

图大模型：图学习遇见大语言模型

报告人：刘瑜

报告时间：2024.06.14



中国科学院 信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING, CAS

目录

1. 概述

- 背景
- 研究路线
 - LLM as Encoder、LLM as Predictor 、LLM as Aligner

2. LLM as Encoder

- TAPE -- ICLR2024

3. LLM as Predictor

- GraphGPT -- SIGIR2024

目录

1. 概述

- 背景
- 研究路线
 - LLM as Encoder、LLM as Predictor 、LLM as Aligner

2. LLM as Encoder

- TAPE -- ICLR2024

3. LLM as Predictor

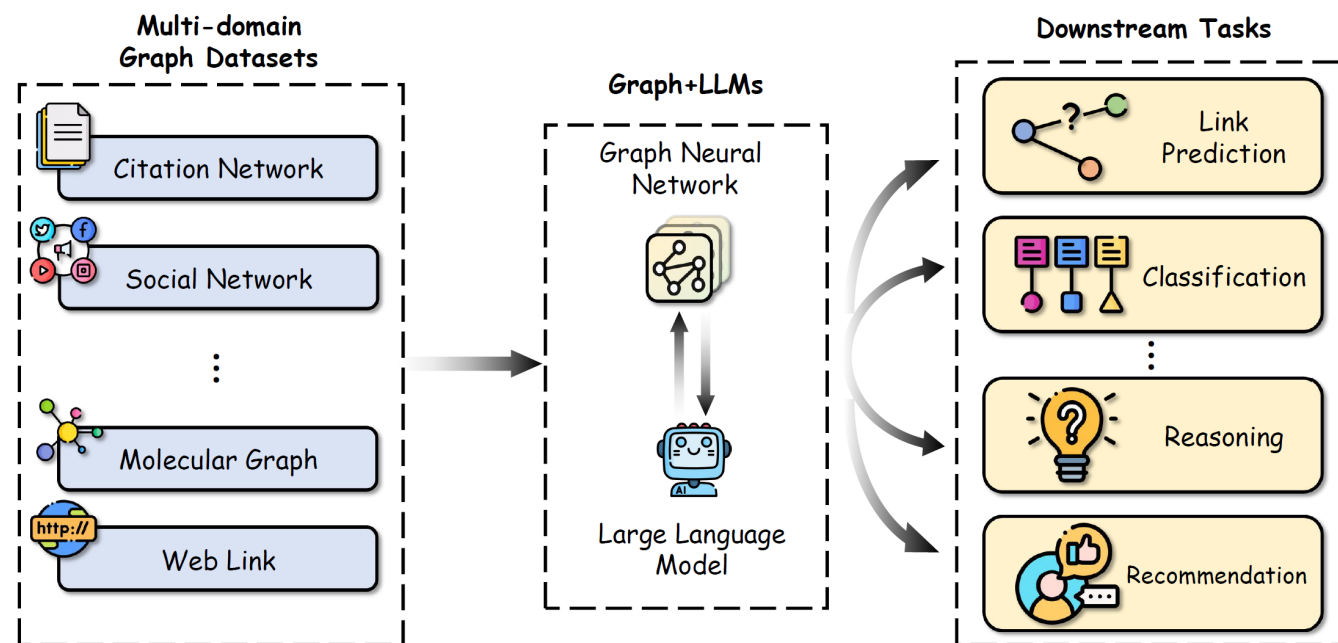
- GraphGPT -- SIGIR2024

1. 概述

1. 背景

目前，大模型发展势不可挡，对图学习的研究产生了很大的冲击。

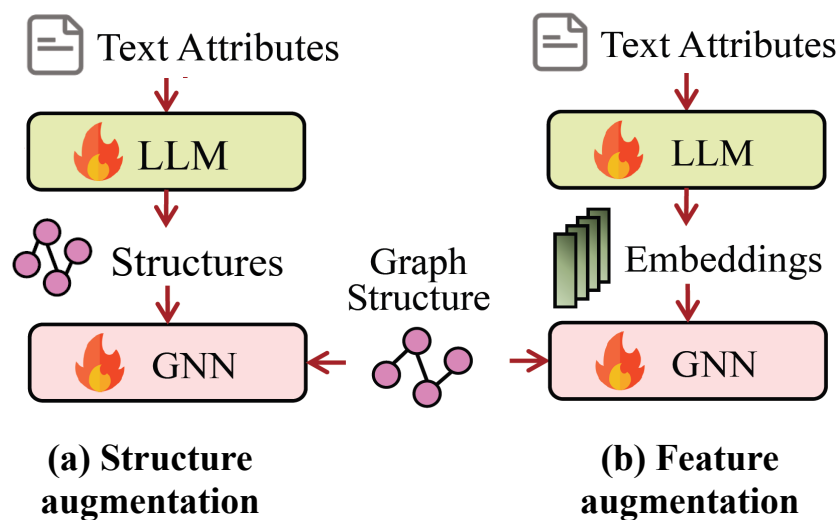
- 图模型不具备大模型的能力
 - 受限的表达能力，过平滑过压缩
 - 没有涌现能力，难以支持多任务
- 大模型解决不了图的问题
 - 大模型难以建模图结构语义
 - 大模型难以处理多样的图任务



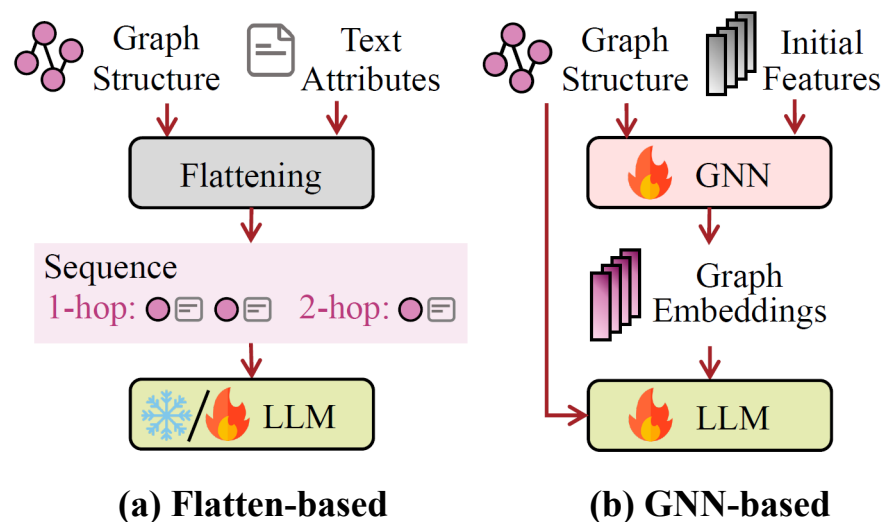
1. 概述

2. 研究路线

- ✓ LLM as Encoder: LLM作为编码器，增强图的结构或特征，GNN作为预测器
- ✓ LLM as Predictor: LLM作为预测器，将图序列化或利用GNN得到图表示
- ✓ LLM as Aligner: LLM和GNN协同作用，包括对齐、迭代和蒸馏。



(1) LLM as Encoder

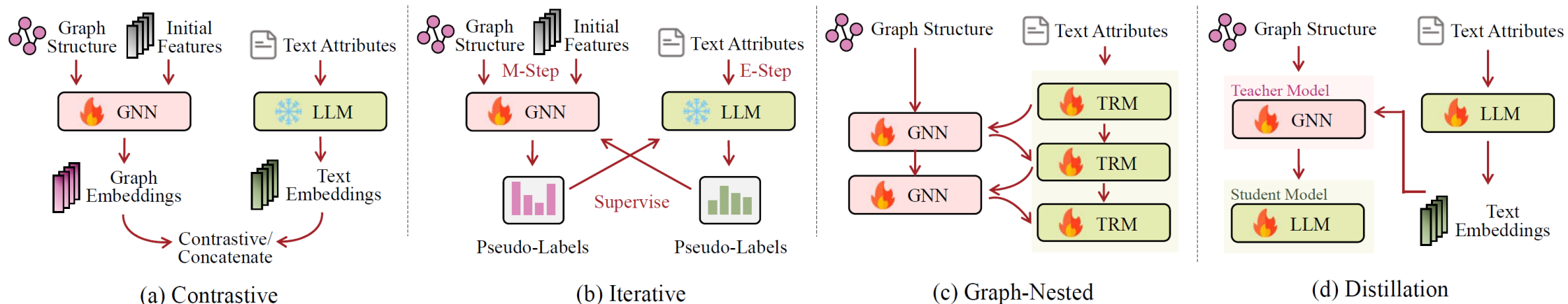


(2) LLM as Predictor

1. 概述

2. 研究路线

- ✓ LLM as Encoder: LLM作为编码器，增强图的结构或特征，GNN作为预测器
- ✓ LLM as Predictor: GNN作为编码器，学习全局结构特征，LLM作为预测器
- ✓ LLM as Aligner: LLM和GNN协同作用，包括对齐、迭代和蒸馏。



(3) LLM as Aligner

目录

1. 概述
 - 背景
 - 研究路线
 - LLM as Encoder、LLM as Predictor 、LLM as Aligner
2. LLM as Encoder
 - TAPE -- ICLR2024
3. LLM as Predictor
 - GraphGPT -- SIGIR2024

2. LLM-as Encoder

Harnessing Explanations: LLM-to-LM Interpreter for Enhanced Text-Attributed Graph Representation Learning

Xiaoxin He¹ Xavier Bresson¹ Thomas Laurent² Adam Perold³ Yann LeCun^{4,5} Bryan Hooi¹

{xiaoxin, xaviercs, bhooi}@comp.nus.edu.sg, tlaurent@lmu.edu
research@provenance.ai, yann@cs.nyu.edu

¹National University of Singapore ²Loyola Marymount University ³Provenance AI

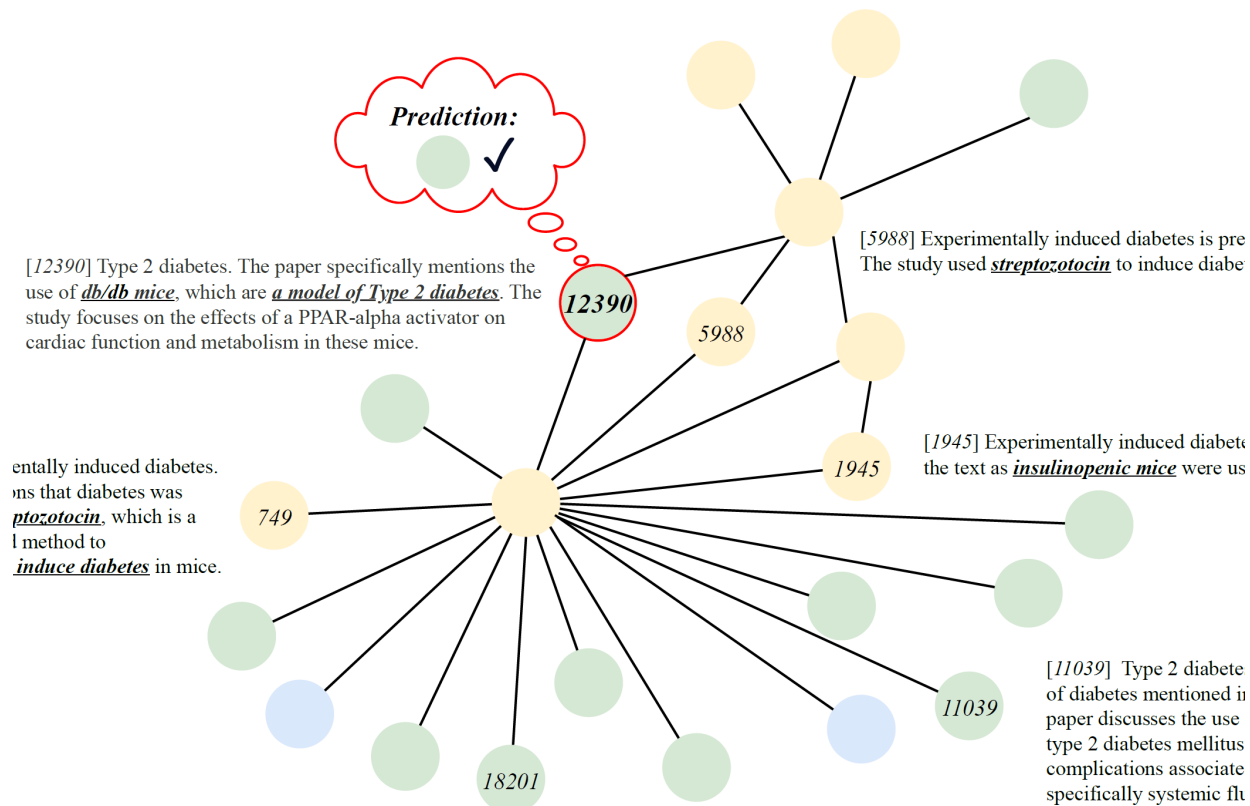
⁴New York University ⁵Meta AI

2. TAPE

1. 背景

文本属性图 (Text-Attributed Graphs, TAGs) 由实体及其关系构成，其中实体拥有文本描述，作为节点的属性。TAG 的一个典型示例是**论文引用图**，其中每篇论文的摘要或正文作为文本描述。

一般地说，文本属性与图拓扑的结合提供了丰富的信息，在多个下游任务中具有重要作用，例如文本分类、推荐系统和社交网络等。



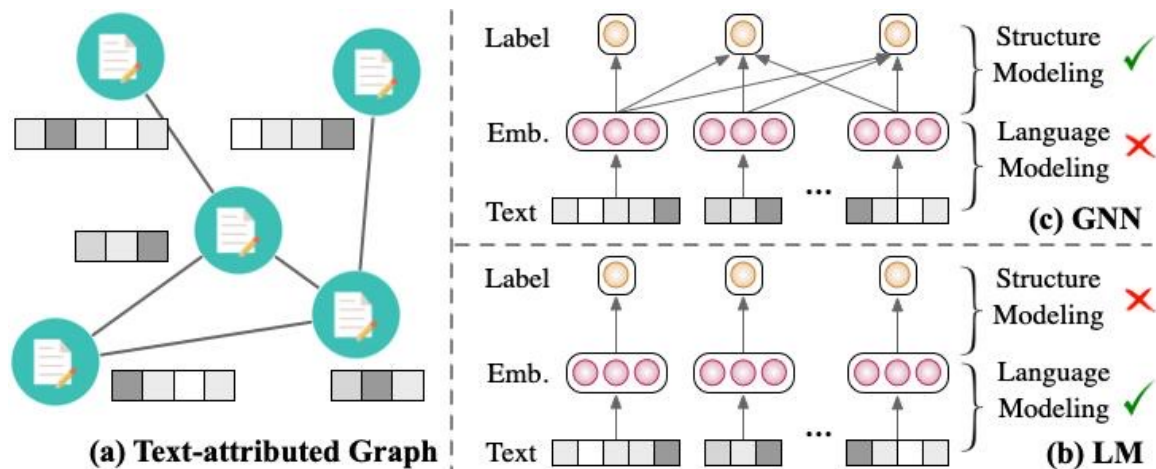
2. TAPE

1. 背景

近年来，**文本属性图TAGs**的**表征学习**一个关键研究问题。现有的工作大致分为两类：

早期，多数方法通过将文本属性转换为浅层或手工制作特征（skip-gram 或词袋特征）来处理它们。

最近，大部分工作通过使用语言模型（LM）增强文本表示能力，但通常需要复杂的设计和大量的计算资源。



2. TAPE

2. 创新

为了解决上述问题，受到大型语言模型（LLM）的冲击，这些模型展示了知识利用和推理的强大能力，因此本文尝试将 LLM 的文本建模能力与 GNN 的结构学习能力相结合的技术。

本文的创新如下：

新颖的表示学习方法。本文提出了第一种与 LMaaS 兼容的方法TAPE，用于利用 LLM 来增强 TAG 上的表示学习。（ Language Model as a Service ）

SOTA 性能。大量实验表明，我们的方法显著提高了各种 GNN 模型在不同数据集上的性能。

数据贡献。针对Cora 和PubMed中缺乏原始文本数据，本文收集并以 TAG 格式发布这些数据集。此外，本文引入了新的 arxiv-2023 引文图数据集。

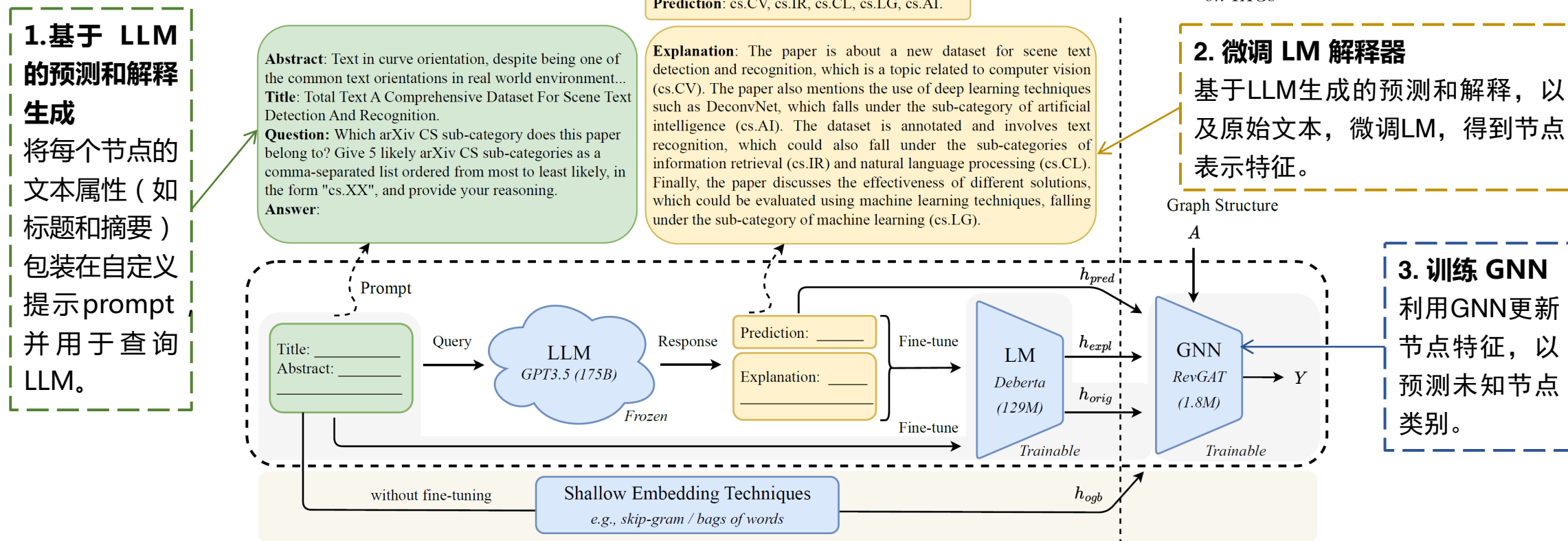
2. TAPE

3. TAPE模型

Overview of TAPE: Title, Abstract, Prediction & Explanation

Step 1: Node Feature Extraction

Step 2: Downstream Tasks on TAGs



2. TAPE

(1) 基于 LLM 的预测和解释生成

其目的在于设计合适的prompt，以指示 LLM 做出多个预测并为其决策提供解释。

具体地，对于每个论文节点 $i \in V$ ，我们生成一个提示prompt，其中包括论文的标题、摘要以及关于论文主题的开放式问题。LLM 会为每篇论文生成一个排名预测列表和文本解释。

Abstract: Text in curve orientation, despite being one of the common text orientations in real world environment...
Title: Total Text A Comprehensive Dataset For Scene Text Detection And Recognition.
Question: Which arXiv CS sub-category does this paper belong to? Give 5 likely arXiv CS sub-categories as a comma-separated list ordered from most to least likely, in the form "cs.XX", and provide your reasoning.
Answer:

Explanation: The paper is about a new dataset for scene text detection and recognition, which is a topic related to computer vision (cs.CV). The paper also mentions the use of deep learning techniques such as DeconvNet, which falls under the sub-category of artificial intelligence (cs.AI). The dataset is annotated and involves text recognition, which could also fall under the sub-categories of information retrieval (cs.IR) and natural language processing (cs.CL). Finally, the paper discusses the effectiveness of different solutions, which could be evaluated using machine learning techniques, falling under the sub-category of machine learning (cs.LG).

Prediction: cs.CV, cs.IR, cs.CL, cs.LG, cs.AI.

2. TAPE

(1) 基于 LLM 的预测和解释生成

Prompt提示问题部分的具体措辞是根据任务和数据集量身定制的，如表所示。

Dataset	Prompt
Cora	Abstract: <abstract text> \n Title: <title text> \n Question: Which of the following sub-categories of AI does this paper belong to: Case Based, Genetic Algorithms, Neural Networks, Probabilistic Methods, Reinforcement Learning, Rule Learning, Theory? If multiple options apply, provide a comma-separated list ordered from most to least related, then for each choice you gave, explain how it is present in the text. \n \n Answer:
Pubmed	Abstract: <abstract text> \n Title: <title text> \n Question: Does the paper involve any cases of Type 1 diabetes, Type 2 diabetes, or Experimentally induced diabetes? Please give one or more answers of either Type 1 diabetes, Type 2 diabetes, or Experimentally induced diabetes; if multiple options apply, provide a comma-separated list ordered from most to least related, then for each choice you gave, give a detailed explanation with quotes from the text explaining why it is related to the chosen option. \n \n Answer:
ogbn-arxiv	Abstract: <abstract text> \n Title: <title text> \n Question: Which arXiv CS sub-category does this paper belong to? Give 5 likely arXiv CS sub-categories as a comma-separated list ordered from most to least likely, in the form “cs.XX”, and provide your reasoning. \n \n Answer:

2. TAPE

(2) 微调 LM 解释器

通过微调一个较小的 LM，它充当 LLM 文本解释的“解释器”，从原始文本和 LLM 的解释和预测文本中提取最有价值的特征，包括原始文本特征、解释特征，以及排序预测特征。

- 原始文本特征和解释特征

首先，将原始文本 s^{orig} 和解释 s^{expl} 分别输入预训练的 LM_{orig} 和 LM_{expl} ，然后，利用 MLP 对 LM 的输出变换，最后利用交叉熵损失训练模型。最终，原始文本特征 h_{orig} 和解释特征 h_{expl} 被用作训练下游 GNN 的丰富特征。

$$h_{\text{orig}} = \text{LM}_{\text{orig}}(s^{\text{orig}}) \in \mathbb{R}^{N \times d}, \quad h_{\text{expl}} = \text{LM}_{\text{expl}}(s^{\text{expl}}) \in \mathbb{R}^{N \times d}.$$

$$y_{\text{orig}} = \text{MLP}_{\text{orig}}(h_{\text{orig}}) \in \mathbb{R}^{N \times C}, \quad y_{\text{expl}} = \text{MLP}_{\text{expl}}(h_{\text{expl}}) \in \mathbb{R}^{N \times C}.$$

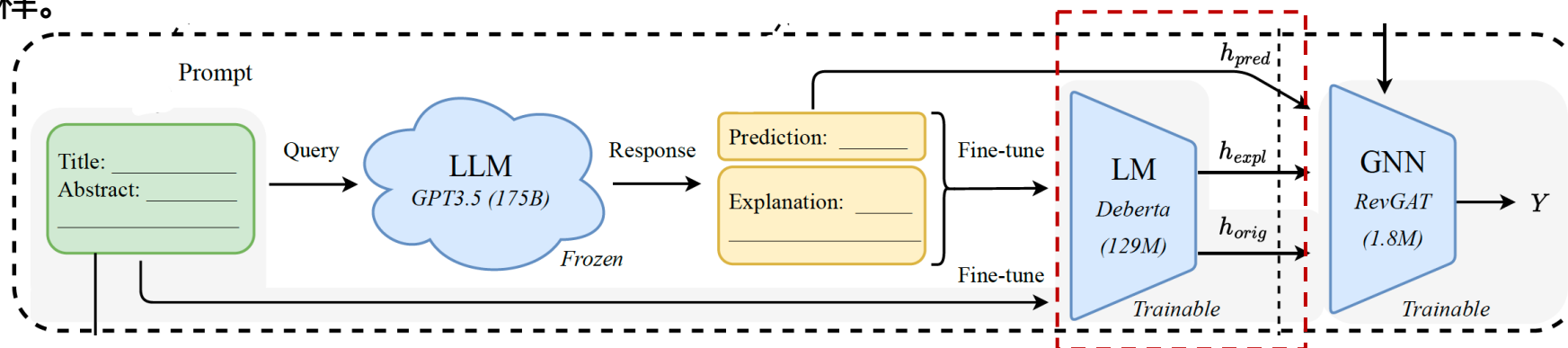
2. TAPE

(2) 微调 LM 解释器

- 排序预测特征

除了解释之外，LLM 还为每个节点提供了排名前 k 的预测列表，这增加了有价值的信息。为了整合这些知识，对于节点 i 的前 k 个预测，首先采用独热编码表示为 one-hot 向量 $p_{i,1}, \dots, p_{i,k} \in \mathbb{R}^C$ ，再将这些向量连接成一个 kC 维向量，然后进行线性变换从而得到排序预测特征 h_{pred} 。

综上，本文将特征表示记为 $h_{\text{TAPE}} = \{h_{\text{orig}}, h_{\text{expl}}, h_{\text{pred}}\}$ ，其中“TAPE”代表每个节点的标题、摘要、预测和解释。



2. TAPE

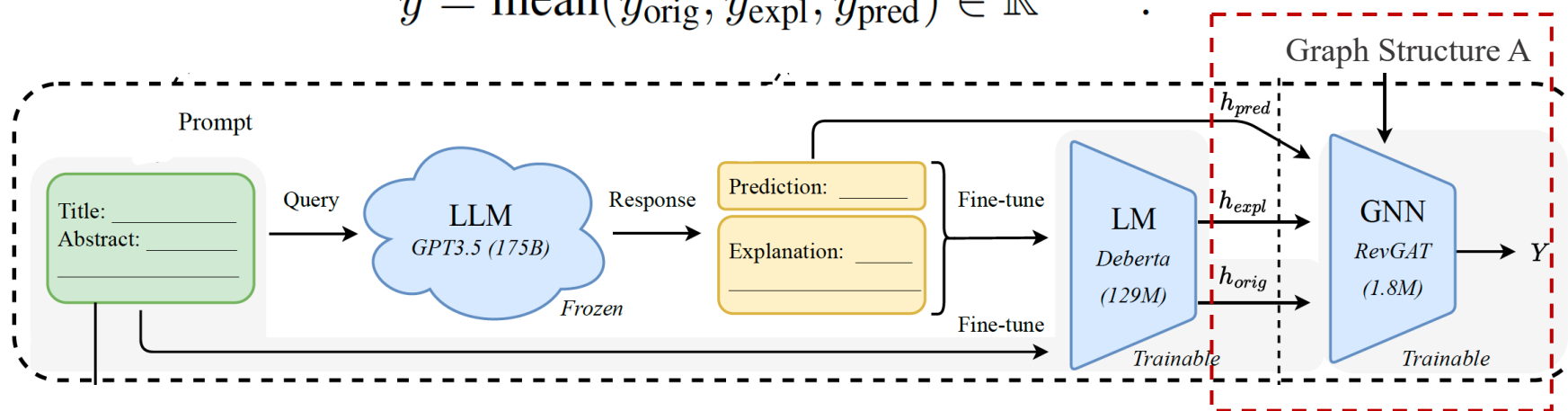
(3) 训练GNN

最后，本文利用上述学到的特征 h_{TAPE} ，训练 GNN。本文采用了集成方法，一种简单有效的特征组合方法。

具体来说，本文在 h_{orig} ， h_{expl} 和 h_{pred} 上分别独立训练 GNN 模型 f_{orig} ， f_{expl} 和 f_{pred} ，以预测真实节点标签，然后取平均值来融合这些预测结果。

$$\hat{y}_{\text{orig/expl/pred}} = f_{\text{orig/expl/pred}}(h_{\text{orig/expl/pred}}, A) \in \mathbb{R}^{N \times C}$$

$$\hat{y} = \text{mean}(\hat{y}_{\text{orig}}, \hat{y}_{\text{expl}}, \hat{y}_{\text{pred}}) \in \mathbb{R}^{N \times C}.$$



2. TAPE

4. 实验结果

- 数据集

5个TAG数据集：Cora、Pubmed、ogbn-arxiv、ogbn-products、arxiv-2023

- 基准模型

GNN模型：GCN、GraphSAGE、RevGAT；

LLM模型：LLM、 LM_{finetune}

Table 4: Statistics of the TAG datasets

Dataset	#Nodes	#Edges	Task	Metric	Augmentation
Cora	2,708	5,429	7-class classif.	Accuracy	✓
Pubmed	19,717	44,338	3-class classif.	Accuracy	✓
ogbn-arxiv	169,343	1,166,243	40-class classif.	Accuracy	
ogbn-products (subset)	54,025	74,420	47-class classif.	Accuracy	
arxiv-2023	46,198	78,548	40-class-classif.	Accuracy	✓

2. TAPE

4. 实验结果

● 主实验

Dataset	Method	GNN			LM			Ours
		h_{shallow}	h_{GIANT}	$G \uparrow$	LLM	$\text{LM}_{\text{finetune}}$	$L \uparrow$	h_{TAPE}
Cora	MLP	0.6388 ± 0.0213	—	37.41%	0.6769	0.7606 ± 0.0378	13.35%	0.8778 ± 0.0485
	GCN	0.8911 ± 0.0015	—	2.33%	0.6769	0.7606 ± 0.0378	16.59%	0.9119 ± 0.0158
	SAGE	0.8824 ± 0.0009	—	5.28%	0.6769	0.7606 ± 0.0378	18.13%	0.9290 ± 0.0307
	RevGAT	0.8911 ± 0.0000	—	4.14%	0.6769	0.7606 ± 0.0378	18.04%	0.9280 ± 0.0275
PubMed	MLP	0.8635 ± 0.0032	—	10.77%	0.9342	0.9494 ± 0.0046	0.75%	0.9565 ± 0.0060
	GCN	0.8031 ± 0.0425	—	17.43%	0.9342	0.9494 ± 0.0046	-0.66%	0.9431 ± 0.0043
	SAGE	0.8881 ± 0.0002	—	8.30%	0.9342	0.9494 ± 0.0046	1.31%	0.9618 ± 0.0053
	RevGAT	0.8850 ± 0.0005	—	8.52%	0.9342	0.9494 ± 0.0046	1.15%	0.9604 ± 0.0047
ogbn-arxiv	MLP	0.5336 ± 0.0038	0.7308 ± 0.0006	42.19%	0.7350	0.7361 ± 0.0004	3.07%	0.7587 ± 0.0015
	GCN	0.7182 ± 0.0027	0.7329 ± 0.0010	4.71%	0.7350	0.7361 ± 0.0004	2.16%	0.7520 ± 0.0003
	SAGE	0.7171 ± 0.0017	0.7435 ± 0.0014	6.98%	0.7350	0.7361 ± 0.0004	4.22%	0.7672 ± 0.0007
	RevGAT	0.7083 ± 0.0017	0.7590 ± 0.0019	9.42%	0.7350	0.7361 ± 0.0004	5.28%	0.7750 ± 0.0012
ogbn-products	MLP	0.5385 ± 0.0017	—	46.3%	0.7440	0.7297 ± 0.0023	7.96%	0.7878 ± 0.0082
	GCN	0.7052 ± 0.0051	—	13.39%	0.7440	0.7297 ± 0.0023	9.58%	0.7996 ± 0.0041
	SAGE	0.6913 ± 0.0026	—	17.71%	0.7440	0.7297 ± 0.0023	11.51%	0.8137 ± 0.0043
	RevGAT	0.6964 ± 0.0017	—	18.24%	0.7440	0.7297 ± 0.0023	12.84%	0.8234 ± 0.0036
arxiv-2023	MLP	0.6202 ± 0.0064	—	35.20%	0.7356	0.7358 ± 0.0006	12.25%	0.8385 ± 0.0246
	GCN	0.6341 ± 0.0062	—	27.42%	0.7356	0.7358 ± 0.0006	8.94%	0.8080 ± 0.0215
	SAGE	0.6430 ± 0.0037	—	30.45%	0.7356	0.7358 ± 0.0006	12.28%	0.8388 ± 0.0264
	RevGAT	0.6563 ± 0.0062	—	28.34%	0.7356	0.7358 ± 0.0006	12.64%	0.8423 ± 0.0256

2. TAPE

4. 实验结果

- 可扩展性

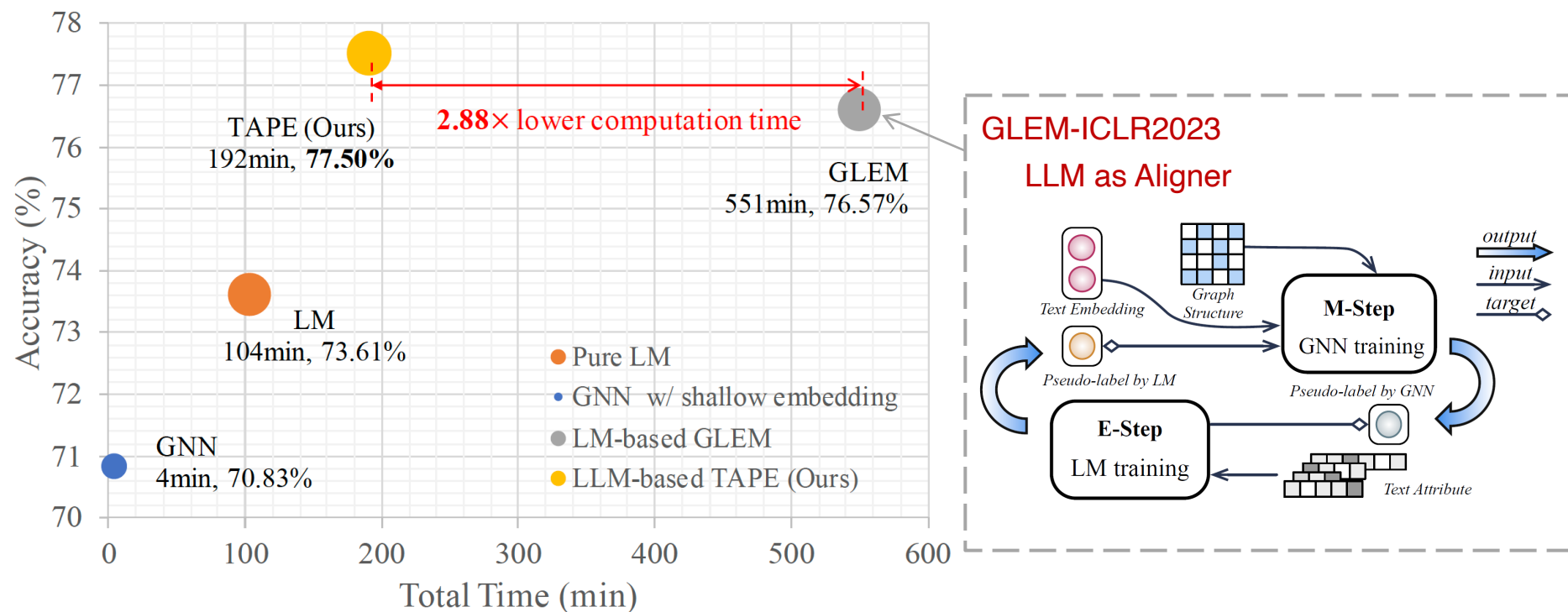


Figure 2: The performance trade-off between node classification accuracy and total training time on ogbn-arxiv [13] for various training approaches that combine language models (LMs) and graph neural networks (GNNs). The experiment employs DeBERTa-base [12] as the LM backbone and RevGAT [17] as the GNN backbone, with the size of the marker indicating the number of parameters.

2. TAPE

4. 实验结果

- 变体实验

Table 3: Ablation study on the ogbn-arxiv dataset, showing the effects of different node features on the performance. Node features include the original text attributes (h_{orig}), the explanations (h_{expl} and predicted h_{pred}) generated by LLM, and the proposed method (h_{TAPE}). Results are averaged over 4 runs with 4 different seeds. The best results are **in bold**.

Method		h_{orig}	h_{expl}	h_{pred}	h_{TAPE}
GCN	val	0.7624 ± 0.0007	0.7577 ± 0.0008	0.7531 ± 0.0006	0.7642 ± 0.0003
	test	0.7498 ± 0.0018	0.7460 ± 0.0013	0.7400 ± 0.0007	0.7520 ± 0.0003
SAGE	val	0.7594 ± 0.0012	0.7631 ± 0.0016	0.7612 ± 0.0010	0.7768 ± 0.0016
	test	0.7420 ± 0.0018	0.7535 ± 0.0023	0.7524 ± 0.0015	0.7672 ± 0.0007
RevGAT	val	0.7588 ± 0.0021	0.7568 ± 0.0027	0.7550 ± 0.0015	0.7785 ± 0.0016
	test	0.7504 ± 0.0020	0.7529 ± 0.0052	0.7519 ± 0.0031	0.7750 ± 0.0012

目录

1. 概述
 - 背景
 - 研究路线
 - LLM as Encoder、LLM as Predictor、LLM as Aligner
2. LLM as Encoder
 - TAPE -- ICLR2024
3. LLM as Predictor
 - GraphGPT -- SIGIR2024

3. LLM prexxx



GraphGPT: Graph Instruction Tuning for Large Language Models

Jiabin Tang¹, Yuhao Yang², Wei Wei², Lei Shi³,
Lixin Su³, Suqi Cheng³, Dawei Yin³ and Chao Huang^{1,2*}

¹Musketeers Foundation Institute of Data Science,

²Department of Computer Science, University of Hong Kong, ³Baidu Inc.

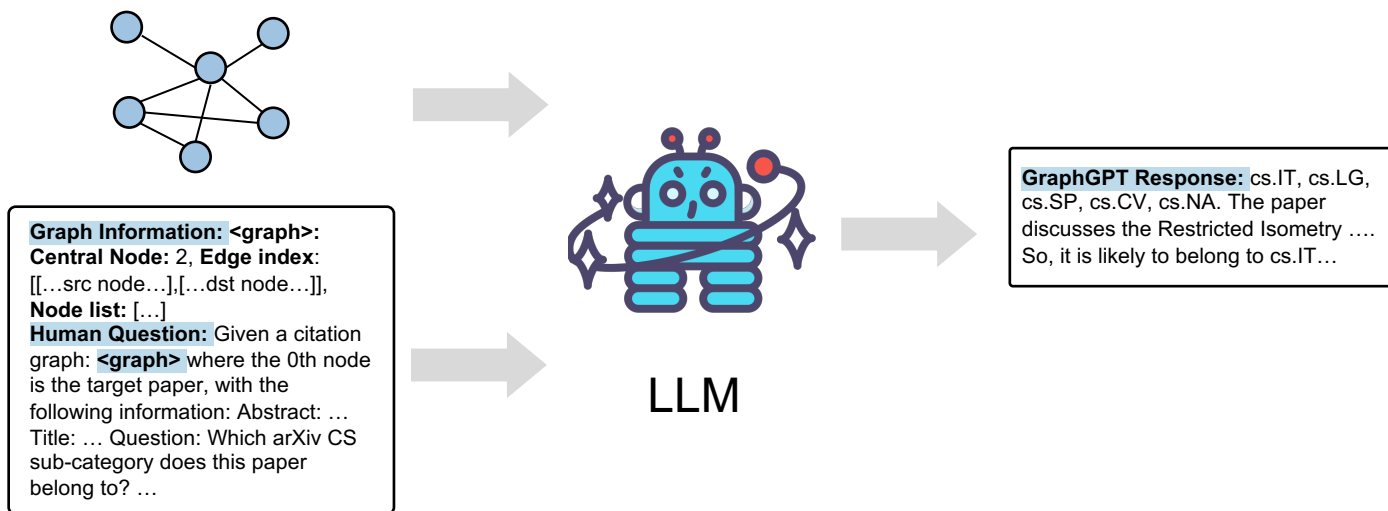
Project Page: <https://GraphGPT.github.io>, **Github:** <https://github.com/HKUDS/GraphGPT>

3. GraphGPT

1. 背景

近年来，**图神经网络**被认为是建模图数据最有力的工具之一。然而，不同图之间的语义信息大不相同，这使得很难用一个统一的图神经网络来建模所有任务。另一方面，**大语言模型**在各种任务上展现出了惊人能力，尤其在从未见过任务上的可迁移能力。

因此，本文尝试**让大语言模型能够理解图结构**，并直接完成图结构相关的下游任务。



3. GraphGPT

2. 创新

然而，如何有效地将 LLM 与 graph 结合，不可回避的挑战和研究问题如下：

- 如何有效地将 graph 输入到 LLM 中，是用自然语言的方式还是其他格式？
- 如何赋予 LLM 理解 graph 的能力，让 LLM 真正理解 graph 结构数据？

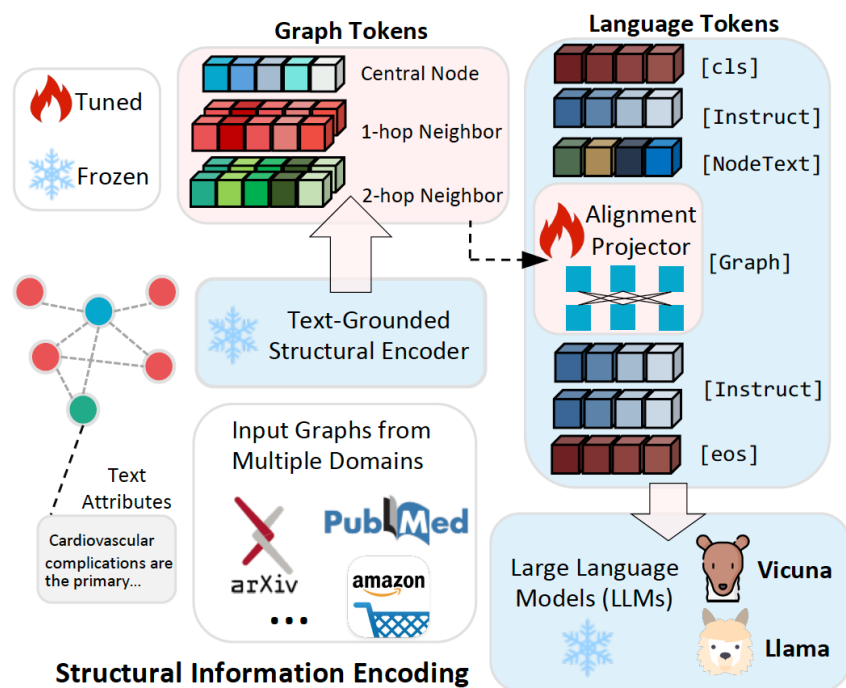
为了解决上述挑战，本文提出了GraphGPT模型，本文的创新如下：

- ✓ 提出基于文本的图结构编码器，将图结构的编码与自然语言空间对齐
- ✓ 提出双阶段指令微调范式，包括自监督的指令微调和特定任务的指令微调

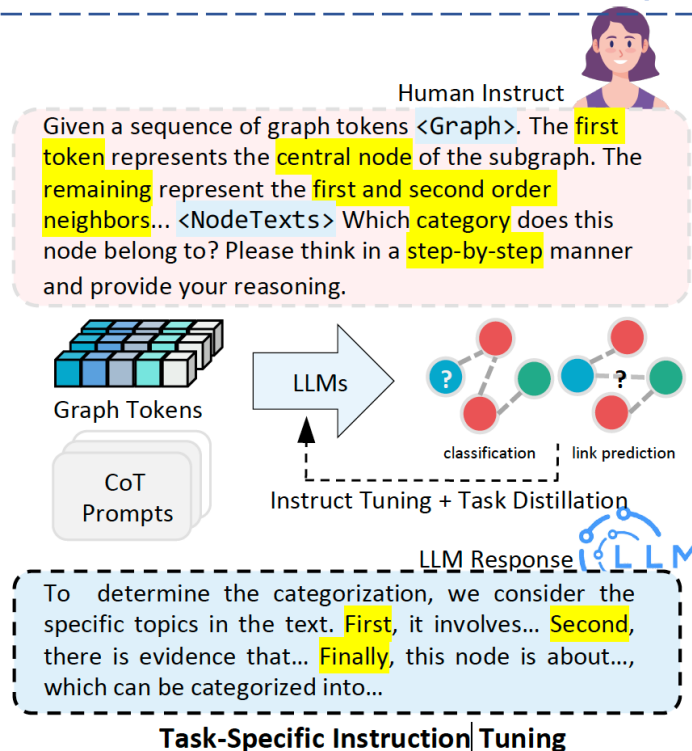
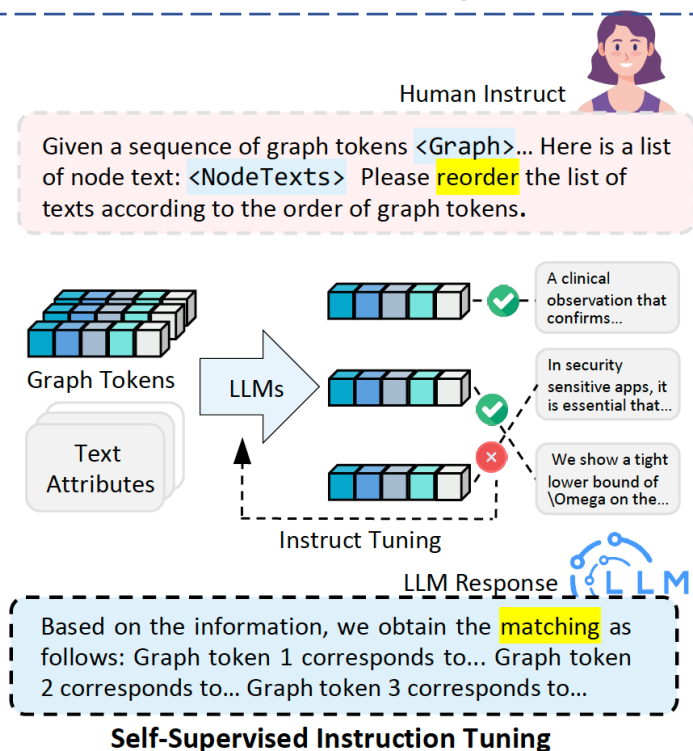
3. GraphGPT

3. GraphGPT模型

基于文本的图结构编码模块



双阶段指令微调范式（自监督的指令微调和特定任务的指令微调）



3. GraphGPT

(1) 图结构编码模块

为了将 graph 结构 输入到 LLM 中，目前主要有两种解决方案：

- 不使用图结构

不提供邻居节点的信息 → 效果欠佳

- 基于文本的图结构

在文本提示中描述图的连接信息 → 输入上下文长度过长

为了解决上述问题，GraphGPT将图编码为 graph tokens。 → 融入图结构，可控的字符长度

<graph_begin> <graph_token 1>, ..., <graph_token n> <graph_end>

Input: (a) ChatGPT with Node Content Only Token Length: 615

Abstract: The use of lower precision has emerged as a popular technique ...

Title: TiM-DNN: Ternary in-Memory accelerator for Deep Neural Networks

Question: Which arXiv CS sub-category does this paper belong to? ...

Output: cs.AR, cs.AI, cs.SY, cs.ET, cs.NE. The paper presents a hardware ... Therefore, the most likely category for this paper is cs.AR ...

Input: (b) ChatGPT with Node Content and Text-based Graph Structure Token Length: 4649

Abstract: The use of lower precision has emerged as a popular technique ...

Title: TiM-DNN: Ternary in-Memory accelerator for Deep Neural Networks

With it as central node (paper 0), a citation graph can be constructed.

The list of neighbors: Paper 1: ..., ..., Paper 102: ...

The citation relations: Paper 0 cites Paper 1, ..., ... cites Paper 102.

Question: Which arXiv CS sub-category does this paper belong to? ...

Output: Based on the title and Abstract, the paper is likely to belong: 1. cs.AR (Hardware Architecture) ...

Input: (c) GraphGPT Token Length: 750

Given a citation graph: <graph> where the 0th node is the target paper, with the following information:

Abstract: The use of lower precision has emerged as a popular technique ...

Title: TiM-DNN: Ternary in-Memory accelerator for Deep Neural Networks

Question: Which arXiv CS sub-category does this paper belong to? ...

Output: Based on the title and abstract, we can identify the following CS sub-categories that are most likely to be relevant: 1. cs.LG ...

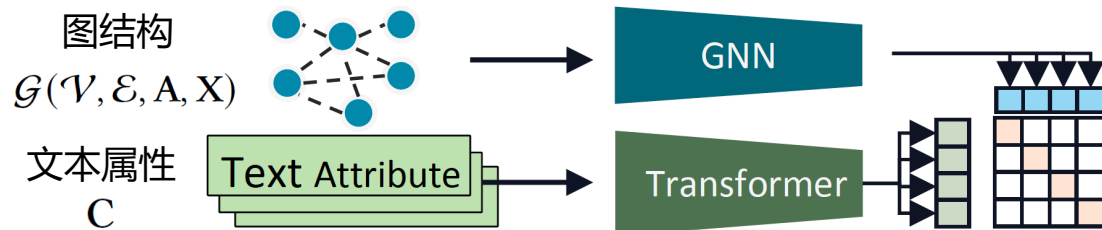
Ground Truth: cs.LG, Machine Learning



3. GraphGPT

(1) 图结构编码模块

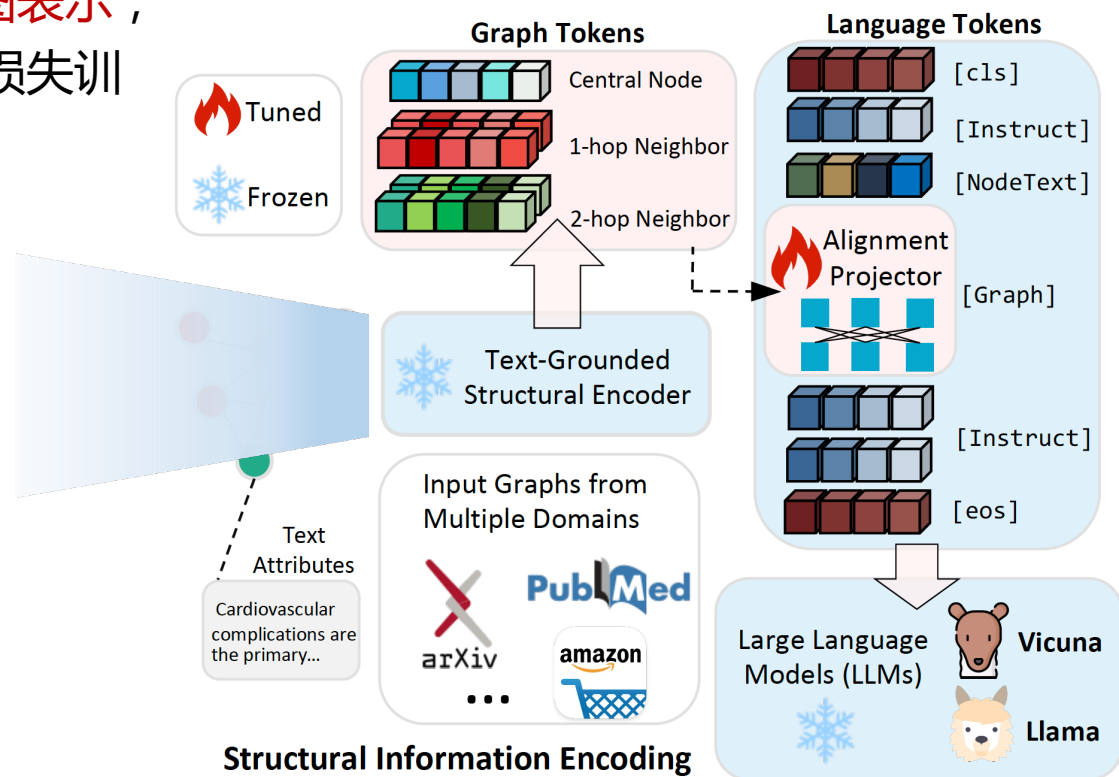
具体地，对于文本属性图，首先利用**GNN学习图表示**，并利用**Transformer学习文本表示**，然后采用对比损失训练模型，统一图空间和文本空间。



$$\mathbf{H} = f_{\mathbf{G}}(\mathbf{X}), \mathbf{T} = f_{\mathbf{T}}(\mathbf{C}), \hat{\mathbf{H}} = \text{norm}(\mathbf{H}), \hat{\mathbf{T}} = \text{norm}(\mathbf{T})$$

$$\Gamma_i = (g_i^{(1)}(\hat{\mathbf{H}})g_i^{(2)}(\hat{\mathbf{T}})^{\top}) \cdot \exp(\tau)$$

$$\mathcal{L} = \sum_i \frac{1}{2} \lambda_i (\text{CE}(\Gamma_i, \mathbf{y}) + \text{CE}(\Gamma_i^{\top}, \mathbf{y}))$$

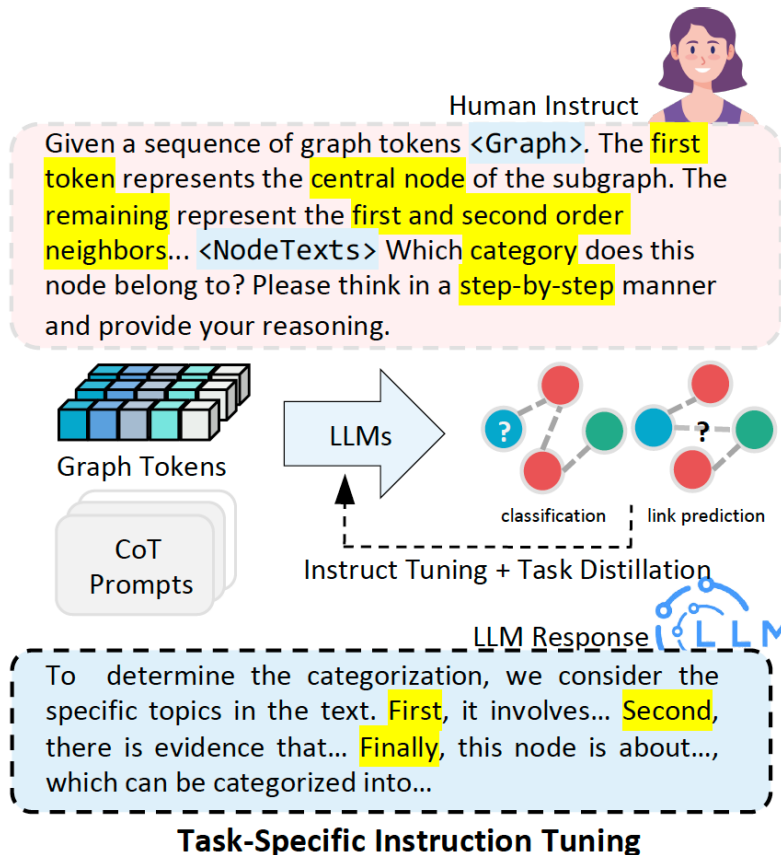
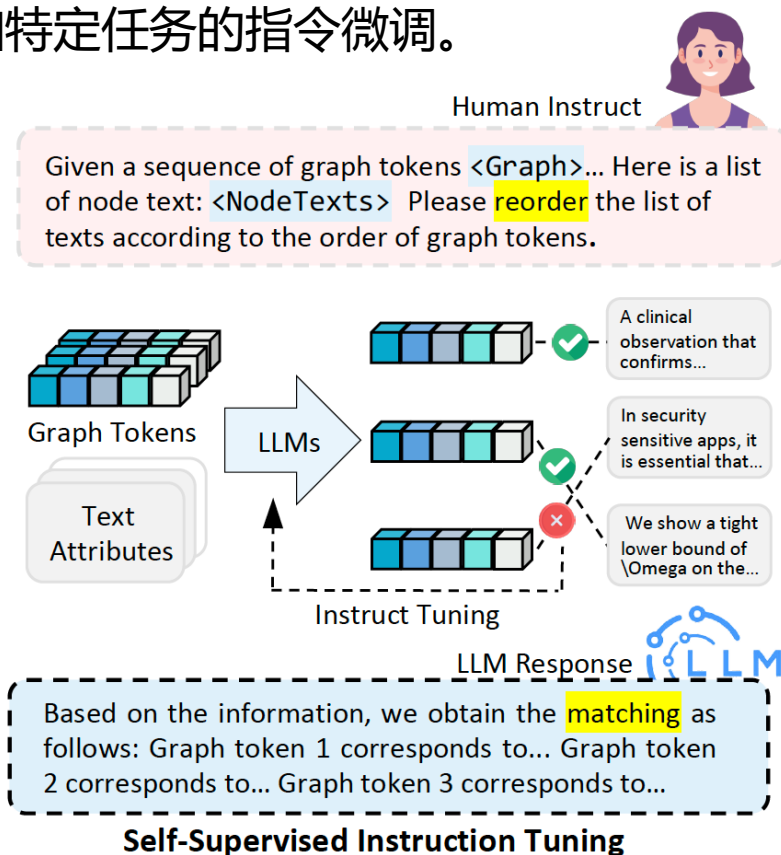


3. GraphGPT

(2) 双阶段指令微调范式

其目的是赋予 LLM 理解 graph 的能力，让 LLM 与 Graph 数据对齐。主要包括两个阶段：自监督的指令微调和特定任务的指令微调。

任务1: graph matching
根据graph tokens的序列对节点文本信息列表进行重新排序。



任务1: Node classification

任务2: Link Prediction

3. GraphGPT

(2) 双阶段指令微调范式

第一阶段：自监督的指令微调

本文设计了一个结构感知图匹配任务，即 **graph matching**，该任务的形式是根据 graph tokens 的序列对节点文本信息列表进行重新排序。

为了实现图匹配任务，重点在于：①设计 Prompt 提示指令；②设计微调策略

Prompt 提示指令由三部分构成：图信息、人类问题、和 GraphGPT 的问答。

Graph Information: <graph>: **Central Node:** 68442, **Edge index:** [[src node...],[dst node...]], **Node list:** [...]

Human Question: Given a sequence of graph tokens <graph> that constitute a subgraph of a citation graph, Here is a list of paper titles: 1. ...2. ..., please reorder the list of papers according to the order of graph tokens.

GraphGPT Response: Based on the given graph tokens and the list of paper titles, we obtain the matching of graph tokens and papers: Graph token 1 corresponds to smt based induction methods for timed systems. Graph token 2 corresponds to ...

3. GraphGPT

(2) 双阶段指令微调范式

第一阶段：自监督的指令微调

示例：

Table 8: Instruction template for the graph matching task.

Graph Matching

Input: Given a sequence of graph tokens `<graph>` that constitute a subgraph of a citation graph, where the first token represents the central node of the subgraph, and the remaining nodes represent the first and second order neighbors of the central node. Each graph token contains the title and abstract information of the paper at this node. Here is a list of paper titles: {1. Paper 1. 2. Paper 2. ... n. Paper n}, please reorder the list of papers according to the order of graph tokens (i.e., complete the matching of graph tokens and papers).

Output: Based on the given graph tokens and the list of paper titles, we obtain the matching of graph tokens and papers as follows: {Graph token 1 corresponds to Paper Q, Graph token 2 corresponds to Paper P, ... Graph token n corresponds to Paper R}.

3. GraphGPT

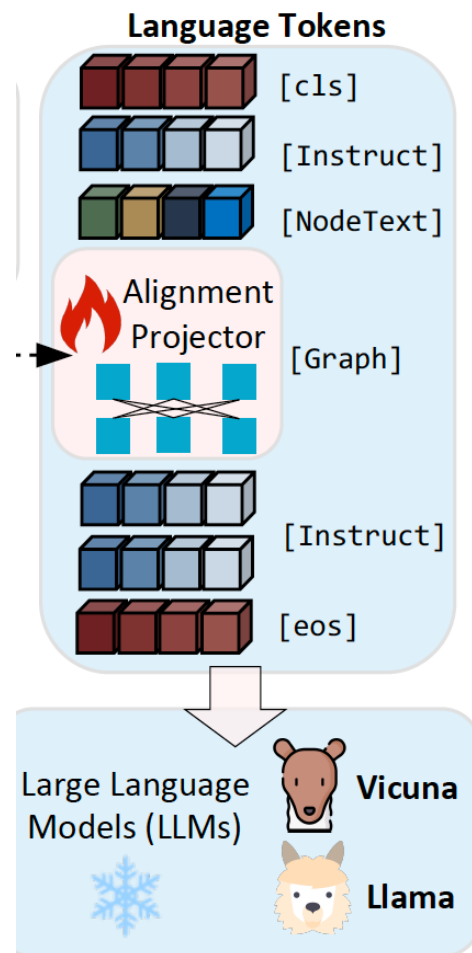
(2) 双阶段指令微调范式

第一阶段：自监督的指令微调

为了实现图匹配任务，重点在于：①设计Prompt提示指令；② 设计微调策略

微调策略：为了更高效的优化微调的过程，提出了一种轻量化对齐策略。在训练过程中，固定 LLM 的参数和 graph encoder 的参数，仅仅关注于优化前文提到的 Projector 的参数。

考虑到 graph matching 的过程是无监督的，这将有可能利用来自不同领域的大量未标注的图数据以增强学习到的 projector 的泛化能力。



3. GraphGPT

(2) 双阶段指令微调范式

第二阶段：特定任务的指令微调

本文在引文网络上设置了两个特定任务，以满足不同图学习任务的特定约束和要求，例如Node classification和Link Prediction任务。

为了实现特定任务，重点在于：①设计Prompt提示指令；② 设计微调策略

Graph Information: <graph>: **Central Node:** 2, **Edge index:** [...src node...],[...dst node...], **Node list:** [...]

Human Question: Given a citation graph: <graph> where the 0th node is the target paper, with the following information: Abstract: ... Title: ... Question: Which arXiv CS sub-category does this paper belong to? ...

GraphGPT Response: cs.IT, cs.LG, cs.SP, cs.CV, cs.NA. The paper discusses the Restricted Isometry So, it is likely to belong to cs.IT...

Graph Information: <graph>: **Central Node 1:** 8471, **Edge index 1:** [...src node...],[...dst node...], **Node list 1:** [...]

<graph>: **Central Node 2:** 19368, **Edge index 2:** [...src node...],[...dst node...], **Node list 2:** [...]

Human Question: Given a sequence of graph tokens: <graph> that constitute a subgraph of a citation graph, Abstract: ... Title: ... and the other sequence of graph tokens: <graph>, ... Abstract: ... Title: ..., are these two central nodes connected? Give me an answer of "yes" or "no".

GraphGPT Response: Yes, they are connected. Based on the first paper, And the second paper proposes

3. GraphGPT

4. 实验结果

- 数据集

3个TAG数据集 : Cora、Pubmed、 OGB-arxiv

- 基准模型

GNN模型 : GCN、 GraphSAGE、 GAT、 Rev GNN、 DGI、 ... ;

LLM模型 : Baichuan-7B, vicuna-7B-v1.1, and vicuna-7B-v1.5

3. GraphGPT

4. 实验结果

Table 1: Performance comparison of various methods on node classification under both supervised and zero-shot settings.

Dataset	Arxiv-Arxiv		Arxiv-PubMed		Arxiv-Cora		(Arxiv+PubMed)-Cora		(Arxiv+PubMed)-Arxiv	
Model	Accuracy	Macro-F1	acc	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
MLP	0.5179	0.2536	0.3940	0.1885	0.0258	0.0037	0.0220	0.0006	0.2127	0.0145
GraphSAGE	0.5480	0.3290	0.3950	0.1939	0.0328	0.0132	0.0132	0.0029	0.1281	0.0129
GCN	0.5267	0.3202	0.3940	0.1884	0.0214	0.0088	0.0187	0.0032	0.0122	0.0008
GAT	0.5332	0.3118	0.3940	0.1884	0.0167	0.0110	0.0161	0.0057	0.1707	0.0285
RevGNN	0.5474	0.3240	0.4440	0.3046	0.0272	0.0101	0.0217	0.0016	0.1309	0.0126
DGI	0.5059	0.2787	0.3991	0.1905	0.0205	0.0011	0.0205	0.0011	0.5059	0.2787
GKD	0.5570	0.1595	0.3645	0.2561	0.0470	0.0093	0.0406	0.0037	0.2089	0.0179
GLNN	0.6088	0.3757	0.4298	0.3182	0.0267	0.0115	0.0182	0.0092	0.3373	0.1115
NodeFormer	0.5922	0.3328	0.2064	0.1678	0.0152	0.0065	0.0144	0.0053	0.2713	0.0855
DIFFormer	0.5986	0.3355	0.2959	0.2503	0.0161	0.0094	0.0100	0.0007	0.1637	0.0234
baichuan-7B	0.0946	0.0363	0.4642	0.3876	0.0405	0.0469	0.0405	0.0469	0.0946	0.0363
vicuna-7B-v1.1	0.2657	0.1375	0.5251	0.4831	0.1090	0.0970	0.1090	0.0970	0.2657	0.1375
vicuna-7B-v1.5	0.4962	0.1853	0.6351	0.5231	0.1489	0.1213	0.1489	0.1213	0.4962	0.1853
GraphGPT-7B-v1.1-cot	0.4913	0.1728	0.6103	0.5982	0.1145	0.1016	0.1250	0.0962	0.4853	0.2102
GraphGPT-7B-v1.5-stage2	0.7511	0.5600	0.6484	0.5634	0.0813	0.0713	0.0934	0.0978	0.6278	0.2538
GraphGPT-7B-v1.5-std	0.6258	0.2622	0.7011	0.6491	0.1256	0.0819	0.1501	0.0936	0.6390	0.2652
GraphGPT-7B-v1.5-cot	0.5759	0.2276	0.5213	0.4816	0.1813	0.1272	0.1647	0.1326	0.6476	0.2854
p-val	$2.26e^{-9}$	$1.56e^{-10}$	$2.22e^{-7}$	$1.55e^{-9}$	$1.04e^{-9}$	$9.96e^{-6}$	$7.62e^{-8}$	$1.97e^{-7}$	$1.5e^{-13}$	$4.63e^{-6}$

3. GraphGPT

4. 实验结果

Table 9: Instruction template for node classification and link prediction on different datasets.

Node Classification (Arxiv in the COT manner)
Input: Given a citation graph: <graph> where the 0th node is the target paper, with the following information: Abstract: {abstract of the central node}. Title: {title of the central node}. Question: Which arXiv CS sub-category does this paper belong to? Give 5 likely arXiv CS sub-categories as a comma-separated list ordered from most to least likely, in the form "cs.XX". Please think about the categorization in a step by step manner and avoid making false associations. Then provide your reasoning.
Output: Based on the information, {reasoning process and answers}.
Arxiv in the standard manner)
Input: Given a citation graph: <graph> where the 0th node is the target paper, and other nodes are its one-hop or multi-hop neighbors, with the following information: Abstract: {abstract of the central node}. Title: {title of the central node}. Question: Which arXiv CS sub-category does this paper belong to? Give the most likely arXiv CS sub-categories of this paper directly, in the form "cs.XX" with full name of the category.
Output: {the ground-truth answer}.

3. GraphGPT

4. 实验结果

Table 2: Performance comparison of various instruction mixtures in supervised learning on the Arxiv dataset and the zero-shot setting on the Cora dataset for node classification.

Dataset	Supervision. on Arxiv		Zero Shot on Cora	
Model	Acc	Macro-F1	Acc	Macro-F1
MLP	0.5179	0.2536	0.0220	0.0006
GraphSAGE	0.5480	0.3290	0.0132	0.0029
GCN	0.5267	0.3202	0.0187	0.0032
GAT	0.5332	0.3118	0.0161	0.0057
RvGNN	0.5474	0.3240	0.0217	0.0016
DGI	0.5059	0.2787	0.0205	0.0011
GKD	0.5570	0.1595	0.0406	0.0037
GLNN	0.6088	0.3757	0.0182	0.0092
NodeFormer	0.5922	0.3328	0.0144	0.0053
DIFformer	0.5986	0.3355	0.0100	0.0007
baichuan-7b	0.0946	0.0363	0.0405	0.0469
vicuna-7B-v1.1	0.2657	0.1375	0.1090	0.0970
vicuna-7B-v1.5	0.4962	0.1853	0.1489	0.1213
Arxiv-std + PubMed-std	0.6390	0.2652	0.1501	0.0936
Arxiv-cot + PubMed-cot	0.6476	0.2854	0.1647	0.1326
Arxiv-mix + PubMed-mix	0.6139	0.2772	0.1544	0.1048
Arxiv-std + PubMed-std + Link	0.5931	0.2238	0.1847	0.1579
Arxiv-mix + Pubmed-mix + Link	0.6874	0.3761	0.1836	0.1494

Table 3: Module ablation study under both supervised and zero-shot settings to analyze the individual contributions.

Dataset	Arxiv-Arxiv		Arxiv-PubMed		Arxiv-Cora	
Variant	Acc	Mac-F1	Acc	Mac-F1	Acc	Mac-F1
w/o GS	0.4962	0.1853	0.6351	0.5231	0.1489	0.1213
w/o LR	0.5807	0.2462	0.2523	0.1925	0.0050	0.0016
ours	0.6258	0.2622	0.7011	0.6491	0.1813	0.1272

Note: GS (Graph Structural), LR (Llm Reasoning)

Table 4: Study on the time and space efficiency of our GraphGPT during both the training and inference stages.

Variants	Training Time	Tuned Parameters	GPU Occupy
Stage-1-tune	OOM	6,607,884,288	OOM
Stage-1-freeze	22:53:33	131,612,672	39517.75
improvement	-	↓ × 50.21	-
Stage-2-tune	OOM	6,607,884,288	OOM
Stage-2-freeze	03:44:35	131,612,672	38961.75
improvement	-	↓ × 50.21	-

总结

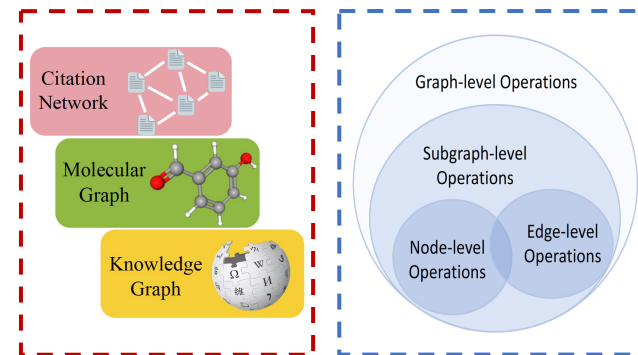
Graph+LLM

1. 挑战

来自不同领域的图数据；图上的任务呈现多样化；

2. 研究路线

- Pretrained GNNs
 - GraphPrompt、GraphMAE、Graph-GNN
- LLM+GNNs
 - LLM as Encoder、LLM as Predictor、LLM as Aligner
- Graph Foundation Models
 - OFA、Ultra、OpenGraph



总结

Graph Foundation Model

1. 未来趋势

图基础模型/ 图大模型，是图领域的一个新兴研究课题，一种能够跨不同图和任务进行泛化的图模型，但是目前**范式尚不清晰**。

2. 关键问题

- Graph Representation
 - 构建图和文本对齐数据、将图转换为自然语言、寻找中间桥梁
- Graph Data
 - 图数据少规模小，缺乏公认的基准，如测试数据和指标
- Graph Models
 - 骨干架构（GNN/ Graph Transformer/ LLM）、图预训练、图微调

论文列表

- LLM for Graphs
 - Harnessing Explanations: LLM-to-LM Interpreter for Enhanced Text-Attributed Graph Representation Learning
 - Label-free Node Classification on Graphs with Large Language Models (LLMs)
 - GraphGPT: Graph Instruction Tuning for Large Language Models
 - GraphTranslator: Aligning Graph Model to Large Language Model for Open-ended Tasks
 - Can Language Models Solve Graph Problems in Natural Language?
 - Train Your Own GNN Teacher: Graph-Aware Distillation on Textual Graphs
 - Distilling Large Language Models for Text-Attributed Graph Learning
 - Large Language Models on Graphs- A Comprehensive Survey
 - A Survey of Graph Meets Large Language Model: Progress and Future Directions
 - Towards Graph Foundation Models: A Survey and Beyond
 - OpenGraph- Towards Open Graph Foundation Models