学习将随机噪声映射为合成图,鉴别器则负责区分真实图与生成图.生成器和鉴别器采用对抗式最小最大目标函数进行训练.为应对图结构数据的独特挑战,人们提出各种基于 GAN 的模型.例如, GraphGAN^[137] 通过对抗学习来建模顶点连通性,而 NetGAN^[22] 采用基于循环神经网络的生成器和鉴别器来建模有偏随机游走的分布,从而生成图结构,并通过 Wasserstein GAN 目标函数对模型进行优化. CONDGEN^[138] 通过将变分方法与对抗训练相结合,进一步推进了这一方向,实现了基于语义上下文的条件图生成.尽管取得了这些进展,基于 GAN 的图模型仍然面临着一些限制,例如训练不稳定性、收敛困难和模式崩溃.

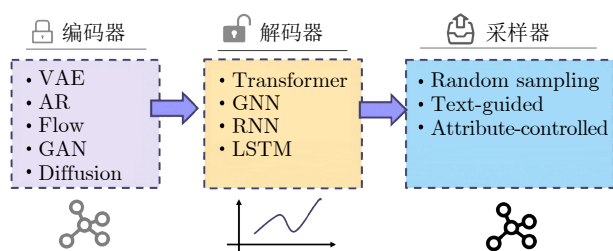


图2 图生成模型的流程图

Fig.2 Flowchart of graph generation models

5) 扩散生成模型. 该模型通过定义一个前向过程来生成数据, 该过程在 T 步内逐渐将高斯噪声添加到输入数据中, 产生一系列噪声样本, 最终状态趋近于标准高斯分布. 为从数据分布中采样, 可以通过学习近似逆向转换来反转该过程, 这通常是通过训练神经网络以匹配条件后验来实现的. 目前已出现三种主要的扩散模型范式, 每种范式都通过不同的方法论来处理生成模型. 以 EDP-GNN^[139] 为代表的朗之万动力学分数匹配模型 (score-matching Langevin dynamics, SMLD) 训练一个分数函数来估计对数数据密度的梯度, 并采用朗之万动力学进行采样, 从而优先考虑高维空间中的效率. 以 Di-Gress^[130] 为代表的去噪扩散概率模型 (denoising diffusion probabilistic model, DDPM) 借鉴非平衡热力学思想, 将图生成过程建模为逐步去噪的逆过程, 从而通过迭代去噪生成目标图数据. 以 Graph-GDP^[140]、GDSS^[23] 以及 CDGS^[25] 为代表的基于分数的生成模型 (score-based generative models, SGMs), 通过随机微分方程将离散时间扩散推广到连续时间公式, 为复杂的时间和空间依赖关系建模提供了更高的灵活性. 其中, CDGS^[25] 和 TACS^[26] 在生成过程中引入属性引导机制, 将样本朝目标分子性质方向优化的同时, 保持结构有效性与生成过程的稳定性.

● 解码器. 在编码器的概率建模基础上, 解码器部分采用不同的神经网络架构进行概率建模的参数化. 概率建模和神经网络架构构成了图生成模型的互补支柱. 概率框架定义了底层生成过程, 而神经网络架构则是学习表征和捕获图数据中复杂依赖关系的骨干. 其中, 主要包括四类神经网络架构:

1) 前馈神经网络. 也称为多层感知器 (multilayer perceptron, MLP), 是神经网络的最基本形式. 它们对于可以表示为固定大小向量的输入数据 (例如节点或边的属性) 特别有效. 像 NeVAE^[20] 和 GraphVAE^[132] 这样的模型使用前馈神经网络来学习图的潜在表征, 然后对其进行采样以生成新的图.

2) 图神经网络. 作为专为处理图结构数据而设计的模型, 利用消息传递机制聚合来自相邻节点的信息, 能够对节点和边之间的复杂关系进行建模. GRAN^[131]、Graphite^[133] 和 MiCaM^[134] 等早期研究将 GNN 集成为自回归、变分自编码器或流生成模型中的解码器. 对于基于生成对抗模型的图生成, GCPN^[141] 利用 GNN 来建模生成器中的子图拓扑和支架交互, 并使用判别器验证动作 (例如节点连接、边类型、终止信号) 以确保结构可行性.

3) 循环神经网络. 该模型能够通过隐藏状态动

态处理顺序数据, 因此特别适用于具有顺序表示的图生成任务. GraphRNN^[19] 通过双层架构体现了这一范式: 图级循环神经网络维护不断增长的图的状态并生成新节点, 而边级 RNN 则为新生成的节点生成边. 在此基础上, MolecularRNN^[142] 扩展了 GraphRNN, 用于生成具有节点和边类型的图. Tigger^[143] 通过将时间交互图分解为时间随机游走, 采用自回归范式进行动态图生成.

4) 长短期记忆网络 (long short-term memory, LSTM). 是 RNN 的一种特殊变体, 它通过解决序列建模中固有的长期依赖性挑战, 进一步推进了图生成. 与标准 RNN 类似, LSTM 将图处理为逐步序列, 但由于其记忆单元设计, 在大规模图的可扩展性方面表现出色. 对于边序列表示, GEEL^[144] 使用 LSTM 来建模图中边的生成, 通过间隙编码和带宽限制降低了词汇复杂度. 它通过基于语法的自适应技术进一步扩展到属性图, 确保生成与边计数一致的可扩展且结构化的图. 对于节点顺序表示, OLR^[145] 使用 LSTM 来学习节点依赖关系, 并引入正则化项, 强制隐藏状态对节点顺序排列的不变性. 此外, LSTM 已应用于动态图生成任务, 例如, TG-GAN^[146] 使用 LSTM 来帮助 GAN 模型生成连续型动态图, 方法是对截断时间随机游动及其组合的深度生成过程进行建模, 它通过联合捕获边序列、时间序列来建模截断时间随机游动.

● 采样器. 在图生成中, 存在两种主要的采样策略: 随机采样和条件采样. 随机采样涉及从学习到的分布中随机生成图, 而无需明确控制其结构或属性. 相反, 条件采样旨在对潜在代码进行采样, 最终尝试生成具有所需属性的新图, 这些约束可以包含语义目标、图级属性或节点/边级属性.

1) 语义条件涉及引导生成朝着所需含义发展的自然语言提示. 早期研究侧重于从文本中提取图结构化表示^[147-148] 和事件时间图^[149]. 最近, 预训练语言模型已被集成到条件生成框架中, GraphGPT-1^[150] 和 LGGM^[151] 对文本提示进行编码, 并将其与图嵌入对齐, 以在生成过程中强制保持语义一致性.

2) 属性条件通过纳入分类属性来扩展结构约束. 深度学习模型通常通过损失函数将这些属性嵌入训练目标中. 例如, GGDiff^[152] 将条件扩散视为随机控制问题, 动态调整采样过程以生成目标属性的图.

3) 边/节点属性对离散或连续的节点/边特征 (例如标签、嵌入) 施加约束. GraphMaker^[153] 通过将属性与结构生成解耦, 将其推广到大型属性图. 然而, 大多数框架难以处理高维属性 (例如文本嵌

入), 通常会截断维度以适应模型限制. 由于现实世界的图数据通常具有高维属性^[154], 例如节点/边的文本特征, 因此生成此类图仍然是一个悬而未决的问题.

2.1.2 基于语言模型的图生成

近年来, 基于 Transformer 架构的大语言模型在自然语言理解与生成方面取得突破, 为增强图数据的文本属性提供了强大工具. 早期研究主要聚焦于数据增广, 通过生成节点或边级别的文本信息来丰富图的语义表达, 从而增强下游任务性能. 随着 LLM 能力的逐步提升, 研究逐渐从简单的文本补全转向语义-结构协同建模.

● 图数据增广. 早期工作集中于利用大语言模型增强图中节点或边的文本特征. 例如, LLM4NG^[28]提出一种轻量级的即插即用的范式, 通过 LLM 生成语义一致的节点文本以增强文本属性图. 如图 3 所示, S_i 表示句子包中的第 i 个句子; U_i 表示对应的句子编码表示; β_i 表示句子级权重; c_i 表示由约束图编码得到的约束表示; n 为句子包中的句子数. 该方法首先利用 LLM 从类别标签中提取语义信息, 并生成符合该类语义的合成文本样本; 随后, 通过一个边缘预测器建模原始图中的结构依赖关系, 将生成的节点以结构一致的方式集成到图中. 这一过程无需修改原始图结构, 即可无缝引入带标签的合成节点与边, 在少样本节点分类任务中显著提升性能. 在此基础上, SAVE-TAG^[155] 针对长尾分布下的文本属性图提出一种语义邻近风险最小化框架. 该方法利用 LLM 在文本空间中进行语义插值, 生成语义上合理且多样化的节点/边文本内容. 更重要的是, SAVE-TAG 设计了一种基于图拓扑的边分配机制, 将生成的文本节点通过拓扑邻近性原则连接至现有图结构中, 从而确保新生成样本在语义与

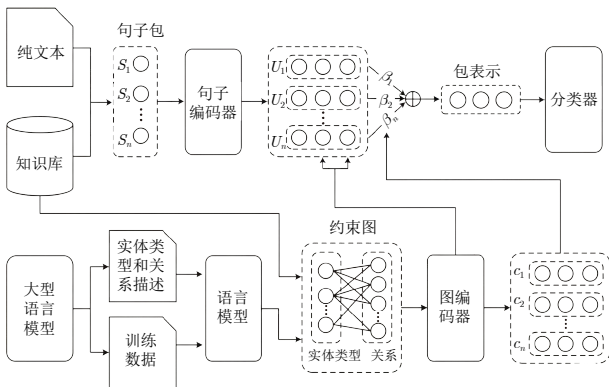


图 3 LLM-CG 的关系抽取任务框架

Fig. 3 The LLM-CG framework for relation extraction task

结构上的双重一致性, 有效缓解了长尾类别中的数据稀疏问题. 从文本属性图的数据增广角度, 主要存在三个应用:

1) 关系抽取任务. 通过将非结构化文本与知识库对齐自动标注训练数据, 但常引入噪声标签. 为此, LLM-CG^[29] 提出一种基于 LLM 的约束图增强框架, 用于提升远程监督关系抽取的鲁棒性与语义准确性. 该方法将实体类型作为中介, 构建一个连接多种关系的语义约束图, 显式建模关系间的逻辑依赖. LLM 被用于生成约束图中各节点 (即关系或实体类型组合) 的初始语义表示, 随后由图神经网络编码器进一步融合语义与结构信息, 实现对噪声标签的过滤与关系语义的精细化建模. 该框架有效利用 LLM 的外部知识, 增强了模型对复杂语义关系的理解能力.

2) 生成鲁棒的图拓扑结构. 面对图结构可能遭受的对抗攻击或噪声干扰, 如何生成或恢复鲁棒的边结构成为关键挑战. LLM4RGNN^[30] 提出一种基于 LLM 的图结构推理框架, 旨在通过大模型的认知能力识别并修复异常结构. 该框架融合了一个局部 LLM 模块, 用于识别潜在的恶意或虚假边 (如对抗性攻击引入的边), 以及一个基于语言模型的边预测器, 用于推断可能缺失的重要连接. 通过联合执行“去噪”与“补全”, LLM4RGNN 能够重建更具鲁棒性和语义一致性的图结构, 从而提升下游任务 (如节点分类、链接预测) 在对抗环境下的稳定性.

3) 生成式预训练. GPT-GNN^[31] 提出一种自监督的属性图生成预训练框架, 旨在通过生成式任务提升 GNN 对图结构与节点属性联合分布的建模能力. 如图 4 所示, θ 为 GPT-GNN 的模型参数. 该方法将图的生成概率分解为两个子任务: a) 节点属性生成; b) 图边生成. 通过对这两个生成过程进行联合建模, GPT-GNN 能够捕捉节点属性与图结构之间的内在依赖关系. 经过预训练后, 模型仅需少量

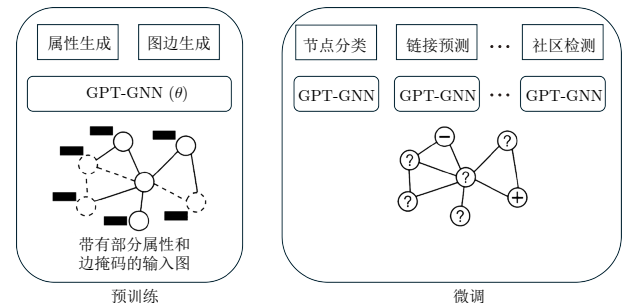


图 4 GPT-GNN 的生成式预训练框架

Fig. 4 The generative pre-training framework in GPT-GNN

标注数据进行微调, 即可在节点分类、链接预测等下游任务中取得优异性能, 展现出强大的小样本学习能力。

● 图结构生成. 随着 LLM 在人类行为模拟与自主规划方面表现出较强能力^[156], 其逐渐成为跨领域模拟研究的新范式, 并被应用于社会动力学^[157]、经济学^[158]等领域. 尤其值得一提的是, 基于大语言模型的多智能体系统在模拟复杂的社交互动方面表现出色. 这推动了 LLM 驱动的图生成器的发展, 这些图生成器用于生成动态的文本文属性图, 其中时间节点和边的形成自然地源于智能体的行为^[159]. LLM 驱动的图生成早期研究侧重于捕捉基本属性, 例如度幂律分布、小世界现象和社群结构^[33], 其中 LLM4-GraphGen^[34] 等框架将图合成形式化为文本提示任务. 后续研究, 包括 IGDA^[35], 引入了迭代推理机制来建模图演化中的因果关系, 而 Chang 等^[36] 则探索了政治领域的其他社交图特征. LLM-Graph-Gen^[39] 注入网络科学领域的专业知识来优化生成网络的网络属性. 然而, 这些方法往往缺乏现实性, 因为交互模式过于简化、可扩展性不足, 通常节点数仅限于 100. 最近, GAG^[37] 采用基于大语言模型的智能体, 通过大规模多代理模拟生成富文本动态社交图. GraphMaster^[38] 基于 GAG 将图生成扩展到社交网络之外, 提出一个多智能体框架, 其中包含专门的智能体 (管理、感知、增强、评估), 可协同优化图合成过程. 这项工作通过为图基础模型 (graph foundation models, GFMs) 开发提供高保真、领域无关的图数据生成, 解决了大规模、语义丰富的图数据集稀缺问题。

2.2 图生成模型应用

本文根据针对文本文属性图的生成式建模方法, 对图生成模型的实际应用进行分类和讨论. 本节探讨可扩展图生成的三个关键应用: 社会网络分析 (social network analysis, SNA)、用户交互行为建模以及分子蛋白质设计. 这些进展凸显了图生成模型在解决特定领域约束下的高维组合挑战中的关键作用。

2.2.1 社会网络分析

社会网络分析一直是计算社会科学的基石, 对于社会网络的模拟生成已成为捕捉现实世界网络复杂动态的有力工具. 在三个关键研究方向上, 基于模拟的图模型展现了重要应用:

1) 网络结构理解. 统计生成模型能够发现现实世界网络中涌现的拓扑模式, 并深入了解其形成机制. Barabási-Albert (BA) 模型^[160] 采用优先连接

(新节点更有可能连接到中心节点) 来生成无标度网络, 这反映了社交网络增长、网页链接和引文动态等现象. 类似地, 边复制模型^[161] 模拟了生物网络中的基因复制、引文网络中的链接继承以及万维网演化等过程, 反映了通过节点邻域复制实现的结构守恒。

2) 网络演化建模. 图演化模拟有助于研究动态过程, 包括流行病传播、交通流量和经济互动^[162]. 最近的研究整合了基于大语言模型的智能体, 以模拟类似人类的行为, 用于大规模社交网络演化, 尤其是在在线环境 (例如社交平台、推荐系统) 中. 研究利用真实数据^[163-164] 或合成用户资料^[37] 来初始化网络. 诸如 OASIS^[165] 之类的混合方法将有限的现实世界关系与广泛的合成数据相结合, 而其他方法则采用同质性驱动的假设来连接相似的用户^[166]. 除了网络结构复制之外, 一些模型还模拟影响力传播, 以研究信息传播和对事件的动态态度转变, 例如意见领袖动态^[167]、谣言传播^[168]. 在经济领域, SRAP-Agent^[169] 提出模拟资源分配策略下的决策框架, 从而通过合成网络交互进行策略评估。

3) 网络外推. 现实世界网络 (例如互联网和社交网络) 的动态增长提出了一个关键问题: 当前的互联网协议在五年后是否仍然可行? 应对该挑战需要能够将现有图谱外推到未来规模的可扩展算法. 然而, 现实世界图谱的庞大规模 (例如数十亿个节点) 对计算复杂度和内存效率施加了严格的限制. 存在两种方法论范式: 生成式方法^[170-171] 通过建模结构特征来合成更大的图谱, 非生成式方法应用粗化/去粗化技术^[172]. 另外, 最近的工作提出利用基于大语言模型的智能体驱动的迭代扩展策略进行网络扩展^[37].

2.2.2 用户交互行为建模

基于模拟的图生成方法还广泛应用于推荐系统中的用户交互行为建模. 推荐系统的核心任务是预测用户与物品之间的交互. 这个问题可以自然地建模为二分图, 其中用户和物品构成两个独立的节点集, 而诸如评分、购买或点击之类的交互则表示为边^[173]. 基于大语言模型的智能体的兴起为用户-物品二分图建模创造了新的机遇. 这些模拟驱动的方法构建了可控的环境, 其中合成的用户和物品可以动态交互, 使研究人员能够研究推荐系统中那些传统方法难以解决的关键挑战. 1) 解决稀疏性和冷启动问题. 传统的协同过滤方法在交互数据有限的情况下会遇到困难, 尤其是对于新用户或历史反馈很少的物品. 基于模拟的方法通过为未知的用户-物品对生成真实的交互来帮助解决这个问题. 例如,

Lusifer^[174] 使用基于大语言模型的智能体动态更新用户状态并生成新商品的反馈, 直接解决了冷启动问题. 2) 时间动态和偏好演变建模. 静态二分图无法捕捉用户偏好和商品受欢迎程度随时间的变化——这是实际应用的一个关键限制. 模拟框架通过将时间模式直接嵌入图生成中来解决这个问题. SUBER^[175] 使用基于大语言模型的智能体模拟不断发展的用户行为轨迹, 生成动态二分图来捕捉偏好随时间的变化. 与孤立地对单个用户进行建模不同, RecAgent^[176] 融入了社交互动, 模拟用户间的相互影响及其偏好变化, 从而反映真实的社交动态. RecUserSim^[177] 扩展了这种方法, 在对话环境中融入记忆和行为模块, 以建模跨对话交互的细粒度演化. Agent4Rec^[178] 展示了具有个性化配置文件和记忆的生成代理如何模拟多样化的长期用户行为, 从而有效地创建时间一致的用户-物品交互模式.

2.2.3 分子蛋白质设计

在分子与蛋白质设计中, 通常采用自然语言提示作为语义条件, 以引导模型生成具有特定文本描述的图结构. 早期研究主要聚焦于从文本中抽取图结构表示^[147], 或根据给定的语义描述生成相应的图结构^[24, 148]. 在分子图生成任务中, 这类方法常以分子性质 (如亲水性、毒性、合成可及性或构象能等) 作为生成条件, 用以指导目标分子的构建^[25-26], 在药物发现与材料设计等领域展现出重要应用价值. 近期, 预训练语言模型被进一步整合到条件图生成框架中. 例如, GraphGPT-1^[150] 与 LGGM^[151] 通过将文本提示编码并与图嵌入对齐, 在生成过程中显式地维持语义一致性, GraphGPT-1 的整体框架如图 5 所示.

3 图预测

近年来, 图预测任务在处理异质结构、建模长程依赖以及应对标注稀缺等方面面临诸多挑战. 该类任务通常要求模型能够融合局部邻域与全局拓扑

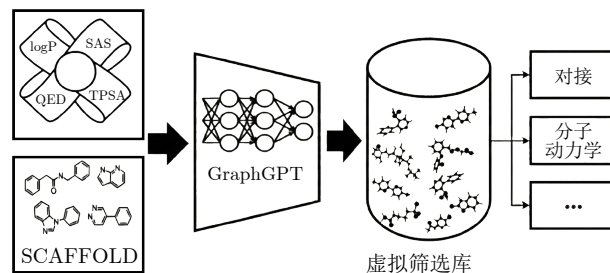


图 5 GraphGPT-1 的分子生成框架
Fig.5 The molecular generation framework in GraphGPT-1

结构、文本属性与语义先验等多源信息, 完成节点、边、整图级别的语义推理. 在文本属性图 $G = (V, E, \{s_i\}, \{t_{ij}\})$ 中, 图预测任务通常需要综合利用节点的文本属性、边的语义信息以及图结构等多种信息, 以提升建模的准确性. LLM 在语言理解与推理方面展现出强大的泛化能力, 为图预测提供了新的建模范式. 一方面, LLM 能通过自然语言提示构建灵活的任务表达, 降低对结构化监督的依赖; 另一方面, 其上下文建模能力有助于刻画远程关系和隐式语义. 然而, 直接将通用 LLM 应用于图预测任务仍面临诸多挑战, 例如缺乏图的拓扑归纳偏置, 难以高效处理大规模图结构等. 针对这些问题, 当前关于图场景下大模型建模的研究, 根据 LLM 在图任务流水线中的功能定位大体可归纳为三类范式 (其核心差异与优缺点总结如表 1 所示). 第一类将 LLM 作为预测器 (LLM as predictor), 由大模型直接完成节点、边、子图的分类与评分等预测任务, 而图神经网络或其他图编码器主要负责将图结构压缩为适配 LLM 的输入表示. 第二类将 LLM 作为增强器 (LLM as enhancer), 利用其强大的语言理解与生成能力进行文本扩写、关系描述生成、伪标签构造或规则挖掘, 以此丰富或重塑图上的语义与结构特征, 下游决策仍主要由图模型完成. 第三类将 LLM 作为对齐器 (LLM as aligner), 通过对比学习、

表 1 三类范式的核心定位与优缺点对比
Table 1 Comparison of core roles and advantages/disadvantages of three paradigms

范式	核心角色定位	优点	缺点
LLM 作为预测器	LLM 直接作为主要预测器, 通过提示词完成任务	1) 无需显式 GNN 建模, 端到端简单 2) 零样本/少样本能力强 3) 天然支持跨任务、跨图泛化	1) 难以高效建模复杂图结构关系 2) 推理成本高, 难以规模化 3) 预测不稳定, 易受提示词和上下文长度影响
LLM 作为增强器	LLM 用于增强节点语义, 最终预测由 GNN 完成	1) 充分保留 GNN 的结构建模优势 2) 推理阶段无需或极少依赖 LLM, 效率高、成本低 3) 更适合大规模图数据与工业场景 4) 结构归纳偏置明确, 预测稳定性强	1) LLM 能力仅作为“辅助”, 未充分发挥其推理能力 2) 若协同训练不足, 语义与结构对齐仍可能受限
LLM 作为对齐器	LLM 用于对齐文本、结构、标签等多模态语义空间	1) 可统一不同图、不同任务的表示空间 2) 提升跨图、跨域泛化能力 3) 有利于构建通用预训练模型	1) 训练流程复杂, 通常需要大规模预训练 2) 对数据多样性和设计假设依赖性较强 3) 对资源和工程能力要求高

知识蒸馏等机制显式对齐图表示与语言表示空间, 在统一的语义坐标系中联合建图结构与自然语言, 使不同模态的表示落入更加一致、可对比的共享语义空间. 这三类范式的差异主要体现在 LLM 在整个建模流程中所扮演的角色和介入的阶段深度上, 呈现出从直接预测到语义增强以及模态对齐的功能演化. 下文将围绕这三类范式对相关工作进行系统梳理, 并进一步讨论其在不同任务背景下的适用性.

3.1 LLM 作为预测器

相较于另外两类更侧重表示增强和模态对齐的范式, LLM 作为预测器范式的目标最为直接: 以 LLM 作为主要预测器, 直接输出下游任务结果. 该方法将图的结构信息、节点或边的属性以及监督信号组织为自然语言提示, 如局部子图、候选边描述或标签说明, 并输入 LLM 执行条件生成或分类预测. 如图 6 所示, 该范式不依赖专门的图神经网络训练, 建模的核心在于提示词和上下文示例的构造, 以及必要时的检索与重排机制, 用于缩小候选空间. 围绕图结构与语言模型的耦合方式, 现有研究可大致分为基于输入侧的图序列化方法与基于架构侧的结构融合方法两类.

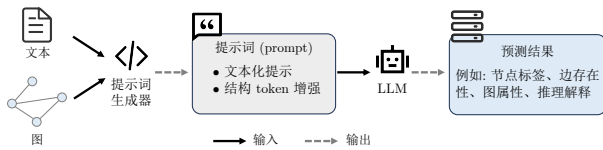


图 6 LLM 作为预测器

Fig. 6 LLM as predictor

3.1.1 基于序列化图输入的语言模型预测

将图结构转换为序列是当前最为简便的建模路径之一. 该方法通过将结构信息编码为序列形式, 结合节点文本一并输入 LLM, 从而复用其已有的上下文理解与推理能力, 实现零样本或小样本迁移, 具备良好的通用性与扩展性. 具体地, 模型通常构造以目标节点 v 为中心的自中心子图 (ego-graph) $\hat{G}_v$, 并通过图到序列转换函数 $\text{Graph2Seq}(\cdot)$ 将其转化为结构 token 序列 $H_{\hat{G}_v}$. 随后, 将结构序列 $H_{\hat{G}_v}$ 与节点文本 s_v 一同输入 LLM 进行预测建模. 整体过程可形式化为:

$$H_{\hat{G}_v} = \text{Graph2Seq}(\hat{G}_v), \quad h_v = \text{LLM}([s_v; H_{\hat{G}_v}])$$

其中, h_v 表示由 LLM 输出的目标节点 v 的节点表示向量. 该范式的核心在于将结构信息从图的“输入侧”注入语言模型, 避免修改网络结构或重新训

练, 从而轻量化地实现图预测任务 (如节点分类、链接预测、图分类等). 根据结构序列的构造方式不同, 现有方法大致可分为两类: 规则驱动方法 (rule-based) 与结构编码驱动方法 (GNN-based).

● 规则驱动方法. 通过预设模板将图的局部结构 (如节点邻居、边关系等) 线性化为自然语言片段, 从而构造出结构化提示 (prompt), 交由 LLM 进行自然语言理解与生成. 例如, 对于一个以节点 v_i 为中心的子图 g_v , 可设计如下提示:

“Center v_i ; 1-hop neighbors: $v_j, v_k, \dots$; What is the category of the node?”

这类方法通常不引入额外图模型, 仅依靠语言模板完成图结构到文本的映射, 具有实现简便、无需额外参数改造的优势, 并且易于在零样本或小样本场景中迁移. 代表性工作 InstructGLM^[40] 采用纯语言化的图表示, 将节点/边属性与局部结构描述拼接为自然语言提示, 在多提示 (multi-instruction) 指令微调下以负对数似然 (negative log-likelihood, NLL) 损失优化模型参数, 让 LLM 直接生成节点分类与链接预测答案, 并辅以自监督链接预测任务强化其结构理解能力, 从而以纯生成式管线在多图基准上超越多种 GNN 和 Graph Transformer 模型. 类似地, GPT4Graph^[41] 将多种图任务 (包括结构理解与语义理解) 统一建模为自然语言问答或文本生成问题. 该方法提出图描述语言 (graph description language, GDL), 将图结构及属性转写为可读文本, 并结合精心设计的提示工程, 使 LLM 能够直接生成多类任务的答案, 如节点标签预测、边存在性判断、图直径或聚类系数计算以及 GQL/Cypher 查询语句生成等, 从而摆脱了仅以嵌入向量供下游模型使用的限制. GraphText^[42] 则进一步提出图-句法树 (graph-syntax tree) 的构造机制: 将节点的多模态特征和多种关系 (如高阶连通、PageRank、最短路径等) 统一组织为可遍历的层次化自然语言提示, 再将节点分类、链接预测、图分类等任务统一改写为文本生成问题, 让同一个预训练 LLM 在共享文本空间中完成多类图预测任务, 并支持上下文学习或少量指令调优直接生成预测答案与自然语言解释. LangGFM^[43] 将各种图任务统一转写为自然语言形式, 在标准图文本格式基础上直接对 LLM 进行指令微调, 让 LLM 在纯语言空间内同时完成节点分类、链接预测和图分类等任务, 而不再显式依赖 GNN 等专门图模块. 总体而言, 规则驱动方法实现成本低、可迁移性强, 但其线性展开过程易丢失拓扑结构与位置不变性, 且邻域结构序列较长时推理与训练成本将显著上升.

● 结构编码驱动方法. 在输入 LLM 之前, 先使用图神经网络等图编码器对节点的邻居子图进行结构化编码, 得到一组结构向量 (通常作为“特殊结构 token”), 再将这些结构 token 与节点文本拼接后输入 LLM 中, 由 LLM 基于融合后的图文序列完成节点分类、链接预测等任务. 其核心思想是: 以 GNN 作为结构特征抽取器, LLM 作为任务预测器, 实现两者的解耦式协作. 常见的图编码器可采用 GraphSAGE^[7]、GAT^[6] 或 GraphFormer^[179] 等结构, 以节点为中心采样其 k 阶邻居, 通过消息传递和注意力聚合生成节点的图结构嵌入表示 H_g , 并将其作为特殊 token 插入文本序列中, 与节点文本一同输入 LLM 进行联合建模. 代表性工作 GraphGPT^[44] 采用图结构编码与文本联合建模的双阶段指令微调框架: 首先利用 GNN 对节点邻居子图进行编码以生成图结构向量, 再与节点的自然语言描述拼接后送入 LLM, 以标准 NLL 目标在多任务指令数据上微调, 使得 LLM 能够解释本场景下的结构语境并完成节点分类、链接预测等任务, 从而在不修改 LLM 架构的前提下显著提升其对图结构的理解能力. LLaGA^[45] 则在 GraphGPT 框架^[44] 上进一步提升了泛化性与统一性: 其目标是在不改动 LLM 参数的前提下, 让 LLM 能直接理解图结构并生成任务答案 (分类、链接预测与解释), 同时保持多任务/多数数据集一致性与零样本泛化能力. LLaGA 采用统一的“问题-答案” (question-answer) 式指令数据, 将节点嵌入与问题文本拼接为统一输入, 并最大化在给定节点信息与问题条件下生成正确答案的似然概率. 通过这种方式, LLaGA 能够同时学习图结构与文本之间的对应关系, 并具备基于图结构信息进行自然语言推理与解释的能力. 结构编码驱动方法结合图模型与语言模型的优势, 能够生成更紧凑、语义更丰富的结构表示, 尤其适用于建模长距离依赖或复杂结构关系的任务. 然而, 该类方法也带来了更高的实现成本: 不仅需要预训练图编码器, 还需处理图表示空间与 LLM 输入空间之间的对齐问题, 从而增加了模型参数规模与系统集成的工程复杂度.

3.1.2 基于结构融合的语言模型内部建模

此类方法在 Transformer 各层中显式注入图结构归纳偏置, 在语言建模流水线内部实现结构信息与文本表示的深度耦合. 不同于仅在输入或输出侧叠加独立图编码器的框架设计, 结构融合范式直接改造 LLM 的层级结构, 使模型在保持原有自然语言建模能力的同时, 能够逐层感知图中的长程依赖与细粒度邻域交互, 从而在结构主导的图预测任务

上展现出更强的表示与推理能力.

以 GOFA^[46] 为代表, 其核心思想是将 token 级 GNN 消息传递算子嵌入 LLM 的编码栈中, 形成“语言-图结构”交替耦合的编码流程, 其整体框架如图 7 所示. 具体而言, 模型首先将文本属性图中节点与边的文本编码为一串 text tokens, 并为每个节点/边附加由 LLM 压缩器引入的 memory tokens, 共同构成图语言编码器的输入. 随后在若干 LLM 压缩层与 token 级 GNN 层之间交替堆叠: LLM 压缩层在序列维度上聚合和压缩长文本, 将句子压缩为紧凑的语义记忆; GNN 层则把这些 memory tokens 作为节点/边特征, 在局部子图上通过基于 Transformer 的注意力消息传递在 token 级执行邻域聚合与结构推理, 对节点的 memory 表示进行更新并反馈到后续语言流中. 最终得到的结构感知 memory embedding 被送入 LLM 解码器, 作为前缀通过统一的语言生成接口完成节点/边级分类、链接相关任务以及自由形式问答等多种图任务. GOFA 进一步在这一框架下设计了以句子补全为核心的多种自监督预训练任务, 以同时提升语言流畅性与对图结构的建模能力.

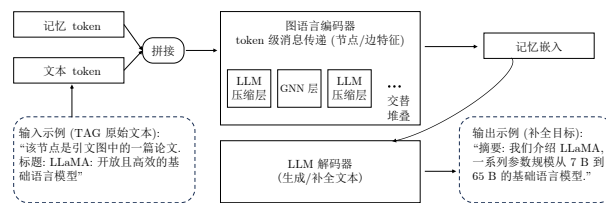


图 7 GOFA 的整体框架示意图

Fig.7 Schematic diagram of the overall framework of GOFA

TEA-GLM^[47] 则沿着另一条思路, 先用图神经网络预训练并将节点表示对齐到大模型的词向量空间, 再通过统一指令模板让大模型在无需任务特定微调的情况下完成跨数据集、跨任务的零样本节点分类与链接预测, 体现了结构编码与语言空间对齐相结合的设计.

尽管结构融合型方法在建模能力上具有显著优势, 但通常需要对原有语言模型进行架构层级的改造与联合预训练, 对数据规模与算力提出更高要求. 此外, 该类方法与闭源或商用语言模型的兼容性较差, 训练过程稳定性不足, 部署与复现成本较高, 在一定程度上限制了其在实际应用中的推广落地.

3.1.3 面向特定任务的 LLM 预测建模方法

除了通用的图预测框架外, 近期也出现了一些面向特定任务场景 (如链接预测、动态图推理) 的

LLM 预测建模方法, 探索如何在这些更具挑战性的设定下发挥 LLM 的推理与泛化能力. LinkGPT^[48] 针对文本属性图中的链接预测任务, 旨在使 LLM 能够直接预测图中缺失的边. 其整体框架包括三个核心模块: 首先, 通过节点编码与成对编码构建节点及节点对的语义表示; 其次, 采用包含邻居预测与链接预测的双阶段指令微调流程, 显式引导模型学习图结构与语言之间的对齐关系; 最后, 在推理阶段采用检索-重排序机制, 先筛选候选节点, 再由 LLM 对候选项进行精排, 从而在保持准确率的同时显著降低推理开销. 通过上述三部分的协同设计, LinkGPT 能在充分发挥 LLM 推理能力的基础上, 有效建模图结构与文本语义之间的复杂关联, 实现高效且具扩展性的链接预测能力. LLM4DyG^[49] 则面向动态图任务, 提出首个空间-时间推理基准, 将包括 when link, when connect, what neighbors, check triadic closure, temporal path, shortest temporal path, sort edges by time, degree ranking 和 temporal motif counting 在内的九类典型问题统一改写为自然语言问答, 例如要求模型回答“何时可以从节点 A 沿时间序列路径到达节点 B”. 此外提出 DST2 提示策略, 通过引导模型按“先时间、后结构”的解耦思维流程推理, 显著提升了 LLM 在动态图上的时序推理能力. GAD^[50] 进一步将上述思路扩展到动态文本属性图上, 利用多智能体框架使多个协同 LLM 在全局与局部视角下直接预测未来的边与节点状态. 此外, HiGPT^[51] 在异构图中通过指令微调和上下文内图分词器 (in-context graph tokenizer), 把不同类型节点与关系统一映射到语言空间, 从而直接用大模型完成节点与边级别的预测与推理任务.

3.2 LLM 作为增强器

与将 LLM 直接用于任务预测的做法不同, LLM 作为增强器的核心思路在于借助其强大的生成能力, 对图中节点或边的属性进行语义增强, 从而为后续的图模型学习提供更具表达力的输入表示. 具体而言, LLMs 首先对节点的原始文本属性进行预处理, 将其转化为更丰富的语义特征表示, 再作为节点特征输入 GNN 中进行下游任务学习. 根据增强的粒度不同, 这类方法可以在特征层 (feature level) 或文本层 (text level) 进行增强, 如图 8 所示. 特征层增强直接将文本属性编码为节点向量, 作为 GNN 的输入特征; 文本层增强则通过 LLM 生成扩展后的文本描述, 再将其编码为增强特征, 从而提升节点表示的表达力. 这一范式有效地将

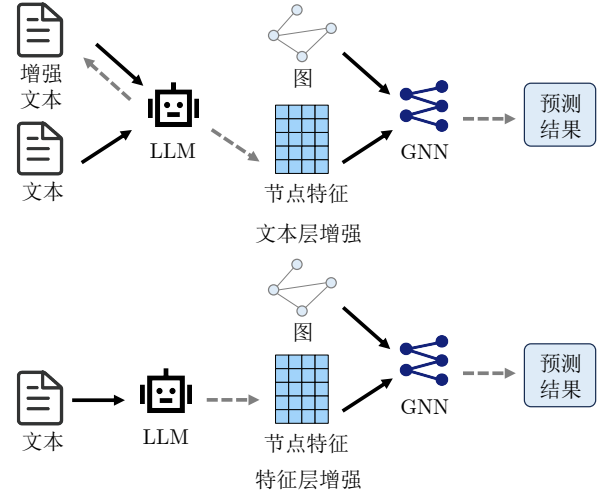


图 8 LLM 作为增强器

Fig.8 LLM as enhancer

LLM 的语言知识注入图表示学习过程中, 提升 GNN 的预测性能^[16].

3.2.1 特征层增强

特征层增强的核心思路是在向量空间中直接注入语言知识. 具体而言, 首先使用可输出嵌入的 LLM 将节点的原始文本属性 s_i 编码为语义表示 $x_i \in \mathbf{R}^a$, 再将这些向量作为节点的初始特征输入 GNN 进行结构化学习. 为有效衔接语言表示与图表示, 现有方法主要采用两类整合范式: 1) 级联式建模范式, 先由 LLM 编码文本, 再由 GNN 聚合结构, 二者串联建模; 2) 协同训练范式, 通过辅助损失或伪标签机制, 使两类模型在训练中相互优化, 逐步对齐语义空间. 两类方法虽然实现路径不同, 但本质上均旨在不改变图学习主干结构的前提下, 以更具语义性的特征初始化 GNN, 从而提升下游预测性能.

● 级联式建模范式. 首先使用 LLM 独立编码节点的文本特征, 随后将获得的语义向量作为初始表示交由 GNN 聚合邻域结构信息, 实现语言表示与结构建模在流程上的解耦. 这类方法具有实现简单、模块替换灵活等优势, 是早期探索 LLM 与图结构融合的代表路径. GIANT^[52] 提出一种基于图结构自监督的语言模型预训练方法, 将“从节点文本预测其多尺度邻居”建模为极端多标签分类 (extreme multi-label classification, XMC) 任务, 并结合 XR-Transformers 的分层标签树结构与多分辨率训练机制, 对大规模文本图中节点的邻域分布进行建模. 具体而言, GIANT 首先利用节点真实邻域构建稀疏多标签监督信号, 再基于标签共现相似度进行层次聚类与递归优化, 使 Transformer 编码器

在层级式 (coarse-to-fine) 训练过程中逐步对齐文本语义与结构邻域之间的对应关系. 训练完成后, 模型可为每个节点生成具备结构感知能力的语义向量, 相较于传统词袋表示或原始 BERT 编码, 其表征效果更优, 可直接作为 GNN 或 MLP 的输入, 在不增加推理开销的前提下显著提升节点分类与链接预测等任务的性能. SimTeG^[53] 则针对文本图中节点表示质量较弱的问题, 提出一种极简的两阶段级联建模框架: 首先对预训练语言模型进行参数高效微调 (parameter-efficient fine-tuning, PEFT), 生成任务相关的节点语义表示; 随后将该语义表示作为节点特征输入 GNN, 进一步聚合结构信息并完成下游任务训练. Gao 等^[54] 在冻结的 LLM 之后插入一个轻量的注意力式传输模块 (attention-based transfer module, AT module), 自适应地从多种提示下的 token 级输出中选择并融合关键信息, 作为节点初始特征输入 GNN, 从而优化 LLM 与 GNN 之间的信息传递. 在更大范围的跨任务场景下, OFA^[55] 来自多域的图数据统一转换为文本属性图, 并用单一 LLM 将节点/边描述与任务说明映射到共享嵌入空间, 随后在插入感兴趣节点 (nodes-of-interest, NOI) 子图与 prompt 子图后交由一个通用 GNN 进行结构化推理, 从而实现跨图多任务分类能力. 该方法在不引入模型结构改动的前提下, 有效提升了节点分类与链接预测的性能. 然而, 该类方法在实现解耦的同时, 也可能导致语义表示与结构聚合之间缺乏显式对齐, 无法在文本编码阶段引入结构归纳偏置, 且在面对任务目标或数据分布变化时, 适应能力相对有限.

● 协同训练范式. 为缓解这些局限, 近年的研究逐渐转向协同训练范式, 通过伪标签生成、辅助结构损失或参数交替优化等机制, 使语言模型与图模型在训练过程中相互传递监督信号、逐步对齐语义空间, 从而实现结构与语义的深度融合. GraphFormers^[56] 通过在 Transformer 层中嵌入 GNN, 实现了节点文本与图结构的迭代融合, 较传统串联方式更能精确建模跨节点的语义依赖关系. PATTON^[60] 则进一步提出结构感知的连续预训练策略, 以提升语言模型对多种下游任务 (如分类、检索、重排序、链接预测) 的适应能力. 其方法是在 Transformer 层之间引入 GNN 模块, 构建交替嵌套的“Transformer+GNN”结构, 使节点文本表示在多层语义建模过程中逐步融合邻居结构信息, 从而获得更鲁棒的图感知表征. Heterformer^[57] 面向异构且部分节点无文本的图数据, 在 Transformer 各层引入虚拟邻居 token 与类型感知注意力机制, 实现结构与语义的层内融合建模, 并通过无监督链接预测

学习通用的节点表征. Edgeformers^[58] 则进一步关注边文本与图结构交互, 利用虚拟节点与非对称注意力实现结构感知的边嵌入表示, 并以边分类与对比学习联合驱动优化. 此外, 面对跨图零样本泛化等高难度任务, 研究者也尝试通过“统一语义空间+局部结构提示”机制增强泛化鲁棒性. UniGLM^[59] 在多张文本属性图上采用统一的语言模型, 进行结构感知的对比预训练. 其核心在于通过自适应正样本选择机制挖掘跨图稳健的语义-结构对齐关系, 并结合惰性对比机制提升训练效率, 从而获得一个具有跨域迁移能力的通用节点表征模型. ZeroG^[61] 在此背景下提出统一的预训练语言模型, 对所有数据集中的节点文本与类别描述进行编码, 映射至共享的语义空间, 从根本上消除不同图之间的特征维度差异. ZeroG 设计了基于提示 (prompt) 的子图采样策略以注入结构信息与全局语义. 具体而言, 模型在源图中抽取满足类别多样性约束的 k -hop 子图 (要求子图中包含的类别数不小于全图类别数的一半), 并在每个子图中引入一个携带数据集领域描述的提示节点 (prompt node), 该节点与子图内所有节点全连接, 用以提供全局语义提示. 之后, 模型在初始节点表示 $H^{(0)}$ 上执行多轮邻居聚合以将局部结构信息注入语言模型产生的节点表征中:

$$H^{(l+1)} = \tilde{A}H^{(l)}, \quad \tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$$

其中, D 为度数矩阵. 此外, 为避免全参数预训练带来的高资源开销与零样本场景下的过拟合风险, ZeroG 仅在语言模型的少量注意力层中插入 LoRA 参数进行轻量化适配, 并以节点表示 h_i 与其真实类别表示 c_{y_i} 之间的相似度作为监督信号进行优化, 损失函数 $\mathcal{L}$ 定义如下:

$$\mathcal{L} = -\log \frac{\exp\left(\frac{\text{sim}(h_i, c_{y_i})}{\tau}\right)}{\sum_y \exp\left(\frac{\text{sim}(h_i, c_y)}{\tau}\right)}$$

其中, τ 为温度系数, 用于调节相似度分布的平滑程度, 通常取值范围为 $0 < \tau \leq 1$, 较小的 τ 会使得分布更加尖锐, 较大的 τ 会使得分布更加平滑; $\text{sim}(h_i, c_{y_i})$ 表示节点特征 h_i 与类别 c_{y_i} 的相似度, 通常通过余弦相似度计算. 在推理阶段则直接通过余弦相似度最大化来预测目标节点的类别. 为进一步提升在大语言模型上的适配效率, 部分研究选择保留预训练语言模型结构, 仅通过轻量参数注入结构信息, 以兼顾模型性能与训练成本. GraphAdapter^[63] 提出一种面向十亿级参数 LLM 的高效图建模框架, 将冻结的 LLM 作为语义编码器, 将 GNN 作为轻量适配器接入 LLM 隐层, 仅在模型最后一层通过

可训练的融合模块注入结构表示. 在预训练阶段, GraphAdapter 采用“语言-结构”联合训练范式, 以节点文本的下一词预测作为统一监督信号, 并通过残差平均机制 (即融合结构分布与原始语言分布的均值) 引导 GNN 聚焦于纠正结构相关 token 的预测误差. 下游阶段则采用 Prompt-aware 微调策略, 在融合表示上接入线性预测头完成节点分类等任务. 后续工作 GAGA^[62] 则在少量代表性节点上调用 LLM 生成标签与解释, 构建额外的标注图, 并通过对比学习让 GNN 在原始图和标注图之间对齐, 从而在训练阶段协同更新 LLM 与 GNN. 此外, SFGL^[64] 先利用 GNN 学习节点的伪标签, 再据此微调文本编码器, 使其产生的文本向量更加贴合图结构, 然后在该向量空间上训练 GNN 完成下游节点分类等任务. 进一步地, OpenGraph^[65] 提出一种面向跨图预训练的 LLM 引导结构生成范式, 模型利用 LLM 对节点文本属性的语义编码进行相似度打分, 并结合采样机制构建边结构, 进而获得覆盖多域的训练图集合, 用于预训练统一的图 Transformer 基础模型并支持零样本图任务迁移.

3.2.2 文本层增强

文本层增强方法则侧重于在自然语言层面对节点描述进行改写与扩展. 具体而言, 利用 LLM 对原始文本属性 s_i 进行标准化、上下文补全或语义重述, 生成更具信息量的增强文本, 如多视角释义、歧义消解与语义补全等. 随后, 将增强后的文本再编码为向量特征并输入 GNN 进行结构化建模. 相较于直接生成嵌入向量, 该方法将增强重点放在语言内容本身, 因而在原始文本噪声较大或信息稀疏的场景下, 能够显著提升可用语义信息. 在实际应用中, 常将“原始文本特征+增强文本特征”进行融合或集成, 以兼顾初始表达与新增语义, 从而提升节点表达的表力与鲁棒性. ENGINE^[66] 提出一种“侧路式”协同框架, 在冻结 LLM 的条件下, 将轻量可训练的 GNN 模块 (G-Ladders) 挂载于 LLM 各层之间, 构成并联的结构侧路. 模型前向过程中, 首先由 LLM 编码节点文本, 提取 token 表示后聚合为节点级表示, 再由每一层的 G-Ladder 注入结构信息. 整个过程中仅需训练侧路参数与最终分类头, 既保留了 LLM 编码质量, 又避免反向传播至 LLM 主体, 显著降低了显存与训练成本. 与此不同, TAPE^[67] 选择将 LLM 作为静态的知识源, 而非直接作为结构建模器. 该方法首先通过 prompt 调用 LLM 获取节点的预测结果及自然语言解释, 并引入一个用于将 LLM 输出解释转化为语言模型特征的解器 (LLM-to-LM interpreter) 将这些文本化知识转化

为向量特征 (包括原始文本表示、解释表示和预测分布), 再交由 GNN 聚合. 该方法无需对 LLM 本体进行训练, 完全绕开梯度传递, 极大简化了工程实现与部署流程. 进一步地, KEA^[17] 将增强重点从任务相关解释扩展至外部知识引入, 而 Pan 等^[68] 则从知识蒸馏角度利用 LLM 生成多粒度文本理据, 并将其蒸馏进本地模型, 无需在线调用 LLM 即可继承文本增强带来的收益. 针对真实场景中文本和边稀疏的特点, UltraTAG-S^[69] 将上述增强文本特征与基于 LLM 的节点选择、边重构策略协同使用, 使得 GNN 即便在严重稀疏的文本文属性图中也能学到可靠的节点表示. 这类方法通常通过引导 LLM 生成与节点相关的补充文本 (如实体信息、解释性描述等), 并与原始节点属性拼接或联合编码, 从而在不依赖 LLM 直接预测的前提下, 显著提升输入语义的覆盖范围与区分能力.

3.3 LLM 作为对齐器

尽管前述两种范式已在图任务中展现出良好的预测性能, 但它们大多隐含一个前提: 语言语义空间与图结构语义空间可以天然对齐. 然而, 图数据往往具有更强的结构归纳偏置和局部性, 若缺乏显式的对齐机制, LLM 难以准确捕捉图的语义特征, 导致模型泛化能力受限. 为解决这一挑战, 近期研究提出了 LLM 作为对齐器的范式, 旨在通过对比学习、知识蒸馏等机制, 将 LLM 与 GNN 的表示空间在训练过程中协同优化, 如图 9 所示. 这类方法不再局限于表示增强或任务预测, 而是关注模态统一与语义对齐, 以实现更深层次的图语义理解与推理建模. 代表性方法之一是 GLEM^[70], 其采用期望最大化 (expectation-maximization, EM) 框架交替训练 LLM 与 GNN, 在不引入结构标注的情况下完成联合建模, 其整体框架如图 10 所示. 设 $\mathcal{L}$ 和 $\mathcal{U}$ 分别表示有标签节点集合和无标签节点集合; α 与 β 分别为 E 步与 M 步目标函数中的权重系数, 用于平衡有标签节点监督信号和无标签节点伪标签信号.

● 在 E 步中, 固定 GNN, 使用 LLM 对所有节点生成伪标签 $\hat{y}_n$, 并与真实标签一起优化 GNN 的训练目标. 此时, 仅使用文本信息建模预测分布:

$$\mathcal{O}(q) = \sum_{v_n \in \mathcal{L}} E_{q_\theta} [\log q_\theta(y_n | s_n)] + (1 - \alpha) \sum_{v_n \in \mathcal{U}} \log q_\theta(\hat{y}_n | s_n)$$

其中, $q_\theta(\cdot)$ 表示由 LLM 参数 θ 建模的文本感知预测分布; $\mathcal{O}(q)$ 表示 E 步中关于 LLM 预测分布 q_θ 的

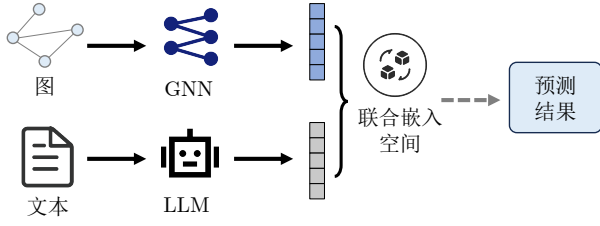


图9 LLM 作为对齐器

Fig.9 LLM as aligner

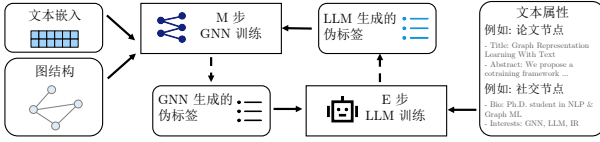


图10 GLEM 的整体框架示意图

Fig.10 Schematic diagram of the overall framework of GLEM

优化目标.

• 在 M 步中, 固定 LLM, 使用 LLM 生成的节点表示 h_v 和伪标签 $\hat{y}_n$ 作为 GNN 的输入信号, 学习结构感知分布:

$$\mathcal{O}(p) = \beta \sum_{v_n \in \mathcal{L}} \log p_\phi(y_n | h_v, A) + (1 - \beta) \sum_{v_n \in \mathcal{U}} \log p_\phi(\hat{y}_n | h_v, A)$$

其中, $\mathcal{O}(p)$ 表示 M 步中关于 GNN 预测分布 p_ϕ 的优化目标.

该方法通过在结构信息注入前使用 LLM 编码语义, 并借助伪标签对齐 LLM 与图结构的表示空间, 从而在训练过程中逐步融合两类模态的知识, 避免直接端到端联合训练带来的不稳定性与任务干扰.

ConGraT^[71] 进一步采用对比学习作为对齐优化目标, 引导 LLM 生成的表示靠近其对应的 GNN 表征. 该方法设计了跨模态对比损失, 使得同一节点在 LLM 空间与图空间中的表示保持一致, 同时与其他节点区分开, 从而强化语义与结构的耦合关系, 提升模型的表示鲁棒性与泛化能力. 此外, G2P2^[72] 针对低资源文本分类任务, 指出仅优化文本或图结构往往会割裂两者的语义关联, 难以充分发挥图结构对语义理解的支持作用. 该方法基于一个关键观察: 文本通常锚定于某种超链接或交互图中, 图结构可为文本建模提供有力补充. 为此, G2P2 首先设计三种图交互视角 (文本-节点、文本-摘要、节点-摘要), 并在预训练阶段引入多视角对比学习, 以将文本编码器与图编码器对齐至共享嵌

入空间. 随后, 在下游任务中, 模型采用基于图的提示策略进行零样本或少样本分类. 其中, 零样本设置使用离散模板提示, 少样本场景则使用连续可学习的提示词, 并结合图上下文信息进行初始化, 以提升表示能力与泛化效果. 在实体对齐任务中, ChatEA^[73] 则让大模型直接充当跨图实体匹配器: 先通过知识图谱代码化转换 (knowledge graph code translation, KD-Code translation) 将多源知识图谱结构与实体特征转写为代码式/自然语言描述, 再采用“候选收集-推理与重思”的两阶段策略, 让 LLM 在对话式多步推理中综合名称、结构与时间等证据以判断实体是否对齐, 显著提升高异构 KG 上的对齐精度. 进一步地, LKD4DyTAG^[74] 先由教师 LLM 对包含时间信息的边文本进行编码, 获得富含语义的边表示, 再通过知识蒸馏将其对齐到轻量级时空 GNN 的边表示中, 从而在同时建模结构、时间与文本的前提下提升未来链接预测和边分类等动态图任务的性能.

3.4 图预测模型应用

本节围绕 LLM 与 GNN 结合建模范式下图预测模型的典型应用展开, 选取欺诈检测、推荐系统与股票预测三类场景作为代表, 分别对应节点分类、链接预测与排序/回归等任务设置, 概述其如何利用文本侧信息与图结构关系提升对真实业务信号的刻画与预测效果.

3.4.1 欺诈检测

在电商/金融反欺诈等场景中, 实体 (账号、设备、商户等) 之间形成复杂关系网络, 同时伴随用户画像、交易备注、申诉/理赔描述等大量文本. 为同时建模结构依赖与文本语义, LLM 与 GNN 的结合成为一种有效路径: GNN 侧重从关系链路中捕捉潜在的团伙结构与风险传播模式; LLM 则从非结构化文本中抽取可区分的语义线索, 弥补纯结构建模对文本信息利用的不足. 具体而言, FLAG^[180] 首先在目标节点的 k -hop 邻域内引入语义相似邻居采样, 优先保留与目标节点语义更相关的邻居, 以缓解“欺诈节点混入正常邻居”导致的邻域伪装与特征稀释; 随后由 LLM 从节点原始文本中提炼与标签更相关的判别性文本 (并区分残余/无关部分), 编码为更具判别力的节点表示; 最后将增强后的文本表示与子图结构一并输入带跳跃连接的 GNN (skip-GNN) 进行训练与风险预测, 从而提升对欺诈节点与团伙行为的识别效果. 此外, MLED^[181] 从“类型级 (type-level)”与“关系级 (relation-level)”两个层面引入 LLM 增强: 一方面为不同节点类型生成并

编码更具区分性的语义表示, 强化欺诈者与正常实体的差异; 另一方面建模不同关系在欺诈识别中的贡献并进行加权融合. 最终将原始节点表征与上述多层增强表征联合输入图模型完成分类, 以提升整体检测效果.

3.4.2 推荐系统

在个性化推荐场景中, 用户-物品之间的点击、收藏与购买等行为会自然形成大规模交互图, 与此同时, 平台还沉淀了标题、属性、评论与问答等丰富文本, 能够更细粒度地刻画物品特征与用户偏好, 但难以被纯图模型充分利用. GNN 与 LLM 的结合将交互结构信号与文本语义信号进行统一表征与联合建模: GNN 侧重从交互链路中学习稳定的行为关联与相似性模式; LLM 则从非结构化文本中抽取可区分的语义表示, 补足结构信号难以覆盖的属性与意图信息, 从而提升推荐的个性化与语义敏感性. 基于这一思路, RecMind^[182] 将 LLM 作为“偏好先验”引入图推荐框架: 模型冻结 LLM 参数, 并借助轻量适配器生成物品与用户的语义表示; 同时采用 LightGCN 从用户-物品交互图中学习结构表示. 随后, RecMind 通过跨视角对比学习对齐两类表示空间, 并在图消息传递过程中引入门控融合机制, 实现对语义与结构信号的自适应权衡——在冷启动与长尾样本上更依赖语义信息, 在交互数据充足时更多利用结构信号, 从而获得语义更敏感且结构一致的排序结果. 进一步地, LLMRec^[183] 利用 LLM 对在线平台的可用内容进行补全与推断, 围绕交互图提出三类增强策略: 强化用户-物品交互边、生成/完善物品节点属性、构建用户节点画像; 在此基础上, 图推荐模型 (如基于 LightGCN^[184] 的协同过滤) 在增强后的图上进行消息传递与表示学习, 将新增的边与属性转化为可用的结构信号与节点表征, 用于后续的排序预测. 此外, 在在线招聘/工作推荐场景中, 平台除岗位描述与简历等文本外, 还会沉淀候选人与岗位之间的多类型行为交互 (如沟通、面试、投递/匹配等), 从而形成带有明确语义的异构行为图. CLRec^[185] 将行为图中的高阶交互关系组织为元路径, 并通过元路径提示构造器将其转写为自然语言提示输入 LLM, 使模型在推荐决策时显式利用结构化行为线索; 同时引入路径增强与选择机制 (如路径顺序扰动、软选择加权) 以缓解提示的顺序/位置偏差, 并结合轻量参数高效微调提升对岗位匹配与潜在偏好的刻画能力, 从而在工作推荐中实现更具个性化与可迁移性的排序与判断.

3.4.3 股票预测

在股票预测中, 市场价格与基本面信号之外,

新闻流、公告与研报等文本能够快速反映公司事件与宏观预期, 是重要的外生信息来源; 同时, 股票之间还存在由行业链条、主题共振与收益联动等诱发的关系依赖, 使得单只股票的变化常伴随跨标的影响扩散. 因而, LLM 与 GNN 的结合在该任务中具有自然优势: LLM 负责从非结构化文本中提取可迁移的语义表征与事件线索, GNN 则在股票关系图上建模联动结构与信息传播, 从而将“文本驱动的语义冲击”与“跨资产的结构关联”统一用于收益预测与后续的选股决策. 基于这一思路, Guo 等^[186] 提出以微调大语言模型直接完成从新闻到收益的预测: 将每只股票在回看窗口内的新闻序列输入预训练 LLM 得到文本表征, 并在其上叠加预测模块输出未来收益估计; 同时系统比较仅编码器与仅解码器两类 LLM 在该任务中的差异, 并针对 LLM 产生的 token 级表示提出瓶颈压缩或聚合表征的融合方式以对接预测头. LED-GNN^[187] 进一步将新闻中隐含的跨股票关系显式转成随时间变化的动态图: 用 LLM 从新闻中抽取“股票 A-关系-股票 B”的关系三元组, 并在每个时间快照据此构建动态新闻图; 再结合由历史价格/成交量构建的市场图, 一起输入动态 GNN 做时序消息传递与融合, 从而利用“关系随事件更新”的动态图结构提升预测效果.

4 图推理

近年来, 大语言模型展现出强大的语义和常识理解能力, 在各种任务中取得了巨大成功. 随着 CoT^[117]、ToT^[121]、GoT^[122] 等推理模式的提出, 大语言模型能够进行更深层次、更类似人类思考方式的推理. 最近, 随着慢思考的提出, OpenAI o 系列模型、DeepSeek-R1 等模型通过强化学习增长大模型的思考时间, 使模型自动出现反思、验证等行为, 进一步增强了大模型的推理能力. 然而, 由于图结构的复杂性, 大模型在图上的推理还亟须探索. 图推理问题涉及从图结构数据中提取、推断或预测有意义的关系和属性. 给定一个由节点和边组成的图, 目标是推理图中编码的底层结构、连通性或语义关系. 这通常需要利用节点的局部邻域信息和图的全局拓扑结构. 该问题涵盖广泛的应用, 包括经典图算法问题、节点分类、链接预测、图分类, 并且是社交网络、知识图谱和分子生物学等领域任务的基础. 一般地, 图推理分为经典图推理任务和知识图谱推理任务.

4.1 经典图推理任务

经典图推理任务指最短路径、最大流、拓扑排

序等各种经典图算法问题,它们是解决各种现实图问题的基础,考察了大模型对图结构本身的理解能力.在一般情况下,当讨论大语言模型解决图推理问题时,输入是 (G, Q) 图问对.其中 $G = (\mathcal{V}, \mathcal{E}, \{s_i\}, \{t_{ij}\})$, $\mathcal{V}$ 是节点集, $\mathcal{E}$ 是边集.对于每一个节点 $v_i \in \mathcal{V}$,有一个对应的序列节点文本属性 s_i ,同样对于每条连边 $e_{ij} \in \mathcal{E}$,有序列边文本属性 t_{ij} .图结构 G 一般用边集或邻接表表示法描述为自然语言, Q 一般是问题描述或问题相关的指令.现有大模型驱动图推理的方法一般分为基于语言的图推理和代码增强的图推理方法,如图11所示,不同经典图推理方法的特点与对比如表2所示.

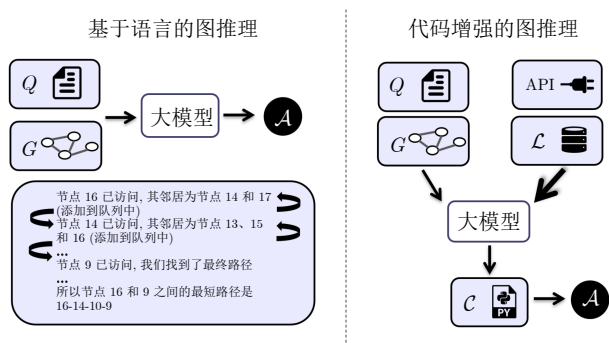


图 11 经典图推理任务的两种方法

Fig. 11 Two approaches for classical graph reasoning tasks

4.1.1 基于语言的图推理

现有大多数方法都是基于语言的,即完全依靠

自然语言进行推理.具体地,大模型 $\mathcal{M}$ 会处理 (G, Q) 对,然后将答案 $\mathcal{A}$ 以文本形式输出:

$$\mathcal{M}(G, Q) \rightarrow \mathcal{A}$$

● 提示工程方法. NLGraph^[75] 基准涵盖二部图匹配、哈密顿路径、图神经网络等八个经典图推理任务,是最早被用来评估大语言模型图推理能力的数据集. LLMtoGraph^[188]、GPT4Graph^[41]、LLM4-DyG^[49] 和 Talk like a graph^[76] 是一系列早期基于提示工程的工作. 这些工作主要探索各种图的表示形式,如边集、邻接表、邻接矩阵等传统表示方法,以及图建模语言 (graph modeling language, GML)^[189] 和图标记语言 (graph markup language, GraphML)^[190] 等更精心设计的图表示语言在以文本描述图结构方面的差异. 此外,研究者们还探究了不同的提示策略如 Zero-Shot、Few-Shot 和 CoT 等对大语言模型理解图结构和进行图推理的影响. MAGMA^[77] 通过提示让大语言模型显式地用自然语言复现图算法的具体执行过程,并比较了多个大语言模型在复现深度/广度优先搜索、迪杰斯特拉等多种图算法上的表现. PSEUDO^[78] 通过将任务对应图算法的伪代码作为上下文,让模型用语言逐行执行伪代码来解决图推理问题. 这些方法所面临的一个共性问题图结构过于复杂而难以被大语言模型记忆. 随着图的规模增大或变得更稠密,图结构会变得非常复杂,需要大量的令牌来描述. 然而,大语言模型在处理长文本时会出现明显的性能下降^[191],同时推理时间也会显著增加. 当单机器面临资源受

表 2 经典图推理方法对比

Table 2 Comparison of classical graph reasoning methods

方法	指令微调	推理路径	代码增强	图编码方法
NLGraph ^[75]	×	×	×	自然语言描述
GPT4Graph ^[41]	×	×	×	自然语言描述
LLM4DyG ^[49]	×	×	×	自然语言描述
Talk like a graph ^[76]	×	×	×	多种自然语言描述
GraphAgent-Reasoner ^[79]	×	√	×	自然语言描述
PSEUDO ^[78]	×	√	×	伪代码增强
GraphLLM ^[80]	√	×	×	GNN 编码 (graph token)
GraphToken ^[81]	√	×	×	GNN 编码 (soft token)
GraphWiz ^[82]	√	√	×	自然语言描述
GUNDAM ^[83]	√	√	×	自然语言描述
GITA ^[84]	√	×	×	多模态 (视觉+文本)
GraphTeam ^[85]	×	×	√	自然语言描述
GCoder ^[86]	√	×	√	自然语言描述
PIE ^[87]	×	×	√	自然语言描述
CodeGraph ^[88]	×	×	√	自然语言描述

限或任务过于复杂的问题时, 一个自然的想法是分布到多机器上解决. GraphAgent-Reasoner^[79] 提出使用多智能体来协同解决大规模的复杂图推理问题. 其 Master 会针对输入的图问题设计分布式图算法, 将复杂的推理任务分解为以节点为中心的简单子任务, 分配到每一个代理上解决. 其显著提高了模型在处理大规模图推理问题上的正确率, 但由于分布式算法的执行轮数与节点数正相关、大语言模型推理时间长等问题, 该方法往往消耗大量的推理时间和成本.

- 图编码方法. 注意到用自然语言描述图结构会导致上下文过长的问题, GraphLLM^[80] 和 Graph-Token^[81] 分别使用 Graph Transformer 和 GNN 来编码图结构. GraphLLM 以文本和位置编码作为 Graph Transformer 的输入, 将图的文本和结构信息表示为固定长度的前缀令牌, 通过前缀指令微调让大语言模型理解其中蕴含的图信息; Graph-Token 通过 GNN 将图结构编码为图令牌并尝试将图令牌和文本令牌对齐, 从而使大语言模型能够理解图结构. 然而 GraphToken 的图编码器是任务特定的, 即对不同的图任务需要专门训练专属的图编码器. 当将一个图任务的编码器应用于另一个任务上时, 模型会出现明显的性能下降, 减少了该方法的实用性.

- 图指令微调方法. GraphWiz^[82] 构建了首个具有显式推理路径的数据集 GraphInstruct. 随后使用监督微调和直接偏好优化对 Llama2 在该数据集上进行微调, 微调后的模型能够通过显式的推理路径解决图推理问题. GraphWiz 的一个缺陷是其在训练的九个图推理任务上严重过拟合. 面对未见的图推理任务或进行开放域提问时, 其会将训练问题的回答模板错误地套用在这些问题上. 同时, GraphWiz 借助 ChatGPT-4 来生成图推理路径, 其部分路径仍存在缺失、错误等问题. GUNDAM^[83] 通过严格执行图算法生成正确完整的硬推理路径来解决指令微调数据质量差的问题. 同时为防止大语言模型在硬推理路径上过拟合, GUNDAM 使用 GPT-4 对路径进行软化, 从而生成多样且正确的图推理路径. 此外, 该工作从信息论角度为图推理路径如何增强模型能力提供了理论支持. GraphInstruct^[192] 提出更全面的包含节点级别、节点对级别以及图级别共 21 个图任务的图推理数据集. 通过在该数据集上训练, 以最终答案和中间步骤作为监督信号进行微调得到了 GraphSolver 模型, 在三个层级的任务上较基准模型均有较大提升. 受到人类主要是通过视觉这一模态理解图的启发, GITA^[84]

使用视觉大模型来理解图结构并在语言层面进行图推理. 然而 GITA 难以处理大规模的图结构, 当节点和边的数目增加时, 由 GraphViz^[193]、Matplotlib^[194] 等生成的视觉图会变得非常复杂, 从而难以被视觉-语言大模型 (vision-language large models, VLLMs) 理解.

为得到正确的推理结果, 大语言模型需要重复相应图算法的推理路径. 然而, 图算法一般有大量的迭代和回溯操作, 这使得大模型难以在推理阶段完整并正确地输出推理路径. 在图规模增加时这一现象会更加严重. 以时间复杂度为 $O(N^2)$ 的图算法为例 (其中 N 为节点数目), 当图节点数目从 10 增加到 100 时, 所需要输出的推理步的数量级会从 100 增加到 10 000, 这对于大模型来说是不可接受的. 大多数基于语言的方法都面临着同样的问题, 导致较低的正确率和较差的扩展性.

4.1.2 代码增强的图推理

近年来, 意识到纯基于语言的方法在处理图推理问题方面的局限性, 一些工作引入代码来处理图推理中重复的迭代和回溯操作. 模型 $\mathcal{M}$ 会基于图问题生成一个程序 $\mathcal{C}$, 最终的答案将通过执行程序 (即 $\text{exec}(\mathcal{C})$) 得到:

$$\mathcal{M}(G, Q) \rightarrow \mathcal{C}, \text{exec}(\mathcal{C}) \rightarrow \mathcal{A}$$

CodeGraph^[88] 将少样本示例代码作为提示输入给大语言模型, 引导其通过编写代码来处理图问题, 解决了基于语言的模型在执行图算法过程中频繁发生的计算错误问题. 然而 CodeGraph 需要为每一个问题单独编写问题描述和示例代码, 限制了方法的通用性与实用性. GraphTeam^[85] 尝试使用多代理协作来解决图问题. 具体地, 该工作构建一个包含各种图相关 API 文档的知识库和一个会随着程序执行逐渐累积的经验库. 在推理过程中, 搜索代理会先从知识库和经验库中检索相关的 API 文档或执行经验, 然后将这些作为上下文输入给编码代理, 让代理能够编写代码来解决图问题. GCoder^[86] 通过监督微调和基于编译器反馈的强化学习来增强大语言模型利用图 API 的能力. 对于未在训练集中出现的图任务, GCoder 会从代码库中检索最相似的代码作为额外上下文来辅助代码生成. GCoder 存在的一个问题是它专注于学习如何调用不同的图算法 API 而非学习算法本身, 这导致模型并未真正掌握底层图算法逻辑, 其图推理能力的泛化性受限. 此外, 当在非训练集覆盖的图任务上测试 GCoder 时会出现明显的性能下降, 这可能是多任务微调的固定缺陷. PIE^[87] 为每个图推理任务种类收集伪代

码, 与问题描述一同作为上下文, 让大语言模型输出解决问题的可执行代码。

4.2 知识图谱推理任务

知识图谱以一系列三元组存储结构化知识, 一般写为 $KG = (h, r, t) \subseteq \mathcal{N} \times \mathcal{R} \times \mathcal{N}$, 其中 h 为头实体, r 为关系, t 为尾实体, $\mathcal{N}$ 代表实体集合, $\mathcal{R}$ 代表关系集合. 知识图谱推理的主要任务为知识图谱问答 (knowledge graph question answering, KGQA), 即给定知识图谱 KG 来推理询问的三元组 (h_q, r_q, t_q) . 具体地, 需要模型根据知识图谱上的实体关系推理 $(?, r_q, t_q)$, $(h_q, ?, t_q)$, $(h_q, r_q, ?)$ 的缺失部分. 传统的知识图谱推理方法分为基于编码的、基于路径的和基于规则的三种. 知识图谱是一种典型的文本属性图, 其节点和连边上通常有着丰富的语义信息. 观察到大语言模型强大的文本处理能力 & 广泛的通识知识储备, 其被应用于知识图谱推理任务中, 带来了显著的性能提升. 现有的大模型驱动知识图谱推理的工作可以分为大模型作为图表示增强器和图推理器两种.

4.2.1 大模型作为图表示增强器

语言模型通过在大量语料上进行预训练可以得到广泛的知识, 其能通过将序列文本 s_i 编码为文本嵌入 $h_i \in \mathbf{R}^d$ 来实现知识的注入. Lukovnikov 等^[89] 首先将预训练语言模型应用到关系预测中, 其使用 BERT 等大语言模型对实体及连边的文本进行编码, 计算连边属于各种类型的概率, 相较传统的浅层神经网络方法有明显的性能提升. ERNIE^[90] 在文本编码器之上还设计了一种知识编码器, 其将普通文本和实体信息对应融合, 再通过反向传播将知识信息注入到语言表示中. 此外, ERNIE 在预训练过程中还使用了遮罩来帮助模型更好地预测实体三元组的缺失部分. Luo 等^[91] 使用大语言模型来编码问题 q 和关系 r , 通过计算嵌入间的内积 $s(r, q) = \text{Emb}^T(r) \times \text{Emb}(q)$ 来衡量关系与问题的相关程度, 从而检索相关的三元组. 其中, $\text{Emb}(\cdot)$ 表示由大语言模型编码得到的文本嵌入向量. QA-GNN^[92] 的框架如图 12 所示, 其使用大语言模型编码问题, 备选答案对以

及知识图谱中的实体以增强特征且能够互相传递信息. 具体地, 其将问题编码作为一个单独节点与子图的话题实体节点连接, 然后应用图神经网络在联合图上推理, 传递实体信息以得到最终答案. Hu 等^[93] 的工作总结之前的方法, 设计两种基于大语言模型的 KGQA 框架——基于分类的知识图谱框架 (classification-based KGQA framework, KGQACL) 和基于检索与排序的知识图谱框架 (ranking-based KGQA framework, KGQARR). KGQACL 使用语言模型检测出问题中的实体和关系, 并基于问题分别给实体和关系打分为 S_e 和 S_r , 随后加权计算实体与关系组合的分数:

$$S(e, r) = \lambda S_e + (1 - \lambda) S_r$$

其中, λ 为根据验证集结果调整的加权参数. 模型将依次验证实体与关系在知识图谱中是否存在, 然后选择得分最高的实体或关系作为答案. KGQARR 不同于 KGQACL 之处在于其根据与问题的语义相似度来检索关系. 具体地, 其先检测出候选实体, 将候选实体所有相连的关系作为候选关系, 随后计算问题与候选关系之间的语义相似度:

$$S(q, r) = \cos(h_q, h_r)$$

其中, h_q, h_r 分别为语言模型编码的问题和关系表示. Edge 等^[111] 首次提出图检索增强生成 (graph retrieval-augmented generation, GraphRAG) 概念, 将知识图生成、检索增强生成和以查询为中心的查询聚焦摘要相结合, 以支持对整个文本语料库的意义理解. 具体地, 其首先从源文档中识别导出实体知识图, 然后将所有密切相关的实体组织成社区并生成社区摘要. 给定一个问题, 每个社区摘要用于生成部分响应, 然后将所有部分响应汇总以生成对用户的最终响应.

4.2.2 大模型作为图推理器

由于大语言模型具备强大的语言处理和逻辑推理能力, 很多工作将大语言模型作为推理器, 输入检索到的与问题相关的三元组来直接生成最终答案. DEKCOR^[94] 将问题、检索到的三元组和候选答案实体作为上下文输入给大语言模型, 模型将预测候选答案的分数. DRLK^[95] 使用与 DEKCOR 类似的上下文, 并使用大语言模型对上下文进行编码, 基于动态分层推理让上下文嵌入与答案之间进行信息交互, 然后预测答案. Yan 等^[96] 对于每一个候选答案实体 a , 首先从知识图谱中检索最相关的路径 $p_1, \dots, p_n$, 然后以问题 q 和相关路径 $p_1, \dots, p_n$ 作为上下文, 让大语言模型预测假设“ a is the answer of q ”是否成立来判断候选答案实体 a 是否作

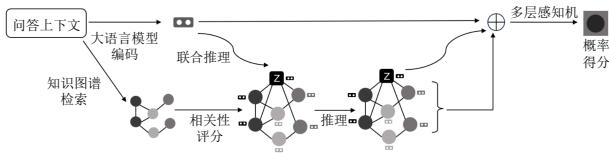


图 12 QA-GNN 的整体框架示意图

Fig. 12 Schematic diagram of the overall framework of QA-GNN

为最终答案. Oreolm^[97] 设计一种知识交互层 (knowledge interaction layer), 其可以与知识图谱推理模块互动, 从而发现不同的推理路径. 知识交互层穿插在大语言模型的 Transformer 层之间, 提供路径信息, 从而最终预测答案. GreaseLM^[98] 利用大语言模型处理文本和图神经网络处理图结构的能力, 将两个模态的信息融合并与单独模态的信息拼接在一起, 通过 MLP 预测答案. UniKGQA^[99] 合并了三元组检索和推理, 其首先根据问题检索语义相关的关系, 然后在关系上进行有向的相关信息传递以更新每个实体的分数, 最后选择分数最高的实体作为答案. ReLMKG^[100] 利用大语言模型对复杂问题和文本路径进行编码, 在不引入额外非结构化文本的情况下利用隐式知识. 大语言模型不同层的输出被用作指导信号, 引导图神经网络逐步执行消息传播和聚合, 充分利用结构化知识图谱中的显式知识. StructGPT^[101] 设计专门的接口, 让大语言模型可以根据问题自主推理, 并调用接口查看各种结构化信息, 模型可以迭代查询直到当前信息足够回答问题. G-Retriever^[102] 首先计算实体和关系与问题之间的相似度, 作为实体和关系的分数, 将子图检索问题转化为带权重的最小生成图问题, 调用 PC-ST 算法求解从而得到与问题相关的子图. 大语言模型以子图及问题作为上下文, 输出问题的答案. Six 等^[103] 在 G-Retriever 的基础上作出改进, 其首先使用推理大语言模型对问题进行分解, 形成若干顺序子问题, 然后依次对子问题执行 PCST 算法得到相关的子图, 最后将子图拼接起来作为上下文提供给大语言模型作答. RoG^[104] 利用大语言模型生成可能帮助解决问题的关系路径, 然后根据检索的关系从知识图谱中提取推理路径, 将其作为上下文让大语言模型输出答案. ToG^[3] 基于大语言模型迭代进行逐跳的检索和推理, 其框架如图 13 所示. 具体

地, ToG 从一系列话题节点出发进行束搜索 (beam search), 其会根据问题选择有价值的关系作为探索方向, 将其余关系剪枝. ToG 会持续搜索直到当前信息足够回答问题或达到搜索深度上限. ToG2.0^[107] 在 ToG 的基础上设计一种混合 RAG 框架, 将文档与知识图谱紧密耦合在一起, 解决了知识图谱本身信息不完整的问题. ToG2.0 一方面利用知识图谱作为导航地图, 通过实体间的关系引导文档检索以挖掘深层线索; 另一方面利用检索到的文档作为上下文来动态剪枝和优化知识图谱上的图搜索路径. Plan-on-Graph^[108] 的思想与 ToG 相似, 但额外引入了自我修正等机制, 其框架包含四个核心组件: 任务分解、路径探索、记忆更新和评估. 其中任务分解通过将问题分解为包含约束条件的子目标来指导自适应探索; 通过记忆模块记录子图、推理路径和子目标完成状态, 为反思提供动态证据; 最后大语言模型基于记忆内容判断是否需要自我修正推理路径, 并决定回溯到哪个实体重新探索. MindMap^[109] 是一种即插即用 (plug-and-play) 的 Prompting 框架以激发模型的图思维能力, 包含证据子图挖掘、证据聚合以及基于思维导图推理三个阶段. 具体地, MindMap 首先提取基于路径和邻居的证据子图, 然后引导模型合并构建推理子图并最终遍历得出答案. MindMap 通过生成可视化的思维导图揭示了推理路径, 极大地提升了复杂决策任务的可解释性与鲁棒性. GNN-RAG^[105] 分别利用图神经网络处理图结构和大语言模型处理文本的能力来进行知识图谱问答. 其利用图神经网络来检索可能的答案实体, 计算答案实体集与问题节点之间的最短路径形成子图, 然后让大语言模型对子图进行推理. SubgraphRAG^[106] 进一步将检索与推理解耦, 使用更轻量和高效率的 MLP 来进行检索, 通过消融实验证明了轻量检索的有效性. 具体地, 传统的图检索关注语义信息, SubgraphRAG 在此基础上额外考虑了结构信息——与话题实体间的距离, 由此检索到的三元组对正确答案的覆盖率相较传统图检索方法有显著提高, 进而提高了最终的问答表现. 受人类通过视觉直观理解图结构的启发, GraphVis^[110] 将检索子图通过 GraphViz^[193] 转化为视觉图, 输入给 VLLMs 进行视觉知识图谱推理. 为了让视觉模型更好理解图结构, GraphVis 进行视觉图理解微调 and 知识图谱问答微调两个阶段的课程学习. 实验显示 GraphVis 不仅在 KGQA 上表现良好, 在视觉问答任务上相较基线模型也有明显提升. PathMind^[112] 针对现有方法会检索无关路径引入噪声或需要迭代推理有较高计算成本的问题进行改进, 提出一种“检

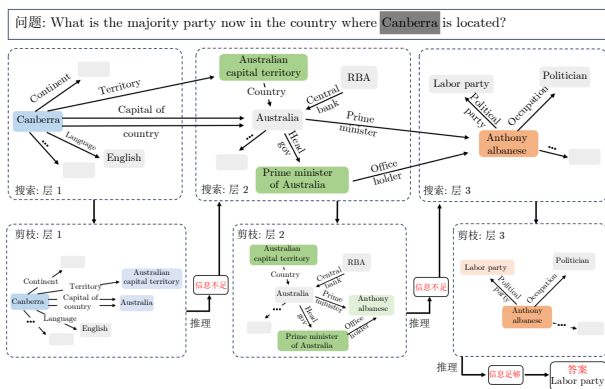


图 13 ToG 的流程示意图

Fig.13 ToG process diagram

索-优先路径-推理”的范式. 具体地, 其通过语义感知路径优先级机制, 结合问题到节点的累积成本和估计节点到答案的未来成本来识别重要节点, 从而构成推理路径, 并采用任务特定指令微调 and 路径偏好对齐双阶段训练, 来引导大语言模型更好地利用推理路径进行忠实和可解释的推理.

4.3 图推理模型应用

图推理模型聚焦于提升模型理解和处理图结构的能力, 是解决各种复杂现实图问题的核心能力. 本节探讨图推理模型在生物医学、推荐系统和金融风险三个领域的应用.

4.3.1 生物医学

GIMLET^[195] 通过给分子 token 添加广义位置编码让大语言模型可以感知分子间距离, 使得模型可以同时处理图结构和文本而无需引入新的模块; 同时其在注意力机制中解耦图结构与指令文本, 增强了图特征在未见任务上的泛化能力. 通过在海量分子任务上进行预训练, GIMLET 在分子属性预测等任务上表现出色, 贴近经过监督学习的图神经网络的性能. KGT^[196] 提供了首个生物医学的知识图谱推理基准 PcQA. 对于用户输入的自然语言问题, 其首先从查询问题中提取关键信息, 在知识图谱的模式图上进行图推理以获得最优关系路径, 基于最优关系进行检索, 并将得到的关系链反馈给大语言模型进行最终推理. OpenTCM^[197] 从《中华医典》数据库的 68 部妇科典籍中提取超过 373 万字, 构建了超 4.8 万实体、15.2 万关系的知识图谱. 通过在知识图谱上检索推理, OpenTCM 可以进行各种中医药信息和诊断问答.

4.3.2 推荐系统

尽管基于大语言模型的推荐系统取得了显著进展, 但相关模型仍然面临大语言模型固有的幻觉以及缺乏特定领域知识的问题. K-RagRec^[198] 从知识图谱中检索相关的高质量结构化信息来增强推荐. 具体地, K-RagRec 首先分别从粗粒度和细粒度层面对物品知识图谱构建子图索引, 形成知识向量数据库; 然后设计自适应检索策略, 根据物品流行度选取物品进行子图检索; 最后对检索到的知识子图进行重排序以优化检索质量, 实现知识增强推荐. COMPASS^[199] 首先通过图实体描述预训练对齐结构化知识图谱与自然语言, 然后通过知识感知的指令微调优化对用户的偏好推理, 使大语言模型学会从对话历史和知识图谱增强的上下文中推理并总结用户偏好, 提升了推荐性能和可解释性. CRAVE^[200] 引入辩论者、评论者等多智能体来进行协作图推理.

各智能体维护独立的经验库并通过开放式辩论实现对话式推荐, 最终从历史对话轨迹中检索与偏好相关的语言化经验来增强推荐.

4.3.3 金融风险

Pirmorad^[201] 探索了大语言模型在金融图谱反洗钱检测中的图推理能力. 其首先从金融知识图谱中提取目标实体周围的 k -hop 邻域子图并序列化为文本图结构表示, 然后将八种典型的洗钱拓扑结构作为上下文提供给大语言模型以评估可疑程度并生成推理依据. 实验结果表明大语言模型在结构化程度较高的模式如 fan-out (资金从单一账户分散至多个账户)、fan-in (多个账户资金汇聚至单一账户) 上表现出色, 反映了图推理模型在反洗钱领域的应用潜力. He 等^[202] 利用大语言模型分析金融文本数据, 从中自动化提取关键信息构建风险知识库, 实现风险因子的关联挖掘和动态演化分析. 实验结果表明该模型能够有效捕获产业链中的风险信号, 实现潜在金融风险的提前预警. Khanvilkar 等^[203] 提出一种面向银行业的实时交易监控与合规解释框架. 其首先构建动态交易图提取结构特征和上下文特征, 利用图神经网络对可疑行为进行分类检测. 针对每笔被标记的可疑交易, 模型检索相关监管条款并生成引用法规的自然语言解释, 从而为合规人员提供可审计的决策依据.

5 数据集与任务评测

5.1 图生成任务的数据集与评测方法

5.1.1 图生成数据集

在图生成领域, 针对不同的应用问题, 本文总结了常用的图数据集, 主要参考的代表性工作包括文献 [37, 204–206] 等, 如表 3 所示. 表中各列依次给出了数据集名称、节点数、边数、所属领域、节点类型、边类型以及代表性使用文献. 其中, 对于二分图数据集, 节点数以“用户数/物品数”的形式表示; “无”表示该数据集中未包含对应的边类型信息. 节点数和边数大于等于 1000 时, 分别以千 (K) 或百万 (M) 为单位表示.

5.1.2 图生成质量评价

在网络生成质量的评估方面, 现有方法主要分为三类指标: 一是基于统计特征的指标, 旨在评估生成网络在拓扑结构、连接模式等统计特性上与真实图的相似性; 二是基于神经网络的指标, 旨在利用神经网络捕捉生成图和参考图的相似度; 三是面向下游任务的指标, 关注生成网络是否能够有效提升下游任务 (如节点分类、链接预测或图分类) 的性

表 3 基于语言模型的图生成常用数据集
Table 3 Common datasets in graph generation models based on language models

数据集	节点数	边数	领域	节点类型	边类型	文献
WARRIOR ^[205]	100.0 K	285.0 K	社交	原始文本	原始文本	[205]
IMDB-text ^[204]	125.714 K	1.5 M	社交	原始文本	原始文本	[204]
Cora-text ^[10]	48.8 K	110.8 K	社交	原始文本	原始文本	[37, 204]
WeiboTech ^[204]	20.767 K	109.3 K	社交	原始文本	原始文本	[204, 206]
WeiboDaily ^[204]	66.5 K	354.1 K	社交	原始文本	原始文本	[204, 206]
Metoo ^[164]	1 K	32.0 K	社交	原始文本	无	[164, 207]
Roe ^[208]	1 K	121.5 K	社交	原始文本	原始文本	[164, 207]
MovieLens-1M ^[209]	6.0 K/3.9 K	32.0 M	推荐	原始文本	原始文本	[176, 178, 210–211]
Amazon review ^[212]	54.5 M/48.2 M	571.54 M	推荐	原始文本	原始文本	[178, 211, 213]
Steam ^[214]	2.6 M/15.5 K	7.793 M	推荐	原始文本	原始文本	[178, 210]
Lastfm ^[215]	1.9 K/17.6 K	266.0 K	推荐	原始文本	无	[216–218]

能, 从而反映其语义合理性和实用性.

● 基于统计特征的指标. 对于分布学习或随机生成, 基于统计特征的指标通过比较测试图与生成图之间的属性分布来衡量生成质量. 当前的评估主要取决于有关图形的先验知识 (即图形统计、属性), 比较生成图的各项属性与测试图的差距, 评估生成图是否具有良好的质量. 通用的图属性主要包括一个节点与整个网络中其他节点的连接数的概率分布 (度分布)、在单个网络内甚至跨多个网络重复出现的子图 (motif 计数), 以及图中节点聚集在一起的紧密程度 (聚类系数)、轨道记数等. 而针对不同的领域则存在图属性. 如社交网络图, 主要的属性包括网络的爪子图记数 (claw count)、图上的楔子图记数 (wedges count)、图的最大连通分量的大小 (largest connected component, LCC)、图的幂律分布的指数 (power-law exponent, PLE) 以及连通分量的数量 (N-component). 对于基于结构化的指标, 通常使用最大平均差异 (maximum mean discrepancy, MMD) 进行生成分布与测试分布的距离测试. 具体来说, 假定生成结果图集合为 $\mathcal{G}_{\text{gen}}$, 训练图集合为 $\mathcal{G}_{\text{train}}$, 测试图集合为 $\mathcal{G}_{\text{test}}$. 假设特征映射函数 $f(G)$, 确保其能够有效捕捉图的结构特性, $F_{\text{gen}} = \{f(G)|G \in \mathcal{G}_{\text{gen}}\}$, $F_{\text{test}} = \{f(G)|G \in \mathcal{G}_{\text{test}}\}$, 此时采用 MMD 计算 F_{gen} 与 F_{test} 的距离:

$$\begin{aligned} \text{MMD}(F_{\text{gen}}, F_{\text{test}}) = & \text{E}[k(F_{\text{gen}}, F_{\text{gen}})] - \\ & 2\text{E}[k(F_{\text{gen}}, F_{\text{test}})] + \\ & \text{E}[k(F_{\text{test}}, F_{\text{test}})] \end{aligned}$$

该值越小, 表示生成图与真实图的分布越接近. 其中, k 是核函数, 通常选择高斯核或多项式核.

● 基于神经网络的指标. 然而, 基于统计特征的评估指标存在三个关键局限: 1) 其输出通常为多

个独立指标, 而非单一标量评分, 导致模型选择困难; 2) 多数方法仅关注拓扑结构, 忽视了节点与边的属性特征, 难以全面反映图的语义信息; 3) 求解高阶统计量 (如子图计数、谱距离) 往往计算复杂度高, 执行效率低下. 为解决这些问题, 近期的研究提出使用基于神经网络的指标 (neural-based metric) 进行网络生成质量的评价. 未经训练的神经网络最近已用于静态图信息提取^[219–220]. 基于这些进展, 后续工作^[221–222] 提出使用此类随机网络作为静态图领域生成模型评估的可行方法, 它们证明了随机图神经网络可用作特征提取器, 用于嵌入合成图和真实图. 然后, 对提取的特征向量使用标准距离度量 (例如 Fréchet 距离^[223] 或 MMD) 比较这些嵌入, 以生成标量度量. 它们通过实证研究证明了这些基于神经网络的指标能够有效捕捉图的拓扑结构和特征, 从而得出统一的标量分数.

近期, 动态图逐渐成了研究热点. 对于动态图生成模型而言, 早期的动态图生成指标通常涉及将图的持续演化离散化为静态快照, 然后应用传统的图相似性度量. 这种方法存在一些局限性: 1) 它将时间相关的事件建模为独立同分布样本, 无法捕捉动态图的非均匀演化; 2) 缺乏对特征和拓扑结构都敏感的统一度量; 3) 无法提供标量指标, 需要多个指标, 且两者之间没有明显的优劣之分; 4) 需要显式实例化每个静态快照, 导致运行时需求不切实际, 阻碍了大规模评估. 随着动态图生成成为研究热点, 如何评估时序图模型的性能也受到广泛关注. 早期方法通常将动态图的演化过程离散化为一系列静态快照, 并在每个时间步上应用静态图相似性度量. 然而, 此类方法将时间相关的演化过程视为独立同分布样本, 忽略了动态图的非均匀与时序依赖特性. 为此, 后续工作^[224] 提出了 JL-metric, 一种面向

动态图生成模型的新型评估指标. 该方法利用动态图神经网络作为时序编码器, 直接捕捉节点与拓扑的联合演化模式, 从而在不显式分解为快照的前提下, 实现对生成动态图整体质量的端到端评估.

● 面向下游任务的指标. 除了针对生成图拓扑结构本身的评估指标外, 近年来部分研究从生成图的实际应用效能出发, 提出面向具体下游任务的评估范式. 这类方法不直接比较图的结构相似性, 而是通过衡量生成图在典型图学习任务中的实用性, 间接评估其质量. 主要应用场景包括节点分类和链接预测, 相关工作大致可分为两大流派: 1) 侧重于结构效用一致性评估. 旨在研究生成图是否“可替代”原始图, 关注生成图是否能在下游任务中复现原始图的训练效果. 代表性工作由文献 [225] 提出, 其引入一种基于判别模型的评估框架: 首先, 在原始图和生成图上分别训练链接预测与节点分类模型; 随后, 在原始图的测试集上统一评估这两个模型的性能, 得到两个任务的准确率或曲线下面积 (area under curve, AUC) 等指标. 若生成图训练的模型性能与原始图训练的模型性能之比接近 1, 则认为生成图具有与原始图相当的“任务效用”, 即其拓扑与特征信息足以支持相似的模型学习过程. 后续工作如文献 [153] 也采用了这一评估范式, 进一步验证了其在衡量生成图功能性质量方面的有效性. 该方法强调生成图与原始图在功能层面的一致性, 而非拓扑结构的逐点对齐, 能够反映生成图在真实任务中的可用性与泛化能力. 2) 将图生成任务视为自监督预训练机制, 旨在通过在大规模或跨域图数据上进行生成式训练, 提升模型在目标测试集上的下游任务性能. 此类方法通常面临两种迁移场景: 一是时间迁移, 预训练使用较早时间窗口的数据, 微调与测试则在后续时间段进行; 二是领域迁移, 预训练与下游任务来自不同领域 (如社交网络迁移学术网络), 评估模型的泛化能力^[31, 151]. 早期工作 VGAE^[226] 提出基于变分图自编码器的无监督图表示学习框架, 通过重构图结构进行生成式训练, 并在链接预测任务上验证其有效性, 奠定了生成式预训练的基础. 随着大语言模型与富文本图的发展, 近期研究进一步拓展了该范式. 例如, GPT-GNN 首先在预训练图数据集上训练一个 GNN 模型, 随后将预训练权重用于初始化下游任务模型, 并结合任务特定解码器在微调集上进行微调, 最终在测试集上评估节点分类与链接预测性能. 类似地, GraphMaster^[38] 也采用了“预训练-微调”范式, 验证生成式学习对下游任务的增益. 此外, LGGM^[151] 进一步探索了在测试图上同时评估生成质量与下游性能的

联合评价机制.

5.2 图预测任务的数据集与评测方法

在图预测任务 (如节点分类、链路预测、图分类) 中, 常用的数据集大体可分为两类: 传统图数据集与面向 LLM 的文本图数据集. 早期广泛使用的 Cora、Citeseer、Pubmed、ogbn-arxiv 等传统图数据集大多仅提供经过预处理的稀疏词袋 (bag-of-words, BoW), 缺乏原始文本内容, 因而难以充分发挥大语言模型对自然语言语义的建模能力. 随着 LLM 在图任务上的应用逐渐兴起, 研究者们开始专门收集与构建包含完整原始文本描述的文本属性图, 并进一步发展出同时为节点和边提供自然语言描述的文本边图 (textual-edge graphs, TEGs), 代表性工作包括 GLBench^[8] 与 TEG-DB^[18], 显著提升了 LLM 在图预测任务上的适用性和评测的一致性. 这些新型数据集涵盖学术网络、社交网络、电商推荐等多领域场景, 为研究图与语言模型的联合学习提供了统一而丰富的基准环境. 表 4 总结了当前图预测任务中常用的代表性数据集及其规模、领域与文本属性特点, 便于后续方法比较.

在图预测任务中, 主流研究通常涵盖节点分类、链接预测和图分类三类子任务. 基于大模型的图预测方法在评测上基本沿用了传统图神经网络方法的标准指标, 因此本文不再对相关计算公式进行详细推导, 而仅对常用指标作简要概述. 对于节点分类任务, 常见指标包括 Accuracy、Macro-F1 和 Micro-F1, 用于衡量模型在节点层面分类的准确性与类别均衡性. 对于链接预测任务, 传统上主要使用 AUC 和 AP (average precision) 等二分类指标来评估模型区分正负样本边的能力. 近年来, 由于链接预测与推荐任务的相关性不断增强, 研究者也开始借鉴推荐系统的排序类指标 MRR (mean reciprocal rank)、Hit Rate 和 Recall@ k , 用于衡量模型对候选边的排序质量. 对于图分类任务, 模型需要对整张图进行类别预测, 常用于分子性质预测、社交网络分类等场景, 评测时通常采用 Accuracy、Precision、Recall 和 F1 等分类指标, 用于衡量模型在图级别上的整体分类性能.

5.3 图推理任务的数据集与评测方法

在图推理领域, 本文分别总结了常用的知识图谱推理数据集和经典图推理数据集, 如表 5 及表 6 所示.

● 知识图谱推理. 知识图谱推理主要考察对结构化知识的理解和复杂语义推理能力, 如实体关系抽取、多跳推理、时间推理等. 数据集通常基于真实

表 4 常用图预测数据集统计信息
Table 4 Statistics of commonly used graph prediction datasets

数据集	节点数	边数	类别数	任务类型	领域	特征类型
Cora ^[227]	2.7 K	5.4 K	7	节点分类	学术	词袋 (BoW)
Citeseer ^[227]	3.3 K	4.7 K	6	节点分类	学术	词袋 (BoW)
Pubmed ^[227]	19.7 K	44.3 K	3	节点分类	学术	词袋 (BoW)
Coauthor-CS ^[227]	18.3 K	81.9 K	15	节点分类	学术	词袋 (BoW)
Coauthor-Physics ^[227]	34.5 K	248 K	5	节点分类	学术	词袋 (BoW)
Amazon-Photo ^[227]	7.5 K	119 K	10	节点分类	电商	词袋 (BoW)
Amazon-Computers ^[227]	13.4 K	245.8 K	8	节点分类	电商	词袋 (BoW)
ogbn-Product ^[6]	54 K	74.4 K	47	节点分类	电商	词袋 (BoW)
WikiCS ^[227]	11.7 K	216.1 K	10	节点分类	Wikipedia	GloVe
ogbn-Arxiv-TA ^[227]	169.3 K	1.2 M	40	节点分类	学术	原始文本
Books-Children ^[227]	76.9 K	1.6 M	24	节点分类	电商	原始文本
Books-History ^[227]	41.6 K	358.6 K	12	节点分类	电商	原始文本
Electronics-Computers ^[227]	87.2 K	721.1 K	10	节点分类	电商	原始文本
Electronics-Photo ^[227]	48.3 K	500.9 K	12	节点分类	电商	原始文本
Sports-Fitness ^[227]	173.1 K	1.8 M	13	节点分类	电商	原始文本
CitationV8 ^[227]	1.1 M	6.1 M	—	链接预测	学术	原始文本
GoodReads ^[227]	676.1 K	8.6 M	—	链接预测	图书	原始文本
MAG-Mathematics ^[48, 60]	19.9 K	34.7 K	—	链接预测	学术	原始文本
MAG-Geology ^[48, 60]	20.5 K	51.5 K	—	链接预测	学术	原始文本
Reddit ^[8]	33.4 K	198.4 K	2	节点分类	社交	原始文本
Instagram ^[8]	11.3 K	144 K	2	节点分类	社交	原始文本
Stackoverflow ^[58]	129.3 K	281.7 K	—	链接预测	技术社区	原始文本
Twitter ^[16]	176.3 K	2.4 M	2	节点分类, 链接预测	社交	原始文本

表 5 知识图谱推理数据集
Table 5 Knowledge graph reasoning datasets

数据集	知识图谱源	问题数	多跳	多实体	年份
WebQuestions ^[228]	Freebase	5 810	√	×	2013
SimpleQuestions ^[229]	Freebase	108 442	×	×	2015
ComplexQuestions ^[230]	Freebase	2 100	√	√	2016
GraphQuestions ^[231]	Freebase	5 166	√	√	2016
WebQuestionsSP ^[232]	Freebase	4 737	√	×	2016
The 30M Factoid QA ^[233]	Freebase	30 M	×	×	2016
SimpleQuestionsWikidata ^[234]	Wikidata	21 957	×	×	2017
LC-QuAD 1.0 ^[235]	DBpedia	5 000	√	√	2017
ComplexWebQuestions ^[236]	Freebase	34 689	√	√	2018
QALD-9 ^[237]	DBpedia	558	√	√	2018
PathQuestion ^[238]	Freebase	7 106	√	×	2018
MetaQA ^[239]	WikiMovies	407 513	√	×	2018
SimpleDBpediaQA ^[240]	DBpedia	43 086	×	×	2018
LC-QuAD 2.0 ^[241]	Wikidata	30 000	√	√	2019
FreebaseQA ^[242]	Freebase	28 348	×	×	2019
Event-QA ^[243]	EventKG	1 000	√	√	2020
GrailQA ^[244]	Freebase	64 331	√	√	2021

表 6 经典图推理基准数据集
Table 6 Classical graph reasoning benchmark datasets

基准	任务数量	节点数	测试样本数
NLGraph ^[75]	8	约 10^1	1000
Talk like a graph ^[76]	8	约 10^1	8000
GraphInstruct ^[192]	21	约 10^1	5100
GraphWiz ^[82]	9	约 10^2	3600
GraphArena ^[245]	10	约 10^1	10000

的知识库构建,如 Freebase、DBpedia、Wikidata 等,具有丰富的语义信息和复杂的关系结构.由于知识图谱的规模庞大和关系的多样性,知识图谱推理数据集的问题数量差异较大,从数百个到数十万个不等,涵盖简单事实查询到复杂的多步推理问题.表 5 总结了当前知识图谱推理常用的数据集,这些数据集为评估模型在真实知识场景下的推理能力提供了重要基准.其中多跳指回答一个问题需要跨越多个关系进行推理,即需要通过多步关系路径才能从问题中的已知实体到达答案实体;多实体指一个问题中涉及或提及多个实体节点,这些实体可能是问题的约束条件、比较对象或需要同时满足的条件.常见评测指标包括 Accuracy、Macro-F1、Micro-F1、Hit、Hit@1,用于衡量模型对多个候选答案选择的整体准确性、类别均衡性和排序质量.

● 经典图推理.经典图推理数据集主要考察对图结构及各种图算法问题的理解,如拓扑排序、最短路径、最大流算法等.数据集一般为人造数据集,首先基于 ER 算法生成随机图,然后运行相关图算法生成答案,套用模板自动化生成问题文本,构建相对简单.由于图结构的复杂性,经典图推理数据集的数据规模较小,往往不超过 100 个节点,图也较为稀疏.表 6 总结了当前经典图推理常用的数据集,覆盖简单的节点连边计数及较难的 MCP、TSP 等 NP 完全问题.由于图问题的输出多样,包含 yes/no、小数、列表、集合等多种形式,不同图问题在评测时需要单独设计评测函数,从模型的自然语言输出中通过正则表达式等方法提取格式输出,再按特定规则和答案进行比较.评测指标一般为正确率.

6 结束语

图生成、图预测与图推理作为图学习的三大核心任务,近年来在大模型驱动下取得了显著进展.然而,现有方法仍面临一系列共性挑战.主要包括对图结构与属性的低效表达、缺乏对生成或推理过程的细粒度可控性、评估体系不健全、可扩展性受限等问题,严重制约了通用图基础模型的发展.本文系统梳理图生成、图预测与图推理三个方向当前

面临的核心瓶颈,并分别提出未来发展的关键路径,旨在为构建可扩展、可信赖、可解释且任务通用的图智能系统提供方向指引.

6.1 图生成的未来方向

图生成模型在多个方向上仍具有重要发展潜力,包括提升模型可扩展性,丰富生成方法以及构建更鲁棒、通用的评估体系.

● 可扩展性.当前基于深度学习的图生成模型虽能捕捉复杂高阶依赖,但普遍存在超线性时间复杂度问题,难以扩展至大规模图,仅有少数方法达到线性复杂度^[27, 246].而基于仿真的模型虽可线性扩展至百万级节点,却依赖过度简化的假设,导致生成图拓扑失真.基于大语言模型仿真的生成方法因效率优势主导大规模生成,而深度学习方法在复现真实图统计特性上更具优势.未来需发展兼顾可扩展性与表达能力的混合框架.

● 生成方法匮乏.现有模型的生成控制能力仍局限于基础图级统计量.然而,在基于文本的图生成任务中,社交网络生成需对齐用户行为模式^[34, 167].当前此类方法在数量与表达力上均明显不足,亟须引入可编码领域先验知识的控制接口.此外,针对节点和边属性的语义与功能性控制,现有研究亦存在显著缺口.

● 评估体系.与图像和语言生成类似,图生成缺乏唯一真值,评估本身具有内在困难.基于神经网络的指标难以扩展,基于统计的指标则忽略重要结构特性^[221],许多场景还需昂贵或领域定制的验证手段^[247].因此,亟须建立更高效、鲁棒且标准化的评估协议.

6.2 图预测的未来方向

LLM 参与图预测的发展需要在表示形式、推理机制与系统设计层面进行更深层次的重构:既要解决图到语言/token 表达与不确定性建模上的基础瓶颈,又要系统性地提升在开放环境下的鲁棒性与安全性,并从更广义的“多角色协同”视角重新审视 LLM 在图学习流程中的位置与功能.

● 结构表达与置信度建模.当前在 LLM 参与图预测时仍面临两方面挑战:一方面,图到语言/token 表达要么丢失关键结构信息,要么过于冗长,导致在大图场景下推理成本高、稳定性差;另一方面,LLM 的输出缺乏良好的不确定性估计与概率校准,难以直接支持下游决策.未来有必要设计更加结构感知、压缩友好的图到语言/token 表达方式,以及分层、可分解的推理机制,使 LLM 在大规模图上参与决策时既能聚焦关键结构模式,又不被冗余

邻域与过长上下文拖累。同时, 还需要引入可靠的不确定性估计与概率校准手段, LLM 在图任务中不再只是“黑盒打分器”, 而是能够输出可信度可控、可直接用于下游决策的预测结果。

● 鲁棒性与安全性. 现有 LLM-GNN 协同框架在面对恶意扰动和数据污染时仍然缺乏系统性的鲁棒性设计: LLM 侧容易受到幻觉、偏见和对提示词的攻击性操控, 轻微的文本改写、诱导性指令甚至有意构造的外部知识即可显著改变其输出; 图模型则长期暴露在对抗扰动、投毒与结构噪声之下, 使得整个协同系统形成更大的联合攻击面。未来需要从统一视角建模 LLM 与图结构的联合威胁与防御: 在 LLM 侧引入针对 prompt 与外部知识的鲁棒融合和约束机制^[248-249], 在图侧结合对抗训练、结构去噪与异常检测等技术^[250], 并在系统层面构建端到端的鲁棒性评估与防护框架, 从而使 LLM-GNN 协同系统在对抗环境和分布偏移下依然保持可预期、可依赖的行为。

● 角色扩展. 除了前述三类主流范式之外, 未来还可以从多种角色维度进一步挖掘 LLM 在图预测任务中的作用: LLM 作为控制器 (controller) 动态控制 GNN 的邻域采样策略、聚合方式与算子选择, 甚至被纳入 GNN 神经架构搜索 (graph neural architecture search, GNAS) 框架, 由 LLM 参与结构和算子的自动设计与筛选; LLM 作为数据设计器 (data designer) 自动设计对比样本与学习路径, 或从规则文档中抽取约束并转化为损失项, 近期如 DMA^[251] 将 LLM 视为有噪声的打标预言机 (noisy oracle), 结合图结构设计查询策略, 本质上由 LLM 与 GNN 共同决定“哪些节点先学”; LLM 作为编辑器 (editor) 从文本和日志中构建图、补边与修边; LLM 作为评论者 (critic) 对 GNN 预测进行打分、纠错并形成偏好或一致性信号; LLM 作为多子图、多代理之间的协调者 (coordinator), 通过自然语言在局部 GNN 之间传递高层规划与推理依据。将这些潜在角色及其已有的初步探索与上述三类主流范式统一到系统的 LLM-GNN 协同框架中, 有望为后续方法设计提供更全面的视角与新的切入点。

6.3 图推理的未来方向

大模型在处理图推理任务方面仍处于初步探索阶段, 在模型及数据集方面均有需改进的地方, 包括提升模型记忆和处理图的规模、增强对图结构的理解能力以及构建更具多样性和挑战性的基准等。

● 可扩展性. 当前大模型处理图的规模有限, 往往只支持 20 ~ 30 个节点的稀疏小图且正确率较

低^[75]。这源于大语言模型在处理长上下文时会出现明显的性能下降, 而以自然语言形式描述图结构会耗费极长的上下文, 即使最先进的大语言模型如 GPT-5、Claude4 等也无法准确记忆 50 个节点的较稠密图结构。此外, 现有方法解决图推理问题的方式主要是通过自然语言重复图算法的执行过程来获得答案, 对于算法复杂度较高的图推理问题需要生成极长的推理路径, 导致推理速度慢、易在中间步骤出错而最终错误等问题。未来需要设计高效的图编码方式, 将重复迭代的操作交给程序运行, 让模型专注于图推理和算法设计本身。

● 图结构理解. 现有研究表明大语言模型在处理语言图结构表示时并未将其理解为图结构而是理解为序列文本^[252], 去掉图结构信息并不会明显影响模型的表现, 这些都暗示大语言模型并未隐式构建图结构, 也没有真正理解图。GITA^[84] 探究以二维视觉图的直观形式表示图结构, 这贴合人理解图结构的方式, 但更大规模图的展示以及非平面图的处理是待解决的问题。未来需要设计模型或进行更多图文的训练以增强模型对图结构的理解能力。

● 评估体系. 目前经典图推理基准 (如表 6 所示) 中的问题多为已有成熟算法的固定图任务, 相关算法也可能已出现在大模型的预训练语料中。因此, 模型在测试中可能只是复现既有算法的执行过程, 难以真正评估其图算法设计能力。此外, 现有基准训练集和测试集的图任务是同一类型, 模型可能只学到了固定算法的固定模式, 并不能真正泛化。现有知识图谱推理数据集 (如表 5 所示) 的跳数仍较少 (几乎在 4 跳以内), 对于长路径推理及长程依赖捕获能力的考察不足。此外, 知识图谱内的推理均局限于常识性的推理, 如简单的从属、并列、共有关系等, 缺少对局部图结构的考察。同时每对知识图谱问答往往只有单独的答案实体而缺少推理路径, 难以评估模型的推理过程是否正确, 也无法提供细粒度的指导。因此, 构建具有挑战性和预训练中未涉及的图推理问题来考察模型的图推理和算法设计能力, 是未来需要努力的方向。

参考文献

- 1 Song X R, Lian J X, Huang H, Wu M Q, Jin H, Xie X. Friend recommendations with self-rescaling graph neural networks. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, USA: Association for Computing Machinery, 2022. 3909-3919
- 2 Xie Q Q, Zhu Y T, Huang J M, Du P, Nie J Y. Graph neural collaborative topic model for citation recommendation. *ACM Transactions on Information Systems (TOIS)*, 2022, 40(3): Article No. 48
- 3 Sun J S, Xu C J, Tang L M Y, Wang S Z, Lin C, Gong Y Y, et al. Think-on-graph: Deep and responsible reasoning of large

- language model on knowledge graph. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
- 4 Réau M, Renaud N, Xue L C, Bonvin A M J J. DeepRank-GNN: A graph neural network framework to learn patterns in protein-protein interfaces. *Bioinformatics*, 2023, **39**(1): Article No. btac759
 - 5 Zhang S, Tong H H, Xu J J, Maciejewski R. Graph convolutional networks: Algorithms, applications and open challenges. In: Proceedings of the 7th International Conference on Computational Data and Social Networks. Shanghai, China: Springer, 2018. 79–91
 - 6 Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: OpenReview.net, 2018.
 - 7 Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 1025–1035
 - 8 Li Y H, Wang P S, Zhu X, Chen A C, Jiang H Y, Cai D, et al. GLBench: A comprehensive benchmark for graph with large language models. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 1340
 - 9 Hu W H, Fey M, Zitnik M, Dong Y X, Ren H Y, Liu B W, et al. Open graph benchmark: Datasets for machine learning on graphs. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2020. Article No. 1855
 - 10 Sen P, Namata G, Bilgic M, Getoor L, Gallagher B, Eliassirad T. Collective classification in network data. *AI Magazine*, 2008, **29**(3): 93–106
 - 11 Miaschi A, Dell’Orletta F. Contextual and non-contextual word embeddings: An in-depth linguistic investigation. In: Proceedings of the 5th Workshop on Representation Learning for NLP. Virtual Event: Association for Computational Linguistics, 2020. 110–119
 - 12 Ethayarajh K. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 55–65
 - 13 Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv: 2307.09288, 2023.
 - 14 Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, et al. PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 2023, **24**(1): Article No. 240
 - 15 Li Y G, Liang F, Zhao L C, Cui Y F, Ouyang W L, Shao J, et al. Supervision exists everywhere: A data efficient contrastive language-image pre-training paradigm. In: Proceedings of the 10th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2022.
 - 16 Jin B W, Liu G, Han C, Jiang M, Ji H, Han J W. Large language models on graphs: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*, 2024, **36**(12): 8622–8642
 - 17 Chen Z K, Mao H T, Li H, Jin W, Wen H Z, Wei X C, et al. Exploring the potential of large language models (LLMs) in learning on graphs. *ACM SIGKDD Explorations Newsletter*, 2023, **25**(2): 42–61
 - 18 Li Z F, Gou Z X, Zhang X N, Liu Z Y, Li S R, Hu Y T, et al. TEG-DB: A comprehensive dataset and benchmark of textual-edge graphs. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 1949
 - 19 You J X, Ying R, Ren X, Hamilton W L, Leskovec J. GraphRNN: Generating realistic graphs with deep auto-regressive models. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018. 5708–5717
 - 20 Samanta B, De A, Jana G, Chattaraj P K, Ganguly N, Rodriguez M G. NeVAE: A deep generative model for molecular graphs. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI, 2019. 1110–1117
 - 21 Luo Y Z, Yan K Q, Ji S W. GraphDF: A discrete flow model for molecular graph generation. In: Proceedings of the 38th International Conference on Machine Learning. Virtual Event: PMLR, 2021. 7192–7203
 - 22 Bojchevski A, Shchur O, Zügner D, Günnemann S. NetGAN: Generating graphs via random walks. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018. 610–619
 - 23 Jo J, Lee S, Hwang S J. Score-based generative modeling of graphs via the system of stochastic differential equations. In: Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA: PMLR, 2022. 10362–10383
 - 24 Zhang M Y, Cai Z A, Pan L, Hong F Z, Guo X Y, Yang L, et al. MotionDiffuse: Text-driven human motion generation with diffusion model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, **46**(6): 4115–4128
 - 25 Huang H, Sun L L, Du B W, Lv W F. Conditional diffusion based on discrete graph structures for molecular graph generation. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI, 2023. 4302–4311
 - 26 Jung H, Park Y, Schmid L, Jo J, Lee D, Kim B, et al. Conditional synthesis of 3D molecules with time correction sampler. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 2417
 - 27 Dai H J, Nazi A, Li Y J, Dai B, Schuurmans D. Scalable deep generative modeling for sparse graphs. In: Proceedings of the 37th International Conference on Machine Learning. Virtual Event: PMLR, 2020. 2302–2312
 - 28 Yu J X, Ren Y X, Gong C H, Tan J Q, Li X, Zhang X C. Leveraging large language models for node generation in few-shot learning on text-attributed graphs. In: Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI, 2025. 13087–13095
 - 29 Liu B, Qi G L. LLM-CG: Large language model-enhanced constraint graph for distantly supervised relation extraction. *Neurocomputing*, 2025, **655**: Article No. 131426
 - 30 Zhang Z J, Wang X, Zhou H C, Yu Y, Zhang M M, Yang C, et al. Can large language models improve the adversarial robustness of graph neural networks? In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Toronto, Canada: Association for Computing Machinery, 2025. 2008–2019
 - 31 Hu Z N, Dong Y X, Wang K S, Chang K W, Sun Y Z. GPT-GNN: Generative pre-training of graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event: Association for Computing Machinery, 2020. 1857–1867
 - 32 de Marzo G, Pietronero L, Garcia D. Emergence of scale-free networks in social interactions among large language models. arXiv preprint arXiv: 2312.06619, 2023.
 - 33 Papachristou M, Yuan Y. Network formation and dynamics among multi-LLMs. arXiv preprint arXiv: 2402.10659, 2025.
 - 34 Yao Y, Wang X, Zhang Z Y, Qin Y J, Zhang Z W, Chu X, et al. Exploring the potential of large language models in graph gen-

- eration. arXiv preprint arXiv: 2403.14358, 2024.
- 35 Havrilla A, Alvarez-Melis D, Fusi N. IGDA: Interactive graph discovery through large language model agents. arXiv preprint arXiv: 2502.17189, 2025.
 - 36 Chang S, Chaszczewicz A, Wang E, Josifovska M, Pierson E, Leskovec J. LLMs generate structurally realistic social networks but overestimate political homophily. In: Proceedings of the 19th International AAAI Conference on Web and Social Media. Copenhagen, Denmark: AAAI, 2025. 341–371
 - 37 Ji J R, Lei R L, Bi J L, Wei Z W, Chen X, Lin Y K, et al. LLM-based multi-agent systems are scalable graph generative models. In: Proceedings of the Findings of the Association for Computational Linguistics. Vienna, Austria: Association for Computational Linguistics, 2025. 1492–1523
 - 38 Du E J, Li X K, Jin T, Zhang Z H, Li R H, Wang G R. GraphMaster: Automated graph synthesis via LLM agents in data limited environments. In: Proceedings of the 39th Conference on Neural Information Processing Systems. Mexico City, Mexico: Curran Associates Inc., 2025.
 - 39 Wang Y, Rossi R A, Park N, Ahmed N K, Koutra D, Dernoncourt F, et al. Demystifying the power of large language models in graph generation. In: Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025. New Mexico, USA: Association for Computational Linguistics, 2025. 8189–8204
 - 40 Ye R S, Zhang C Q, Wang R H, Xu S Y, Zhang Y F. Language is all a graph needs. In: Proceedings of the Findings of the Association for Computational Linguistics: EACL 2024. St. Julian's, Malta: Association for Computational Linguistics, 2024. 1955–1973
 - 41 Guo J Y, Du L, Liu H Y, Zhou M Y, He X Y, Han S. GPT4Graph: Can large language models understand graph structured data? An empirical evaluation and benchmarking. arXiv preprint arXiv: 2305.15066, 2023.
 - 42 Zhao J N, Zhuo L, Shen Y K, Qu M, Liu K, Bronstein M, et al. GraphText: Graph reasoning in text space. arXiv preprint arXiv: 2310.01089, 2023.
 - 43 Lin T Q J, Yan P W, Song K S, Jiang Z R, Kang Y Y, Lin J, et al. LangGFM: A large language model alone can be a powerful graph foundation model. arXiv preprint arXiv: 2410.14961, 2024.
 - 44 Tang J B, Yang Y H, Wei W, Shi L, Su L X, Cheng S Q, et al. GraphGPT: Graph instruction tuning for large language models. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington, USA: Association for Computing Machinery, 2024. 491–500
 - 45 Chen R J, Zhao T, Jaiswal A K, Shah N, Wang Z Y. LLaGA: Large language and graph assistant. In: Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: PMLR, 2024. 7809–7823
 - 46 Kong L C, Feng J R, Liu H, Huang C S, Huang J X, Chen Y X, et al. GOFA: A generative one-for-all model for joint graph language modeling. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.
 - 47 Wang D, Zuo Y, Li F Z, Wu J J. LLMs as zero-shot graph learners: Alignment of GNN representations with LLM token embeddings. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 193
 - 48 He Z M, Zhu J, Qian S Y, Chai J, Koutra D. LinkGPT: Teaching large language models to predict missing links. arXiv preprint arXiv: 2406.04640, 2024.
 - 49 Zhang Z Y, Wang X, Zhang Z W, Li H Y, Qin Y J, Zhu W W. LLM4DyG: Can large language models solve spatial temporal problems on dynamic graphs? In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: Association for Computing Machinery, 2024. 4350–4361
 - 50 Lei R L, Ji J R, Ding H P, Yi L, Wei Z W, Liu Y C, et al. Exploring the potential of large language models as predictors in dynamic text-attributed graphs. arXiv preprint arXiv: 2503.03258, 2025.
 - 51 Tang J B, Yang Y H, Wei W, Shi L, Xia L, Yin D W, et al. HiGPT: Heterogeneous graph language model. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: Association for Computing Machinery, 2024. 2842–2853
 - 52 Chien E, Chang W C, Hsieh C J, Yu H F, Zhang J, Milenkovic O, et al. Node feature extraction by self-supervised multi-scale neighborhood prediction. In: Proceedings of the 10th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2022.
 - 53 Duan K Y, Liu Q, Chua T S, Yan S C, Ooi W T, Xie Q Z, et al. SimTeG: A frustratingly simple approach improves textual graph learning. arXiv preprint arXiv: 2308.02565, 2023.
 - 54 Gao H, Huang W X, Wu F G, Zhao J S, Zheng C W, Liu H P. LLM enhancers for GNNs: An analysis from the perspective of causal mechanism identification. In: Proceedings of the 42nd International Conference on Machine Learning. Vancouver, Canada: PMLR, 2025. 18585–18611
 - 55 Liu H, Feng J R, Kong L C, Liang N Y, Tao D C, Chen Y X, et al. One for all: Towards training one graph model for all classification tasks. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
 - 56 Yang J H, Liu Z, Xiao S T, Li C Z, Lian D F, Agrawal S, et al. GraphFormers: GNN-nested Transformers for representation learning on textual graph. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event: Curran Associates Inc., 2021. Article No. 2206
 - 57 Jin B W, Zhang Y, Zhu Q, Han J W. Heterformer: Transformer-based deep node representation learning on heterogeneous text-rich networks. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach, USA: Association for Computing Machinery, 2023. 1020–1031
 - 58 Jin B W, Zhang Y, Meng Y, Han J W. Edgeformers: Graph-empowered Transformers for representation learning on textual-edge networks. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
 - 59 Fang Y, Fan D Z, Ding S R, Liu N H, Tan Q Y. UniGLM: Training one unified language model for text-attributed graphs embedding. In: Proceedings of the 18th ACM International Conference on Web Search and Data Mining. Hannover, Germany: Association for Computing Machinery, 2025. 973–981
 - 60 Jin B W, Zhang W T, Zhang Y, Meng Y, Zhang X Y, Zhu Q, et al. PATTON: Language model pretraining on text-rich networks. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada: Association for Computational Linguistics, 2023. 7005–7020
 - 61 Li Y H, Wang P S, Li Z X, Yu J X, Li J. ZeroG: Investigating cross-dataset zero-shot transferability in graphs. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: Association for Computing Machinery, 2024. 1725–1735
 - 62 Xie H Y, Hu L J, Yu L, Huang T H, Li L F, Li M, et al. Efficient text-attributed graph learning through selective annotation and graph alignment. arXiv preprint arXiv: 2506.07168, 2025.
 - 63 Huang X W, Han K Q, Yang Y, Bao D Z, Tao Q J, Chai Z W, et al. Can GNN be good adapter for LLMs? In: Proceedings of the ACM Web Conference. Singapore: Association for Computing Machinery, 2024. 893–904

- 64 Lu J L, Liu Y X, Zhang Y T, Fu Y. Scale-free graph-language models. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.
- 65 Xia L H, Kao B, Huang C. OpenGraph: Towards open graph foundation models. In: Proceedings of the Findings of the Association for Computational Linguistics. Miami, USA: Association for Computational Linguistics, 2024. 2365–2379
- 66 Zhu Y, Wang Y K, Shi H Z, Tang S L. Efficient tuning and inference for large language models on textual graphs. In: Proceedings of the 33rd International Joint Conference on Artificial Intelligence. Jeju, South Korea: ijcai.org, 2024. Article No. 634
- 67 He X X, Bresson X, Laurent T, Perold A, LeCun Y, Hooi B. Harnessing explanations: LLM-to-LM interpreter for enhanced text-attributed graph representation learning. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
- 68 Pan B, Zhang Z, Zhang Y F, Hu Y T, Zhao L. Distilling large language models for text-attributed graph learning. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. Boise, USA: Association for Computing Machinery, 2024. 1836–1845
- 69 Zhang Z H, Li X K, Li R H, Li Z J, Zhou B, Wang G R. Toward general and robust LLM-enhanced text-attributed graph learning. In: Proceedings of the International Conference on Multimedia Retrieval. Amsterdam, The Netherlands: Association for Computing Machinery, 2026. 1477–1485
- 70 Zhao J N, Qu M, Li C Z, Yan H, Liu Q, Li R, et al. Learning on large-scale text-attributed graphs via variational inference. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
- 71 Brannon W, Kang W N J N, Fulay S, Jiang H, Roy B, Roy D, et al. ConGraT: Self-supervised contrastive pre training for joint graph and text embeddings. In: Proceedings of the Text-Graphs-17: Graph-based Methods for Natural Language Processing. Bangkok, Thailand: Association for Computational Linguistics, 2024. 19–39
- 72 Wen Z H, Fang Y. Augmenting low-resource text classification with graph-grounded pre-training and prompting. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei, China: Association for Computing Machinery, 2023. 506–516
- 73 Jiang X H, Shen Y H, Shi Z C, Xu C J, Li W, Li Z X, et al. Unlocking the power of large language models for entity alignment. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, Thailand: Association for Computational Linguistics, 2024. 7566–7583
- 74 Roy A, Yan N, Mortazavi M. LLM-driven knowledge distillation for dynamic text-attributed graphs. arXiv preprint arXiv: 2502.10914, 2025.
- 75 Wang H, Feng S B, He T X, Tan Z X, Han X C, Tsvetkov Y. Can language models solve graph problems in natural language? In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 1345
- 76 Fatemi B, Halcrow J, Perozzi B. Talk like a graph: Encoding graphs for large language models. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
- 77 Taylor A K, Cuturrufo A, Yathish V, Ma M D, Wang W. Are large-language models graph algorithmic reasoners? arXiv preprint arXiv: 2410.22597, 2024.
- 78 Skianis K, Nikolentzos G, Vazirgiannis M. Graph reasoning with large language models via pseudo-code prompting. arXiv preprint arXiv: 2409.17906, 2024.
- 79 Hu Y W, Lei R L, Huang X Y, Wei Z W, Liu Y C. Scalable and accurate graph reasoning with LLM-based multi-agents. arXiv preprint arXiv: 2410.05130, 2025.
- 80 Chai Z W, Zhang T J, Wu L, Han K Q, Hu X H, Huang X W, et al. GraphLLM: Boosting graph reasoning ability of large language model. *IEEE Transactions on Big Data*, 2026, **12**(2): 475–483
- 81 Perozzi B, Fatemi B, Zelle D, Tsitsulin A, Kazemi M, Al-Rfou R, et al. Let your graph do the talking: Encoding structured data for LLMs. arXiv preprint arXiv: 2402.05862, 2024.
- 82 Chen N, Li Y H, Tang J H, Li J. GraphWiz: An instruction-following language model for graph computational problems. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: Association for Computing Machinery, 2024. 353–364
- 83 Ouyang S, Hu Y L, Chen G, Liu Y. GUNDAM: Aligning large language models with graph understanding. arXiv preprint arXiv: 2409.20053, 2024.
- 84 Wei Y B, Fu S, Jiang W S, Zhang Z J, Zeng Z X, Wu Q, et al. GITA: Graph to visual and textual integration for vision-language graph reasoning. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 2
- 85 Li X, Chu Q Z, Chen Y B, Liu Y, Liu Y Q, Yu Z K, et al. GraphTeam: Facilitating large language model-based graph analysis via multi-agent collaboration. arXiv preprint arXiv: 2410.18032, 2025.
- 86 Zhang Q F, Hong X B, Tang J H, Chen N, Li Y H, Li W Z, et al. GCoder: Improving large language model for generalized graph problem solving. arXiv preprint arXiv: 2410.19084, 2024.
- 87 Gong C, Bian W R, Zhang Z J, Zheng W G. Pseudocode-injection magic: Enabling LLMs to tackle graph computational tasks. arXiv preprint arXiv: 2501.13731, 2025.
- 88 Cai Q L, Wang Z W, Diao S Z, Kwok J, Song Y Q. CodeGraph: Enhancing graph reasoning of LLMs with code. arXiv preprint arXiv: 2408.13863, 2024.
- 89 Lukovnikov D, Fischer A, Lehmann J. Pretrained Transformers for simple question answering over knowledge graphs. In: Proceedings of the 18th International Semantic Web Conference on the Semantic Web. Auckland, New Zealand: Springer, 2019. 470–486
- 90 Zhang Z Y, Han X, Liu Z Y, Jiang X, Sun M S, Liu Q. ERNIE: Enhanced language representation with informative entities. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 1441–1451
- 91 Luo D, Su J D, Yu S S. A BERT-based approach with relation-aware attention for knowledge base question answering. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). Glasgow, UK: IEEE, 2020. 1–8
- 92 Yasunaga M, Ren H Y, Bosselut A, Liang P, Leskovec J. QAGNN: Reasoning with language models and knowledge graphs for question answering. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Virtual Event: Association for Computational Linguistics, 2021. 535–546
- 93 Hu N, Wu Y K, Qi G L, Min D H, Chen J Y, Pan J Z, et al. An empirical study of pre-trained language models in simple knowledge graph question answering. *World Wide Web*, 2023, **26**(5): 2855–2886
- 94 Xu Y C, Zhu C G, Xu R C, Liu Y, Zeng M, Huang X D. Fusing context into knowledge graph for commonsense question answering. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Virtual Event: Association for Computational Linguistics, 2021. 1201–1207
- 95 Zhang M, Dai R F, Dong M, He T T. DRLK: Dynamic hierarchical reasoning with language model and knowledge graph for question answering. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Lin-

- guistics, 2022. 5123–5133
- 96 Yan Y M, Li R M, Wang S R, Zhang H Z, Zan D G, Zhang F Z, et al. Large-scale relation learning for question answering over knowledge bases with pre-trained language models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. 3653–3660
 - 97 Hu Z N, Xu Y C, Yu W H, Wang S H, Yang Z Y, Zhu C G, et al. Empowering language models with knowledge graph reasoning for open-domain question answering. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. 9562–9581
 - 98 Zhang X K, Bosselut A, Yasunaga M, Ren H Y, Liang P, Manning C D, et al. GreaseLM: Graph REASONing enhanced language models for question answering. In: Proceedings of the 10th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2022.
 - 99 Jiang J H, Zhou K, Zhao X, Wen J R. UniKGQA: Unified retrieval and reasoning for solving multi-hop question answering over knowledge graph. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
 - 100 Cao X, Liu Y. ReLMKG: Reasoning with pre-trained language models and knowledge graphs for complex question answering. *Applied Intelligence*, 2023, **53**(10): 12032–12046
 - 101 Jiang J H, Zhou K, Dong Z C, Ye K M, Zhao X, Wen J R. StructGPT: A general framework for large language model to reason over structured data. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2023. 9237–9251
 - 102 He X X, Tian Y J, Sun Y F, Chawla N V, Laurent T, LeCun Y, et al. G-Retriever: Retrieval-augmented generation for textual graph understanding and question answering. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 4224
 - 103 Six V, Dufraisse E, de Chalendar G. Compositional reasoning for graph retrieval with large language models. arXiv preprint arXiv: 2506.13380, 2025.
 - 104 Luo L H, Li Y F, Haffari G, Pan S R. Reasoning on graphs: Faithful and interpretable large language model reasoning. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
 - 105 Mavromatis C, Karypis G. GNN-RAG: Graph neural retrieval for efficient large language model reasoning on knowledge graphs. In: Proceedings of the Findings of the Association for Computational Linguistics. Vienna, Austria: Association for Computational Linguistics, 2025. 16682–16699
 - 106 Li M F, Miao S Q, Li P. Simple is effective: The roles of graphs and large language models in knowledge-graph-based retrieval-augmented generation. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.
 - 107 Ma S J, Xu C J, Jiang X H, Li M Z, Qu H R, Yang C H, et al. Think-on-graph 2.0: Deep and faithful large language model reasoning with knowledge-guided retrieval augmented generation. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.
 - 108 Chen L Y, Tong P R, Jin Z M, Sun Y, Ye J P, Xiong H. Plan-on-Graph: Self-correcting adaptive planning of large language model on knowledge graphs. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 1189
 - 109 Wen Y L, Wang Z F, Sun J M. MindMap: Knowledge graph prompting sparks graph of thoughts in large language models. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, Thailand: Association for Computational Linguistics, 2024. 10370–10388
 - 110 Deng Y H, Ye C C, Huang Z J, Ma M D, Kou Y W, Wang W. GraphVis: Boosting LLMs with visual knowledge graph integration. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 2155
 - 111 Edge D, Trinh H, Cheng N, Bradley J, Chao A, Mody A, et al. From local to global: A graph RAG approach to query-focused summarization. arXiv preprint arXiv: 2404.16130, 2025.
 - 112 Liu Y, Lin X X, Shang Y M, Li Y X, Wang S, Cao Y N. PathMind: A retrieve-prioritize-reason framework for knowledge graph reasoning with large language models. In: Proceedings of the 40th AAAI Conference on Artificial Intelligence. Singapore: AAAI, 2026. 15386–15393
 - 113 Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2020. Article No. 159
 - 114 Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman F L, et al. GPT-4 technical report. arXiv preprint arXiv: 2303.08774, 2024.
 - 115 Kaplan J, McCandlish S, Henighan T, Brown T B, Chess B, Child R, et al. Scaling laws for neural language models. arXiv preprint arXiv: 2001.08361, 2020.
 - 116 Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, et al. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 2022: 1–30
 - 117 Kojima T, Gu S S, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2022. Article No. 1613
 - 118 Wang X Z, Wei J, Schuurmans D, Le Q V, Chi E H, Narang S, et al. Self-consistency improves chain of thought reasoning in language models. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
 - 119 Xie Y X, Kawaguchi K, Zhao Y R, Zhao J X, Kan M Y, He J X, et al. Self-evaluation guided beam search for reasoning. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 1802
 - 120 Long J Y. Large language model guided tree-of-thought. arXiv preprint arXiv: 2305.08291, 2023.
 - 121 Yao S Y, Yu D, Zhao J, Shafran I, Griffiths T L, Cao Y, et al. Tree of thoughts: Deliberate problem solving with large language models. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 517
 - 122 Besta M, Blach N, Kubicek A, Gerstenberger R, Podstawski M, Gianinazzi L, et al. Graph of thoughts: Solving elaborate problems with large language models. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2024. 17682–17690
 - 123 Yin S K, Fu C Y, Zhao S R, Li K, Sun X, Xu T, et al. A survey on multimodal large language models. arXiv preprint arXiv: 2306.13549, 2024.
 - 124 Liu H T, Li C Y, Wu Q Y, Lee Y J. Visual instruction tuning. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 1516
 - 125 Huang R J, Li M Z, Yang D C, Shi J T, Chang X K, Ye Z H, et al. AudioGPT: Understanding and generating speech, music, sound, and talking head. In: Proceedings of the 38th AAAI

- Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2024. 23802–23804
- 126 OpenAI. Video Generation Models as World Simulators, Technical Report, OpenAI, USA, 2024.
- 127 Bonifati A, Holubová I, Prat-Pérez A, Sakr S. Graph generators: State of the art and open challenges. *ACM Computing Surveys (CSUR)*, 2021, **53**(2): Article No. 53
- 128 Eigenschink P, Reutterer T, Vamosi S, Vamosi R, Sun C, Kalcher K. Deep generative models for synthetic data: A survey. *IEEE Access*, 2023, **11**: 47304–47320
- 129 Zang C X, Wang F. MoFlow: An invertible flow model for generating molecular graphs. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event: Association for Computing Machinery, 2020. 617–626
- 130 Vignac C, Krawczuk I, Siraudin A, Wang B H, Cevher V, Frossard P. DiGress: Discrete denoising diffusion for graph generation. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
- 131 Liao R J, Li Y J, Song Y, Wang S L, Hamilton W L, Duvenaud D, et al. Efficient graph generation with graph recurrent attention networks. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019. Article No. 383
- 132 Simonovsky M, Komodakis N. GraphVAE: Towards generation of small graphs using variational autoencoders. In: Proceedings of the 27th International Conference on Artificial Neural Networks. Rhodes, Greece: Springer, 2018. 412–422
- 133 Grover A, Zweig A, Ermon S. Graphite: Iterative generative modeling of graphs. In: Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: PMLR, 2019. 2434–2444
- 134 Geng Z J, Xie S F, Xia Y C, Wu L J, Qin T, Wang J, et al. De Novo molecular generation via connection-aware motif mining. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
- 135 Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using real NVP. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: OpenReview.net, 2017.
- 136 Rezende D J, Mohamed S. Variational inference with normalizing flows. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015. 1530–1538
- 137 Wang H W, Wang J, Wang J L, Zhao M, Zhang W N, Zhang F Z, et al. GraphGAN: Graph representation learning with generative adversarial nets. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI, 2018. 2508–2515
- 138 Yang C, Zhuang P Y, Shi W H, Luu A, Li P. Conditional structure generation through graph variational generative adversarial nets. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019. Article No. 120
- 139 Yang L, Zhang Z L, Zhang W T, Hong S D. Score-based graph generative modeling with self-guided latent diffusion. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
- 140 Huang H, Sun L L, Du B W, Fu Y J, Lv W F. GraphGDP: Generative diffusion processes for permutation invariant graph generation. In: Proceedings of the IEEE International Conference on Data Mining (ICDM). Orlando, USA: IEEE, 2022. 201–210
- 141 You J X, Liu B W, Ying Z T, Pande V S, Leskovec J. Graph convolutional policy network for goal-directed molecular graph generation. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018. 6412–6422
- 142 Popova M, Shvets M, Oliva J, Isayev O. MolecularRNN: Generating realistic molecular graphs with optimized properties. arXiv preprint arXiv: 1905.13372, 2019.
- 143 Gupta S, Manchanda S, Bedathur S, Ranu S. TIGGER: Scalable generative modelling for temporal interaction graphs. In: Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual Event: AAAI, 2022. 6819–6828
- 144 Jang Y, Lee S, Ahn S. A simple and scalable representation for graph generation. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
- 145 Cohen-Karlik E, Rozenberg E, Freedman D. Overcoming order in autoregressive graph generation for molecule generation. arXiv preprint arXiv: 2402.03387, 2024.
- 146 Zhang L M, Zhao L, Qin S, Pfoser D. TG-GAN: Deep generative models for continuously-time temporal graph generation. arXiv preprint arXiv: 2005.08323, 2020.
- 147 Wang C, Xue N W, Pradhan S. A transition-based algorithm for AMR parsing. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver, USA: Association for Computational Linguistics, 2015. 366–375
- 148 Chen B, Sun L, Han X P. Sequence-to-action: End-to-end semantic graph generation for semantic parsing. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia: Association for Computational Linguistics, 2018. 766–777
- 149 Tan X W, Zhou Y X, Pergola G, He Y L. Set-aligning framework for auto-regressive event temporal graph generation. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Mexico City, Mexico: Association for Computational Linguistics, 2024. 3872–3892
- 150 Lu H, Wei Z Q, Wang X Z, Zhang K, Liu H. GraphGPT: A graph enhanced generative pretrained Transformer for conditioned molecular generation. *International Journal of Molecular Sciences*, 2023, **24**(23): Article No. 16761
- 151 Wang Y, Rossi R A, Park N, Chen H Y, Ahmed N K, Trivedi P, et al. A large-scale training paradigm for graph generative models. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.
- 152 Tenorio V M, Zilberstein N, Segarra S, Marques A G. Graph guided diffusion: Unified guidance for conditional graph generation. arXiv preprint arXiv: 2505.19685, 2025.
- 153 Li M F, Kreacic E, Pothuru V K, Li P. GraphMaker: Can diffusion models generate large attributed graphs? arXiv preprint arXiv: 2310.13833, 2024.
- 154 Zhang D C, Yang M L, Ying R, Lauw H W. Text-attributed graph representation learning: Methods, applications, and challenges. In: Proceedings of the ACM Web Conference. Singapore: Association for Computing Machinery, 2024. 1298–1301
- 155 Wang L Y, Wang Y, Ni B, Zhao Y, Wang H Y, Ma Y, et al. SAVE-TAG: Semantic-aware vicinal risk minimization for long-tailed text-attributed graphs. arXiv preprint arXiv: 2410.16882, 2025.
- 156 Gao C, Lan X C, Li N, Yuan Y, Ding J T, Zhou Z L, et al. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 2024, **11**(1): Article No. 1259
- 157 Park J S, O'Brien J, Cai C J, Morris M R, Liang P, Bernstein M S. Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. San Francisco, USA: Association for Computing Machinery, 2023. Article No. 2
- 158 Li N, Gao C, Li M Y, Li Y, Liao Q M. EconAgent: Large language model-empowered agents for simulating macroeconomic

- activities. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, Thailand: Association for Computational Linguistics, 2024. 15523–15536
- 159 Mou X Y, Ding X W, He Q, Wang L, Liang J C, Zhang X N, et al. From individual to society: A survey on social simulation driven by large language model-based agents. *ACM Computing Surveys*, 2026, **58**(11): Article No. 280
- 160 Barabási A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, **286**(5439): 509–512
- 161 Kleinberg J M, Kumar R, Raghavan P, Rajagopalan S, Tomkins A S. The web as a graph: Measurements, models, and methods. In: Proceedings of the 5th Annual International Conference on Computing and Combinatorics. Tokyo, Japan: Springer, 1999. 1–17
- 162 Ioannides Y M. Random graphs and social networks: An economics perspective [Online], available: <https://www.researchgate.net/publication/24137958>, July 21, 2026
- 163 Gao C, Lan X C, Lu Z H, Mao J Z, Piao J, Wang H D, et al. S3: Social-network simulation system with large language model-empowered agents. arXiv preprint arXiv: 2307.14984, 2025.
- 164 Mou X Y, Wei Z Y, Huang X J. Unveiling the truth and facilitating change: Towards agent-based large-scale social movement simulation. In: Proceedings of the Findings of the Association for Computational Linguistics. Bangkok, Thailand: Association for Computational Linguistics, 2024. 4789–4809
- 165 Yang Z Y, Zhang Z B, Zheng Z R, Jiang Y X, Gan Z Y, Wang Z Y, et al. OASIS: Open agent social interaction simulations with one million agents. arXiv preprint arXiv: 2411.11581, 2025.
- 166 Rossetti G, Stella M, Cazabet R, Abramski K, Cau E, Citraro S, et al. Y social: An LLM-powered social media digital twin. arXiv preprint arXiv: 2408.00818, 2024.
- 167 Zhang X Q, Chen X Y, Liu Y H, Wang J Z, Hu Z X, Yan R. A large scale time-aware agents simulation for influencer selection in digital advertising campaigns. arXiv preprint arXiv: 2411.01143, 2024.
- 168 Liu Y H, Song Z R, Zhang J T, Zhang X Q, Chen X Y, Yan R. The stepwise deception: Simulating the evolution from true news to fake news with LLM agents. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Suzhou, China: Association for Computational Linguistics, 2025. 26176–26192
- 169 Ji J R, Li Y, Liu H T, Du Z C, Wei Z W, Qi Q, et al. SRAP-agent: Simulating and optimizing scarce resource allocation policy with LLM-based agent. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP. Miami, USA: Association for Computational Linguistics, 2024. 267–293
- 170 Leskovec J, Chakrabarti D, Kleinberg J, Faloutsos C, Ghahramani Z. Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research*, 2010, **11**: 985–1042
- 171 Park H, Kim M S. TrillionG: A trillion-scale synthetic graph generator using a recursive vector model. In: Proceedings of the ACM International Conference on Management of Data. Chicago, USA: Association for Computing Machinery, 2017. 913–928
- 172 Gutfraind A, Safro I, Meyers L A. Multiscale network generation. In: Proceedings of the 18th International Conference on Information Fusion. Washington, USA: IEEE, 2015. 158–165
- 173 Gao C, Zheng Y, Li N, Li Y F, Qin Y R, Piao J H, et al. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems*, 2023, **1**(1): Article No. 3
- 174 Ebrat D, Paradalis E, Rueda L. Lusifer: LLM-based user simulated feedback environment for online recommender systems. In: Proceedings of the 4th International Conference on Computing and Machine Intelligence (ICMI). MI, USA: IEEE, 2025. 1–6
- 175 Corecco N, Piatti G, Lanzendörfer L A, Fan F X, Wattenhofer R. SUBER: An RL environment with simulated human behavior for recommender systems. In: Proceedings of the 27th European Conference on Artificial Intelligence. Santiago de Compostela, Spain: IOS Press, 2024. 2210–2217
- 176 Wang L, Zhang J S, Yang H, Chen Z Y, Tang J K, Zhang Z Y, et al. User behavior simulation with large language model-based agents. *ACM Transactions on Information Systems*, 2025, **43**(2): Article No. 55
- 177 Chen L Y, Dai Q Y, Zhang Z Y, Feng X Y, Zhang M Y, Tang P C, et al. RecUserSim: A realistic and diverse user simulator for evaluating conversational recommender systems. In: Proceedings of the Companion Proceedings of the ACM on Web Conference. Sydney, Australia: Association for Computing Machinery, 2025. 133–142
- 178 Zhang A, Chen Y X, Sheng L H, Wang X, Chua T S. On generative agents in recommendation. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington, USA: Association for Computing Machinery, 2024. 1807–1817
- 179 Ying C X, Cai T L, Luo S J, Zheng S X, Ke G L, He D, et al. Do Transformers really perform bad for graph representation? In: Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event: Curran Associates Inc., 2021. Article No. 2212
- 180 Yang C D, Liu H R, Wang D X, Zhang Z Q, Yang C, Shi C. FLAG: Fraud detection with LLM-enhanced graph neural network. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Toronto, Canada: Association for Computing Machinery, 2025. 5150–5160
- 181 Huang T R, Wang Y L, Li Q T, He C L, Gao J L. Can LLMs find fraudsters? Multi-level LLM enhanced graph fraud detection. In: Proceedings of the 33rd ACM International Conference on Multimedia. Dublin, Ireland: Association for Computing Machinery, 2025. 1530–1538
- 182 Xue C, Lu Y W, Yang C, Xing J M. RecMind: LLM-enhanced graph neural networks for personalized consumer recommendations. In: Proceedings of the IEEE International Conference on Consumer Electronics (ICCE). Dubai, United Arab Emirates: IEEE, 2026. 1–6
- 183 Wei W, Ren X B, Tang J B, Wang Q Y, Su L X, Cheng S Q, et al. LLMRec: Large language models with graph augmentation for recommendation. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. Merida, Mexico: Association for Computing Machinery, 2024. 806–815
- 184 He X N, Deng K, Wang X, Li Y, Zhang Y D, Wang M. LightGCN: Simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Virtual Event: Association for Computing Machinery, 2020. 639–648
- 185 Wu L K, Qiu Z P, Zheng Z, Zhu H S, Chen E H. Exploring large language model for graph data understanding in online job recommendations. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2024. 9178–9186
- 186 Guo T, Hauptman E. Fine-tuning large language models for stock return prediction using newsflow. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing: Industry Track. Miami, USA: Association for Computational Linguistics, 2024. 1028–1045
- 187 Xu Z Y, Liu Y, Wang Y C, Bao R H, Harimoto K, Sun X. Modeling interactions between stocks using LLM-enhanced graphs for volume prediction. In: Proceedings of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st

- Workshop on Large Language Models for Finance and Legal (LLMFinLegal). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2025. 153–163
- 188 Liu C, Wu B. Evaluating large language models on graphs: Performance insights and comparative analysis. arXiv preprint arXiv: 2308.11224, 2023.
- 189 Himsolt M. GML: A Portable Graph File Format, Technical Report, Universität Passau, Germany, 1997.
- 190 Brandes U, Eiglsperger M, Lerner J, Pich C. Graph markup language (GraphML) [Online], available: <https://cs.brown.edu/people/rtamassi/gdhandbook/chapters/graphml.pdf>, July 21, 2026
- 191 Liu N F, Lin K, Hewitt J, Paranjape A, Bevilacqua M, Petroni F, et al. Lost in the middle: How language models use long contexts. In: Proceedings of the Transactions of the Association for Computational Linguistics. Cambridge, USA: MIT Press, 2024. 157–173
- 192 Luo Z H, Song X R, Huang H, Lian J X, Zhang C H, Jiang J Q, et al. GraphInstruct: Empowering large language models with graph understanding and reasoning capability. arXiv preprint arXiv: 2403.04483, 2025.
- 193 Gansner E R, North S C. An open graph visualization system and its applications to software engineering. *Software: Practice and Experience*, 2000, **30**(11): 1203–1233
- 194 Tosi S. *Matplotlib for Python Developers: Build Remarkable Publication Quality Plots the Easy Way*. Birmingham: Packt Publishing, 2009.
- 195 Zhao H T, Liu S C, Ma C, Xu H N, Fu J, Deng Z H, et al. GIMLET: A unified graph-text model for instruction-based molecule zero-shot learning. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 257
- 196 Feng Y C, Zhou L, Ma C, Zheng Y K, He R K, Li Y X. Knowledge graph-based thought: A knowledge graph-enhanced LLM framework for pan-cancer question answering. *GigaScience*, 2025, **14**: Article No. giae082
- 197 He J L, Guo Y Q, Lam L K, Leung W, He L X, Jiang Y A, et al. OpenTCM: A GraphRAG-empowered LLM-based system for traditional Chinese medicine knowledge retrieval and diagnosis. arXiv preprint arXiv: 2504.20118, 2025.
- 198 Wang S J, Fan W Q, Feng Y, Lin S R, Ma X Y, Wang S Q, et al. Knowledge graph retrieval-augmented generation for LLM-based recommendation. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics. Vienna, Austria: Association for Computational Linguistics, 2025. 27152–27168
- 199 Qiu Z C, Luo L H, Pan S R, Liew A W C. Unveiling user preferences: A knowledge graph and LLM-driven approach for conversational recommendation. arXiv preprint arXiv: 2411.14459, 2024.
- 200 Zhu Y C, Steck H, Liang D W, He Y H, Kallus N, Li J D. LLM-based conversational recommendation agents with collaborative verbalized experience. In: Proceedings of the Findings of the Association for Computational Linguistics. Suzhou, China: Association for Computational Linguistics, 2025. 2207–2220
- 201 Pirmorad E. Exploring the in-context learning capabilities of LLMs for money laundering detection in financial graphs. In: Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW). Washington, USA: IEEE, 2025. 174–177
- 202 He P L, Meng H N. A fast discovery model for production financial risk using large language model and knowledge graph. *Journal of Circuits, Systems and Computers*, 2025, **34**(9): Article No. 2550206
- 203 Khanvilkar K, Kommuru K. Regulatory graphs and GenAI for real-time transaction monitoring and compliance explanation in banking. arXiv preprint arXiv: 2506.01093, 2025.
- 204 Peng J, Ji J R, Lei R L, Wei Z W, Liu Y C, Hong C T. GDGB: A benchmark for generative dynamic text-attributed graph learning. arXiv preprint arXiv: 2507.03267, 2026.
- 205 Ren R Y, Qiu P, Qu Y Q, Liu J, Zhao W X, Wu H, et al. BASES: Large-scale web search user simulation with large language model based agents. In: Proceedings of the Findings of the Association for Computational Linguistics. Miami, USA: Association for Computational Linguistics, 2024. 902–917
- 206 Zhang X Q, Liu Y H, Wang J Z, Hu Z X, Chen X Y, Yan R. SAGraph: A large-scale social graph dataset with comprehensive context for influencer selection in marketing. In: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval. Padua, Italy: Association for Computing Machinery, 2025. 3733–3742
- 207 Mou X Y, Qian C, Liu W, Yan L, Hu Y, Huang X J, et al. EcoLANG: Efficient and effective agent communication language induction for social simulation. In: Proceedings of the Findings of the Association for Computational Linguistics. Suzhou, China: Association for Computational Linguistics, 2025. 5287–5304
- 208 Chang R C, Rao A, Zhong Q K, Wojcieszak M, Lerman K. #RoeOverturned: Twitter dataset on the abortion rights controversy. In: Proceedings of the 17th International AAAI Conference on Web and Social Media. Limassol, Cyprus: AAAI, 2023. 997–1005
- 209 Harper F M, Konstan J A. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2016, **5**(4): Article No. 19
- 210 Huang X, Lian J X, Lei Y X, Yao J, Lian D F, Xie X. Recommender AI agent: Integrating large language models for interactive recommendations. *ACM Transactions on Information Systems*, 2025, **43**(4): Article No. 96
- 211 Zhang J J, Hou Y P, Xie R B, Sun W Q, McAuley J J, Zhao W X, et al. AgentCF: Collaborative learning with autonomous language agents for recommender systems. In: Proceedings of the ACM Web Conference. Singapore: Association for Computing Machinery, 2024. 3679–3689
- 212 Hou Y P, Li J C, He Z K, Yan A, Chen X S, McAuley J. Bridging language and items for retrieval and recommendation. arXiv preprint arXiv: 2403.03952, 2024.
- 213 Liu Y J, Liu W, Gu X Y, Rui Y, He X D, Zhang Y D. LMAgent: A large-scale multimodal agents society for multi-user simulation. *IEEE Transactions on Multimedia*, DOI: 10.1109/TMM.2026.3651078
- 214 Kang W C, McAuley J. Self-attentive sequential recommendation. In: Proceedings of the IEEE International Conference on Data Mining (ICDM). Singapore: IEEE, 2018. 197–206
- 215 Cantador I, Brusilovsky P, Kuflik T. Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011). In: Proceedings of the 5th ACM Conference on Recommender Systems. Chicago, USA: Association for Computing Machinery, 2011. 387–388
- 216 Ozsoy M G. Multilingual prompts in LLM-based recommenders: Performance across languages. arXiv preprint arXiv: 2409.07604, 2024.
- 217 Liao J Y, Li S H, Yang Z Y, Wu J C, Yuan Y C, Wang X, et al. LLaRA: Large language-recommendation assistant. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington, USA: Association for Computing Machinery, 2024. 1785–1795
- 218 Zheng K, Sun Q F, Xu C, Yu P, Guo Q W. Towards a unified paradigm: Integrating recommendation systems as a new language in large models. arXiv preprint arXiv: 2412.16933, 2024.
- 219 Jiang B, Zhang Z Y, Lin D D, Tang J, Luo B. Semi-supervised

- learning with graph learning-convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 11305–11312
- 220 Morris C, Ritzert M, Fey M, Hamilton W L, Lenssen J E, Rattan G, et al. Weisfeiler and leman go neural: Higher-order graph neural networks. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI, 2019. 4602–4609
- 221 Thompson R, Knyazev B, Ghalebi E, Kim J, Taylor G W. On evaluation metrics for graph generative models. In: Proceedings of the 10th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2022.
- 222 Wang C E, Kantarcioglu M. Graph generative models evaluation with masked autoencoder. In: Proceedings of the Trends and Applications in Knowledge Discovery and Data Mining. Sydney, Australia: Springer, 2025. 137–148
- 223 Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6629–6640
- 224 Hosseini R, Simini F, Vishwanath V, Willett R, Hoffmann H. Quality measures for dynamic graph generative models. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.
- 225 Yoon M, Wu Y, Palowitch J, Perozzi B, Salakhutdinov R. Graph generative model for benchmarking graph neural networks. In: Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: PMLR, 2023. 40175–40198
- 226 Kipf T N, Welling M. Variational graph auto-encoders. arXiv preprint arXiv: 1611.07308, 2016.
- 227 Yan H, Li C Z, Long R S, Yan C, Zhao J N, Zhuang W W, et al. A comprehensive study on text-attributed graphs: Benchmarking and rethinking. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 755
- 228 Berant J, Chou A, Frostig R, Liang P. Semantic parsing on freebase from question-answer pairs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Seattle, USA: Association for Computational Linguistics, 2013. 1533–1544
- 229 Bordes A, Usunier N, Chopra S, Weston J. Large-scale simple question answering with memory networks. arXiv preprint arXiv: 1506.02075, 2015.
- 230 Bao J W, Duan N, Yan Z, Zhou M, Zhao T J. Constraint-based question answering with knowledge graph. In: Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan: The COLING 2016 Organizing Committee, 2016. 2503–2514
- 231 Su Y, Sun H, Sadler B M, Srivatsa M, Gür I, Yan Z H, et al. On generating characteristic-rich question sets for QA evaluation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin, USA: Association for Computational Linguistics, 2016. 562–572
- 232 Yih W T, Richardson M, Meek C, Chang M W, Suh J. The value of semantic parse labeling for knowledge base question answering. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: Association for Computational Linguistics, 2016. 201–206
- 233 Serban I V, García-Durán A, Gulcehre C, Ahn S, Chandar S, Courville A C, et al. Generating factoid questions with recurrent neural networks: The 30M factoid question-answer corpus. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: Association for Computational Linguistics, 2016. 588–598
- 234 Diefenbach D, Tanon T P, Singh K D, Maret P. Question answering benchmarks for Wikidata. In: Proceedings of the Posters & Demonstrations and Industry Tracks Co-located With 16th International Semantic Web Conference (ISWC 2017). Vienna, Austria: CEUR-WS.org, 2017.
- 235 Trivedi P, Maheshwari G, Dubey M, Lehmann J. LC-QuAD: A corpus for complex question answering over knowledge graphs. In: Proceedings of the 16th International Semantic Web Conference on the Semantic Web. Vienna, Austria: Springer, 2017. 210–218
- 236 Talmor A, Berant J. The web as a knowledge-base for answering complex questions. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, Louisiana: Association for Computational Linguistics, 2018. 641–651
- 237 Usbeck R, Gusmita R H, Ngomo A N, Saleem M. 9th challenge on question answering over linked data (QALD-9) (invited paper). In: Proceedings of the 4th Workshop on Semantic Deep Learning (SemDeep-4) and NLIWoD4: Natural Language Interfaces for the Web of Data (NLIWoD-4) and 9th Question Answering Over Linked Data Challenge (QALD-9) Co-located With 17th International Semantic Web Conference (ISWC 2018). Monterey, USA: CEUR-WS.org, 2018. 58–64
- 238 Zhou M T, Huang M L, Zhu X Y. An interpretable reasoning network for multi-relation question answering. In: Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, USA: Association for Computational Linguistics, 2018. 2010–2022
- 239 Zhang Y Y, Dai H J, Kozareva Z, Smola A J, Song L. Variational reasoning for question answering with knowledge graph. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI, 2018. 6069–6076
- 240 Azmy M, Shi P, Lin J, Ilyas I F. Farewell freebase: Migrating the SimpleQuestions dataset to DBpedia. In: Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, USA: Association for Computational Linguistics, 2018. 2093–2103
- 241 Dubey M, Banerjee D, Abdelkawi A, Lehmann J. LC-QuAD 2.0: A large dataset for complex question answering over Wikidata and DBpedia. In: Proceedings of the 18th International Semantic Web Conference on the Semantic Web. Auckland, New Zealand: Springer, 2019. 69–78
- 242 Jiang K, Wu D K, Jiang H. FreebaseQA: A new factoid QA data set matching trivia-style question-answer pairs with freebase. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. 318–323
- 243 Costa T S, Gottschalk S, Demidova E. Event-QA: A dataset for event-centric question answering over knowledge graphs. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Virtual Event: Association for Computing Machinery, 2020. 3157–3164
- 244 Gu Y, Kase S, Vanni M, Sadler B M, Liang P, Yan X F, et al. Beyond I.I.D.: Three levels of generalization for question answering on knowledge bases. In: Proceedings of the Web Conference. Ljubljana, Slovenia: Association for Computing Machinery, 2021. 3477–3488
- 245 Tang J H, Zhang Q F, Li Y H, Chen N, Li J. GraphArena: Evaluating and exploring large language models on graph computation. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: ICLR, 2025.
- 246 Chen X H, He J X, Han X, Liu L P. Efficient and degree-guided graph generation via discrete diffusion modeling. In: Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: PMLR, 2023. 4585–4610
- 247 O'Bray L, Horn M, Rieck B, Borgwardt K M. Evaluation metrics for graph generative models: Problems, pitfalls, and prac-

tical solutions. In: Proceedings of the 10th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2022.

248 Li A G, Yu L. Summary factual inconsistency detection based on LLMs enhanced by universal information extraction. In: Proceedings of the Findings of the Association for Computational Linguistics. Vienna, Austria: Association for Computational Linguistics, 2025. 25450–25465

249 Zhang L, Song D D, Wu Z J, Tian Y H, Zhou C Z, Xu J, et al. Detecting hallucination in large language models through deep internal representation analysis. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence. Montreal, Canada: ijcai.org, 2025. 8357–8365

250 Zhang Y, Zhao Y Y, Li Z Q, Cheng X Q, Wang Y, Kotevska O, et al. A survey on privacy in graph neural networks: Attacks, preservation, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2024, **36**(12): 7497–7515

251 Sheng Z A, Guo W Y, Shao Y X, Zhang W T, Cui B. LLMs are noisy oracles! LLM-based noise-aware graph active learning for node classification. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2. Toronto, Canada: Association for Computing Machinery, 2025. 2526–2537

252 Huang J, Zhang X J, Mei Q Z, Ma J Q. Can LLMs effectively leverage graph structural information through prompts, and why? arXiv preprint arXiv: 2309.16595, 2024.



郑艳萍 中国人民大学高瓴人工智能学院博士后. 2024 年获得中国人民大学高瓴人工智能学院博士学位. 主要研究方向为图机器学习.
E-mail: zhengyanping@ruc.edu.cn
(ZHENG Yan-Ping Postdoctor at the Gaoling School of Artificial Intelligence, Renmin University of China. She received her Ph.D. degree from Gaoling School of Artificial Intelligence, Renmin University of China in 2024. Her main research interest is graph machine learning.)



季嘉蕊 中国人民大学高瓴人工智能学院博士研究生. 2023 年获得同济大学学士学位. 主要研究方向为大语言模型, 图学习.
E-mail: jijiarui@ruc.edu.cn
(JI Jia-Rui Ph.D. candidate at the Gaoling School of Artificial Intelligence, Renmin University of China. She received her bachelor degree from Tongji University in 2023. Her research interests include large language models and graph learning.)



胡雨韦 中国人民大学高瓴人工智能学院博士研究生. 2024 年获得中国人民大学学士学位. 主要研究方向为大语言模型, 图学习.
E-mail: huyuweiyisui@ruc.edu.cn
(HU Yu-Wei Ph.D. candidate at the Gaoling School of Artificial Intelligence, Renmin University of China. He received his bachelor degree from Renmin University of China in 2024. His research interests include large language models and graph learning.)



魏哲巍 中国人民大学高瓴人工智能学院教授. 主要研究方向为人工智能, 大数据基础算法. 本文通信作者.
E-mail: zhewei@ruc.edu.cn
(WEI Zhe-Wei Professor at the Gaoling School of Artificial Intelligence, Renmin University of China. His research interests include artificial intelligence and foundational algorithms for big data. Corresponding author of this paper.)