图基础模型应用

Graph Foundation Model Application

SHY (ASCII LAB) 2024/9/20

OUTLINE

1 图基础模型应用方向概述及论文统计

2 图基础模型部分应用的相关工作

- 2.1 Graph Anomaly Detection
- 2.2 Graph Question Answering

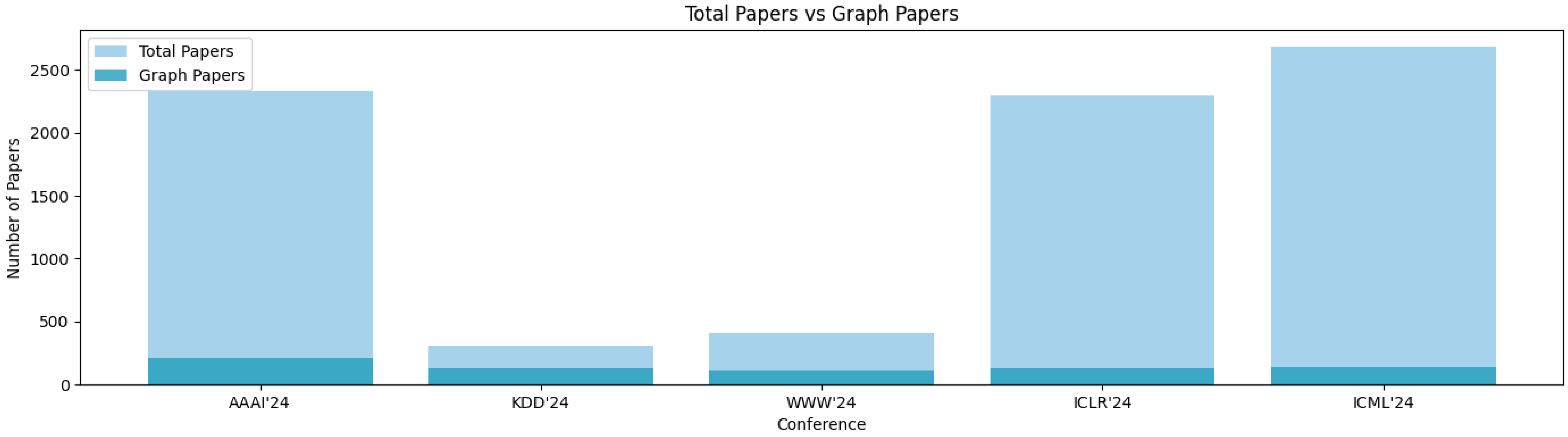
3 总结

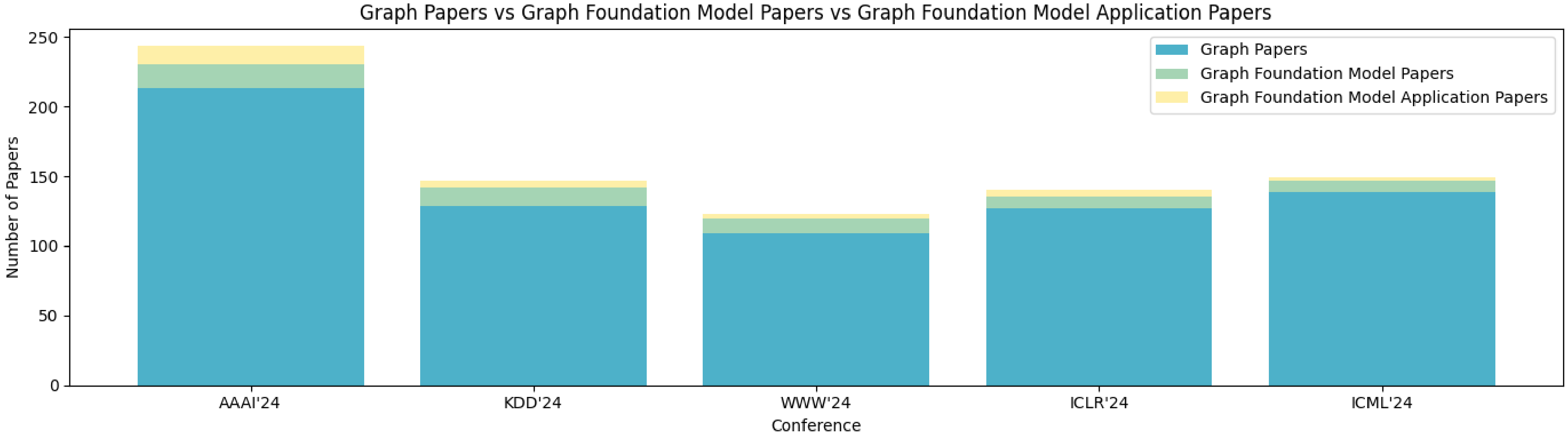
- 3.1 未来研究方向
- 3.2 研究团队

1 图基础模型应用方向概述及论文统计

1.1 图基础模型应用概述

- Graph Foundation Model Application
  - Knowledge Graph Completion
    - Towards Foundation Models For Knowledge Graph Reasoning
    - Making large language models perform better in knowledge graph completion
  - Graph Question Answering
    - Zero-shot Logical Query Reasoning on any Knowledge Graph
    - GRAPHLLM: BOOSTING GRAPH REASONING ABILITY OF LARGE LANGUAGE MODEL
    - Can Language Models Solve Graph Problems in Natural Language?
    - GraphWiz: An Instruction-Following Language Model for Graph Problems
    - Let Your Graph Do the Talking: Encoding Structured Data for LLMs
  - Graph Anomaly Detection
    - AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models
  - Algorithm Reasoning
    - A Generalist Neural Algorithmic Learner
  - In Context Learning
    - PRODIGY: Enabling In-context Learning Over Graphs
  - Drug Discovery
    - A systematic survey of chemical pre-trained models
    - Diet-ODIN: A Novel Framework for Opioid Misuse Detection with Interpretable Dietary Patterns
    - Enhancing Activity Prediction Models in Drug Discovery with the Ability to Understand Human Language
  - Molecule Graphs
    - A molecular multimodal foundation model associating molecule graphs with natural language
    - Can Large Language Models Empower Molecular Property Prediction?
  - Transportation system
    - Building transportation foundation model via generative graph transformer
  - Recommendation
    - GraphPro: Graph Pretraining and Prompt Learning for Recommendation


Total Papers vs Graph Papers
Total Papers
Graph Papers
Number of Papers
2500
2000
1500
1000
500
0
AAAI'24
KDD'24
WWW'24
ICLR'24
ICML'24
Conference


Graph Papers vs Graph Foundation Model Papers vs Graph Foundation Model Application Papers
Graph Papers
Graph Foundation Model Papers
Graph Foundation Model Application Papers
Number of Papers
250
200
150
100
50
0
AAAI'24
KDD'24
WWW'24
ICLR'24
ICML'24
Conference

难做好

容易占坑

2 图基础模型部分应用相关工作

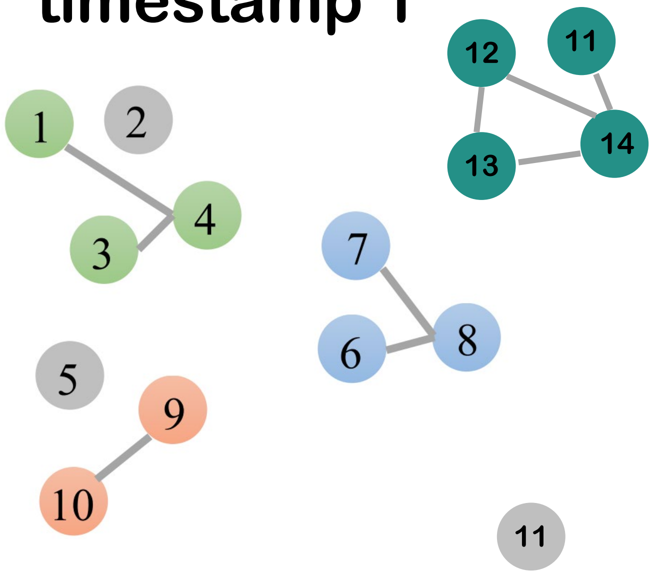
- 2.1 Graph Anomaly Detection
- 2.2 Graph Question Answering

Dynamic Graph

Definition:

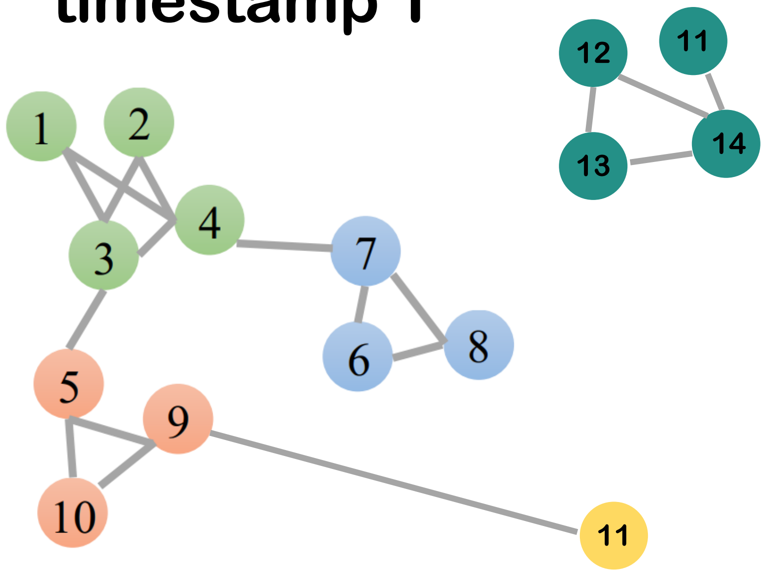
- **动态图：** 动态图是指一系列随时间变化的静态图快照，可以表示为$G = \{\mathcal{G}^1, \dots, \mathcal{G}^T\}$。其中$T$是时间步的总数。每个时间步$t$的快照$\mathcal{G}_t = (\mathcal{V}, \mathcal{E}^t)$是一个加权无向图，包含共享的节点集$V$、边集$\mathcal{E}^t$以及时间步$t$的加权邻接矩阵$A^t$。

timestamp 1



••• •••

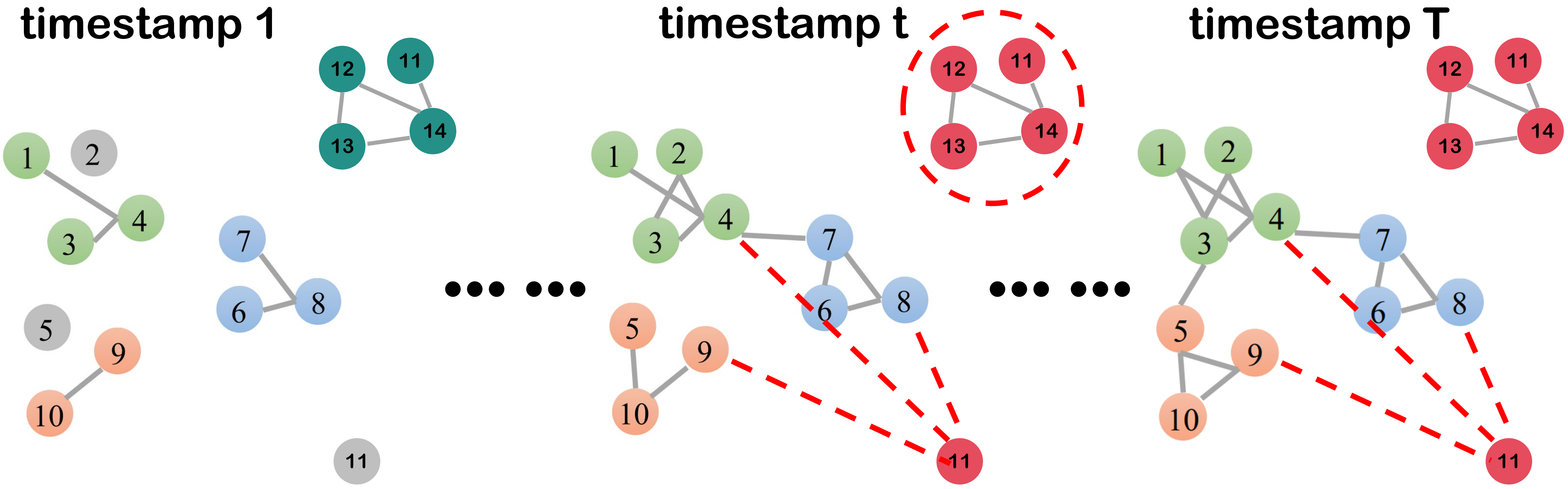
timestamp T



Dynamic Graph Anomaly Detection

Definition:

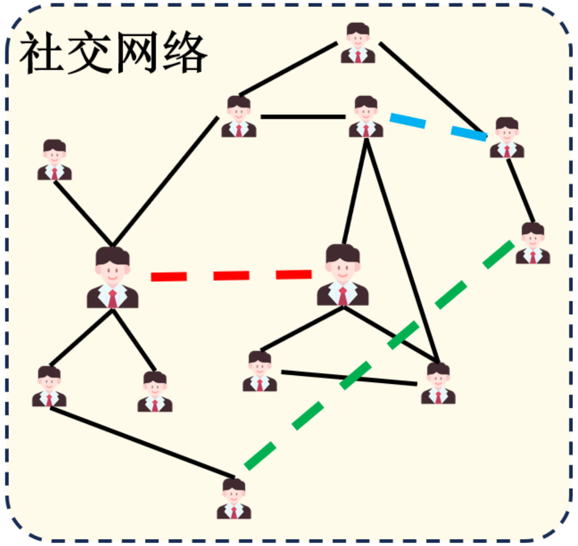
➢ **动态图异常检测：** 动态图异常检测的目标是寻找在**时间步 *t*** 或**整个时间范围**内显著偏离正常行为的节点、边或子图。



Dynamic Graph Anomaly Detection

现有**Dynamic Graph Anomaly Detection**问题:

- 在真实场景中，随着**时间**的推移，异常边缘的类型可能会发生变化，甚至出现**新的异常边缘类型**。现有方法要么被设计为检测随机插入的边缘，要么需要足够的标记数据来进行模型训练。而对于新出现的异常边缘，通常只有**少量**的标记样本可用于模型训练。因此，如何在动态图中利用**小样本**检测多种异常边缘?



三种异常类型边

AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models

Outline

Shuo Liu
Institute of Computing Technology,
Chinese Academy of Sciences

Di Yao*
Institute of Computing Technology,
Chinese Academy of Sciences
yaodi@ict.ac.cn

Lanting Fang
Beijing Institute of Technology

Zhetao Li
Jinan University

Wenbin Li
Institute of Computing Technology,
Chinese Academy of Sciences

Kaiyu Feng
Beijing Institute of Technology

Xiaowen Ji
Southeast University

Jingping Bi*
Institute of Computing Technology,
Chinese Academy of Sciences
bjp@ict.ac.cn

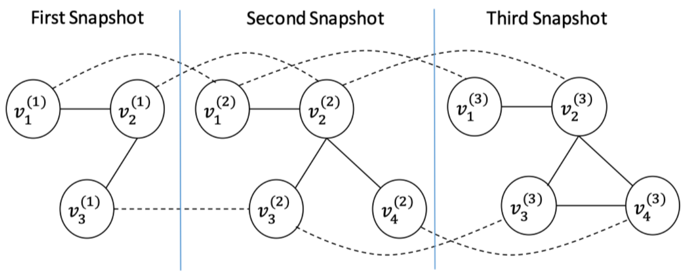
➢ **Thinking**：作者希望利用LLMs的**泛化能力**来解决动态图中**少样本**的异常边缘检测问题。通过**对齐**语言空间与图空间，结合**上下文学习**的方法，解决标注样本不足的问题，从而实现对动态图中不同异常类型边缘的有效检测。

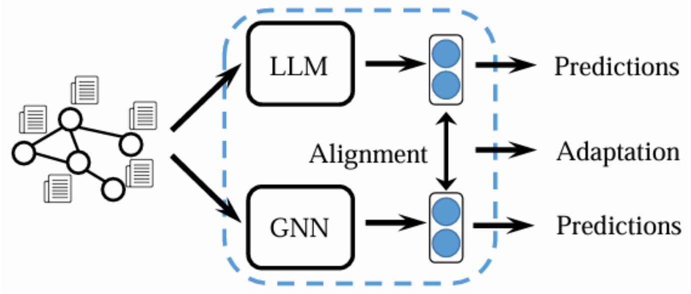
➢ **How**：实现动态图的异常边缘检测主要有以下三个挑战：

1）如何表示动态图的**结构信息**和**时间信息**；

2）如何将Graph和LLM的空间进行**对齐**；

3）如何结合**上下文**的标签信息来识别异常。



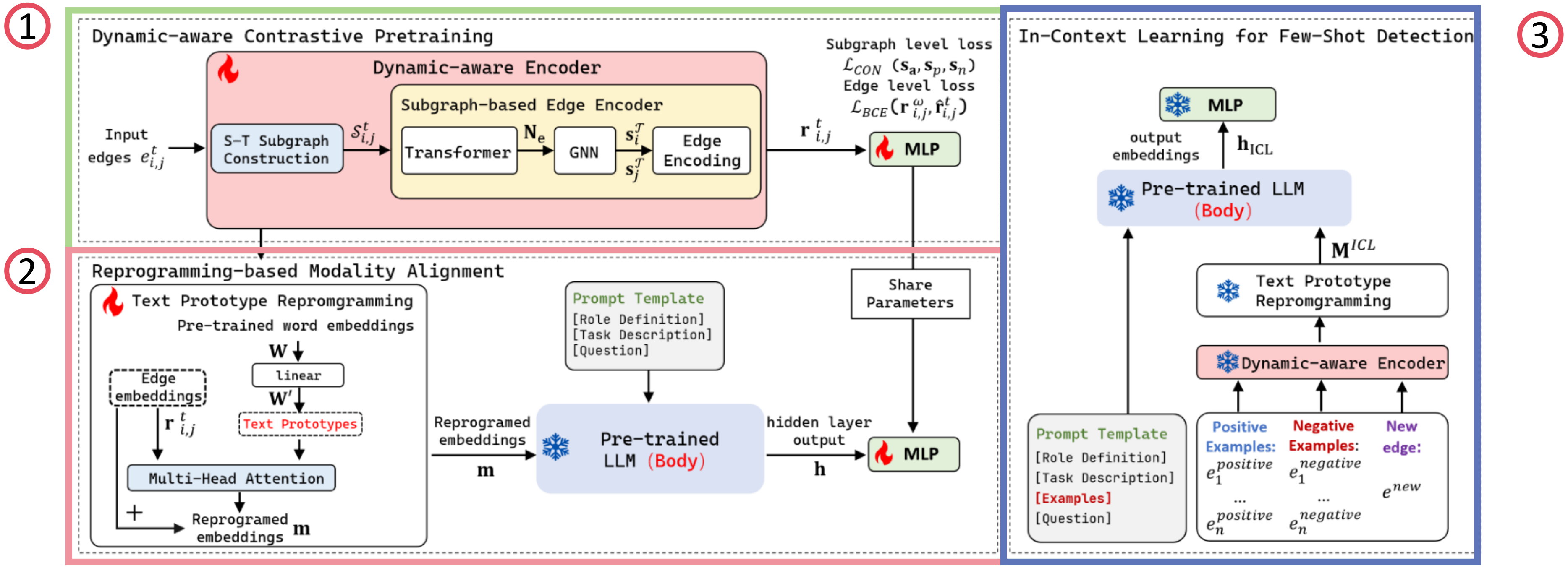


AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models, arxiv’24, ICT & Beijing Institute of Technology

Method

AnomalyLLM由三个关键模块组成：

- 1） Dynamic-aware Contrastive Pretraining
- 2） Reprogramming-based Modality Alignment
- 3） In-Context Learning for Few-Shot Detecion



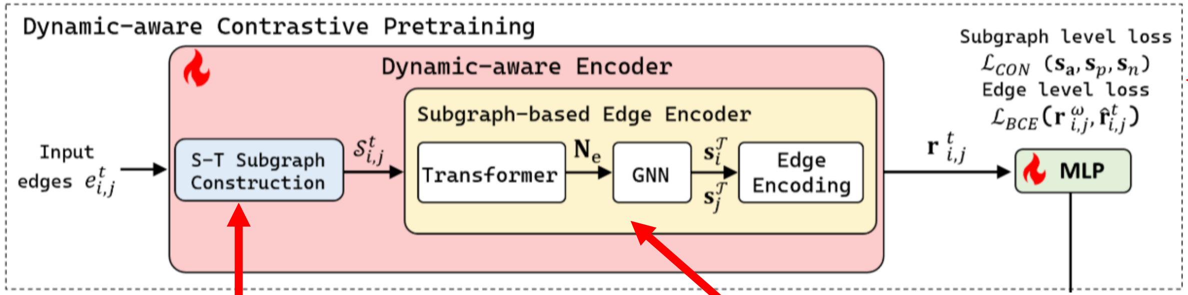
AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models, arxiv'24, ICT & Beijing Institute of Technology

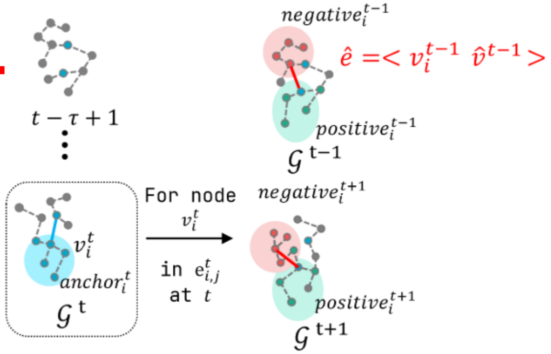
Method

1）Dynamic-aware Contrastive Pretraining

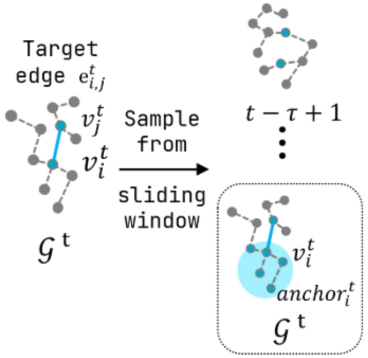
动态感知的对比预训练模块通过系统地利用**结构信息和时间信息**，结合Transformer和GNN来生成边和子图的表示，最后利用**对比学习范式**优化**边**和**子图**的嵌入表示。

Challenge1：如何表示动态图的**结构信息和时间信息**？





$$g_{i,j}^t = [g_i^t, g_j^t]$$



$$\mathcal{S}_{i,j}^t = \{g_{i,j}^\tau\} \quad \text{for } \tau = t-\Gamma+1, \ldots, t$$

$$\mathbf{z}_l = \mathbf{z}_{\text{diff}}(v_l^\tau) + \mathbf{z}_{\text{dist}}(v_l^\tau) + \mathbf{z}_{\text{temp}}(v_l^\tau)$$

$$\mathbf{Z}_e = [[\mathbf{z}_l]_{v_l \in g_{i,j}^\tau}]_{g_{i,j}^\tau \in \mathcal{S}_{i,j}^t}$$

$$\mathbf{N}_e = \text{Transformer}(\mathbf{Z}_e)$$

$$\mathbf{s}_i^\tau \in \mathbb{R}^{(K+1)\times d_{enc}}$$

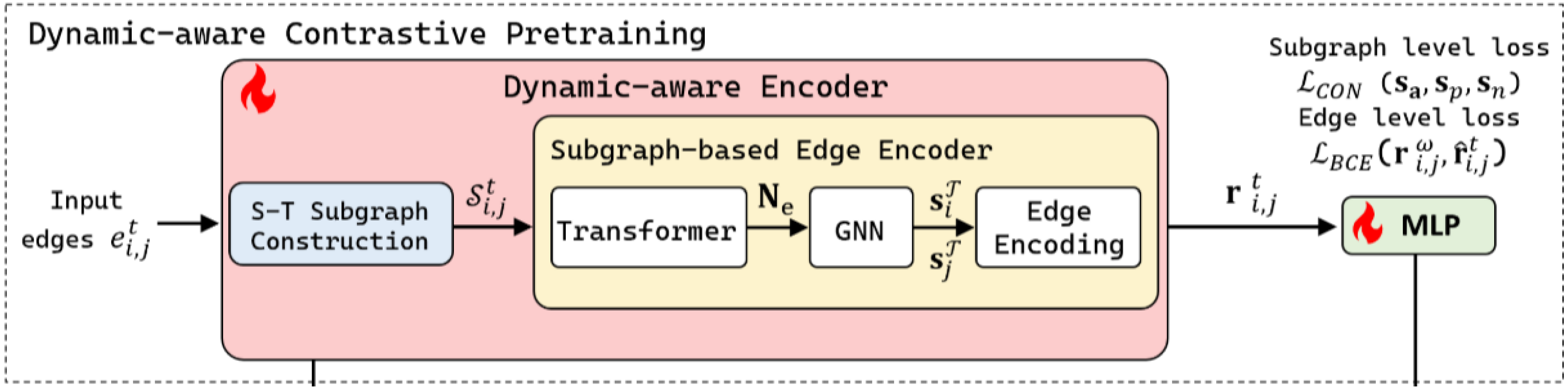
$$\mathbf{r}_{i,j}^t = \text{fc}(\text{concat}(\text{AvgPool}(\mathbf{s}_i^\tau), \text{AvgPool}(\mathbf{s}_j^\tau)))$$

Dynamic-aware Contrastive Pretraining

Q1：为什么利用对比学习范式进行预训练?

1）对比学习不需要**大量标注数据**进行训练，而是通过正负样本对的构造来学习有效的表示。

2）对比学习目标是通过**最小化**正样本距离、**最大化**负样本距离，能够显著增强嵌入表示的区分性。对于不同时间步或不同结构下的图数据，能够有效区分相似与不相似的样本。

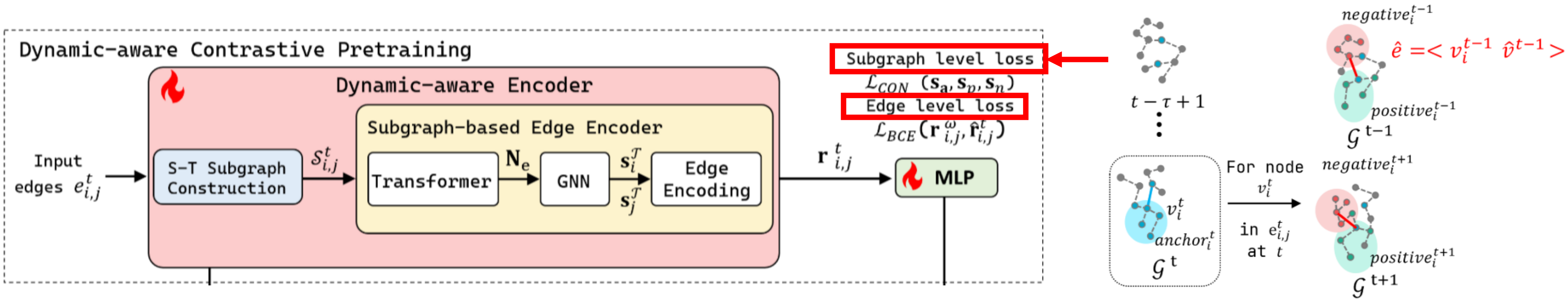


Dynamic-aware Contrastive Pretraining

Q2：对比学习有哪几种对比方式?

1）**边级别对比：**通过构造同一时间步t上，将现有的边表示$\mathbf{r}_{i,j}^{t}$和随机生成的边表示$\hat{\mathbf{r}}_{i,j}^{t}$进行对比

2）**子图级别对比：**通过在不同时间步$\{t-1,t,t+1\}$上，时间步t上的节点v_i子图作为锚点，时间步$\{t-1,t+1\}$上的节点v_i子图作为正样本，时间步$\{t-1,t+1\}$上随机节点$\hat{v}$子图作为负样本进行三元组对比。



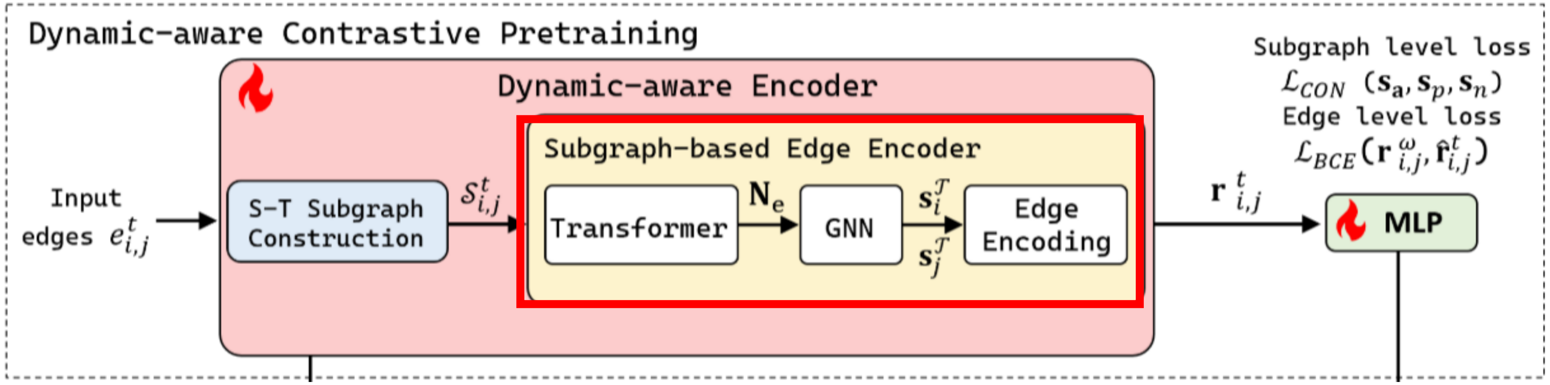
$$\mathcal{L}_{BCE}=-\left(\log\left(1-\mathrm{MLP}\left(\mathbf{r}_{i,j}^{t}\right)\right)+\log\left(\mathrm{MLP}\left(\hat{\mathbf{r}}_{i,j}^{t}\right)\right)\right)$$

$$\mathcal{L}_{con}=-\log\frac{\exp\left(\cos\left(\mathbf{s}_a,\mathbf{s}_p\right)/\delta\right)}{\exp\left(\cos\left(\mathbf{s}_a,\mathbf{s}_p\right)/\delta\right)+\exp\left(\cos\left(\mathbf{s}_a,\mathbf{s}_n\right)/\delta\right)}$$

Dynamic-aware Contrastive Pretraining

Q3：如何对边和子图进行表示呢?

假设此时获得了包含时间和结构信息的子图序列$S_{i,j}^{t}=\{[g_i,g_j]^{\tau}\}$，其中$\tau=t-\Gamma+1,\dots,t$。首先，通过Transformer层对子图序列进行编码，获取当前时间步t中子图中节点i的嵌入表示$\mathbf{N}_i^t$。然后，利用GNN层对节点的结构信息进行聚合，得到当前时间步t中时间窗口τ对于节点i的子图嵌入$\mathbf{s}_i^{\tau}\in\mathbf{s}_i^{t}$。最后，将节点$i$和$j$的子图嵌入表示$\mathbf{s}_i^{\tau}$和$\mathbf{s}_j^{\tau}$进行池化操作，得到边嵌入$\mathbf{r}_{i,j}^{t}$。

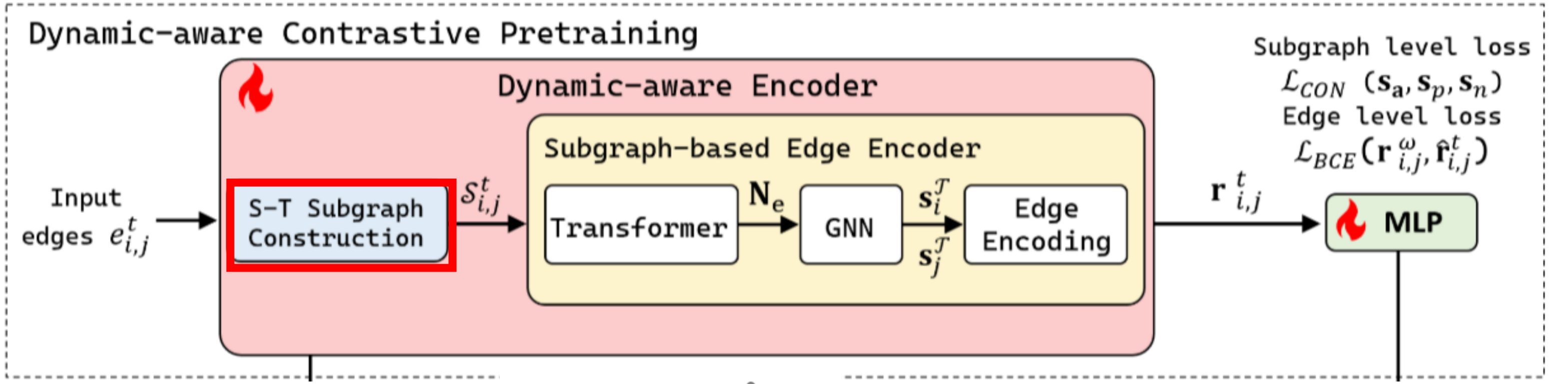


$$\mathbf{r}_{i,j}^{t}=\mathrm{fc}\left(\mathrm{concat}\left(\mathrm{AvgPool}(\mathbf{s}_i^{\tau}),\mathrm{AvgPool}(\mathbf{s}_j^{\tau})\right)\right)\quad \text{for } \tau=t-\Gamma+1,\dots,t$$

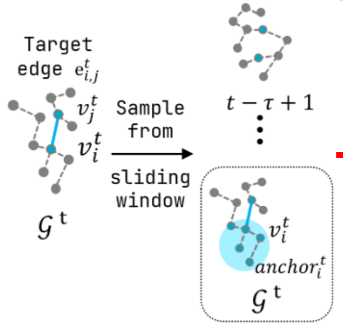
Dynamic-aware Contrastive Pretraining

Q4：如何构建包含时间和结构信息的子图序列?

首先，根据每个节点在时间t上，选择$Top-K$连接权重的节点构建子图g_i^t和g_j^t。此时$g_{i,j}^t=[g_i^t,g_j^t]$已经包含了边$e_{i,j}^t$的结构信息。接着，通过时间窗口获取Γ个时间序列图，使得序列$S_{i,j}^t$不仅包含了结构信息，还融入了时间维度的动态变化。



$$g_{i,j}^t=[g_i^t,g_j^t]$$



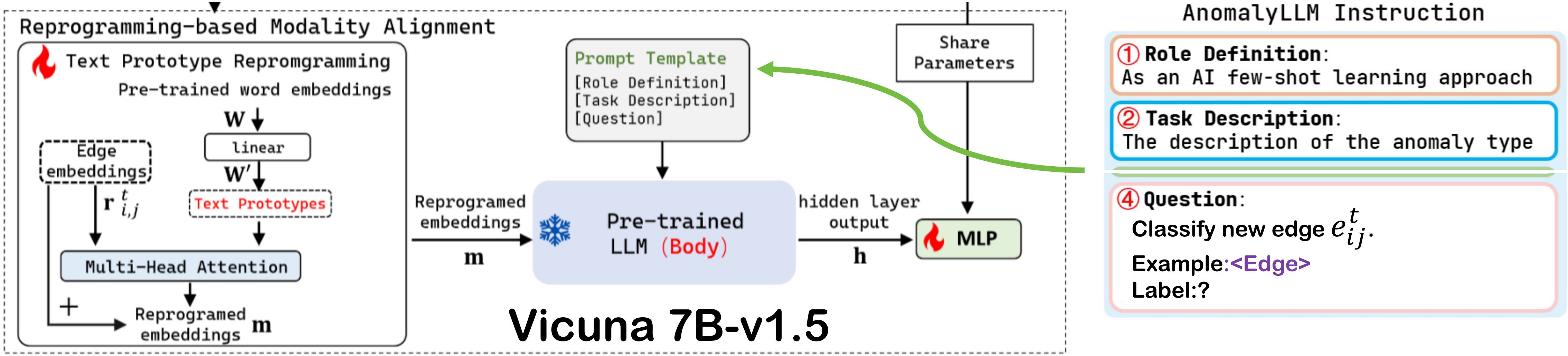
$$\mathcal{S}_{i,j}^t=\{g_{i,j}^{\tau}\} \quad \text{for } \tau=t-\Gamma+1,\ldots,t$$

Method

2）Reprogramming-based Modality Alignment

模态对齐重编码模块首先通过大语言模型生成与动态图相关词，再抽象成**文本原型**。接着，利用**注意力机制**捕捉边嵌入与文本原型之间的相互关系。随后，通过 Prompt 引导LLM生成包含语义信息和图信息的嵌入h，并使用MLP来判断边e_{ij}^{t}是否为随机生成的。

Challenge2：如何将Graph和LLM的空间进行**对齐?**

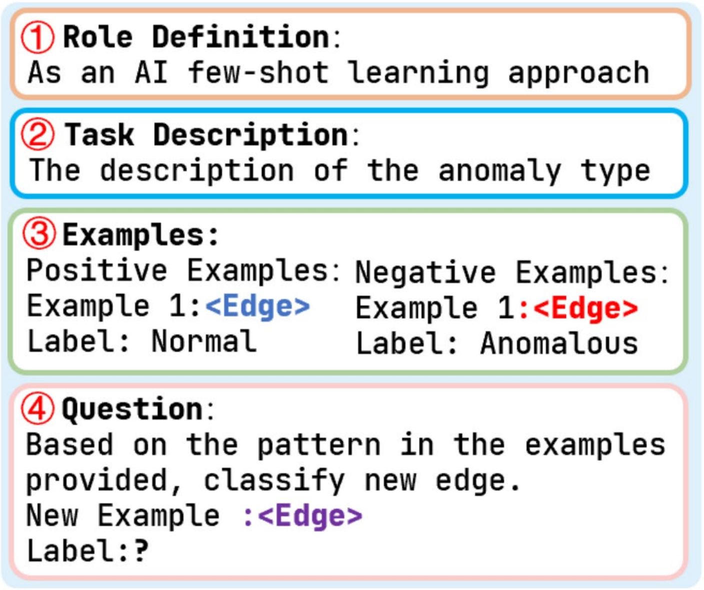


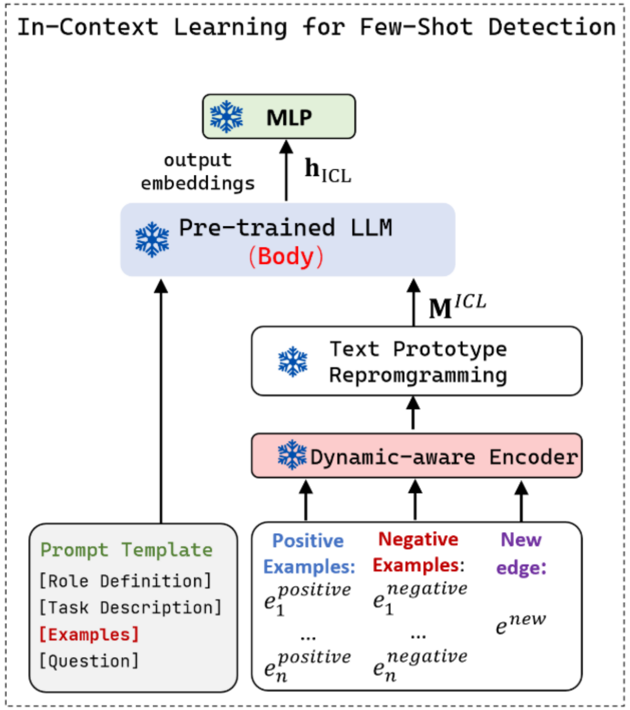
Method

3）In-Context Learning for Few-Shot Detection

由于AnomalyLLM在预训练过程中没有异常类型的信息，因此在下游预测时需要充分利用**异常类型**的标记信息。作者通过从正常边和异常边集合中选取相同数量边样本作为**上下文信息**，并结合需要预测的边e^{new}来共同预测异常边的类型。

Challenge3：如何结合**上下文**的标签信息来识别异常？





Experiment

实验1： 在两个公共动态图数据集上进行测试，注入**三种不同的异常类型**，比较三种少样本异常检测方法的性能。此外，还在两个真实世界的**不同领域**的异常数据集上进行测试。

结果： 在插入边缘异常的数据集和真实数据集都由明显的提升效果。

Table 1: Performance comparison results of few-shot anomaly detection on multiple anomaly types.

Dataset	Model	1-shot			5-shot			10-shot		
		CDA	LPL	HHL	CDA	LPL	HHL	CDA	LPL	HHL
BlogCataLog	StrGNN	0.5891	0.5756	0.5974	0.6018	0.6041	0.6122	0.6222	0.6329	0.6402
	AddGraph	0.5994	0.6023	0.5988	0.6097	0.6033	0.6104	0.6216	0.6238	0.6172
	Deep Walk	0.6102	0.6073	0.6202	0.6113	0.6122	0.6196	0.6155	0.6176	0.6154
	TGN	0.6732	0.6699	0.6919	0.7112	0.7023	0.7118	0.7263	0.7387	0.7311
	GDN	0.6733	0.6795	0.6609	0.6997	0.7051	0.7121	0.7321	0.7311	0.7319
	SAD	0.6841	0.6792	0.6411	0.7002	0.7018	0.6988	0.7342	0.7216	0.7265
	TADDY	0.6892	0.6983	0.6891	0.7148	0.7186	0.7177	0.7258	0.7326	0.7334
	AnomalyLLM	**0.8288**	**0.8334**	**0.8255**	**0.8331**	**0.8319**	**0.8407**	**0.8402**	**0.8456**	**0.8447**
UCI Message	StrGNN	0.6143	0.5956	0.5722	0.6113	0.7132	0.6512	0.6442	0.6724	0.6249
	AddGraph	0.5842	0.5466	0.5647	0.6018	0.6667	0.6321	0.4642	0.5728	0.7001
	Deep Walk	0.6198	0.6187	0.6142	0.6256	0.6263	0.6176	0.6255	0.6209	0.6197
	TGN	0.6521	0.6535	0.6643	0.7098	0.7193	0.7155	0.7335	0.7365	0.7324
	GDN	0.6577	0.6818	0.6611	0.7201	0.7289	0.7255	0.7493	0.7511	0.7546
	SAD	0.6703	0.6587	0.6693	0.7102	0.7146	0.7194	0.7416	0.7453	0.7406
	TADDY	0.6992	0.7078	0.6132	0.7204	0.7237	0.7218	0.7255	0.7278	0.7243
	AnomalyLLM	**0.8414**	**0.8358**	**0.8368**	**0.8446**	**0.8459**	**0.8424**	**0.8488**	**0.8546**	**0.8442**

Table 4: Performance on Real-World Labeled Dataset

Dataset	Method	1-shot	5-shot	10-shot
T-Finance	AddGraph	0.6126	0.6149	0.6277
	TGN	0.6646	0.6701	0.6865
	GDN	0.6672	0.6689	0.6898
	SAD	0.6724	0.6754	0.6876
	AnomalyLLM	**0.8018**	**0.8056**	**0.8087**
T-Social	AddGraph	0.6116	0.6245	0.6221
	TGN	0.6706	0.6754	0.6887
	GDN	0.6694	0.6782	0.6908
	SAD	0.6779	0.6746	0.6805
	AnomalyLLM	**0.8101**	**0.8187**	**0.8206**

vicuna-7B-v1.5

2 图基础模型部分应用相关工作

- 2.1 Graph Anomaly Detection
- 2.2 Graph Question Answering

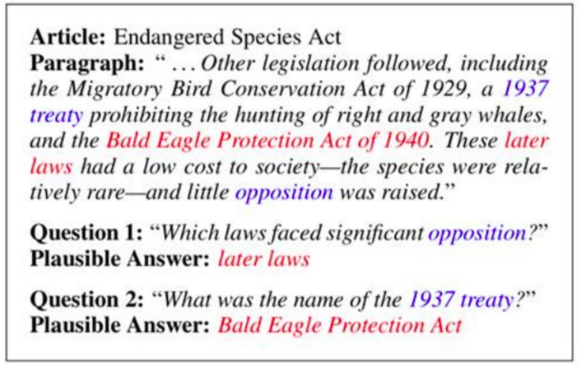
中国科学院信息工程研究所 INSTITUTE OF INFORMATION ENGINEERING,CAS | ASCII

Graph Question Answering

Background:

➢ ①问答任务旨在深入理解用户问题并**提供相应的答案**。在交通预测等实际场景中，可以捕捉用户需求并给出准确的答复。LLMs在处理问答任务方面表现出色，这得益于它们在**理解**和**推理**的能力。②LLMs在一些应用中都表达了很多**隐式的图结构**，但未能有效表达理解图结构数据。③西安交通大学和华盛顿大学首次提出图问答（图基础问答任务）来探讨LLMs能否**显式表达图数据**。

①

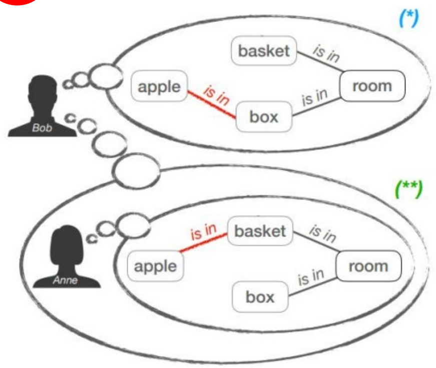


Article: Endangered Species Act
Paragraph: "*... Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*"

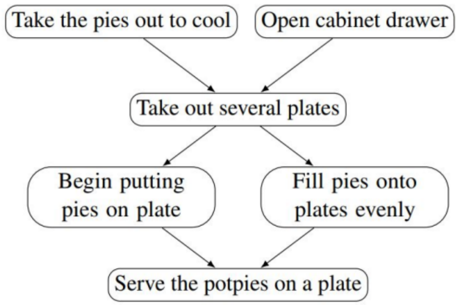
Question 1: "*Which laws faced significant opposition?*"
Plausible Answer: *later laws*

Question 2: "*What was the name of the 1937 treaty?*"
Plausible Answer: *Bald Eagle Protection Act*

②

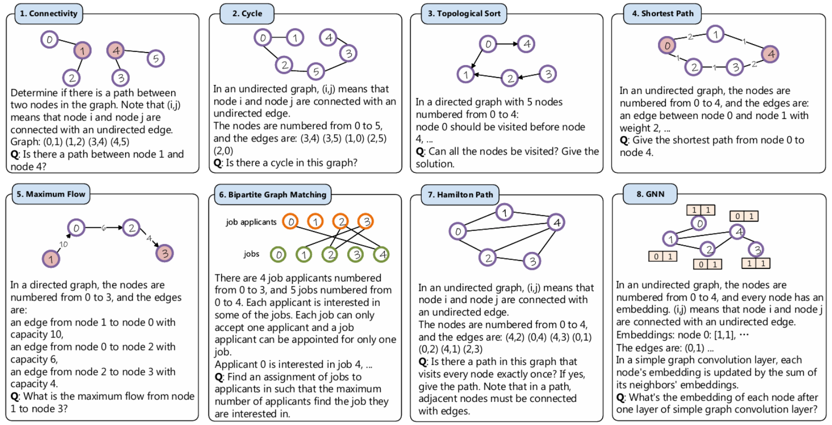


belief state in theory-of-mind



structured commonsense reasoning

③



Can Language Models Solve Graph Problems in Natural Language?

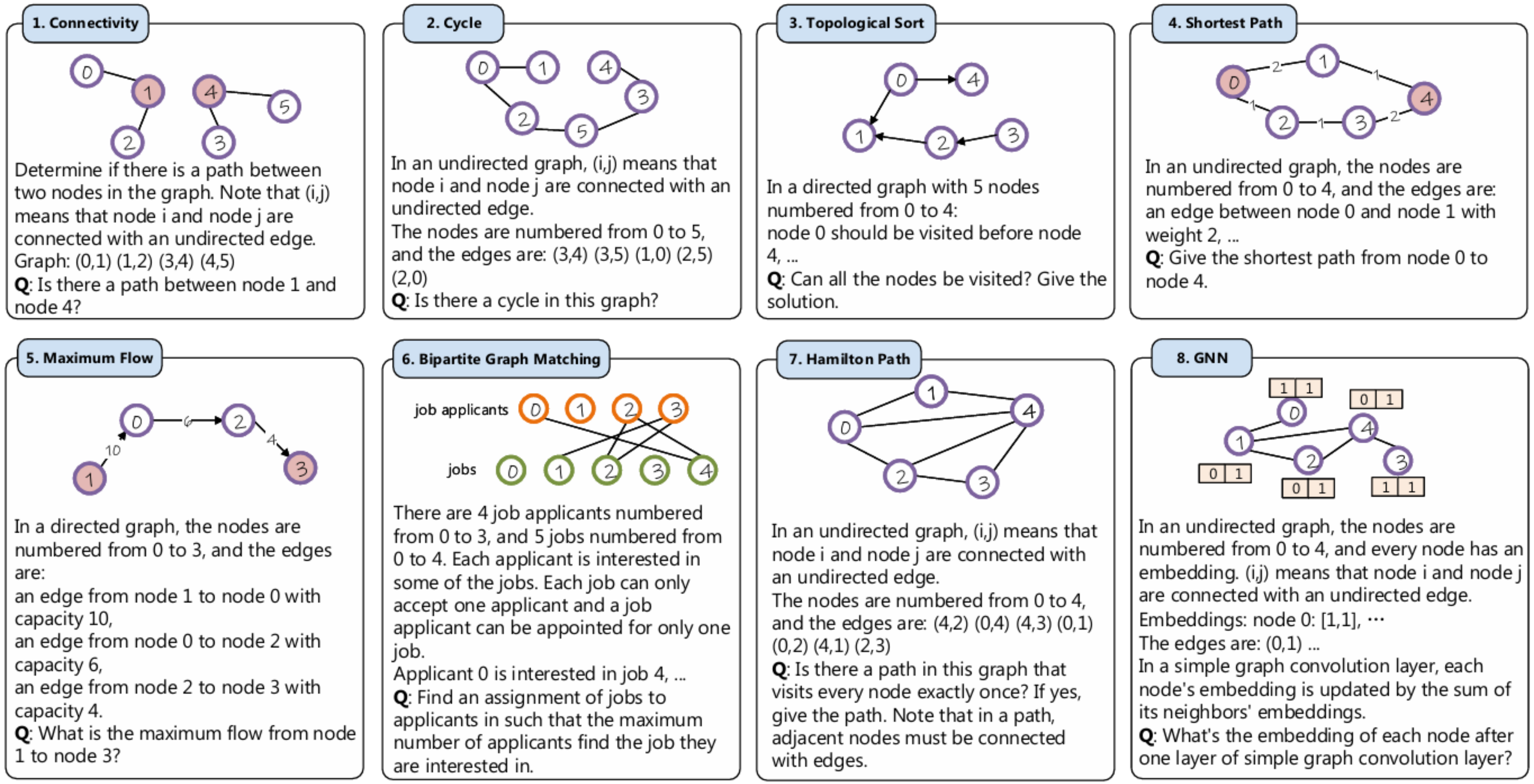
Outline

Heng Wang*[1], Shangbin Feng*[2], Tianxing He[2], Zhaoxuan Tan[3], Xiaochuang Han[2], Yulia Tsvetkov[2]
[1]Xi'an Jiaotong University [2]University of Washington [3]University of Notre Dame
wh2213210554@stu.xjtu.edu.cn, shangbin@cs.washington.edu

- 本文提出了首个使用**自然语言**处理**图基础问答的**BenchMark NLGraph；
- 作者进行实验并统计LLM performance，总结了**一些规律**；
- 作者提出了两种LLM理解图结构数据的改善方法。

BenchMark

➢ 作者提出了自然语言图(NLGraph)基准测试，共分为八种基于图推理问题，5902 个问题。



Experiment

实验1： 通过比较五种方法在**三种简单任务**中**三种难度级别**的结果。

结果： 作者首先发现，在简单的图问答任务中，LLM 取得了令人印象深刻的表现，并展示了初步的**图思维能力**。

Default LLM：TEXT-DAVINCI-003

Method	Connectivity				Cycle				Shortest Path				
	Easy	Medium	Hard	Avg.	Easy	Medium	Hard	Avg.	Easy	Hard	Easy (PC)	Hard (PC)	Avg.
RANDOM	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	6.07	6.69	14.73	13.81	17.81
ZERO-SHOT	83.81	72.75	63.38	71.31	50.00	50.00	50.00	50.00	29.40	21.00	46.00	26.76	30.79
FEW-SHOT	93.75	83.83	76.61	84.73	80.00	**70.00**	**61.00**	**70.33**	31.11	26.00	49.19	35.73	35.51
CoT	**94.32**	82.17	77.21	84.57	**84.67**	63.33	53.25	66.75	63.89	**29.50**	76.84	35.79	51.51
0-CoT	79.55	65.83	68.53	71.30	55.33	57.67	49.00	54.00	8.89	7.50	62.39	**43.95**	32.03
CoT+SC	93.18	**84.50**	**82.79**	**86.82**	82.00	63.67	53.50	66.39	**68.89**	29.00	**80.25**	38.47	**54.15**

Subset	Connect.	Cycle	Topo. Sort	Shortest Path	Max. Flow	Bipartite Graph	Hamilton Path	GNNs
# EASY	352 / 730	150 / 300	180 / 360	180 / 360	150 / 300	300 / 600	150 / 300	100 / 200
SPEC.	n: 5-10	n: 5-10	n: 5-10	n: 5-10	n: 5-10	n: 6-20	n: 5-10	n: 5-8
# MEDIUM	1,200 / 8,580	600 / 1,800	150 / ,350	/	/	/	/	/
SPEC.	n: 11-25	n: 11-25	n: 11-25	/	/	/	/	/
# HARD	680 / 7,090	400 / 2,000	200 / ,200	200 / 1,200	200 / 1,200	210 / 1,260	200 / 600	140 / 840
SPEC.	n: 26-35	n: 26-35	n: 26-35	n: 11-20	n: 11-20	n: 17-33	n: 11-20	n: 9-15

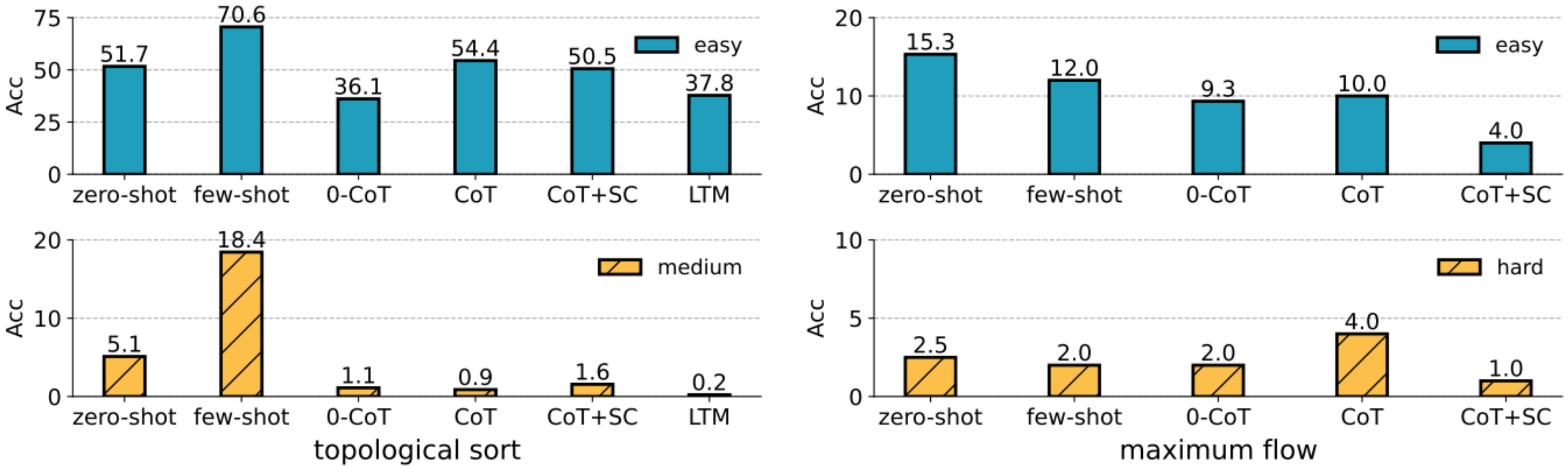
Experiment

实验2： 模拟图神经网络任务的模型性能 + 复杂图问答任务。

结果： 高级Prompt如CoT，SC(self-consistency)在图神经网络任务上提高了准确率，但是在复杂的图推理问题上，高级Prompt如COT, COT+SC, 同时LTM(Least-to-most)的效果反而低于Few-SHOT。

GNN Task

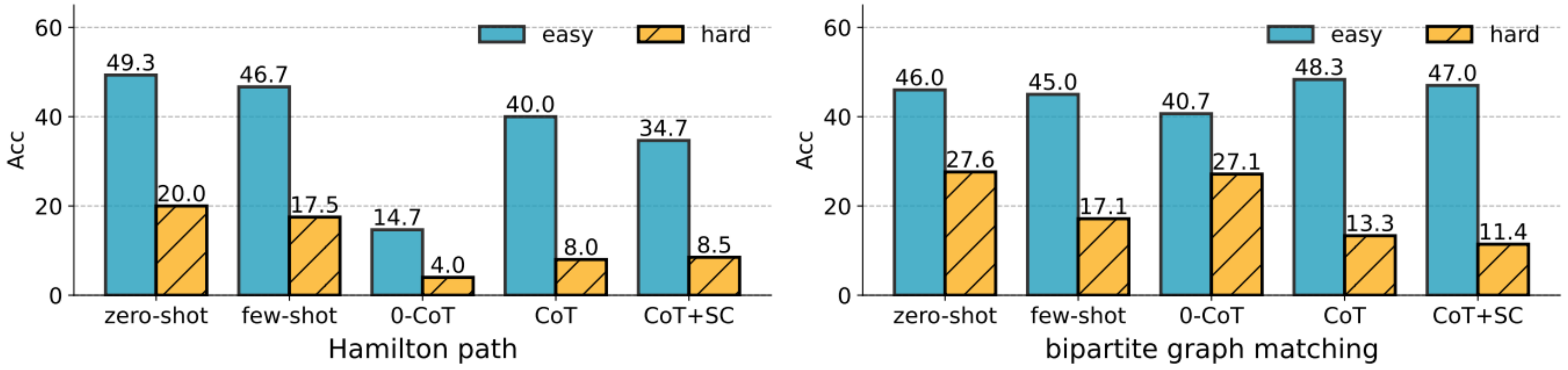
Method	PC (↑)	Acc (↑)	RE (↓)
ZERO-SHOT	13.61	0.00	20.04
FEW-SHOT	20.04	0.00	37.83
CoT	**64.55**	**31.00**	14.34
0-CoT	13.85	0.00	44.55
CoT+SC	63.92	28.00	**13.28**

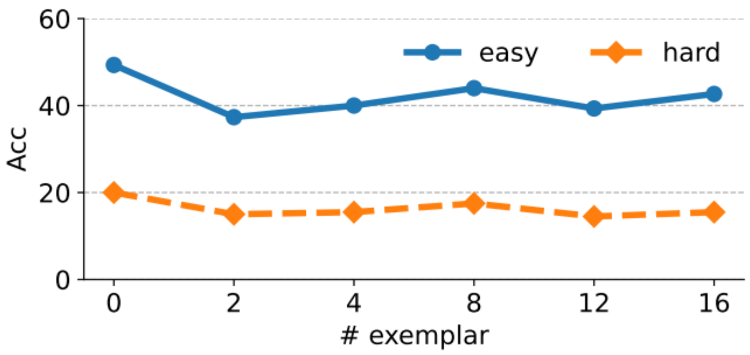


Experiment

实验3： 探讨**上下文学习**在图推理任务中的效果。

结果： 作者发现，在处理复杂的图推理问题时，如哈密顿路径（Hamilton path）和二分图匹配（bipartite graph matching），使用**更多样本**的上下文学习并没有提高模型的性能，反而零样本学习表现得更好。



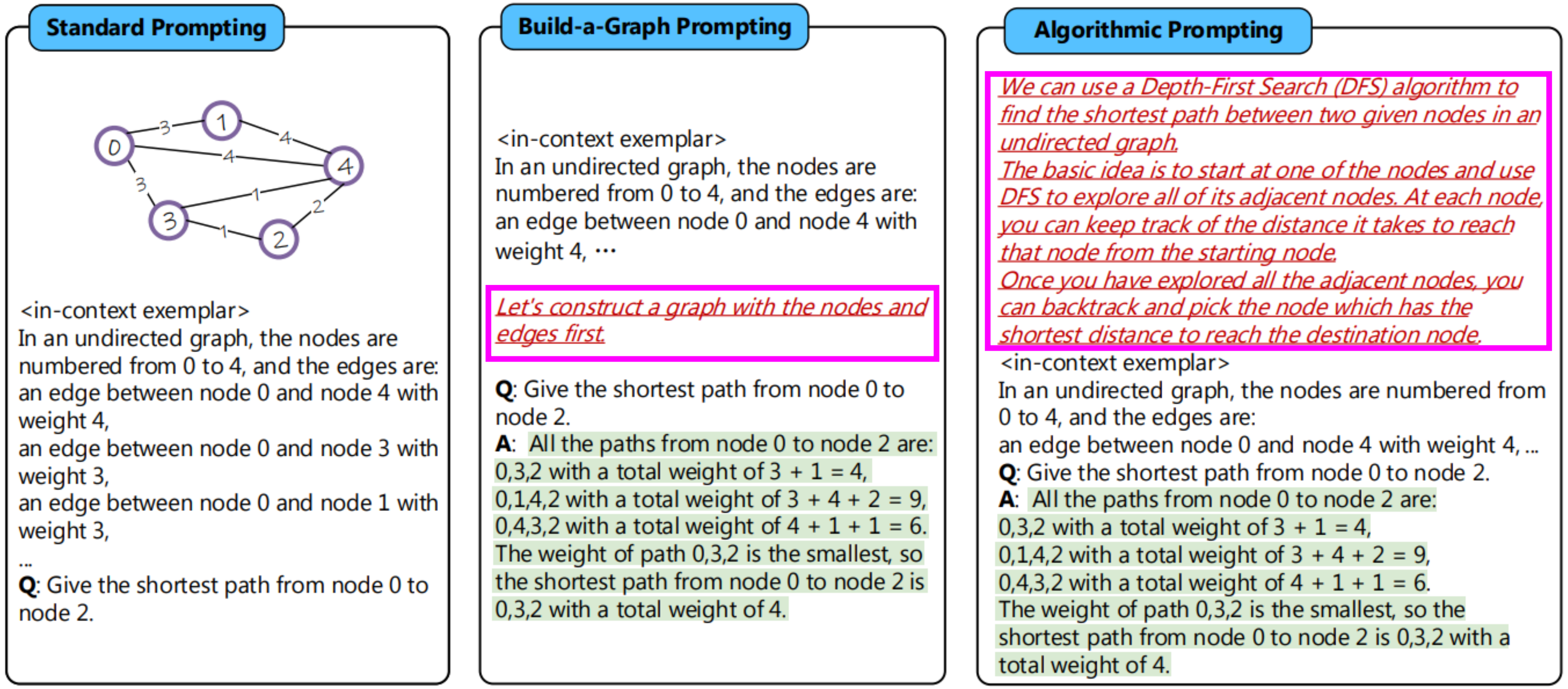


Can Language Models Solve Graph Problems in Natural Language?, NIPS'23, Xi'an Jiaotong University & University of Washington & University of Notre Dame

Method

Build-a-Graph Prompting：作者认为将图的文本描述映射到实际概念空间可能会有帮助。因此加了一句“让我们首先构建一个带有节点和边的图”

Algorithmic Prompting：提示LLM可以使用一些具体的算法解决问题。比如告诉LLMs使用**DFS**或者**BFS**来解决这个问题。



Method

Build-a-Graph Prompting：作者认为将图的文本描述映射到实际概念空间可能会有帮助。因此加了一句“让我们首先构建一个带有节点和边的图”

Algorithmic Prompting：提示LLM可以使用一些具体的算法解决问题。比如告诉LLMs使用**DFS**或者**BFS**来解决这个问题。

Method	Cycle				Shortest Path					Hamilton Path		
	Easy	Medium	Hard	Avg.	Easy	Hard	Easy (PC)	Hard (PC)	Avg.	Easy	Hard	Avg.
CoT	84.67	63.33	53.25	66.75	63.89	29.50	76.84	35.79	51.51	**40.00**	**8.00**	**24.00**
CoT+BaG	**86.00**	69.33	62.00	**72.44**	**67.78**	**33.50**	**79.20**	**42.56**	**55.76**	38.67	6.00	22.34
CoT+ALGORITHM	77.33	**74.00**	**64.00**	71.78	63.89	28.00	76.06	38.70	51.66	36.67	7.50	22.09

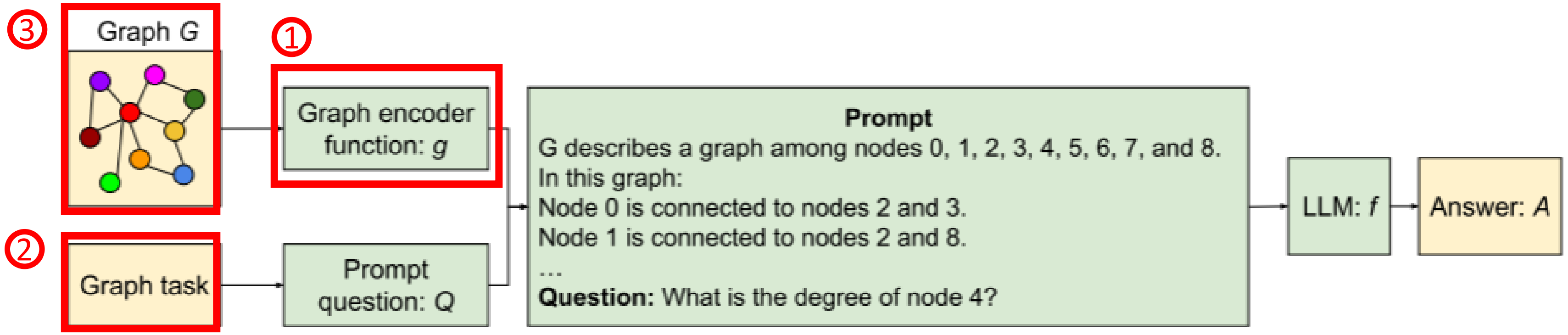
在两个简单的任务上，**Cycle**和**Shortest Path**在**Avg.**分别提高了**5.69%**和**4.25%**；
在汉密尔顿路径等更复杂的图推理任务仍是一个有待研究的问题。

TALK LIKE A GRAPH: ENCODING GRAPHS FOR LARGE LANGUAGE MODELS

Outline

Bahare Fatemi, Jonathan Halcrow, Bryan Perozzi
Google Research
{baharef,halcrow,bperozzi}@google.com

➢ **Motivation**：作者也在探讨LLMs能否显式地表达图数据，使LLMs的图推理成为可能。作者发现LLM在图推理任务上的性能在三个基本层面上有所不同：1）**图编码方法**；2）**图任务本质的性质**；3）**图结构**。



BenchMark

- 作者提出了一组新的benchmark(**GraphQA**)，**一共7种问题**

问题	描述
Edge existence	确定给定的边是否存在于图中。
Node degree	计算图中给定节点的度，即与该节点相连的边的数量。
Node count	计算图中节点的总数。
Edge count	计算图中边的总数。
Connected nodes	找出图中与给定节点直接相连的所有节点。
Cycle check	确定图是否包含环，即是否存在闭合路径。
Disconnected nodes	找出图中与给定节点不相连的所有节点。

Method

①

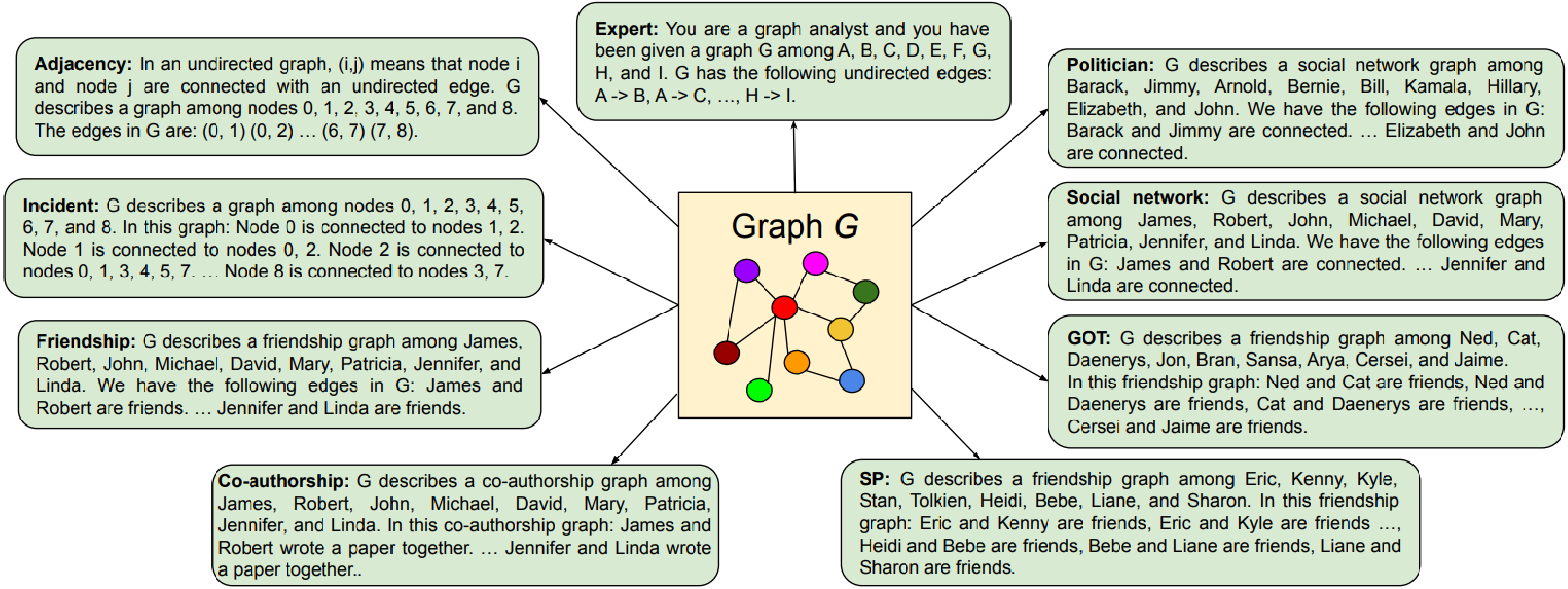
➢ **图编码函数：赋予基础图特定背景**。首先是图中**节点的编码**，其次是节点之间的**边的编码**。

- 编码节点
  - 整数编码
  - 著名的英语名字
  - 电视剧《权力的游戏》和《南方公园》流行角色名字
  - 美国政治家的名字
  - 字母表字母

- 编码边
  - 括号 → （源节点，目标节点）
  - 友谊 → 源节点和目标节点是朋友
  - 合著者 → 源节点和目标节点共同撰写了一篇论文
  - 社交网络 → 源节点和目标节点相互连接
  - 箭头 → 源节点→目标节点
  - 入射 → 源节点连接到目标节点

Method

➢ 作者**结合**节点编码和边编码的方式，形成了九种图编码方式。



TALK LIKE A GRAPH:ENCODING GRAPHS FOR LARGE LANGUAGE MODELS, ICLR’24, Google Research & Waymo Research

Experiment

实验1：作者测试了预训练的 LLM 在图任务上的性能：**Edge existence**、**Node degree**、**Node count**、**Edge count**、**Connected nodes**和**Cycle check**。

Method	Encoding	Edge Existence	Node degree	Node count	Edge count	Connected nodes	Cycle check
	Overall (μ/δ)	44.5 / 9.4	14.0/16.0	21.73 / 8.6	12.4 / 4.8	14.7 / 11.0	76.0 / 13.2
ZERO-SHOT	Adjacency	45.8	12.4	18.8	14.0	19.8	71.6
	Incident	39.6	25.0	15.6	10.6	53.8	68.8
	Co-authorship	44.0	13.8	22.0	11.4	7.6	70.8
	Friendship	46.6	11.2	23.0	10.2	4.0	**82.0**
	SP	46.4	9.0	22.4	15.0	6.2	80.4
	GOT	**49.0**	13.6	22.8	13.2	7.6	79.0
	Social network	43.2	16.0	22.8	10.8	8.2	81.2
	Politician	44.6	15.2	24.2	11.6	8.8	81.0
	Expert	41.2	10.0	24.0	14.8	16.4	69.6
	Overall (μ/δ)	33.5 / 11.6	10.4 / 22.4	14.6 / 9.4	9.4 / 4.8	8.8 / 9.2	32.3 / 23.2
ZERO-COT	Adjacency	34.2	15.4	11.0	12.2	6.0	46.2
	Incident	41.4	26.6	10.0	12.2	35.2	39.0
	Co-authorship	29.8	9.8	15.6	8.2	3.0	28.2
	Friendship	28.4	7.0	19.4	7.4	3.0	31.2
	SP	32.6	9.2	15.6	8.4	5.0	34.8
	GOT	34.6	8.4	16.2	8.4	5.4	33.4
	Social network	30.8	6.6	14.0	9.2	3.8	26.0
	Politician	38.0	4.2	14.6	8.6	3.2	23.0
	Expert	31.6	6.0	14.8	10.0	14.2	28.8
	Overall (μ/δ)	36.8 / 13.8	17.4 / 23.4	25.3 / 35.6	12.0 / 9.0	12.4 / 15.2	37.4 / 24.0
FEW-SHOT	Adjacency	42.8	15.4	47.2	18.6	22.2	47.8
	Incident	38.8	33.6	51.2	14.6	36.6	45.0
	Co-authorship	29.4	15.6	15.6	10.2	9.0	46.8
	Friendship	40.6	12.2	18.4	9.8	6.4	41.4
	SP	34.6	18.0	18.0	12.0	6.8	38.2
	GOT	40.6	17.2	14.2	12.0	3.4	28.6
	Social network	37.4	15.0	21.2	10.2	7.8	34.2
	Politician	38.0	13.4	21.4	9.6	7.8	30.8
	Expert	29.0	16.6	20.4	11.2	11.8	23.8
	Overall (μ/δ)	42.8 / 7.0	29.2 / 60.4	27.6 / 42.4	12.8 / 17.4	13.1 / 18.0	58.0 / 16.4
COT	Adjacency	42.8	71.2	57.0	**25.2**	22.4	56.6
	Incident	41.6	**75.0**	**57.6**	21.4	30.2	62.6
	Co-authorship	43.2	16.4	15.2	8.8	8.4	54.8
	Friendship	46.6	14.6	23.0	7.8	9.6	61.8
	SP	42.6	17.4	17.0	10.6	8.2	59.4
	GOT	44.0	17.8	16.2	11.8	7.2	60.4
	Social network	42.6	16.4	21.6	8.4	8.0	60.6
	Politician	42.2	16.6	22.6	9.2	9.4	59.4
	Expert	39.6	17.4	18.0	12.4	14.4	46.2
	Overall (μ/δ)	37.3 / 16.6	28.0 / 61.8	26.9 / 33.8	12.5 / 17.8	15.8 / 31.8	52.1 / 26.0
COT-BAG	Adjacency	45.8	66.8	48.6	25.0	20.6	56.8
	Incident	45.6	75.2	51.2	21.8	**41.0**	63.0
	Co-authorship	25.0	14.6	17.4	7.2	9.2	37.0
	Friendship	39.0	16.2	21.8	7.4	9.8	52.0
	SP	33.6	17.0	21.6	11.4	11.4	52.2
	GOT	32.6	15.6	18.0	11.0	10.0	54.6
	Social network	44.8	13.4	19.6	9.0	10.0	51.2
	Politician	40.4	17.6	22.8	8.2	10.2	57.2
	Expert	29.2	15.8	20.8	11.6	20.4	45.0

PaLM 62B and PaLM 2

Experiment

结果1：1）LLM在基本的图任务上整体表现不佳；2）简单的提示最适合用于简单的任务；3）图编码函数对LLM推理有重要影响；

Method	Encoding	Edge Existence	Node degree	Node count	Edge count	Connected nodes	Cycle check
ZERO-SHOT	Overall (μ/δ)	44.5 / 9.4	14.0/16.0	21.73 / 8.6	12.4 / 4.8	14.7 / 11.0	76.0 / 13.2
	Adjacency	45.8	12.4	18.8	14.0	19.8	71.6
	Incident	39.6	25.0	15.6	10.6	53.8	68.8
	Co-authorship	44.0	13.8	22.0	11.4	7.6	70.8
	Friendship	46.6	11.2	23.0	10.2	4.0	**82.0**
	SP	46.4	9.0	22.4	15.0	6.2	80.4
	GOT	**49.0**	13.6	22.8	13.2	7.6	79.0
	Social network	43.2	16.0	22.8	10.8	8.2	81.2
	Politician	44.6	15.2	24.2	11.6	8.8	81.0
	Expert	41.2	10.0	24.0	14.8	16.4	69.6
ZERO-COT	Overall (μ/δ)	33.5 / 11.6	10.4 / 22.4	14.6 / 9.4	9.4 / 4.8	8.8 / 9.2	32.3 / 23.2
	Adjacency	34.2	15.4	11.0	12.2	6.0	46.2
	Incident	41.4	26.6	10.0	12.2	35.2	39.0
	Co-authorship	29.8	9.8	15.6	8.2	3.0	28.2
	Friendship	28.4	7.0	19.4	7.4	3.0	31.2
	SP	32.6	9.2	15.6	8.4	5.0	34.8
	GOT	34.6	8.4	16.2	8.4	5.4	33.4
	Social network	30.8	6.6	14.0	9.2	3.8	26.0
	Politician	38.0	4.2	14.6	8.6	3.2	23.0
	Expert	31.6	6.0	14.8	10.0	14.2	28.8
FEW-SHOT	Overall (μ/δ)	36.8 / 13.8	17.4 / 23.4	25.3 / 35.6	12.0 / 9.0	12.4 / 15.2	37.4 / 24.0
	Adjacency	42.8	15.4	47.2	18.6	22.2	47.8
	Incident	38.8	33.6	51.2	14.6	36.6	45.0
	Co-authorship	29.4	15.6	15.6	10.2	9.0	46.8
	Friendship	40.6	12.2	18.4	9.8	6.4	41.4
	SP	34.6	18.0	18.0	12.0	6.8	38.2
	GOT	40.6	17.2	14.2	12.0	3.4	28.6
	Social network	37.4	15.0	21.2	10.2	7.8	34.2
	Politician	38.0	13.4	21.4	9.6	7.8	30.8
	Expert	29.0	16.6	20.4	11.2	11.8	23.8
COT	Overall (μ/δ)	42.8 / 7.0	29.2 / 60.4	27.6 / 42.4	12.8 / 17.4	13.1 / 18.0	58.0 / 16.4
	Adjacency	42.8	71.2	57.0	**25.2**	22.4	56.6
	Incident	41.6	**75.0**	**57.6**	21.4	30.2	62.6
	Co-authorship	43.2	16.4	15.2	8.8	8.4	54.8
	Friendship	46.6	14.6	23.0	7.8	9.6	61.8
	SP	42.6	17.4	17.0	10.6	8.2	59.4
	GOT	44.0	17.8	16.2	11.8	7.2	60.4
	Social network	42.6	16.4	21.6	8.4	8.0	60.6
	Politician	42.2	16.6	22.6	9.2	9.4	59.4
	Expert	39.6	17.4	18.0	12.4	14.4	46.2
COT-BAG	Overall (μ/δ)	37.3 / 16.6	28.0 / 61.8	26.9 / 33.8	12.5 / 17.8	15.8 / 31.8	52.1 / 26.0
	Adjacency	45.8	66.8	48.6	25.0	20.6	56.8
	Incident	45.6	75.2	51.2	21.8	**41.0**	63.0
	Co-authorship	25.0	14.6	17.4	7.2	9.2	37.0
	Friendship	39.0	16.2	21.8	7.4	9.8	52.0
	SP	33.6	17.0	21.6	11.4	11.4	52.2
	GOT	32.6	15.6	18.0	11.0	10.0	54.6
	Social network	44.8	13.4	19.6	9.0	10.0	51.2
	Politician	40.4	17.6	22.8	8.2	10.2	57.2
	Expert	29.2	15.8	20.8	11.6	20.4	45.0

Experiment

结果1：1）LLM在基本的图任务上整体表现不佳；2）简单的提示最适合用于简单的任务；3）图编码函数对LLM推理有重要影响；

Method	Encoding	Edge Existence	Node degree	Node count	Edge count	Connected nodes	Cycle check
	Overall (μ/δ)	44.5 / 9.4	14.0/16.0	21.73 / 8.6	12.4 / 4.8	14.7 / 11.0	76.0 / 13.2
ZERO-SHOT	Adjacency	45.8	12.4	18.8	14.0	19.8	71.6
	Incident	39.6	25.0	15.6	10.6	53.8	68.8
	Co-authorship	44.0	13.8	22.0	11.4	7.6	70.8
	Friendship	46.6	11.2	23.0	10.2	4.0	**82.0**
	SP	46.4	9.0	22.4	15.0	6.2	80.4
	GOT	**49.0**	13.6	22.8	13.2	7.6	79.0
	Social network	43.2	16.0	22.8	10.8	8.2	81.2
	Politician	44.6	15.2	24.2	11.6	8.8	81.0
	Expert	41.2	10.0	24.0	14.8	16.4	69.6
	Overall (μ/δ)	33.5 / 11.6	10.4 / 22.4	14.6 / 9.4	9.4 / 4.8	8.8 / 9.2	32.3 / 23.2
ZERO-COT	Adjacency	34.2	15.4	11.0	12.2	6.0	46.2
	Incident	41.4	26.6	10.0	12.2	35.2	39.0
	Co-authorship	29.8	9.8	15.6	8.2	3.0	28.2
	Friendship	28.4	7.0	19.4	7.4	3.0	31.2
	SP	32.6	9.2	15.6	8.4	5.0	34.8
	GOT	34.6	8.4	16.2	8.4	5.4	33.4
	Social network	30.8	6.6	14.0	9.2	3.8	26.0
	Politician	38.0	4.2	14.6	8.6	3.2	23.0
	Expert	31.6	6.0	14.8	10.0	14.2	28.8
	Overall (μ/δ)	36.8 / 13.8	17.4 / 23.4	25.3 / 35.6	12.0 / 9.0	12.4 / 15.2	37.4 / 24.0
FEW-SHOT	Adjacency	42.8	15.4	47.2	18.6	22.2	47.8
	Incident	38.8	33.6	51.2	14.6	36.6	45.0
	Co-authorship	29.4	15.6	15.6	10.2	9.0	46.8
	Friendship	40.6	12.2	18.4	9.8	6.4	41.4
	SP	34.6	18.0	18.0	12.0	6.8	38.2
	GOT	40.6	17.2	14.2	12.0	3.4	28.6
	Social network	37.4	15.0	21.2	10.2	7.8	34.2
	Politician	38.0	13.4	21.4	9.6	7.8	30.8
	Expert	29.0	16.6	20.4	11.2	11.8	23.8
	Overall (μ/δ)	42.8 / 7.0	29.2 / 60.4	27.6 / 42.4	12.8 / 17.4	13.1 / 18.0	58.0 / 16.4
COT	Adjacency	42.8	71.2	57.0	**25.2**	22.4	56.6
	Incident	41.6	**75.0**	**57.6**	21.4	30.2	62.6
	Co-authorship	43.2	16.4	15.2	8.8	8.4	54.8
	Friendship	46.6	14.6	23.0	7.8	9.6	61.8
	SP	42.6	17.4	17.0	10.6	8.2	59.4
	GOT	44.0	17.8	16.2	11.8	7.2	60.4
	Social network	42.6	16.4	21.6	8.4	8.0	60.6
	Politician	42.2	16.6	22.6	9.2	9.4	59.4
	Expert	39.6	17.4	18.0	12.4	14.4	46.2
	Overall (μ/δ)	37.3 / 16.6	28.0 / 61.8	26.9 / 33.8	12.5 / 17.8	15.8 / 31.8	52.1 / 26.0
COT-BAG	Adjacency	45.8	66.8	48.6	25.0	20.6	56.8
	Incident	45.6	75.2	51.2	21.8	**41.0**	63.0
	Co-authorship	25.0	14.6	17.4	7.2	9.2	37.0
	Friendship	39.0	16.2	21.8	7.4	9.8	52.0
	SP	33.6	17.0	21.6	11.4	11.4	52.2
	GOT	32.6	15.6	18.0	11.0	10.0	54.6
	Social network	44.8	13.4	19.6	9.0	10.0	51.2
	Politician	40.4	17.6	22.8	8.2	10.2	57.2
	Expert	29.2	15.8	20.8	11.6	20.4	45.0

TALK LIKE A GRAPH:ENCODING GRAPHS FOR LARGE LANGUAGE MODELS, ICLR'24, Google Research & Waymo Research

Experiment

②

实验2： 作者使用Friendship作为图编码函数，并使用两种不同的**问题编码器函数**进行实验：**图问题编码器**和**应用问题编码器**。

图问题编码器： 使用编码与图相关的任务，例如：“节点i的度是多少？”。

应用问题编码器： 使用一个更实际的，日常的上下文中解释图问题。边缘存在成为“评估友谊存在”，节点度成为“计算朋友的数量”……

上下文的文本信息很重要！！

Method	Question encoder	LLM	Edge Existence	Node degree	Node count	Edge count	Connected nodes
ZERO-SHOT	Graph	PaLM 2-XXS	42.8	10.8	5.4	**5.6**	1.6
	Application	PaLM 2-XXS	**60.8**	**14.0**	**9.4**	4.4	**11.4**
	Graph	PaLM 62B	46.6	11.2	**23.0**	10.2	4.0
	Application	PaLM 62B	**47.8**	**16.6**	17.8	**13.2**	**6.0**
COT	Graph	PaLM2 XXS	50.4	8.8	8.4	4.2	10.2
	Application	PaLM2 XXS	**56.4**	**12.2**	**8.6**	**5.4**	**11.0**
	Graph	PaLM 62B	**46.6**	14.6	**23.0**	7.8	9.6
	Application	PaLM 62B	38.6	**16.6**	16.0	**12.2**	**10.0**

Experiment

实验3： 作者评估了LLMs在处理**Disconnected nodes**任务中的表现。

结果： LLMs在处理断开节点任务时表现不佳。ZERO-SHOT 提示方法的准确率仅为**0.5%**，而ZERO-COT、FEW-SHOT、COT和COT-BAG方法的准确率几乎为**0.0%**。

分析原因： 这可能是因为图编码函数主要编码了连接节点的信息，而没有**明确编码**未连接节点的信息，导致LLMs在处理连接节点关系时表现较好，但在捕捉连接缺失时表现较差，从而在与**Disconnected nodes**的任务中表现不佳。

Experiment

③

实验4：作者使用了多种图生成算法来创建随机图，展示图结构对LLM的推理性能的影响。

结果：图结构对LLM在图推理任务中的表现有显著影响。

Method	Graph generator	Edge Existence	Node degree	Node count	Edge count	Connected nodes	Cycle check
ZERO-SHOT	Overall	49.1	17.6	23.0	12.1	23.3	75.2
	ER	45.1	13.6	22.1	11.7	14.9	76.3
	BA	50.2	18.0	24.9	13.6	20.1	72.0
	SBM	45.0	13.8	21.9	9.2	13.8	86.5
	Star	58.0	34.0	32.8	31.7	61.7	8.1
	SFN	57.6	23.1	19.9	8.0	38.1	90.0
	Path	**60.9**	14.8	31.9	28.8	26.6	5.9
	Complete	19.8	12.6	20.7	6.2	13.3	**91.7**
COT	Overall	40.4	29.6	31.7	12.2	24.3	59.5
	ER	41.2	28.4	28.8	12.6	12.8	61.2
	BA	40.0	30.0	35.0	14.3	20.8	58.5
	SBM	40.3	26.5	30.2	8.7	13.0	65.8
	Star	40.3	**38.0**	**41.8**	**31.6**	**68.6**	21.3
	SFN	40.2	32.2	30.8	7.1	43.2	66.0
	Path	42.0	35.1	35.3	31.1	27.6	19.7
	Complete	39.6	21.9	28.9	3.9	14.6	69.3

Let Your Graph Do the Talking: Encoding Structured Data for LLMs

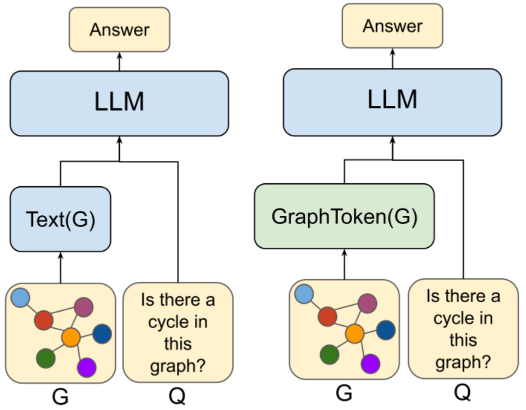
Outline

Bryan Perozzi [1] Bahare Fatemi [1] Dustin Zelle [1] Anton Tsitsulin [1]
Mehran Kazemi [1] Rami Al-Rfou [2] Jonathan Halcrow [1]

现有LLMs表达图结构化数据的方式:

1）将结构化数据转换为适用于LLM嵌入的词汇令牌。

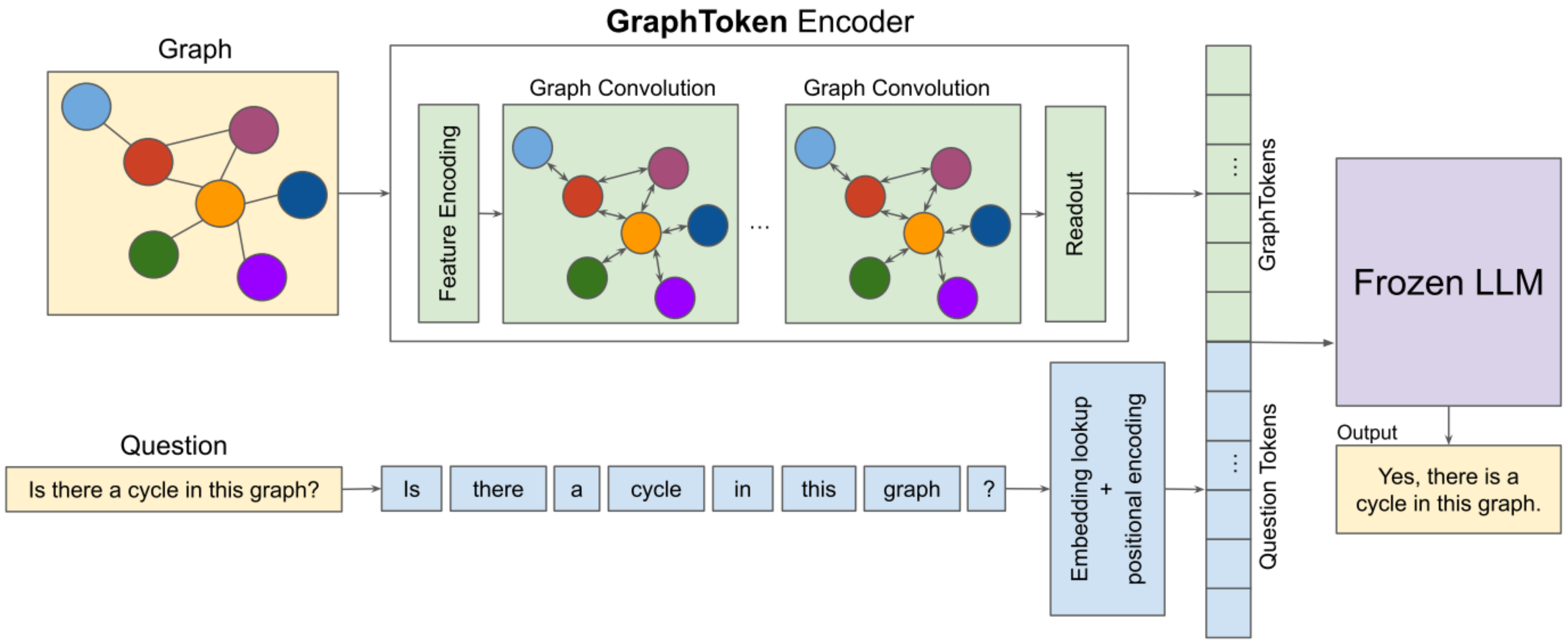
2）使用神经网络直接编码结构化数据为连续向量表示。



➢ **Problem:** 作者他们先前工作忽略了图的一些**全局结构信息**以及**多跳的邻居关系**。

Method

➢ 作者利用**高效微调（Parameter Efficient Fine-Tuning）**的思想，把图结构数据编码成嵌入表示。通过将这些嵌入作为连续且可扩展的Prompt，使LLM能够更好地理解和处理图基础问题。



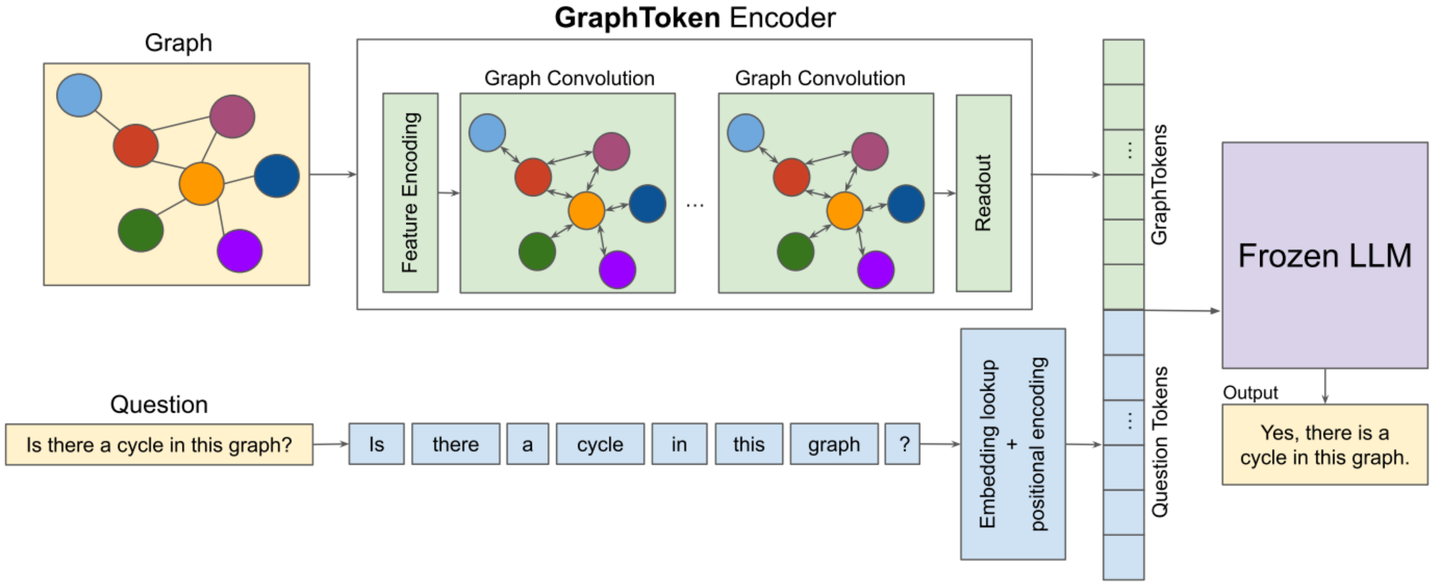
Let Your Graph Do the Talking: Encoding Structured Data for LLMs, arxiv'24, Google Research & Waymo Research

Method

1）**GraphToken Encoder**：把图结构作为输入，利用图位置编码定义节点特征。使用GNN生成图表示。基于任务的不同（Graph、Edge、Node），编码器采用不同的Readout提取图表示。

2）**LLM**：任何能够处理嵌入序列的大型语言模型，作者使用**PaLM 2**大语言模型；

3）**训练过程**：输入三元组（G,T,A），计算扩展查询 $Q=\mathcal{E}(G)||\mathcal{T}(T)$结果，然后通过最小化$\mathcal{L}(A \mid Q)$损失，更新GraphToken的参数，而LLM参数保持不变。



Experiment

实验1：通过准确性比较了GraphToken与图推理任务上的提示工程和软提示方法。

结果：GraphToken 在所有图、节点和边级任务上均显着优于现有方法。 虽然 SOFT-PROMPT 在某些任务上取得了第二好的成绩，但这主要是因为它倾向于预测大多数标签，导致预测结果不稳定。

Method	Graph Tasks				Node Tasks		Edge Tasks		
	Node count	Edge count	Cycle check	Triangle counting	Node degree	Connected nodes	Reachability	Edge existence	Shortest path
ZERO-SHOT	0.217	0.124	0.760	0.015	0.140	0.147	0.849	0.445	0.115
ZERO-COT	0.146	0.094	0.323	0.127	0.104	0.088	0.735	0.335	0.336
FEW-SHOT	0.253	0.120	0.374	0.030	0.174	0.124	0.794	0.368	0.227
COT	0.276	0.128	0.580	0.081	0.292	0.131	0.452	0.428	0.386
COT-BAG	0.269	0.125	0.521	0.081	0.280	0.158	0.452	0.373	0.404
SOFT-PROMPT	0.056	0.018	0.832	0.162	0.098	0.068	0.838	0.544	0.462
GraphToken	**0.996**	**0.426**	**0.956**	**0.348**	**0.962**	**0.264**	**0.932**	**0.738**	**0.638**

Flan

PaLM 2

Experiment

实验2：作者研究了选择**不同的图编码器**如何影响语言模型使用的图表示质量，包括图卷积的选择、网络可用的特征和超参数等。

结果：不同的图编码器适用于**不同的图基础问答任务**。

		Graph Tasks				Node Tasks		Edge Tasks		
	Method	Node count	Edge count	Cycle check	Triangle counting	Node degree	Connected nodes	Reachability	Edge existence	Shortest path
Non-linear	GCN	0.746	0.056	**0.964**	0.208	0.264	**0.264**	0.918	0.68	0.604
	GIN	0.704	0.052	0.898	0.194	0.252	0.18	0.902	0.65	0.586
	MPNN	0.792	0.368	0.956	**0.348**	**0.962**	0.25	0.934	0.648	**0.638**
	HGT	0.252	0.084	0.934	0.234	0.266	0.184	**0.944**	0.718	0.6
	MHA	0.912	0.264	0.962	0.266	0.552	0.244	0.932	**0.738**	0.608
Linear	Node Set	**0.996**	0.080	0.948	0.198	0.19	0.118	0.942	0.596	0.568
	Edge Set	0.618	**0.426**	**0.964**	0.228	0.22	0.096	0.904	0.592	0.568

3 总结

- 3.1 未来研究方向
- 3.2 研究团队

未来研究方向

- 1）**异常类型方向：**现有的异常检测方法大多依赖于**已知的异常类型**，因此在识别新型或未知异常时表现不佳，难以应对实际场景中的**多样化**和复杂性变化。
- 2）**图问答方向：**当前的图问答研究主要集中在**基础图问题**上，对于更复杂的图问题涉及较少。图基础模型在处理**复杂问题**时的推理能力仍需进一步探索和提升。

3 总结

- 3.1 未来研究方向
- 3.2 研究团队

研究团队

石川 Shi Chuan (C. Shi)

博士，教授
图数据挖掘与机器学习实验室的创办人
北京邮电大学 计算机学院
中国北京海淀区西土城路10号
电话：(86) 10-62283698
邮箱：shichuan (at) bupt.edu.cn
微信公众号：北邮 GAMMA Lab
[下载pdf简历] [Turn to English Page]

北邮石川

Quanming Yao (姚权铭) 清华姚权铭

Assistant Professor & Ph.D Advisor

Department of Electronic Engineering, Tsinghua University

Huawei endowed professorship and Bingkun Zhou (周炳琨) Scholar

E-mail: qyaoaa [AT] connect.ust.hk / tsinghua.edu.cn

Office: 11-305 Room, Rohm Building, Tsinghua. Beijing, China, 100084 (MAP)

Github, Google Scholar, Zhihu

Dr. Jian Tang

Assistant Professor,
HEC Montreal
Montreal Institute for Learning Alogorithms (MILA)

Montreal, Canada

I am currently an associate professor at Mila-Quebec AI Institute and HEC Montreal. Prior to that, I was a Postdoc at University of Michigan and Carnegie Mellon University. I also worked at Microsoft Research Asia as an associate researcher between 2014-2016. For more information, please check my CV.

Hiring!! Our group has multiple PhD positions next Fall. In particular, we are looking for students to work on the following projects:

Mila唐建

- Geometric Deep Learning, Graph Neural Networks for Drug Design
- Equivariant Neural Networks for Molecular Simulation
- Knowledge Graph Construction and Reasoning, Natural Language Understanding

Students who are interested in working with me please apply through Mila admission (students working with me will be affiliated with UdeM) or send me an email directly.

Professor Huang, Chao

PhD *Notre Dame*
Assistant Professor
Institute of Data Science Scholar

Fax: (+852) 2559 8447
Email: chuang [AT] cs [DOT] hku [DOT] hk
Homepage: https://sites.google.com/view/chaoh

Chao Huang is an Assistant Professor at the Department of Computer Science at the University of Hong Kong (HKU). His research focuses on developing novel machine learning frameworks to tackle various challenges in Data Mining, Information Retrieval, Spatial-Temporal Data Analytics, User Behavior Modeling, Recommendation, Graph Mining, and Deep Representation Learning. Prior to that, I received my Ph.D. in Computer Science from the University of Notre Dame in USA.

港大黄超

研究团队

密歇根州立大学汤继良

Jiliang Tang

Jiliang Tang is a University Foundation Professor in the computer science and engineering department at Michigan State University. He got one early promotion to associate professor at 2021 and then a promotion to full professor (designated as MSU foundation professor) at 2022. Before that, he was a research scientist in Yahoo Research and got his PhD from Arizona State University in 2015 under Dr. Huan Liu. His research interests include graph machine learning,trustworthy AI and their applications in education and biology. He was the recipient of various awards including 2022 AI's 10 to Watch, 2022 IAPR J. K. AGGARWAL Award, 2022 SIAM/IBM Early Career Research Award, 2021 IEEE ICDM Tao Li Award, 2021 IEEE Big Data Security Junior Research Award, 2020 ACM SIGKDD Rising Star Award, 2020 Distinguished Withrow Research Award, 2019 NSF Career Award, and 8 best paper awards (or runner-ups). His dissertation won the 2015 KDD Best Dissertation runner up and Dean's Dissertation Award. He serves as conference organizers (e.g., KDD, SIGIR, WSDM and SDM) and journal editors (e.g., TKDD, TOIS and TKDE). He has published his research in highly ranked journals and top conference proceedings, which have received tens of thousands of citations with h-index 95 (Google Scholar) and extensive media coverage (Links).

Email: tangjili at msu dot edu
Office: Engineering Building 2148
Mail: 428 S Shaw Ln Rm 3115, East Lansing, MI 48824
Lab: Data Science and Engineering Lab (Webpage, and Twitter)

清华朱文武

Wenwu ZHU

Apr 6, 2021 13:09 3274

Wenwu Zhu

Professor

Phone: +86-10-62790967

OpenAI任鸿宇

Hongyu Ren

Research scientist at OpenAI

Homepage

About me

I'm a research scientist at OpenAI exploring ChatGPT.

I received my Ph.D. in Computer Science from Stanford University and a B.S. with honors in Computer Science from Peking University.

Awards

- Apple PhD Fellowship
- Baidu PhD Fellowship
- Masason Foundation Fellowship

Teaching

- Head TA, CS246 Mining Massive Datasets, Winter 2022
- TA, CS231n Deep Learning for Computer Vision, Spring 2022
- TA, CS224w Machine Learning on Graphs, Fall 2021

Feel free to reach me at hongyu@openai.com (work) or hyren@cs.stanford.edu (personal).

谷歌

Bryan Perozzi

News

- **1/2017** - "Ties That Bind: Characterizing Classes by Attributes and Social Ties" accepted at **WWW'17** (Web Science Track).
- **05/2016** - Defended "Local Modeling of Attributed Graphs: Algorithms and Applications"!
- **05/2016** - "When Recommendation Goes Wrong: Anomalous Link Discovery in Recommendation Networks" accepted at **KDD'16**.
- **05/2016** - "Scalable anomaly ranking of attributed neighborhoods" awarded **Best Paper Runner-up** at SDM'16!
- **05/2015** - "Freshman or Fresher? Quantifying the Geographic Variation of Internet Language" accepted at **ICWSM'16**.
- **01/2016** - "Scalable Anomaly Ranking of Attributed Neighborhoods" accepted at **SDM'16**.

About Me

I am a Research Scientist, working at the intersection of data mining, machine learning, graph theory, and network science. I am particularly interested in local graph algorithms.

top

Research Interests

- Data mining
- Graph Mining
- Graph theory
- Distributed algorithms
- Natural language processing
- Machine learning

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

研究团队

ICT姚迪

中国科学院大学 University of Chinese Academy of Sciences

[中文] [English]

招生信息 教育背景 工作经历 专利与奖励 出版信息

基本信息

姚迪 男 硕导 中国科学院计算技术研究所

电子邮件： yaodi@ict.ac.cn

通信地址： 北京市海淀区科学院南路6号

邮政编码： 100190

招生信息

招生专业

081201-计算机系统结构

招生方向

时空数据挖掘,异常检测,因果机器学习

教育背景

2017-10--2018-10 新加坡南洋理工大学 访问学生

2013-09--2019-07 中国科学院计算技术研究所 研究生/博士

2009-09--2013-07 东北大学 本科/学士

NUS Bryan HOOI

Bryan HOOI

Assistant Professor

PhD in Machine Learning, Carneige Mello ersity, 2019

MSc in Computer Science and BSc in Ma

COM3-02-22

660 12373

bhooi@comp.nus.edu.sg

bhooi.github.io

Bryan HOOI is an assistant professor in the Computer Science Department, School of Computing at the National University of Singapore. He has obtained his PhD degree in Machine Learning from Carnegie Mellon University, USA in 2019, his Master of Science degree in Computer Science and Bachelor with Honours degree in Mathematics from Stanford University, USA in 2014.

His research interests include machine learning, graph mining, anomaly detection, spatiotemporal data, and biomedical applications of AI.

浙大陈华钧

陈华钧 博士

单位 计算机科学与技术学院

教授 | 博士生导师

邮箱 huajunsir at zju 点 edu dot cn

研究方向 人工智能、知识图谱、自然语言处理、AI for Science

Jure Leskovec 斯坦福

Jure Leskovec 关注

Professor of Computer Science, Stanford University

在 cs.stanford.edu 的电子邮件经过验证 - 首页

Data mining Machine Learning Graph Neural Networks Knowledge Graphs Complex Networks

Thanks!