



北京理工大学
信息系统及安全对抗实验中心
信息安全与对抗技术研究所
Beijing Forest Studio



图序列预训练与黎曼图基础模型

www.isclab.org.cn

硕士研究生 代豫

2026年10月08日

- 相关内容
 - 杨景然《属性缺失异质图神经网络》——2025.02.16
 - 杨景然《动态异质图神经网络》——2025.07.06

- 预期收获
- 题目内涵解析与研究目标
- 研究背景与意义
- 研究历史与现状
- 知识基础
- 算法原理
 - GraphGPT
 - R-GFM
- 特点总结与工作展望
- 参考文献

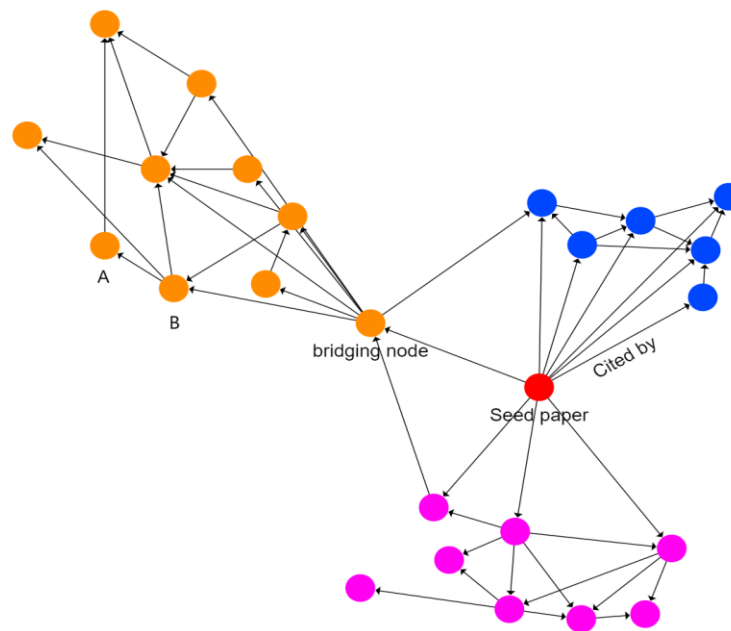
- 预期收获
 - 1. 理解图自监督预训练与下游分类之间的关系
 - 2. 掌握GraphGPT的欧拉序列化、节点身份编码与生成式预训练机制
 - 3. 理解R-GFM的多跳子图建模、Graph-of-Graphs与动态几何专家路由
 - 4. 能够区分单图分类、跨图迁移和少样本分类的实验设置

- 内涵解析

- 图数据：由节点、边及属性组成的关系数据，记为 $G=(V,E,X)$
- 自监督学习：由图结构和属性构造训练目标，预训练阶段不依赖下游人工类别标签
- 分类：依据学习到的表示预测节点或整图的类别
- 图基础模型：通过预训练支持多种图或任务的模型，其迁移范围需要实验验证

- 研究目标

- 学习适合分类的结构与属性表示
- 降低下游任务对标注数据的依赖
- 提升跨数据集迁移与大规模图处理能力

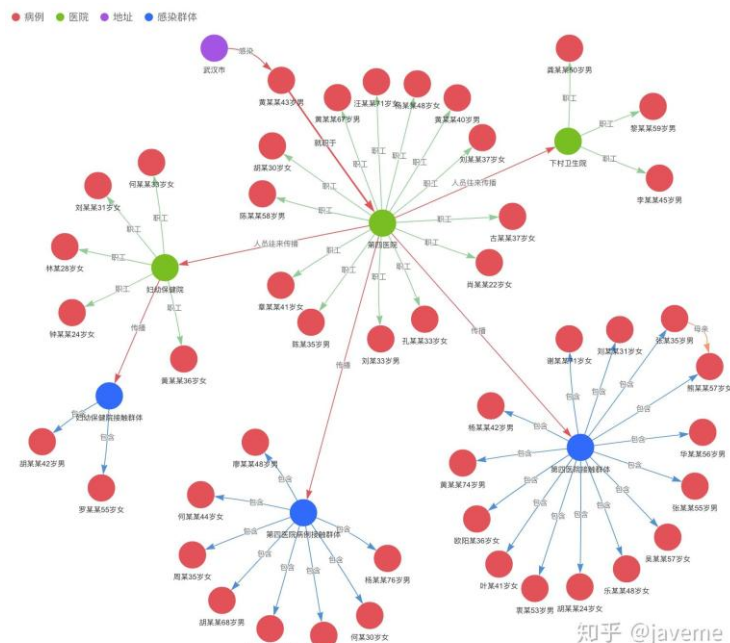


- 研究背景

- 图数据广泛存在于引文网络、社交网络、电子商务和生物网络
- 高质量类别标签获取成本高，无标签关系数据更容易积累
- 不同图的属性语义、结构尺度和拓扑形态存在差异，限制模型迁移

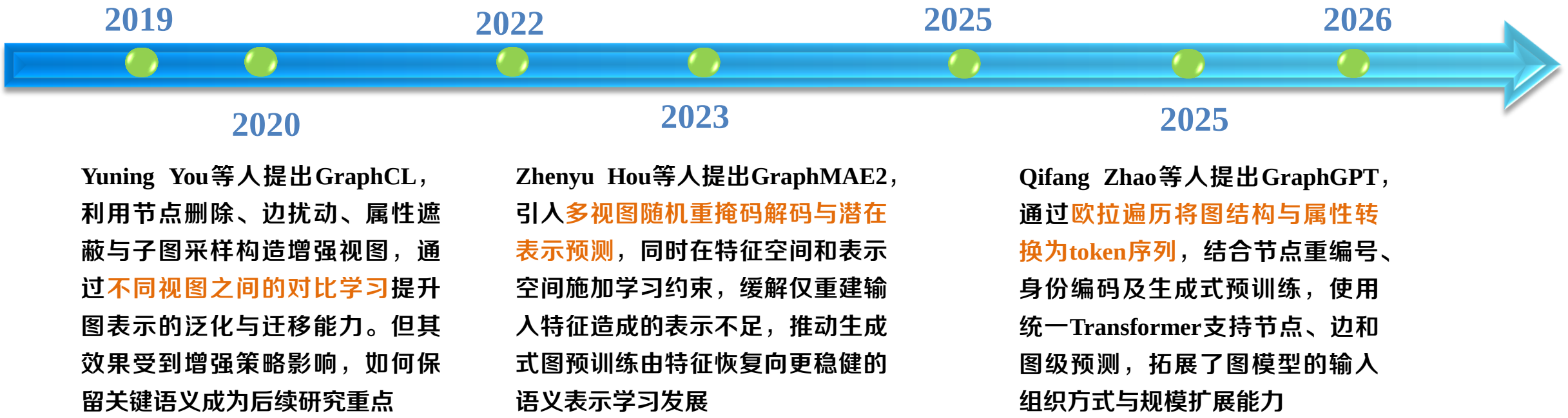
- 研究意义

- 利用无标签图数据训练可复用的表示模型
- 支持低标注条件下的节点分类与图分类
- 为图模型规模扩展、结构泛化和跨域应用提供方法基础



知乎 @javame

Petar Veličković等人提出DGI，通过最大化节点局部表示与图全局摘要之间的互信息，在无标签条件下学习节点表示，为下游节点分类提供特征基础，推动图表示学习由监督训练向自监督预训练发展	Zhenyu Hou等人提出GraphMAE，通过掩码节点特征重建开展生成式图自监督学习，结合重掩码解码与比例余弦误差改善重建训练，减少对负样本和复杂增强策略的依赖，为节点分类与图分类提供有效的预训练表示	Li Sun等人提出RiemannGFM，将树与环作为跨图共享的结构词汇，利用黎曼几何构建适配不同结构的表示空间，学习可跨数据集迁移的结构知识，推动图预训练由单图表示学习向具有结构泛化能力的图基础模型发展	Haokun Liu等人提出R-GFM，将不同跳数子图组织为Graph-of-Graphs，通过相似性建边实现跨尺度交互，并引入动态黎曼几何专家路由，适应不同图的结构尺度与几何差异，提升跨数据集少样本分类与迁移表现
---	--	--	---



- 图神经网络

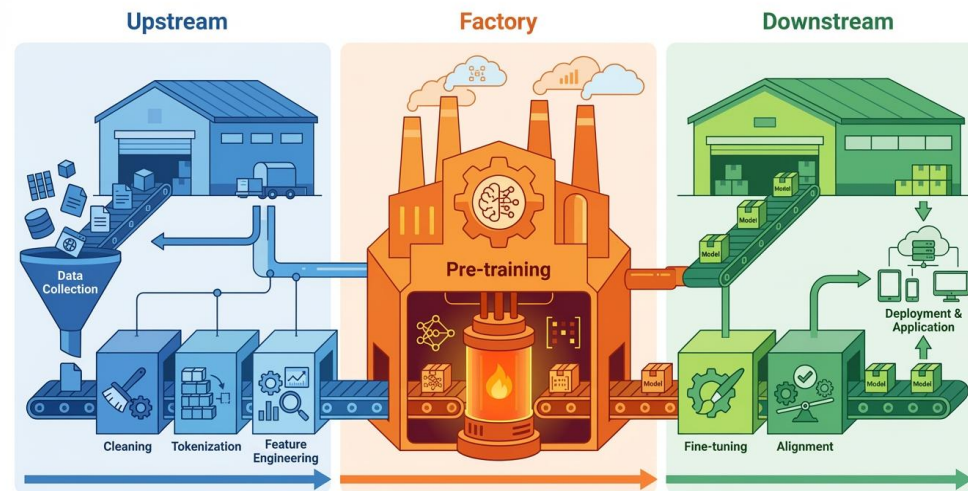
- 认知行为通过**邻居聚合与节点更新**学习表示
- k 层消息传递通常覆盖 k 跳邻域，范围扩大可能同时带来有用关系与结构噪声

- 自监督分类流程

- 构造预测、重建或对比任务，训练图编码器
- 利用少量类别标签训练分类器或微调整个模型
- 在独立测试划分上评估分类结果

- 两类代表目标

- **生成式**：预测被遮蔽信息或后续token
- **对比式**：学习同一样本不同视图之间的一致表示



- Transformer

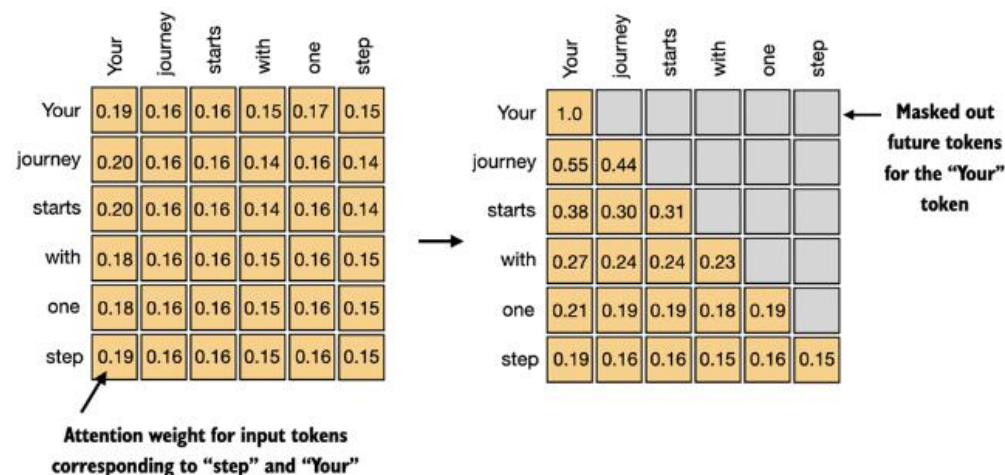
- 通过自注意力建立序列位置之间的依赖
- 位置编码描述序列顺序，图结构需要由输入编码或模型机制提供

- 下一token预测：NTP

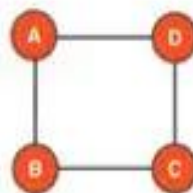
- 根据已有前缀预测下一个token
- 使用因果注意力，避免读取后续答案

- 调度式掩码token预测：SMTP

- 按预设调度策略改变掩码比例，使模型学习不同信息缺失程度下的预测
- 使用双向注意力，根据可见上下文恢复被掩码的token

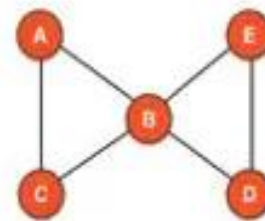


- 欧拉路径
 - 在处理后的图中遍历每条边一次，节点允许重复出现
 - 连通无向图所有节点度数为偶数时存在**欧拉回路**；恰有两个奇度节点时存在开放
- 欧拉化
 - 对不满足条件的图复制部分已有边，使其能够完成欧拉遍历
 - 相邻节点token表达边连接，重复出现的节点维持共享身份
- 图序列化的要求
 - 保留节点、边、方向和属性信息
 - 区分真实关系与用于连接分量的辅助跳转



A - B - C - D - A

既是哈密顿回路，也是欧拉回路



A - B - D - E - B - C - A

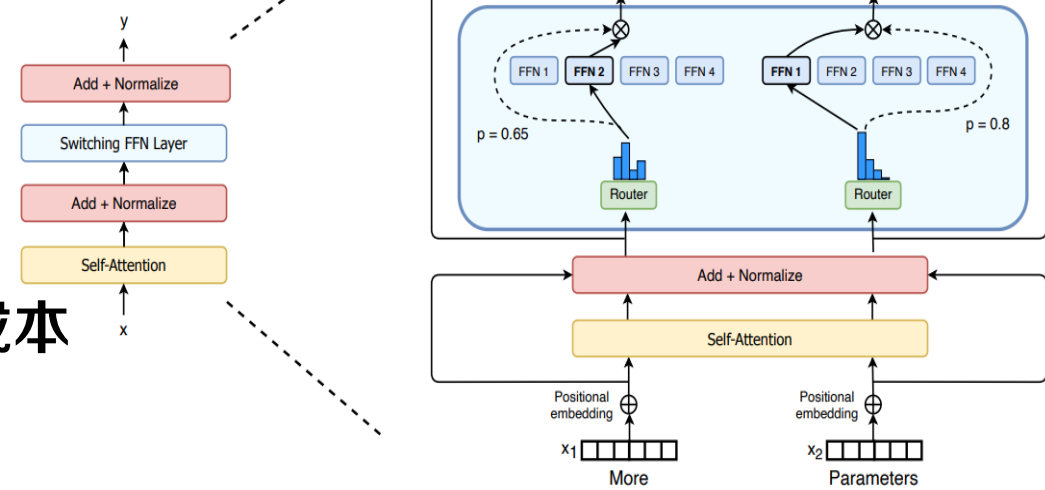
欧拉回路，没有哈密顿回路



- 黎曼几何
 - 研究曲面及更一般空间中的距离、角度与曲率

几何空间	曲率	直观结构特征
双曲空间	$\kappa < 0$	有利于表达树状和层级关系
欧氏空间	$\kappa = 0$	表达相对平坦的几何关系
球面空间	$\kappa > 0$	提供闭合、正曲率的表示空间

- 混合专家MoE
 - 多个专家采用不同表示机制
 - 路由器根据输入产生专家权重
 - 稀疏路由只激活部分专家，在表达能力和成本之间取得平衡





GraphGPT: Generative Pre-trained Graph Eulerian Transformer

T	目标	利用无标签图进行生成式预训练，支持节点、边和图级预测
I	输入	图或子图的节点、边及属性；下游任务训练标签
P	处理	图欧拉化与序列化：通过欧拉遍历，将图结构与属性转为token序列 重编号与身份编码：调整节点编号，编码全局身份，减轻编号偏差 自监督预训练：采用NTP或SMTP，学习图结构与属性关联 任务格式化与微调：附加目标token，利用下游标签微调并预测
O	输出	类别及其概率；链接分数或回归值

P	问题	图缺乏天然顺序，已有编码方式难以统一不同任务与大规模预训练
C	条件	图结构及属性可访问，序列长度适配上下文窗口，具备预训练计算资源
D	难点	结构信息保留、序列长度控制、节点身份保持和预训练知识迁移
L	水平	ICML 2025, CCF A

• 科学问题

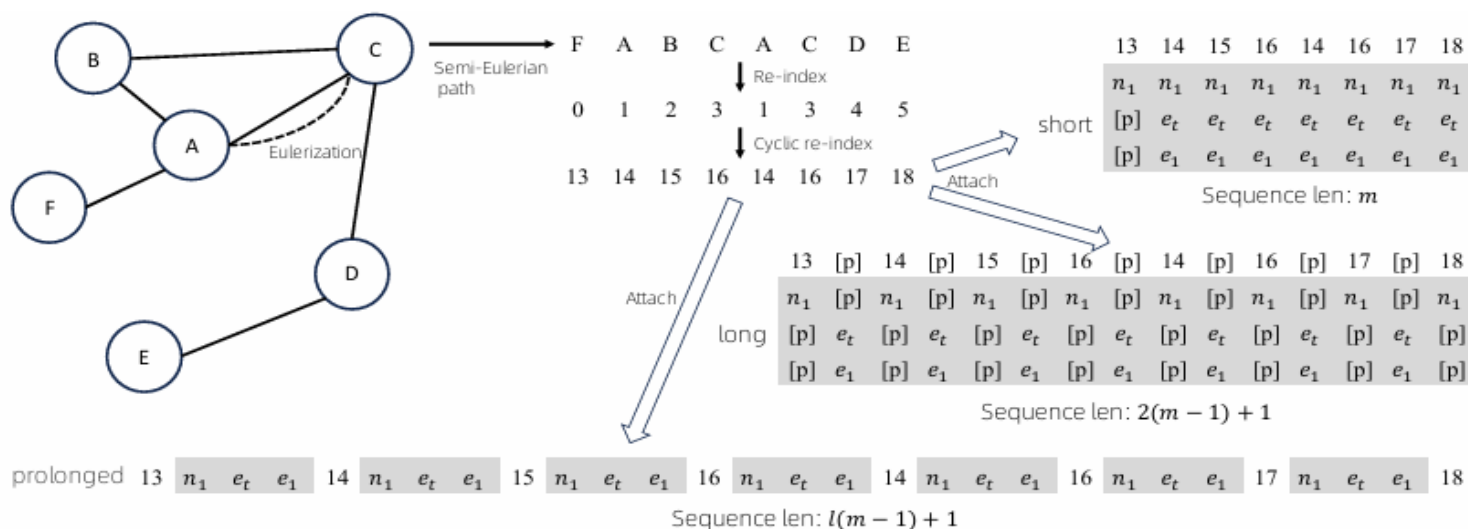
- 无序图如何转换为可恢复结构的token序列
- 怎样让模型学习图模式，减少对某一遍历顺序和节点编号的依赖
- 生成式预训练能否在不同预测层级复用

• 应用问题

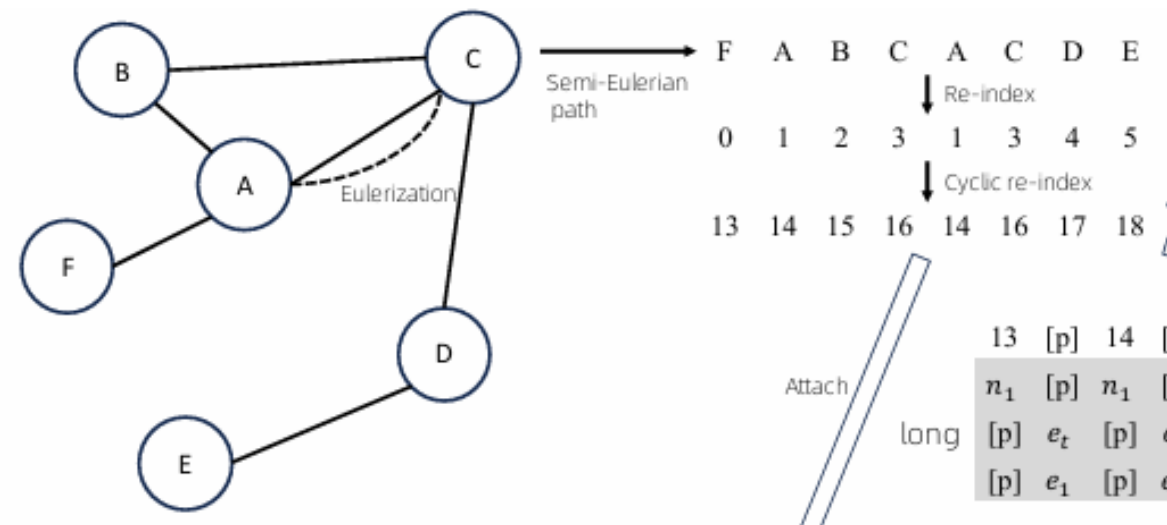
- 现有图Transformer依赖人工设计的结构特征或专用模块，增加不同数据集上的适配成本
- 节点唯一token词表随图规模增长，带来存储成本
- 节点、边与图级任务的输入格式差异影响模型复用

- 欧拉化与路径生成
 - 复制必要的已有边，使图满足欧拉遍历条件
 - 遍历全部边，将连接关系转换为节点序列
- 节点重编号
 - 按节点首次出现顺序分配局部编号
 - 通过循环重编号，减轻编号token的训练偏差

- 属性附加与序列格式
 - 将节点属性、边属性和边方向编码到序列中
 - 提供short、long、prolonged三种输入格式



- 图欧拉化
 - 检查图是否满足欧拉遍历条件
 - 复制必要的已有边，使处理后的图能够完成全边遍历
- 路径生成
 - 随机选择一条有效欧拉路径，节点允许重复出现
- 关键作用
 - 利用**相邻节点**表示边连接，保留图的**拓扑关系**
 - 同一图可生成不同路径，为预训练提供不同序列视图



- 首次出现重编号

- 根据节点在路径中首次出现的顺序依次编号
- 同一节点再次出现时使用相同编号，保留共享身份

- 循环重编号

$$i' = (i + r) \bmod N$$

- i 为局部编号， r 为随机偏移量， N 为预设编号范围
- 通过随机偏移，将局部编号映射到不同编号token

- 关键作用

- 减少输入对原始节点编号的依赖
- 缓解低编号频繁出现、高编号训练不足的问题

F	A	B	C	A	C	D	E
				↓ Re-index			
0	1	2	3	1	3	4	5
				↓ Cyclic re-index			
13	14	15	16	14	16	17	18



- 属性编码
 - 离散属性直接转换为token
 - 连续属性可按数字逐位编码
 - 使用专用token表示边方向

//

long

13	[p]	14	[p]	15	[p]	16	[p]	14	[p]	16	[p]	17	[p]	18
n_1	[p]	n_1	[p]	n_1	[p]	n_1	[p]	n_1	[p]	n_1	[p]	n_1	[p]	n_1
[p]	e_t	[p]	e_t	[p]	e_t	[p]	e_t	[p]	e_t	[p]	e_t	[p]	e_t	[p]
[p]	e_1	[p]	e_1	[p]	e_1	[p]	e_1	[p]	e_1	[p]	e_1	[p]	e_1	[p]

Sequence len: $2(m - 1) + 1$

- 三种序列格式

格式	信息组织	序列长度
short	同一位置组合节点、边与属性信息	m
long	节点与边分别占据序列位置	$2(m-1)+1$
prolonged	各类信息进一步展开为独立token	$l(m-1)+1$

short

13	14	15	16	14	16	17	18
n_1	n_1	n_1	n_1	n_1	n_1	n_1	n_1
[p]	e_t	e_t	e_t	e_t	e_t	e_t	e_t
[p]	e_1	e_1	e_1	e_1	e_1	e_1	e_1

Sequence len: m

- 关键作用
 - 将结构与属性统一为Transformer输入
 - 通过不同格式权衡序列长度与信息组

织方式 prolonged 13 n_1 e_t e_1 14 n_1 e_t e_1 15 n_1 e_t e_1 16 n_1 e_t e_1 14 n_1 e_t e_1 16 n_1 e_t e_1 17 n_1 e_t e_1 18

Sequence len: $l(m - 1) + 1$

• 数据资源及指标

数据集	任务	指标
ogbn-arxiv	节点多分类	Accuracy
ogbn-proteins	节点多标签分类	ROC-AUC
ogbg-molpcba	图多标签分类	AP
Triangles	图结构类别预测	Accuracy
ogbl-ppa	链接预测	Hits@100
ogbl-citation2	链接预测	MRR
PCQM4Mv2	图回归	MAE

• 对比与验证

- 比较GNN、图Transformer与不同规模GraphGPT
- 验证预训练、节点重编号、身份编码和模型规模的影响

- 图回归：GraphGPT-B48的测试MAE为0.0804，优于GPTrans-L的0.0821

	Models	MAE ↓		Params
		Valid	Test	
GNN	GCN ¹	0.1379	0.1398	2.0M
	GIN ²	0.1195	0.1218	3.8M
	GCN ¹ -VN ³	0.1153	0.1152	4.9M
	GIN ² -VN ³	0.1083	0.1084	6.7M
GT	TokenGT ⁴	0.0910	0.0919	48.5M
	Graphformer ⁵	0.0864	N/A	48.3M
	GPS-Deep ⁶	0.0852	0.0862	138.1M
	GPS++ (no 3D) ⁷	0.0818	N/A	40.0M
	GPTrans-L ⁸	<u>0.0809</u>	<u>0.0821</u>	86.0M
Ours	GraphGPT-M	0.0827	N/A	37.7M
	GraphGPT-B ₁₂	<u>0.0807</u>	N/A	113.6M
	GraphGPT-B ₂₄	0.0793	N/A	227.3M
	GraphGPT-B ₄₈	0.0792	0.0804	453.4M

- 图分类：GraphGPT-B24的测试AP为31.81%，表现具有竞争力，但低于GPTrans-L的32.43%

Models	Average Precision (%) $\uparrow$		Params
	Test	Valid	
GCN ¹	20.20 $\pm$ 0.24	20.59 $\pm$ 0.33	0.57M
GIN ²	22.66 $\pm$ 0.28	23.05 $\pm$ 0.27	1.92M
GINE ³ -VN ⁴	29.17 $\pm$ 0.15	30.65 $\pm$ 0.30	6.1M
NGIN ⁵ -VN ⁴	30.07 $\pm$ 0.37	30.59 $\pm$ 0.56	44.19M
PDF ⁶	30.31 $\pm$ 0.26	31.15 $\pm$ 0.20	3.84M
Graphormer-L ^{†7}	31.40 $\pm$ 0.32	<u>32.27$\pm$0.24</u>	119.5M
EGT-Larger ^{†8}	29.61 $\pm$ 0.24	N/A	110.8M
GRPE-Large ^{†9}	31.50 $\pm$ 0.10	N/A	118.3M
GPTrans-L ^{†10}	32.43$\pm$0.22	N/A	86.0M
GraphGPT-M [†]	30.13 $\pm$ 0.25	31.62 $\pm$ 0.24	37.7M
GraphGPT-B ₁₂ [†]	31.28 $\pm$ 0.23	<u>32.27$\pm$0.15</u>	113.6M
GraphGPT-B ₂₄ [†]	<u>31.81$\pm$0.1</u>	32.54$\pm$0.2	227.3M

- 结构泛化：联合预训练后，大图分类准确率达到58.96%，高于Transformer-RWSE的54.76%

Models	Accuracy (%) $\uparrow$		Params
	T-small	T-large	
GIN ¹	71.53 $\pm$ 0.94	33.54 $\pm$ 0.30	0.15M
Transformer ²	12.08 $\pm$ 0.31	10.01 $\pm$ 0.04	0.2M
Transformer-LapPE ³	78.29 $\pm$ 0.25	10.64 $\pm$ 2.94	0.2M
Transformer-RWSE ³	99.40$\pm$0.10	<u>54.76$\pm$7.24</u>	0.2M
Graphormer ⁴	99.09$\pm$0.31	42.34 $\pm$ 6.48	0.2M
GET-B	32.60 $\pm$ 1.86	13.99 $\pm$ 1.78	113.5M
GraphGPT-B ^a	92.16 $\pm$ 0.28	26.51 $\pm$ 1.01	113.5M
GraphGPT-B ^b	81.38 $\pm$ 0.27	37.68 $\pm$ 0.99	113.5M
GraphGPT-B ^c	99.08$\pm$0.14	38.80 $\pm$ 3.60	113.5M
GraphGPT-B ^d	90.93 $\pm$ 0.51	40.79 $\pm$ 1.40	113.5M
GraphGPT-B ^e	64.28 $\pm$ 0.33	17.38 $\pm$ 0.61	113.5M
GraphGPT-B ^f	86.14 $\pm$ 7.38	26.94 $\pm$ 4.80	113.5M
GraphGPT-B ^g	86.57 $\pm$ 2.74	23.45 $\pm$ 1.44	113.5M
GraphGPT-B ^{a+b}	84.83 $\pm$ 0.81	39.62 $\pm$ 1.84	113.5M
GraphGPT-B ^{a+c}	<u>98.68$\pm$0.18</u>	50.07 $\pm$ 3.28	113.5M
GraphGPT-B ^{b+c}	<u>98.26$\pm$0.30</u>	52.33 $\pm$ 2.61	113.5M
GraphGPT-B ^{a+b+d}	89.98 $\pm$ 0.54	33.45 $\pm$ 2.51	113.5M
GraphGPT-M ^{a+b+c}	95.07 $\pm$ 0.67	51.72 $\pm$ 1.12	33.7M
GraphGPT-B ^{a+b+c}	<u>98.63$\pm$0.18</u>	58.96$\pm$1.90	113.5M

- 链接预测：ppa与citation2上的最佳结果分别为76.55%和93.05%，均超过表中基线

Models	ogbl-ppa HR@100 (%) $\uparrow$	ogbl-citation2 MRR (%) $\uparrow$
Common Neighbor	27.65 $\pm$ 0.00	51.47 $\pm$ 0.00
Adamic Adar	32.45 $\pm$ 0.00	51.89 $\pm$ 0.00
Resource Allocation ¹	49.33 $\pm$ 0.00	51.98 $\pm$ 0.00
Node2Vec ²	22.26 $\pm$ 0.83	61.41 $\pm$ 0.11
Matrix Factorization ³	32.29 $\pm$ 0.94	51.86 $\pm$ 4.43
GCN ⁴	18.67 $\pm$ 1.32	84.74 $\pm$ 0.21
GraphSAGE ⁵	16.55 $\pm$ 2.40	82.60 $\pm$ 0.36
SEAL ⁶	48.80 $\pm$ 3.16	87.67 $\pm$ 0.32
AGDN ⁷	41.23 $\pm$ 1.59	85.49 $\pm$ 0.29
SIEG ⁸	63.22 $\pm$ 1.74	90.18 $\pm$ 0.15
MPLP ⁹	65.24 $\pm$ 1.50	90.72 $\pm$ 0.12
RefinedGAE ¹⁰	73.74 $\pm$ 0.92	84.55 $\pm$ 0.15
GraphGPT-M	65.44 $\pm$ 0.43	92.82 $\pm$ 0.27
GraphGPT-B	68.76 $\pm$ 0.67	93.05$\pm$0.20
GraphGPT-XXL	76.55$\pm$0.67	N/A

- 节点分类: proteins与arxiv分别达到85.33%和72.10%，随模型规模增长而提高，但未超过最强GNN基线

Models	ogbn-proteins ROC-AUC (%) $\uparrow$	ogbn-arxiv Accuracy (%) $\uparrow$
GCN ^{1,2}	77.29 $\pm$ 0.46	73.53 $\pm$ 0.12
GraphSAGE ^{1,3}	82.21 $\pm$ 0.32	73.00 $\pm$ 0.28
GAT ^{1,4}	85.01 $\pm$ 0.46	73.30 $\pm$ 0.18
DRGAT ⁵	N/A	74.16$\pm$0.07
AGDN ⁶	88.65$\pm$0.13	73.41 $\pm$ 0.25
DeeperGCN ⁷	85.80 $\pm$ 0.17	71.92 $\pm$ 0.16
GraphGPS ^{1,8}	77.15 $\pm$ 0.64	71.23 $\pm$ 0.59
NAGphormer ^{1,9}	72.17 $\pm$ 0.45	70.88 $\pm$ 0.24
Expformer ^{1,10}	77.62 $\pm$ 0.33	72.32 $\pm$ 0.36
GOAT ^{1,11}	79.31 $\pm$ 0.42	72.76 $\pm$ 0.29
NodeFormer ^{1,12}	77.86 $\pm$ 0.84	67.78 $\pm$ 0.28
SGFormer ^{1,13}	79.92 $\pm$ 0.48	72.76 $\pm$ 0.33
Polynormer ^{1,14}	79.53 $\pm$ 0.67	73.40 $\pm$ 0.22
GraphGPT-S	83.56 $\pm$ 0.16	70.83 $\pm$ 0.33
GraphGPT-M	84.02 $\pm$ 0.21	71.20 $\pm$ 0.34
GraphGPT-B	<u>85.33$\pm$0.10</u>	72.10 $\pm$ 0.30

- 预训练作用：proteins测试ROC-AUC由无预训练Decoder的57.52%提升至SMTP的83.56%

DATASETS	PRE-TRAINING	TEST	VALID
PCQM4Mv2	χ^D	N/A	0.0978
	χ^E	N/A	0.0856
	NTP	N/A	0.0875
	SMTP	N/A	0.0807
OGBG-MOLPCBA	χ^D	12.80	13.31
	χ^E	25.80	26.33
	NTP	23.85	27.77
	SMTP	27.56	28.74
	SMTP*	27.20	28.49
	SMTP* + FT [†]	28.07	29.01
OGBL-PPA	χ^D	41.28	40.14
	χ^E	42.13	41.57
	NTP	55.56	54.87
	SMTP	55.68	54.93
OGBN-PROTEINS	χ^D	57.52	61.19
	χ^E	53.20	56.39
	NTP	75.61	80.47
	SMTP	83.56	87.73

- 节点重编号：两种规模模型的测试AP分别由23.10%→23.85%、22.70%→25.17%
- 身份编码：ppa与proteins的测试指标分别由44.38%→55.56%、60.22%→75.61%

PARAMS	RE-INDEX	PT LOSS	TEST	VALID
4.48M	✗	0.0844	23.10	25.25
	✓	0.0874	23.85	27.77
114.12M	✗	0.0689	22.70	26.21
	✓	0.0750	25.17	28.57

DATASETS	PARAMS	NIE	TEST	VALID
OGBL-PPA	14.75M	✗	44.38	45.08
		✓	55.56	54.87
OGBN-PROTEINS	10.76M	✗	60.22	65.66
		✓	75.61	80.47

- 算法贡献

- 欧拉序列化统一表达图结构、节点和边属性
- NTP / SMTP支持无标签生成式预训练
- 统一输入格式适配不同预测层级，验证模型规模扩展能力

- 算法不足

- 预训练与长序列处理需要较多计算资源
- 依赖数据集语义，跨域迁移能力存在限制
- 部分节点分类任务仍落后于较强GNN



Learning Graph Foundation Models on Riemannian Graph-of-Graphs

TIPO

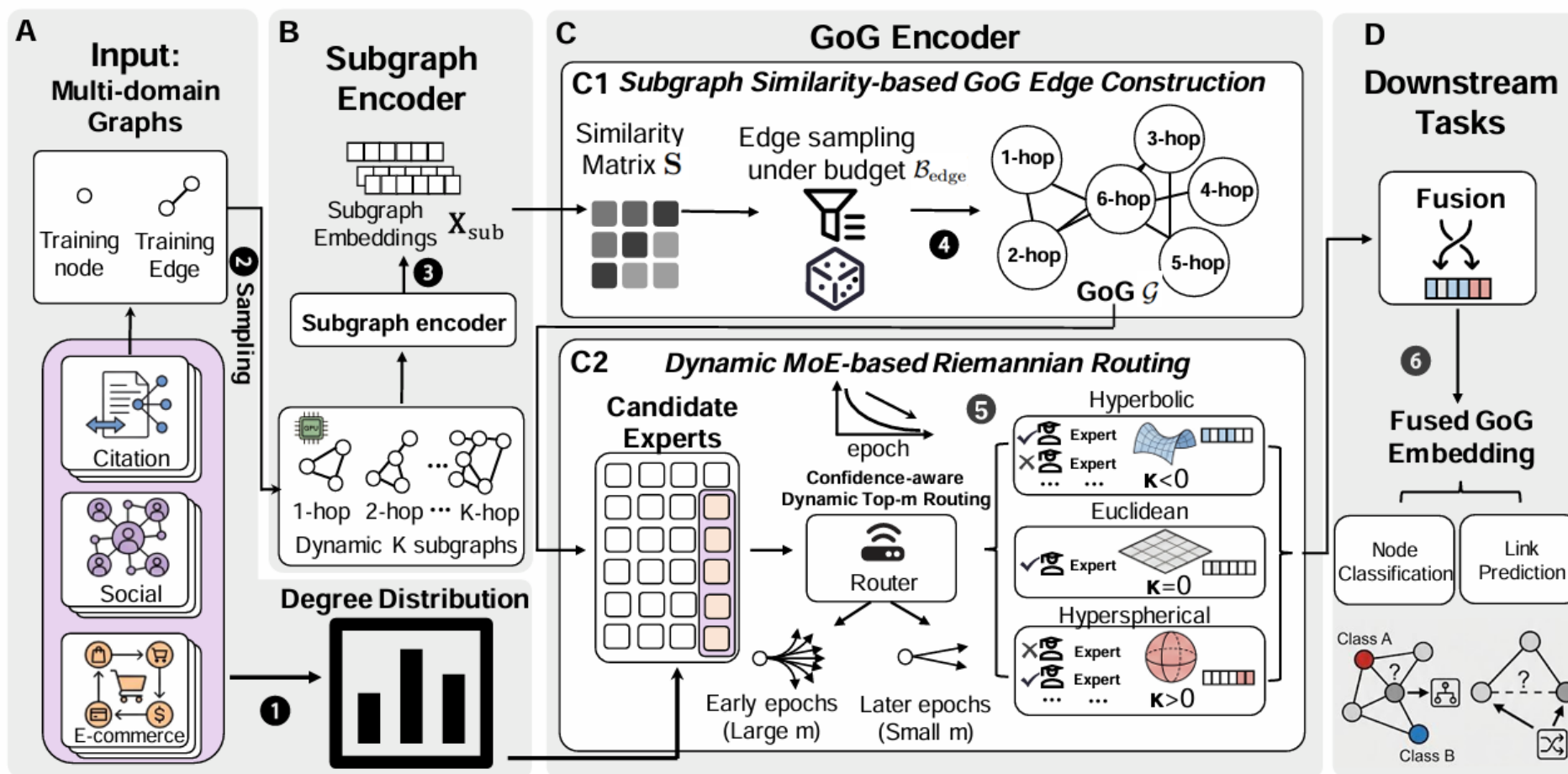
T	目标	学习可迁移的图表示，提高跨域少样本节点分类等任务的性能
I	输入	多源图与目标图的结构、属性；目标域每类1 / 3 / 5个标注样本
P	处理	1. 多跳子图采样：在显存预算内提取不同跳数的邻域子图。 2. 对比预训练：构造增强视图，学习子图表示。 3. GoG构建：以子图为节点，按表示相似性建立连接。 4. 黎曼专家路由：动态选择不同曲率专家，编码跨尺度关系。 5. 融合与任务适配：融合专家表示，利用少量标签完成下游预测。
O	输出	融合图表示；节点类别及其概率或链接预测分数

P	问题	固定跳数和单一表示空间难以适应不同图的结构尺度与几何差异
C	条件	源图结构属性可访问，具有显存预算，目标任务适配标签可用
D	难点	多尺度信息筛选、稀疏GoG构建、专家数量选择和跨域泛化
L	水平	ICML 2026, CCF A

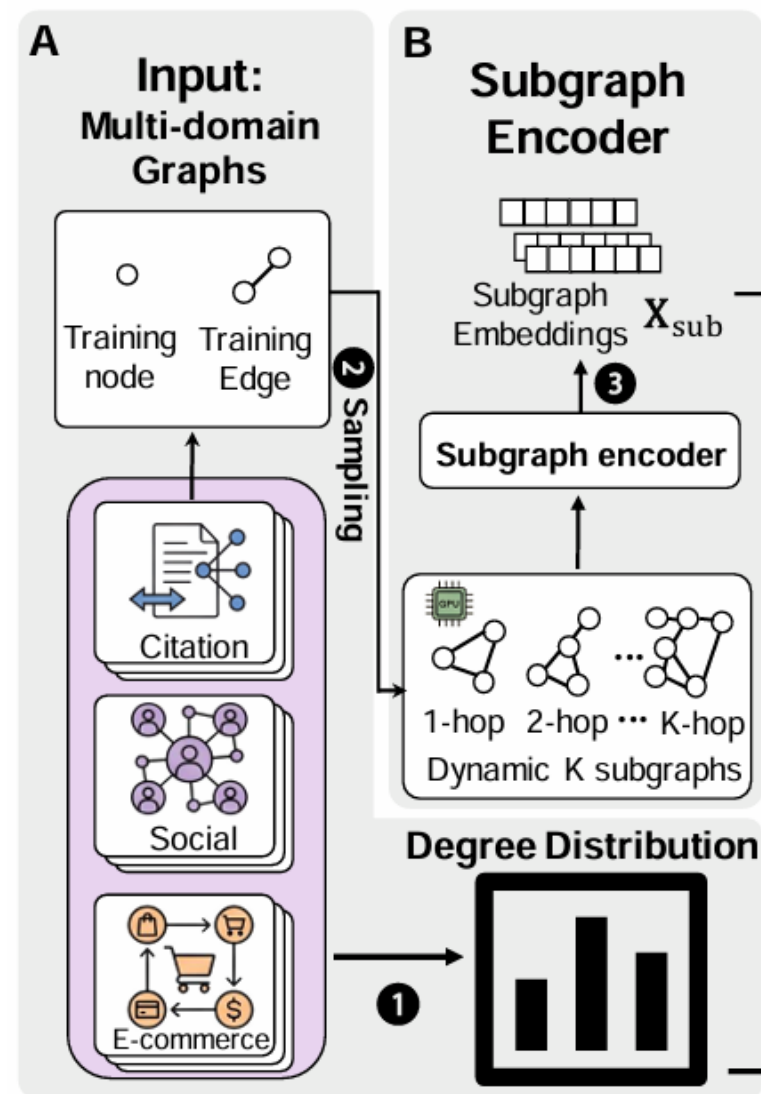
科学问题 应用问题

- 科学问题
 - 如何同时保留局部判别信息与较远范围的结构关系
 - 如何建立不同结构尺度之间的有用交互
 - 怎样依据图结构选择合适的几何表示空间
- 应用问题
 - 现有图模型采用固定邻域范围，难以同时适应引文、网页和商品网络的不同结构
 - 固定跳数可能造成感受野不足，也可能引入无关邻居
 - 新数据集标签稀缺，完整大图训练受到存储和计算限制

- R-GFM通过结构统计与多跳子图编码获取多尺度表示，利用GoG构建与动态黎曼专家路由学习跨尺度关系，最终融合表示完成下游预测



- 结构统计
 - 计算节点度的**变异系数**，描述图结构异质性
 - 根据结构统计确定候选几何专家集合
- 多跳采样
 - 以目标节点为中心，提取1跳至K跳子图
 - 根据显存预算确定最大可行K，获取**不同尺度**的结构信息
- 子图编码
 - 使用子图编码器，将不同跳数子图转换为嵌入
 - 通过增强视图与对比预训练，学习子图表示

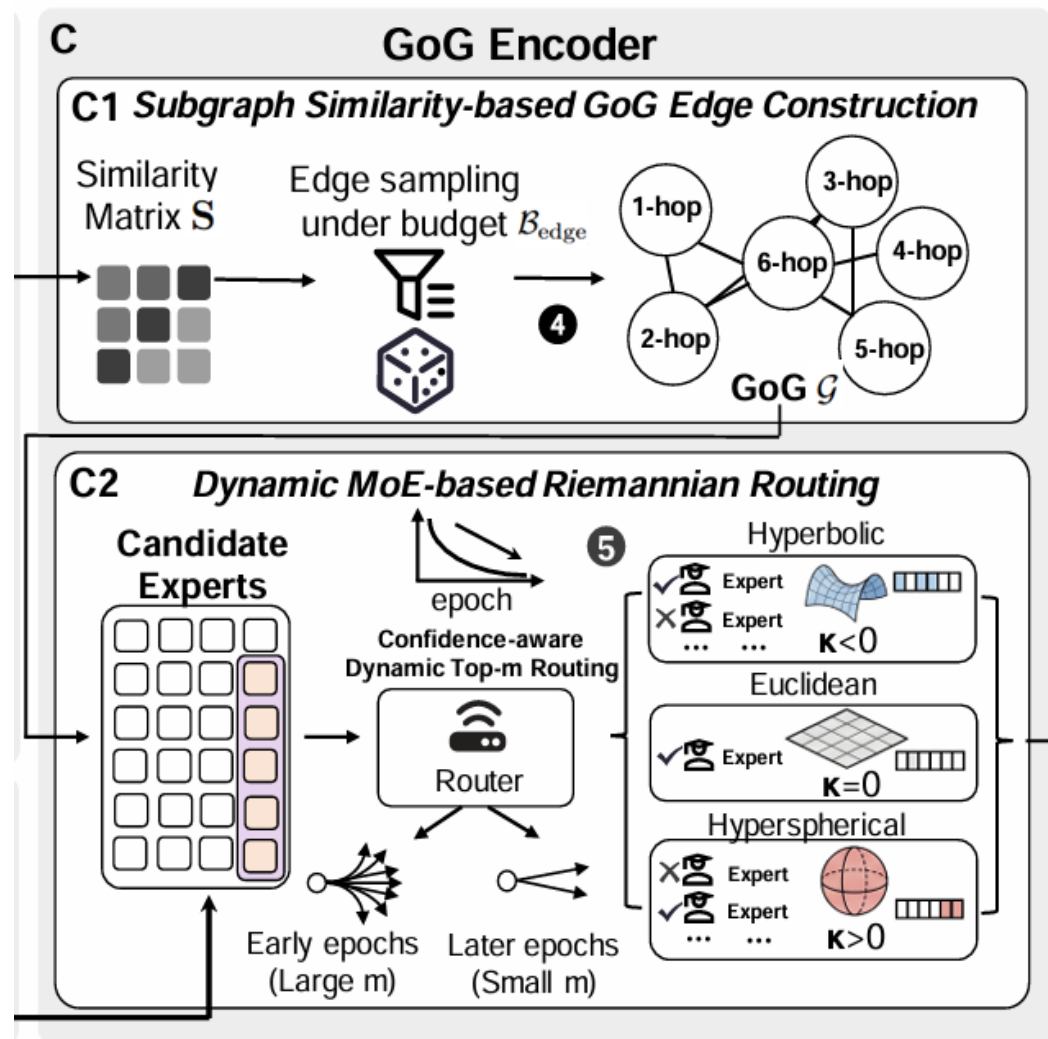


- 相似性GoG构建

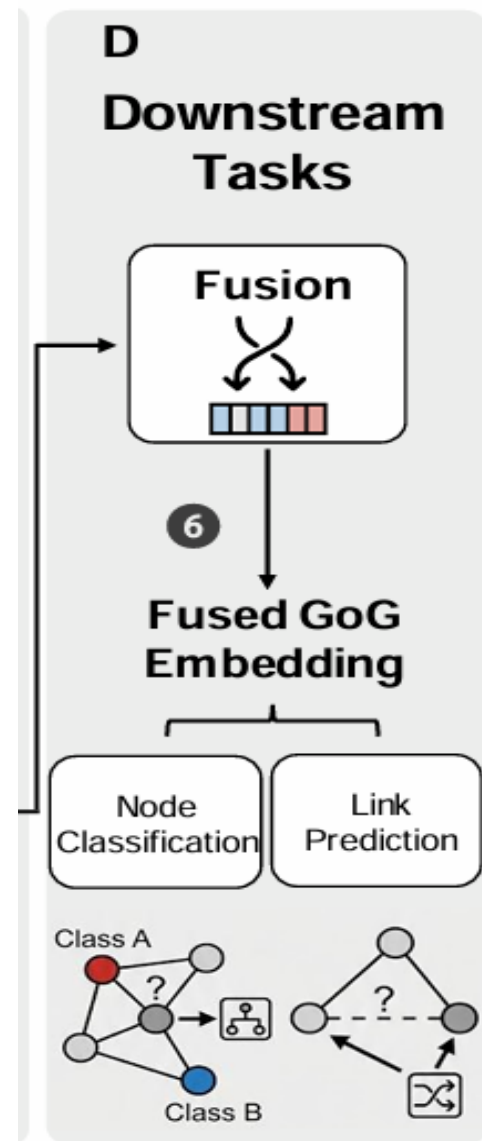
- 将不同跳数的子图嵌入作为高层图节点
- 根据嵌入**相似性**，在边预算内采样连接
- 通过稀疏连接建立不同结构尺度之间的交互多跳采样

- 动态黎曼专家路由

- 使用**双曲**、**欧氏**、**球面**等不同曲率专家编码GoG
- 路由器根据输入选择专家，并动态调整激活数量
- 获得适配不同结构特征的GoG表示



- 表示融合
 - 汇聚选中专家产生的GoG表示
 - 将多尺度结构与不同几何空间的信息融合为统一嵌入
- 任务适配
 - 将融合嵌入送入下游预测模块
 - 利用目标数据集的少量标签进行任务适配
- 模型输出
 - **节点分类**：预测目标节点类别
 - **链接预测**：预测节点之间的连接关系



- 数据资源
 - 引文: Cora、CiteSeer、PubMed
 - 网页: Cornell、Texas、Wisconsin
 - 商品: Amazon-Computers、Amazon-Photo
 - Wikipedia: Chameleon、Squirrel作为主要预训练来源，主结果表不将两者作为目标
- 评价指标
 - **Accuracy**，结果报告均值与标准差
- 对比方法
 - 任务监督GNN、自监督预训练方法、图提示方法、图基础模型

- 1-shot Cora上达到49.50%，高于GCOPE的39.05%与RiemannGFM的43.48%

Methods	Wisconsin	Cornell	Citeseer	Cora	Pubmed	Computers	Photos	Texas
Task-Supervised GNNs								
GCN	17.46 ± 3.35	19.53 ± 5.77	26.89 ± 3.01	31.98 ± 4.11	44.29 ± 2.22	39.43 ± 5.68	50.39 ± 1.19	18.48 ± 4.08
GAT	16.86 ± 3.16	16.51 ± 8.41	25.27 ± 2.10	26.81 ± 4.42	45.11 ± 5.52	38.05 ± 4.47	56.51 ± 2.35	18.36 ± 4.63
H2GCN	20.34 ± 5.77	29.30 ± 10.40	24.47 ± 1.35	23.61 ± 3.92	40.97 ± 1.26	19.29 ± 6.87	42.95 ± 4.32	29.94 ± 16.38
FAGCN	26.19 ± 7.93	28.72 ± 8.12	20.56 ± 1.02	20.75 ± 3.85	41.65 ± 1.46	27.02 ± 2.97	47.91 ± 2.17	31.35 ± 14.17
Self-Supervised Pretraining with Fine-tuning								
DGI	28.24 ± 5.30	29.73 ± 10.11	34.52 ± 8.81	39.88 ± 5.53	47.10 ± 4.95	34.59 ± 5.94	50.66 ± 6.93	29.73 ± 6.34
GraphCL	32.16 ± 8.83	31.35 ± 12.47	27.56 ± 5.01	35.26 ± 8.62	43.66 ± 3.12	34.40 ± 6.93	48.86 ± 5.53	31.35 ± 22.66
GraphACL	29.41 ± 11.09	25.41 ± 9.86	28.34 ± 4.19	36.96 ± 8.71	38.20 ± 10.11	44.42 ± 7.06	56.60 ± 8.90	24.86 ± 9.44
MixHop	29.82 ± 9.32	17.14 ± 5.54	21.35 ± 3.67	21.46 ± 5.95	43.57 ± 4.05	29.77 ± 15.57	32.77 ± 9.26	26.09 ± 15.93
Prompt-based Adaptation								
GPPT	23.53 ± 15.93	25.41 ± 5.27	33.24 ± 8.73	40.60 ± 6.00	41.64 ± 9.29	44.57 ± 13.38	53.94 ± 11.93	18.38 ± 12.00
GPF	23.53 ± 9.36	36.22 ± 9.46	27.16 ± 4.10	34.74 ± 3.48	26.62 ± 10.18	42.15 ± 8.75	59.13 ± 6.05	28.65 ± 17.31
GraphPrompt	20.39 ± 7.92	28.65 ± 5.92	19.80 ± 2.79	33.68 ± 9.83	42.94 ± 1.98	33.34 ± 6.09	42.60 ± 6.75	24.32 ± 9.17
GraphPrompt+	16.86 ± 5.63	29.19 ± 5.24	28.28 ± 3.55	29.76 ± 6.06	38.74 ± 1.70	43.01 ± 6.72	58.23 ± 4.61	19.46 ± 14.84
Graph Foundation Models								
GCOPE	28.69 ± 7.00	25.00 ± 6.58	38.62 ± 2.58	39.05 ± 2.53	44.52 ± 2.51	28.50 ± 4.91	59.17 ± 2.45	19.50 ± 12.18
SAMGPT	21.57 ± 12.09	36.22 ± 5.30	28.78 ± 8.99	43.08 ± 10.40	37.95 ± 9.58	45.80 ± 4.54	49.80 ± 11.89	31.89 ± 11.26
MDGFM	17.65 ± 7.34	26.49 ± 10.73	23.32 ± 15.32	42.88 ± 9.12	38.42 ± 7.92	42.68 ± 2.53	57.78 ± 9.07	31.54 ± 11.78
RiemannGFM	31.20 ± 9.35	31.35 ± 7.76	31.86 ± 6.45	<u>43.48 ± 5.28</u>	44.10 ± 5.05	46.20 ± 8.79	57.20 ± 16.84	29.12 ± 7.44
GPM	29.19 ± 4.69	27.83 ± 11.00	29.88 ± 6.36	38.96 ± 8.18	35.80 ± 8.20	48.58 ± 10.66	37.85 ± 8.17	27.70 ± 12.09
G2PM	34.70 ± 8.73	32.96 ± 5.44	31.04 ± 2.90	40.04 ± 3.13	49.71 ± 8.66	51.92 ± 10.10	59.90 ± 5.02	30.37 ± 9.36
R-GFM	35.41 ± 7.29	36.71 ± 9.92	57.54 ± 9.49	49.50 ± 3.97	49.80 ± 5.38	52.30 ± 3.33	61.08 ± 5.26	32.36 ± 12.10

Results are reported in percent. The best method is bolded, and the runner-up is underlined.

- 3-shot Cora上达到59.26%

Methods	Wisconsin	Cornell	Citeseer	Cora	Pubmed	Computers	Photos	Texas
Task-Supervised GNNs								
GCN	15.51 ± 5.61	20.00 ± 2.04	33.61 ± 4.98	40.12 ± 6.52	40.97 ± 2.16	31.54 ± 5.47	59.05 ± 4.60	21.95 ± 11.22
GAT	18.94 ± 1.99	15.46 ± 5.51	22.07 ± 8.64	29.20 ± 3.13	34.70 ± 7.32	38.92 ± 9.21	47.20 ± 4.62	28.05 ± 12.06
H2GCN	21.32 ± 13.55	31.17 ± 11.47	23.29 ± 2.97	24.94 ± 6.05	42.95 ± 3.02	13.77 ± 4.30	35.97 ± 4.50	16.95 ± 1.36
FAGCN	31.32 ± 5.57	31.90 ± 11.89	20.15 ± 0.65	19.75 ± 6.68	40.09 ± 1.09	25.51 ± 5.75	42.35 ± 2.53	26.95 ± 8.06
Self-Supervised Pretraining with Fine-tuning								
DGI	36.08 ± 3.82	30.27 ± 14.48	45.26 ± 2.85	58.26 ± 2.61	52.94 ± 7.65	45.15 ± 6.63	68.87 ± 9.46	40.00 ± 5.54
GraphCL	36.08 ± 10.61	44.32 ± 13.46	32.02 ± 3.48	45.30 ± 6.52	44.20 ± 1.53	49.62 ± 7.42	62.69 ± 0.12	41.62 ± 10.22
GraphACL	27.45 ± 6.04	27.03 ± 6.04	38.14 ± 5.59	57.10 ± 4.72	45.04 ± 3.37	46.23 ± 8.33	66.85 ± 3.98	37.84 ± 9.56
Prompt-based Adaptation								
GPPT	28.63 ± 3.28	35.68 ± 9.63	43.66 ± 6.31	39.22 ± 7.75	38.06 ± 10.37	49.70 ± 6.78	56.04 ± 8.56	39.46 ± 15.94
GPF	34.12 ± 10.35	28.65 ± 5.01	33.24 ± 6.79	46.22 ± 4.83	34.80 ± 14.11	42.67 ± 7.08	60.57 ± 8.19	41.62 ± 3.24
GraphPrompt	29.02 ± 6.71	32.43 ± 10.81	28.94 ± 2.62	44.42 ± 5.20	51.44 ± 3.68	42.62 ± 7.13	57.73 ± 7.06	30.27 ± 5.86
GraphPrompt+	27.06 ± 4.00	31.35 ± 11.79	35.14 ± 5.81	46.12 ± 5.10	52.16 ± 4.24	54.43 ± 4.53	64.42 ± 7.58	34.59 ± 9.58
Graph Foundation Models								
GCOPE	38.21 ± 4.22	23.44 ± 12.42	37.52 ± 1.48	58.71 ± 1.96	55.26 ± 1.93	53.01 ± 3.31	68.15 ± 2.54	34.90 ± 11.04
SAMGPT	30.98 ± 7.78	29.73 ± 11.21	46.46 ± 10.02	45.76 ± 3.84	48.26 ± 6.58	<u>55.27 ± 6.04</u>	62.98 ± 8.16	25.95 ± 15.06
MDGFM	34.90 ± 6.37	29.73 ± 10.40	47.72 ± 4.63	57.52 ± 6.12	42.86 ± 4.88	48.44 ± 10.53	62.84 ± 4.87	35.14 ± 11.72
RiemannGFM	40.40 ± 5.71	44.86 ± 8.65	47.44 ± 7.81	58.20 ± 2.10	53.60 ± 13.59	40.77 ± 6.98	70.82 ± 3.10	33.51 ± 11.29
R-GFM	43.10 ± 4.03	45.39 ± 3.09	73.98 ± 1.42	59.26 ± 5.04	59.19 ± 3.01	56.02 ± 3.90	71.50 ± 2.80	44.18 ± 7.02

Results are reported in percent. The best method is bolded, and the runner-up is underlined.

- 5-shot Cora上达到63.77%

Methods	Wisconsin	Cornell	Citeseer	Cora	Pubmed	Computers	Photos	Texas
Task-Supervised GNNs								
GCN	27.00 ± 6.60	20.92 ± 6.02	31.61 ± 6.05	39.48 ± 6.08	38.82 ± 2.98	30.06 ± 10.72	66.96 ± 9.21	38.08 ± 13.65
GAT	28.85 ± 8.91	18.30 ± 7.49	23.99 ± 4.10	29.82 ± 2.31	34.92 ± 12.27	37.12 ± 8.21	54.37 ± 4.90	28.21 ± 11.78
H2GCN	28.39 ± 13.19	35.82 ± 10.30	22.29 ± 3.74	20.83 ± 6.56	42.61 ± 3.61	19.67 ± 11.45	47.07 ± 3.78	32.82 ± 14.56
FAGCN	24.61 ± 4.70	25.36 ± 9.00	18.69 ± 1.31	15.82 ± 3.01	35.92 ± 7.71	28.68 ± 1.89	50.83 ± 3.79	46.15 ± 9.94
Self-Supervised Pretraining with Fine-tuning								
DGI	39.61 ± 3.77	38.92 ± 6.78	53.66 ± 6.21	<u>63.72 ± 2.23</u>	57.02 ± 3.47	52.03 ± 5.29	69.74 ± 4.95	42.70 ± 9.04
GraphCL	39.61 ± 8.25	42.70 ± 5.54	39.26 ± 3.63	<u>49.34 ± 4.94</u>	44.88 ± 2.40	57.78 ± 3.96	69.39 ± 6.46	47.03 ± 13.05
GraphACL	30.98 ± 6.56	29.73 ± 1.91	44.68 ± 6.36	62.90 ± 5.69	50.10 ± 5.39	51.60 ± 6.46	65.22 ± 8.34	34.59 ± 7.74
Prompt-based Adaptation								
GPPT	35.29 ± 7.72	44.86 ± 4.52	45.98 ± 5.97	48.78 ± 5.49	48.88 ± 8.25	30.14 ± 12.77	64.89 ± 2.38	<u>48.11 ± 10.71</u>
GPF	33.73 ± 3.37	33.51 ± 6.96	38.34 ± 6.35	49.98 ± 5.98	43.00 ± 3.48	47.46 ± 6.54	70.89 ± 3.81	<u>30.81 ± 12.86</u>
GraphPrompt	31.37 ± 3.10	35.14 ± 8.55	35.96 ± 1.97	48.24 ± 3.43	58.56 ± 2.40	48.43 ± 4.50	64.34 ± 3.99	27.03 ± 3.82
GraphPrompt+	27.84 ± 5.46	32.43 ± 10.11	41.60 ± 5.29	51.62 ± 2.00	58.76 ± 6.22	57.04 ± 5.29	69.34 ± 2.48	45.95 ± 10.81
Graph Foundation Models								
GCOPE	34.38 ± 3.93	35.92 ± 9.30	47.32 ± 1.59	57.33 ± 2.08	60.17 ± 1.81	55.76 ± 3.86	68.28 ± 2.45	26.34 ± 10.63
SAMGPT	39.22 ± 9.20	41.08 ± 11.13	<u>59.44 ± 5.70</u>	62.76 ± 3.93	47.08 ± 9.10	<u>58.81 ± 7.14</u>	72.08 ± 4.20	31.35 ± 6.07
MDGFM	36.08 ± 7.40	41.08 ± 7.72	<u>57.08 ± 3.71</u>	61.86 ± 4.40	48.84 ± 9.86	<u>49.68 ± 4.93</u>	70.37 ± 7.36	21.08 ± 5.77
RiemannGFM	<u>40.40 ± 8.71</u>	<u>44.86 ± 4.05</u>	54.68 ± 3.67	62.92 ± 5.67	<u>61.48 ± 6.54</u>	48.93 ± 5.75	<u>73.47 ± 6.42</u>	41.08 ± 12.01
R-GFM	47.75 ± 4.97	47.97 ± 2.82	74.59 ± 1.87	63.77 ± 2.47	63.39 ± 1.85	59.55 ± 4.17	73.62 ± 1.15	51.64 ± 6.31

Results are reported in percent. The best method is bolded, and the runner-up is underlined.

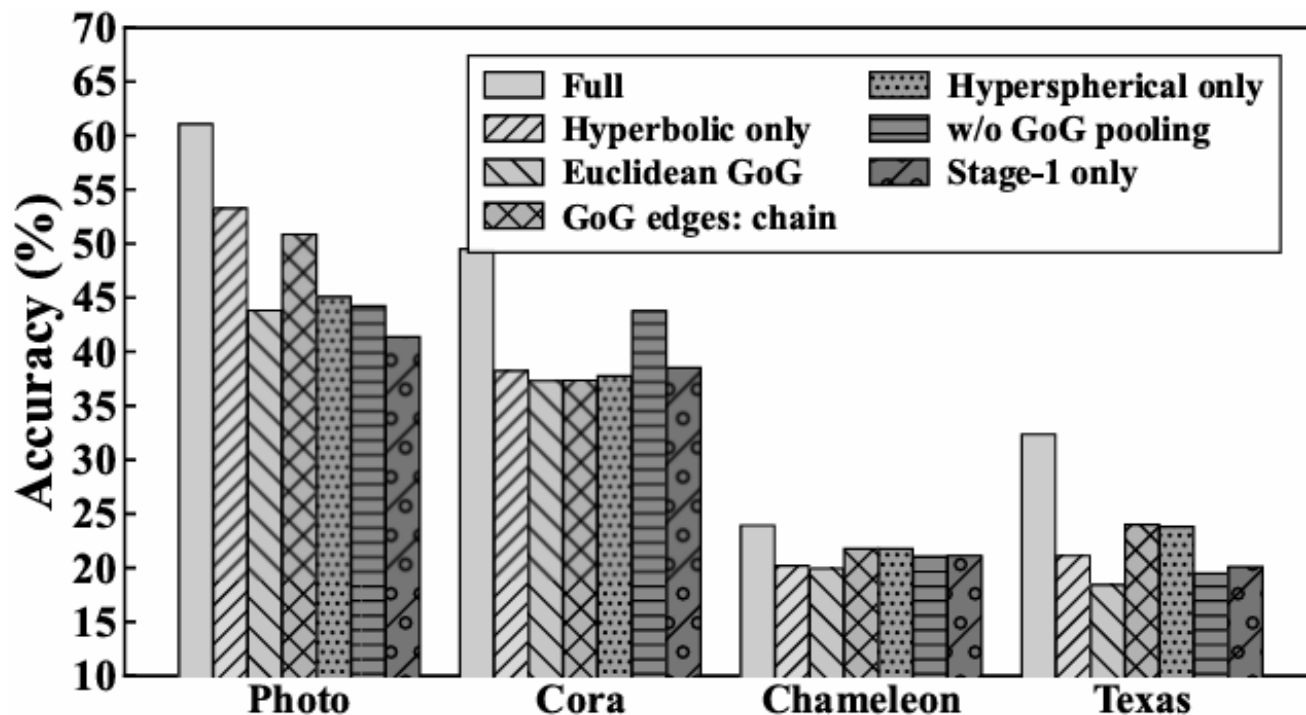
- 跨域迁移：统一语言特征设置下，Cora准确率达到61.78%，高于RiemannGFM的50.92%

Methods	Books-History	Ele-Computers	Cora	Instagram
GCOPE	<u>30.67 ± 7.90</u>	23.10 ± 3.38	37.80 ± 2.58	<u>59.17 ± 7.40</u>
SAMGPT	OOM	OOM	OOM	OOM
MDGFM	OOM	OOM	OOM	OOM
RiemannGFM	17.66 ± 9.12	<u>24.56 ± 8.97</u>	<u>50.92 ± 3.58</u>	46.60 ± 9.87
R-GFM	37.05 ± 11.06	30.73 ± 5.31	61.78 ± 7.77	59.49 ± 4.46

- 链接预测：Cora、PubMed、Photo、Texas的ROC-AUC分别为89.27%、88.66%、81.53%、87.94%，均高于表中基线

Methods	Cora	Pubmed	Photos	Texas
GCOPE	<u>88.71 ± 0.57</u>	84.99 ± 0.06	78.45 ± 5.51	<u>80.51 ± 1.75</u>
SAMGPT	88.62 ± 3.51	84.92 ± 0.40	80.59 ± 0.33	51.20 ± 4.42
MDGFM	88.41 ± 2.12	84.78 ± 2.09	81.24 ± 5.80	65.22 ± 4.05
RiemannGFM	88.54 ± 0.11	<u>86.37 ± 0.33</u>	<u>81.44 ± 0.95</u>	72.41 ± 5.96
R-GFM	89.27 ± 0.64	88.66 ± 5.16	81.53 ± 0.83	87.94 ± 0.96

- GoG构建: 仅保留第一阶段编码或去除GoG汇聚, 分类性能下降
- 连接方式: 将相似性连接替换为链式连接, 削弱跨尺度建模效果
- 几何编码: 将黎曼编码器替换为欧氏编码器, 多个数据集上的结果下降
- 专家组合: 仅使用双曲或球面空间均不及完整模型, 支持混合几何专家的有效性



• 算法贡献

- 将不同跳数子图组织为GoG，显式学习跨尺度结构关系
- 以结构统计确定候选几何专家，并动态调整激活数量
- 在跨数据集少样本分类、扰动和消融实验中验证设计

• 算法不足

- 子图构建、GoG编码和几何专家增加实现复杂度
- 度分布统计只能概括部分结构特征
- 显存可行的最大K不一定是语义最优K
- 迁移结论受到源域覆盖、特征对齐和目标标签条件限制



特点总结与未来展望

- 特点总结

- GraphGPT: 将图结构与属性转化为统一序列, 支持生成式图预训练
 - 欧拉遍历与节点重编号保留连接关系, 结合属性编码形成适配Transformer的输入
 - 通过NTP或SMTP学习结构与属性关联, 并以统一任务格式支持节点、边和图级预测
- R-GFM: 突破固定结构尺度与单一几何表示空间的限制
 - 多跳子图编码与相似性GoG构建, 建立不同结构尺度之间的有效交互
 - 通过结构统计确定候选几何专家, 结合动态路由与表示融合提升跨域少样本预测能力

- 未来展望

- 探索图序列预训练与多尺度几何建模的结合, 提升分类表示的判别性与迁移能力
- 加强跨域、少样本与结构扰动条件下的泛化验证, 兼顾预测性能与计算资源需求

- [1] ZHAO Q, REN W, LI T, et al. GraphGPT: Generative Pre-trained Graph Eulerian Transformer[C]//Proceedings of the 42nd International Conference on Machine Learning. PMLR, 2025, 267: 77630–77656.
- [2] LIU H, DING Z, XIE X. Learning Graph Foundation Models on Riemannian Graph-of-Graphs[C]//Proceedings of the 43rd International Conference on Machine Learning. PMLR, 2026, 306: 75862–75890.

知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者，寿。

谢谢！

