

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



属性缺失异质图神经网络

硕士研究生 杨景然

2025年02月16日

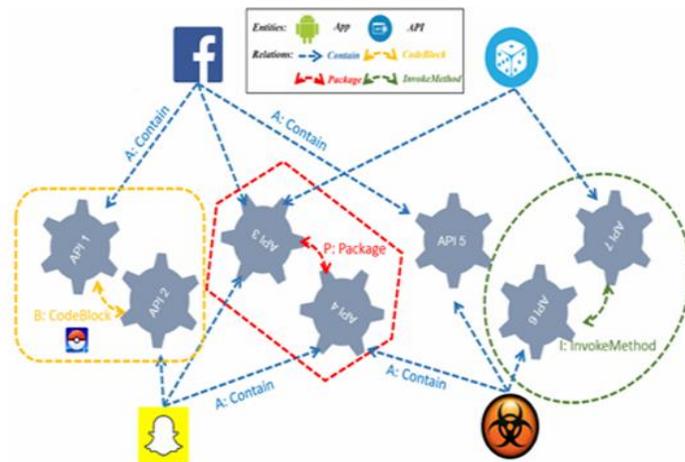
- **总结反思**
 - 讲解语速较快，时间较短，语气词较多
 - 内容详略不当，算法部分讲解深度不足、浅尝辄止
 - 互动部分较少
- **相关内容**
 - 2024.06.02 夏志豪 《图神经网络的反事实解释方法》
 - 2022.09.25 李新帅 《异质图神经网络》

- 预期收获
- 内容引入
- 知识基础
- 内涵解析与研究目标
- 研究背景与研究意义
- 算法原理
 - AC-HEN
 - RA-HGNN
- 特点总结与工作展望
- 参考文献

- 预期收获
 - 掌握属性图、异质图和属性缺失异质图的基本概念
 - 了解属性缺失异质图神经网络发展历程
 - 理解属性缺失异质图神经网络的基本原理
 - 明确属性缺失异质图神经网络的应用和前沿发展

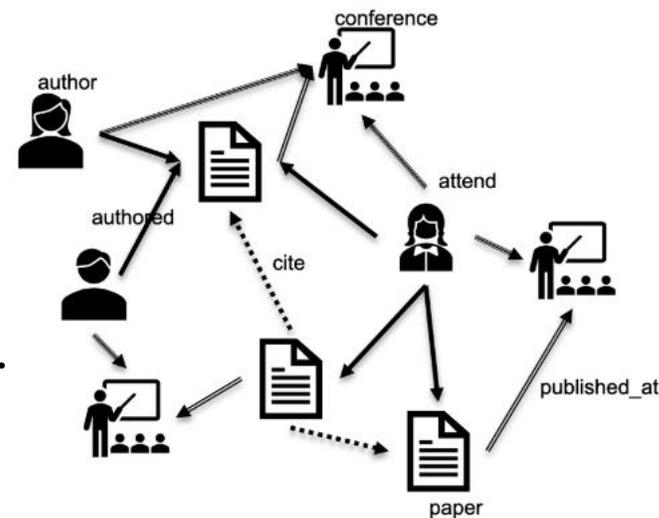
- Deepseek网络安全事件分析过程中我们关注什么？

- 攻击方、被攻击方
- 恶意行为、攻击手段、是否有防御行为
- 攻击路径
- 更多的：恶意行为造成的损失……



- 研究领域调研过程中我们关注什么？

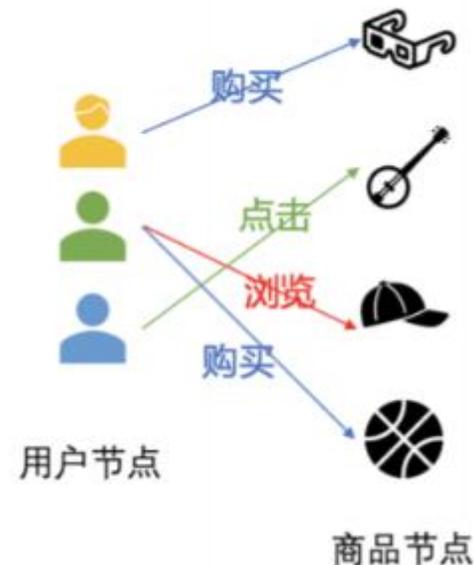
- 领域大牛、杰出作者有哪些？
- 他们发表了哪些优秀论文？
- 优秀文章、开山之作都引用了哪些文献？
- 更多的：领域大牛的研究兴趣、优秀论文的核心思想……



异质图

- 基本概念

- 图数据：由顶点的有穷非空集合 $V(G)$ 和顶点之间边的集合 $E(G)$ ，通常表示为 $G = (V, E)$
- 同质图：图中的所有节点、边都是同一种类型
- 异质图：图中包含多种不同类型的节点、边
 - 节点类型+边类型 > 2
 - $G = (V, E, y, s)$ ， y 是节点类型的集合， s 是边类型的集合
 - 将不同类型的数据合到一起，加入更多的语义信息数据
 - 结合node和relation信息来提高表征能力



异质图还有什么不足？

• 基本概念

– 属性异质图：异构图的一种扩展，不仅关注节点和边的类型，还引入了节点属性

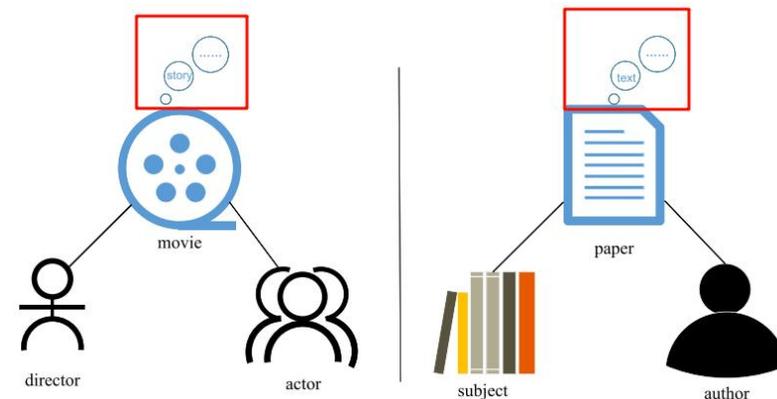
- 节点属性：用户的信息、商品的价格、**文章的关键词**等
- 边属性：用户与商品之间的**购买金额**、评论的星级等
- **节点属性**更具普遍性

– 优势

- 通过引入属性异构图，可以实现**细粒度**的建模，使得图神经网络或其他图学习方法能够利用这些丰富的信息

– 现实问题

- 部分属性获得**成本高昂**或无法获得
- 例如：社交网络或电商网络中多数用户不填写自己的身份信息等



属性缺失异质图

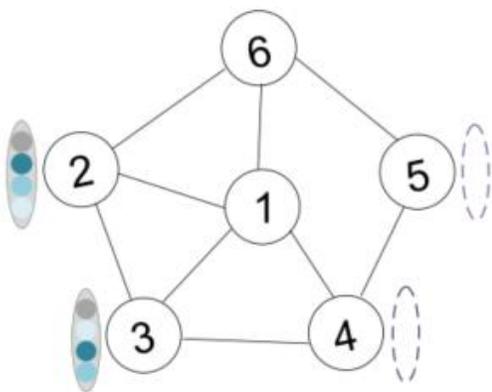
节点属性缺失类型:

- 属性完全缺失
- 属性部分缺失

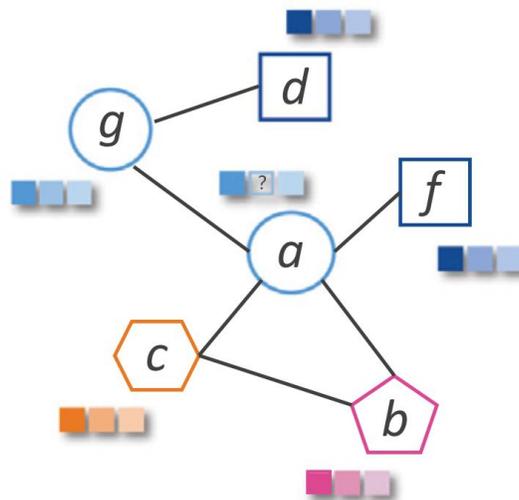
属性部分缺失更具普遍性

定义: 通常表示为 $G = (V, E, X, M)$

X : 节点属性矩阵 M : 缺失属性节点标识矩阵



节点属性完全缺失



节点属性部分缺失



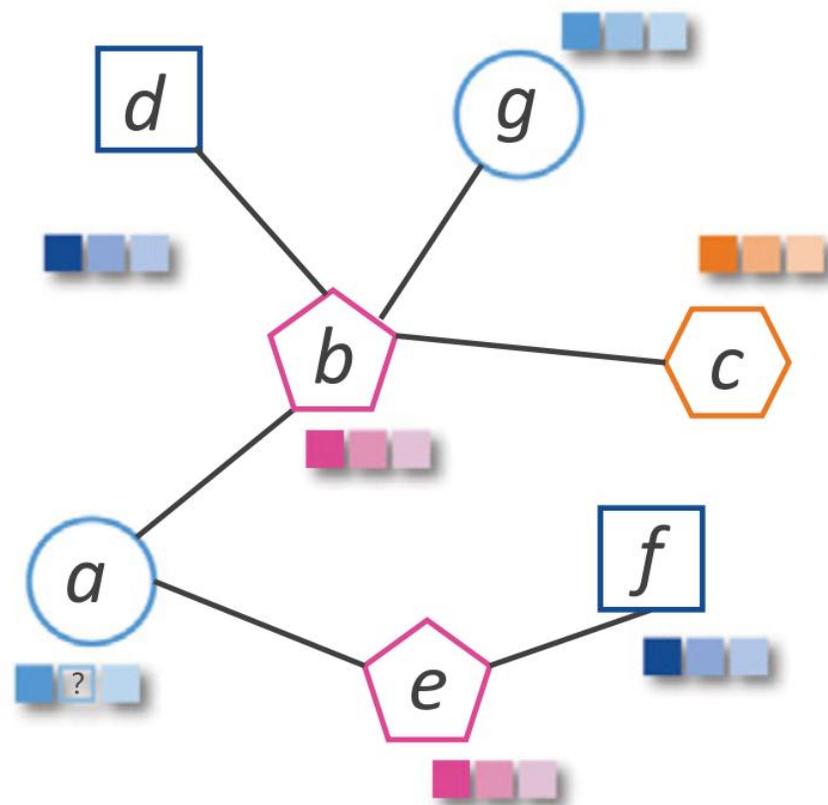
你好端端地缺失干嘛?

- 内涵解析

- 属性缺失异质图：在一个由不同类型节点和边构成的异质图中，**部分节点或边**的属性信息缺失
- 图神经网络是**学习图结构数据**的深度学习网络

- 研究目标

- 以**属性缺失异质图**数据为研究对象
- **结合残差神经网络、图神经网络、注意力机制**
- 探索异质图中的节点之间的**信息传递机制**
- 提取和发掘图结构数据中的特征和模式
- 满足聚类、分类、预测、分割、生成等下游任务



- 研究背景

- 异质图的广泛应用：社交、学术、电商网络
- 属性缺失的**普遍性**：数据缺失是常态，尤其是在涉及多模态、跨域数据的异质图中
- 图神经网络（GNN）的发展：GNN通过信息传播机制，能够有效地从图结构中学习节点和边的表示

- 研究意义

- 提高异质图数据的利用效率
- 跨领域应用潜力
- 为**下游任务**（节点分类、节点聚类提供帮助）



研究历史 属性缺失异质图神经网络

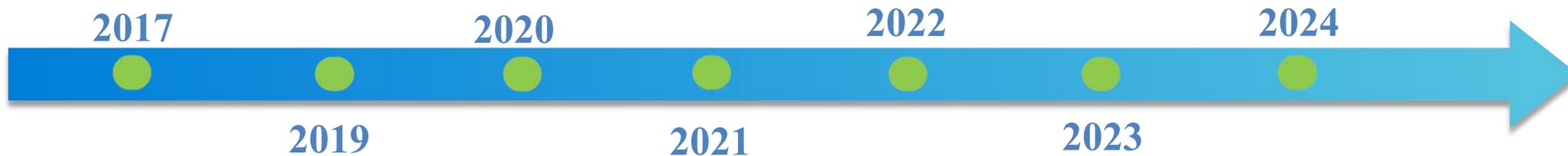


xiao等人提出metapath2vec方法使用通过元路径引导的随机游走生成节点表示，奠定了异质图表示学习的基础

Fu等人提出了MAGNN方法：元路径基础上聚合节点属性，增强节点分类任务的性能

Wang等人提出一个通用属性补全框架，设计基于注意力的高阶邻居聚合，更明确地探索高阶邻居节点对属性补全的作用

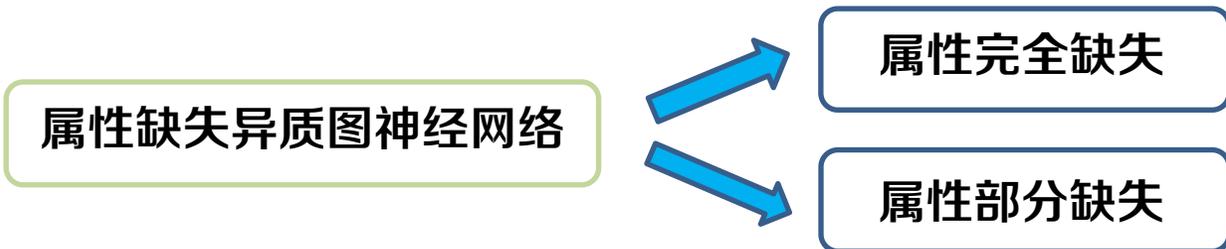
Zhao等人为解决多维度节点特征整合困难，引入类型转换矩阵优化节点嵌入。利用残差注意力机制，聚合有属性节点的信息为无属性节点补全属性



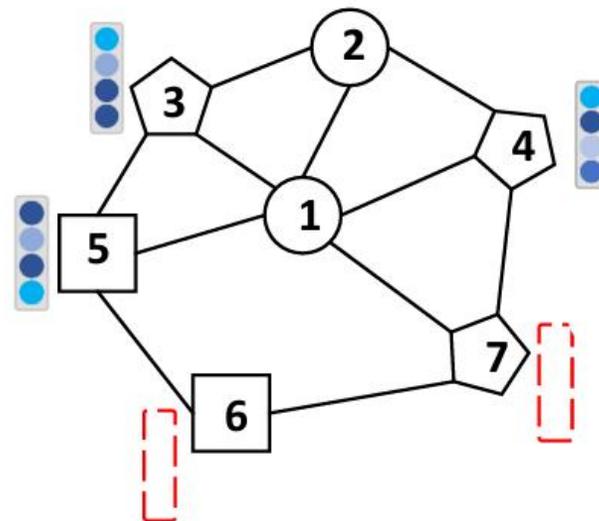
2017
2019
Shao等人提出HAN方法，引入注意力机制动态学习不同元路径的重要性，增强异质图节点表征

2020
2021
Shao等人首次系统提出异质图属性补全问题，通过拓扑先验和注意力机制动态补全缺失属性，并与下游任务端到端联合优化，显著提升节点分类性能

2022
2023
Zhang等人针对属性缺失场景的抗干扰补全方法。通过引入注意力净化器，动态屏蔽对抗攻击导致的噪声属性，确保补全过程的可靠性



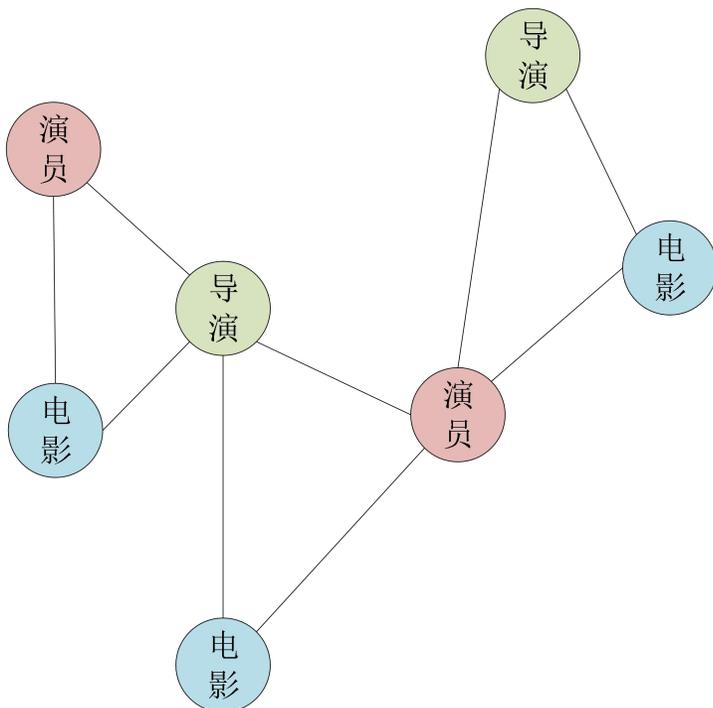
- 属性补全与图学习的联合优化
 - 补全与图神经网络联合优化，通过端到端训练提升整体性能
 - 早期方法将属性补全视为独立的数据预处理任务（如矩阵填充、插值法）
- 元路径与上下文感知补全
 - 元路径（meta-path）是异质图的核心语义结构
 - 自适应加权策略（图融合过程中各视图权重）
 - 半监督和弱监督学习
 - 注意力机制与高阶邻居的探索
 - 例如，HGNN-AC（2021）利用双重注意力机制分别捕获结构依赖和属性相关性，实现上下文感知的属性补全





元路径

- 连接不同类型节点的一条路径
- 不同的元路径会有不同的路径类型，通常用节点类型路径来表示

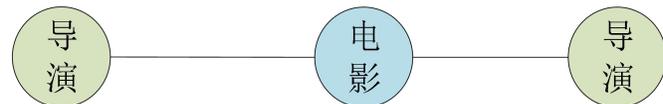


电影 (M) 导演 (D) 演员 (A)

MAM



DMD



DMAMD



AMA



AMDMA



评价指标

- 属性补全的评价指标

- Heat Kernel和Pearson相关系数都是**直接计算**补全后的属性与真实属性之间的**相似性**

- x_u^c 是补全后的节点 u 的属性值（即估计值）， x_u^g 是 u 的真实属性值

- Heat Kernel

- $Heat\ Kernel = e^{-\frac{\|x_u^c - x_u^g\|^2}{t}}$

- 设置 $t = 2$

- Pearson相关系数

- $P = \frac{cov(X_u^c, X_u^g)}{\sigma_{x_u^c} \sigma_{x_u^g}}$

- $cov(X_u^c, X_u^g)$ 是 x_u^c 和 x_u^g 的协方差， $\sigma_{x_u^c}, \sigma_{x_u^g}$ 是方差



Heterogeneous graph neural network for attribute completion

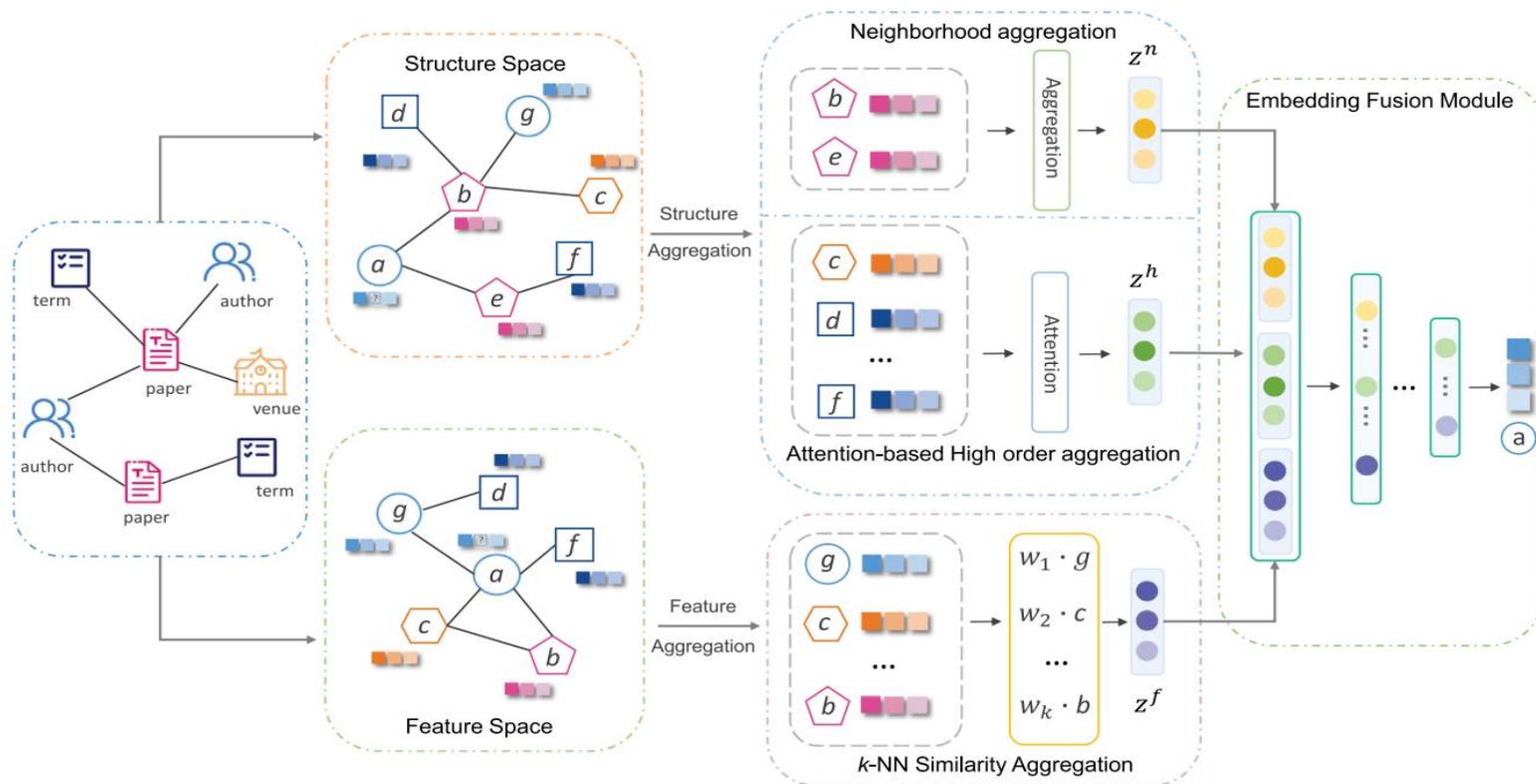
TIPO

T	目标	完成基于GNN的异构图模型中的 节点属性 补全
I	输入	部分属性缺失 的图数据 DBLP (14,328篇论文、4,057位作者、20个会议/期刊、8,789个术语) ACM (4019篇论文、7167位作者、60个主题) IMDB (780部电影, 5841名演员和2269名导演)
P	处理	1.特征聚合: 结合KNN与余弦相似度, 聚合节点特征, 得到 特征嵌入 2.结构聚合: 使用单层GCN实现加权的领域聚合, 得到 结构嵌入 3.嵌入融合: 将得到的两种嵌入进行 融合 4.属性预测: 使用MLP将融合后的传递给多层感知机, 得到目标的预测
O	输出	完成缺失属性填充后的图数据
P	问题	1.只考虑拓扑结构, 忽略节点相似性 2.忽略高阶邻居
C	条件	需要存在属性完整的节点
D	难点	避免聚合高阶邻居过程中的过度平滑和梯度消失问题
L	水平	Knowledge-Based Systems2022 (SCI一区)

算法原理图

• 算法原理图

- 特征聚合：结合KNN与余弦相似度，聚合节点特征，得到特征嵌入
- 结构聚合：使用单层GCN实现加权的领域聚合，得到结构嵌入
- 嵌入融合：将得到的两种嵌入进行融合



三个模块相互协同!

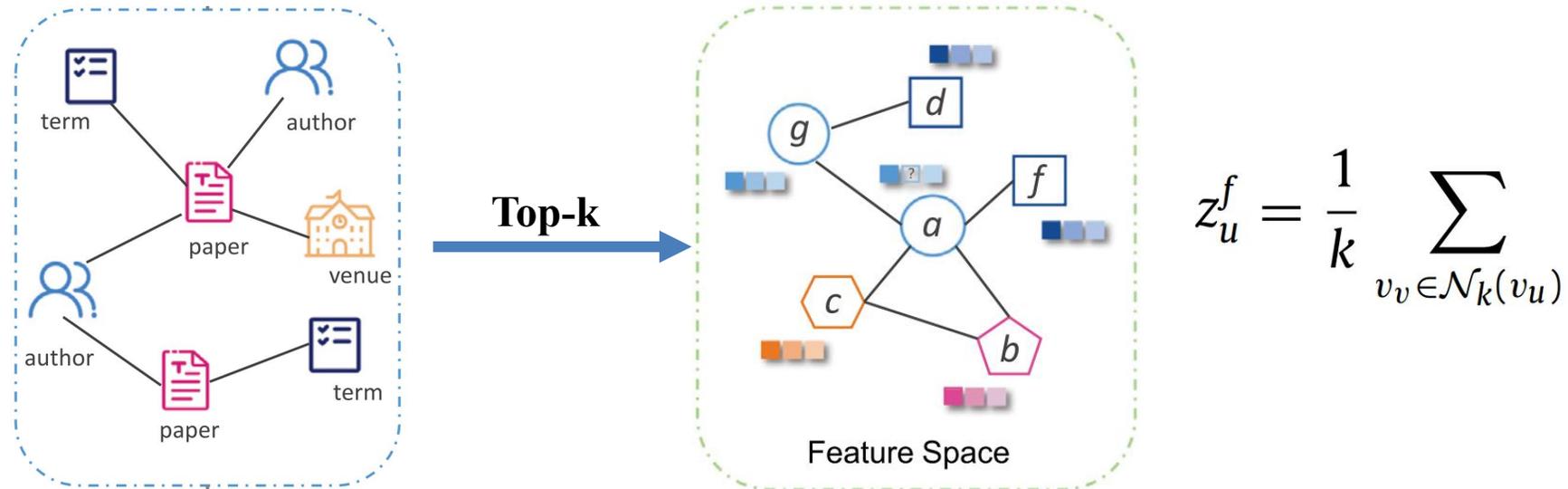
创新分析

- 现有方法存在问题
 - 利用网络拓扑结构进行节点属性补全，未探索特征空间中节点相似性的作用
 - 对于节点属性部分缺失，特征空间中的相似性可以有效地帮助填充属性
 - 但由于**相同类型的节点不一定有边**，GNN有时无法在异质图上实现基于相似性的属性填充
- AC-HEN的解决方法
 - 将节点进行**分类**：具有不完整属性的目标节点和具有完整属性的源节点
 - 目标节点又可分为缺失属性和存在属性
 - 使用KNN**选择源节点**来生成目标节点的特征嵌入
 - 根据目标节点的现有属性计算源节点与目标节点之间的**相似度得分**

解决方法

– 为每个目标节点选择Top-k个相似的源节点

- 使用余弦相似度计算相似性
- 需考虑因素：两个节点是否直连、节点间边权值不同等
- 使用k个参数来动态调整目标节点的前k个源节点的权重

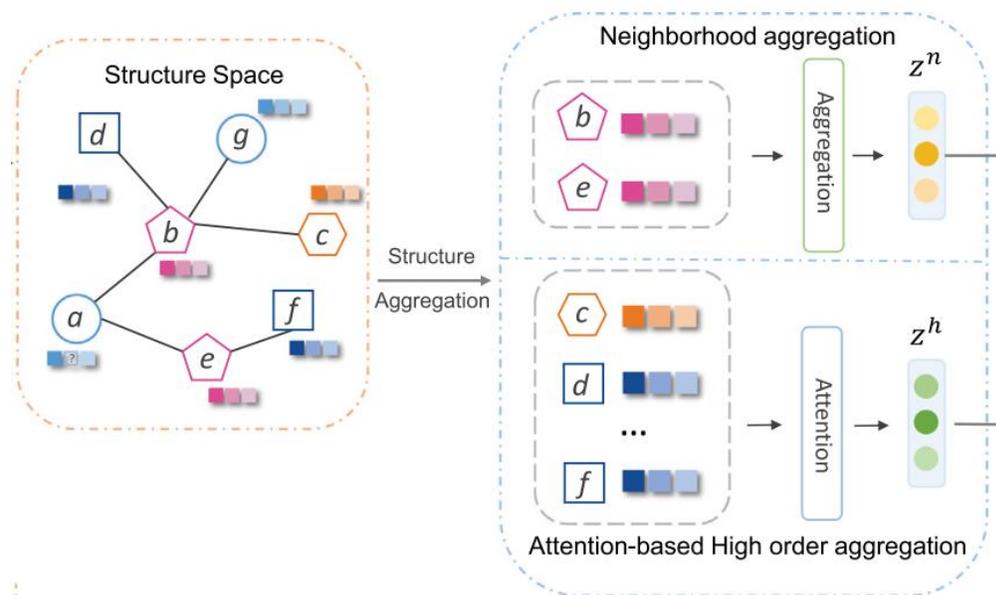
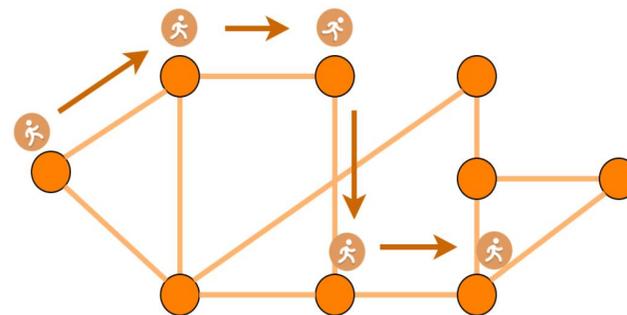


• 现有方法存在问题

- 只考虑**一阶邻居**或忽略不同阶邻居对属性补全的**重要度差异**
- 只采用**平均聚合**方法，未探索权重

• AC-HEN的解决方法

- 基于注意力的高阶邻居聚合，探索高阶邻居节点对目标节点属性补全的不同影响
 - 随机游走采样高阶邻居，并结合注意力机制进行加权聚合
 - 聚合高阶邻居，而非**多层堆叠GCN**，保留节点特征差异性



• 数据资源

– 数据集:

数据集	节点类型1	节点类型2	节点类型3	节点类型4
DBLP	论文*14328	作者*4057	期刊/会议*20	术语*8798
ACM	论文*4019	作者*7167	主题*60	-
IMDB	电影*780	演员*5841	导演*2269	-

Datasets	Nodes	Edges	Attributes
DBLP	#author(A): 4057 #paper(P): 14328 #term(T): 7723 #venue(V): 20	# A-P: 19645 # P-T: 85810 # P-V: 14328	A: missing
ACM	#paper(P): 4019 #author(A): 7167 #subject(S): 60	# P-P: 9615 # P-A: 13407 # P-S: 4019	A: missing
IMDB	#movie(M): 4278 #director(D): 2081 #actor(A): 5257	# M-D: 4278 # M-A: 12828	M: missing

• 评价指标

- Heat Kernel指数
- Pearson相关系数
- 节点分类准确率

• 对比方法

- EM (2002)
- MC (2009)
- SVR (2015)
- MLP (2017)
- HGAT (2019)
- AGCN (2020)
- AC-HGNN (2021)

传统方法

基于GNN

数据集设置

- 对目标节点（例如DBLP数据集中的作者），使用 **50% 的节点**进行训练
 - 当在DBLP数据集上训练模型时，选择20%的节点（例如，作者节点）并随机丢弃50%的属性；验证集和测试集采用相同的处理方式

其他设置

- 嵌入维数设置为256
- 多头注意力机制中注意力头的数量设置为 8
- 在特征聚合中，选择了 4 个最近的邻居节点
- 随机游走的步长为 15（每次游走最多会经过15个节点）

• 实验结果

- 基于余弦相似度来聚合特征空间中top-k个相似邻居的属性，使Pearson相关性和Heat Kernel得到很大的提高
- 基于GNN的方法可以有效地通过网络结构捕获相似的特征信息，以及传播信息

Results of attribute completion in terms of heat kernel and Pearson correlation coefficient on three datasets.

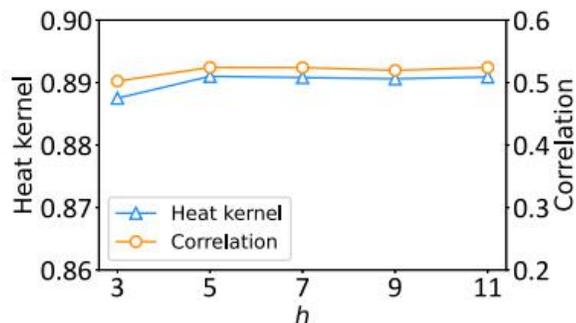
Method	ACM		DBLP		IMDB	
	Heat Kernel	Correlation	Heat Kernel	Correlation	Heat Kernel	Correlation
MC	0.6747 ± 0.0032	0.2369 ± 0.0051	0.3641 ± 0.0045	0.0924 ± 0.0002	0.5158 ± 0.0195	0.1168 ± 0.0079
EM	0.7534 ± 0.0112	0.2446 ± 0.0121	0.2611 ± 0.0102	0.1824 ± 0.0104	0.4037 ± 0.0112	0.1489 ± 0.0104
MLP	0.7739 ± 0.0091	0.3119 ± 0.0102	0.2987 ± 0.0131	0.1486 ± 0.0124	0.3724 ± 0.0095	0.1429 ± 0.0072
SVR	0.7376 ± 0.0074	0.3029 ± 0.0079	0.2819 ± 0.0201	0.1504 ± 0.0156	0.4372 ± 0.0212	-0.0121 ± 0.0011
HGAT	0.7931 ± 0.0101	0.3723 ± 0.0108	0.3821 ± 0.0141	0.3541 ± 0.0150	0.5815 ± 0.0991	0.2965 ± 0.0105
AGCN	0.7832 ± 0.0021	0.4433 ± 0.0024	0.3942 ± 0.0045	0.3872 ± 0.0051	0.5721 ± 0.0045	0.2954 ± 0.0051
HGNN-AC	0.8057 ± 0.0052	0.4632 ± 0.0061	0.4241 ± 0.0054	0.4098 ± 0.0048	0.5874 ± 0.0021	0.3219 ± 0.0027
AC-HEN	0.8903 ± 0.0028	0.5327 ± 0.0045	0.4556 ± 0.0031	0.4691 ± 0.0029	0.6171 ± 0.0018	0.3754 ± 0.0012

- 实验处理
 - 将填充后的图输入MAGNN得到节点嵌入，将嵌入信息输入SVM进行分类
- 下游任务实验结果
 - AC-HEN算法能有效为下游分类任务提供帮助

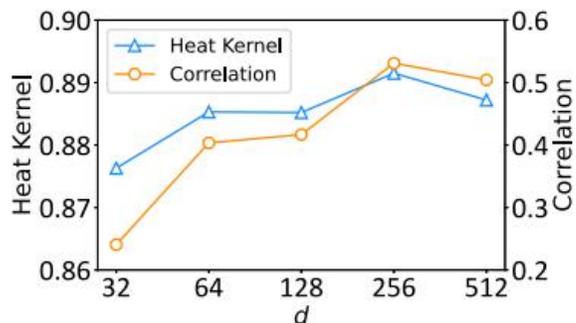
Results of node classification task in terms of Macro-F1 and Micro-F1 on three datasets.

Method	ACM		DBLP		IMDB	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
MC	89.37 ± 0.21	89.62 ± 0.22	91.08 ± 0.05	91.22 ± 0.08	57.49 ± 0.12	58.07 ± 0.21
EM	89.39 ± 0.10	89.68 ± 0.11	91.22 ± 0.21	91.97 ± 0.28	53.75 ± 0.06	54.58 ± 0.05
MLP	90.99 ± 0.54	91.03 ± 0.43	92.27 ± 0.12	92.82 ± 0.12	57.39 ± 0.61	57.94 ± 0.52
SVR	90.13 ± 0.10	90.25 ± 0.09	92.45 ± 0.31	93.03 ± 0.27	56.84 ± 0.59	57.47 ± 0.62
HGAT	91.19 ± 0.21	91.30 ± 0.17	92.52 ± 0.14	93.20 ± 0.11	57.96 ± 0.31	58.55 ± 0.37
AGCN	91.53 ± 0.17	91.62 ± 0.19	92.51 ± 0.09	93.17 ± 0.11	58.74 ± 0.45	58.86 ± 0.39
HGNN-AC	91.69 ± 0.11	91.75 ± 0.13	93.07 ± 0.25	93.68 ± 0.33	59.29 ± 0.18	59.74 ± 0.11
AC-HEN	91.94 ± 0.21	92.31 ± 0.17	94.39 ± 0.22	94.99 ± 0.21	60.28 ± 0.15	60.88 ± 0.17

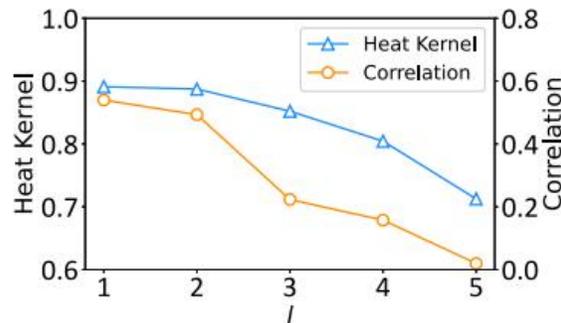
• 参数敏感性实验



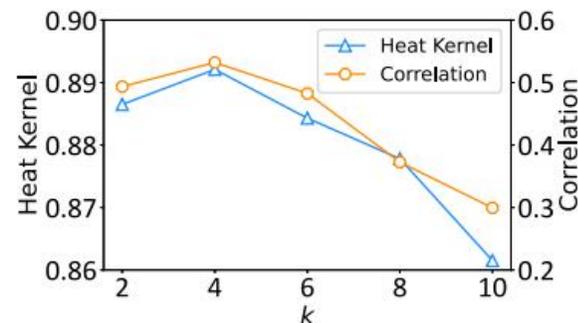
(a) #high-order neighbors



(b) Dimension d



(c) #layers l



(d) #nearest neighbors k

• 复杂度实验

Run time comparison of HGNN-AC with GNN baselines.

Method	ACM	DBLP	IMDB
HGAT	2295.66	3869.32	1032.01
AGCN	8269.55	9133.40	3232.65
HGNN-AC	4528.36	7018.00	2203.32
AC-HEN	3155.47	6176.97	1644.61

首评总结

- 算法贡献
 - 考虑特征空间中的**特征聚合**和网络结构空间中的一阶和**高阶邻居聚合**，提出了一种基于GNN的异构网络属性补全框架
 - 明确探索了高阶邻居节点对目标节点属性补全的不同影响
 - 减少了特征过度平滑和梯度消失的影响
- 算法不足
 - 节点的属性完全缺失时，无法计算该节点的邻居节点相似度，后续算法难以进行
 - 属性补全的性能提升有限
 - **复杂度**仍然较高





Attribute completion of heterogeneous graph neural networks based on residual attention mechanism

TIPO

T	目标	利用带属性节点补全 非属性节点 的属性
I	输入	属性完全缺失 的图数据 DBLP (14,328篇论文、4,057位作者、20个会议/期刊、8,789个术语) ACM (4019篇论文、7167位作者、60个主题) IMDB (780部电影, 5841名演员和2269名导演)
P	处理	1.将不同类型节点首先映射到 相同的特征空间 , 并计算节点嵌入 2.通过注意力层增加边缘残差连接, 计算节点之间的注意力分数 3.挑选最佳邻居进行加权聚合 4.主动删除部分带有属性节点的属性, 增加重构损失
O	输出	完成属性填充后的图数据
P	问题	1.现有方法在降低模型复杂度的同时难以提高模型精度 2.忽略了跨类型信息融合的重要性
C	条件	需要存在属性完整的节点
D	难点	如何同时兼顾模型复杂度和模型效果
L	水平	ESWA2024 (SCI一区)

算法原理图

算法流程

节点类型转换

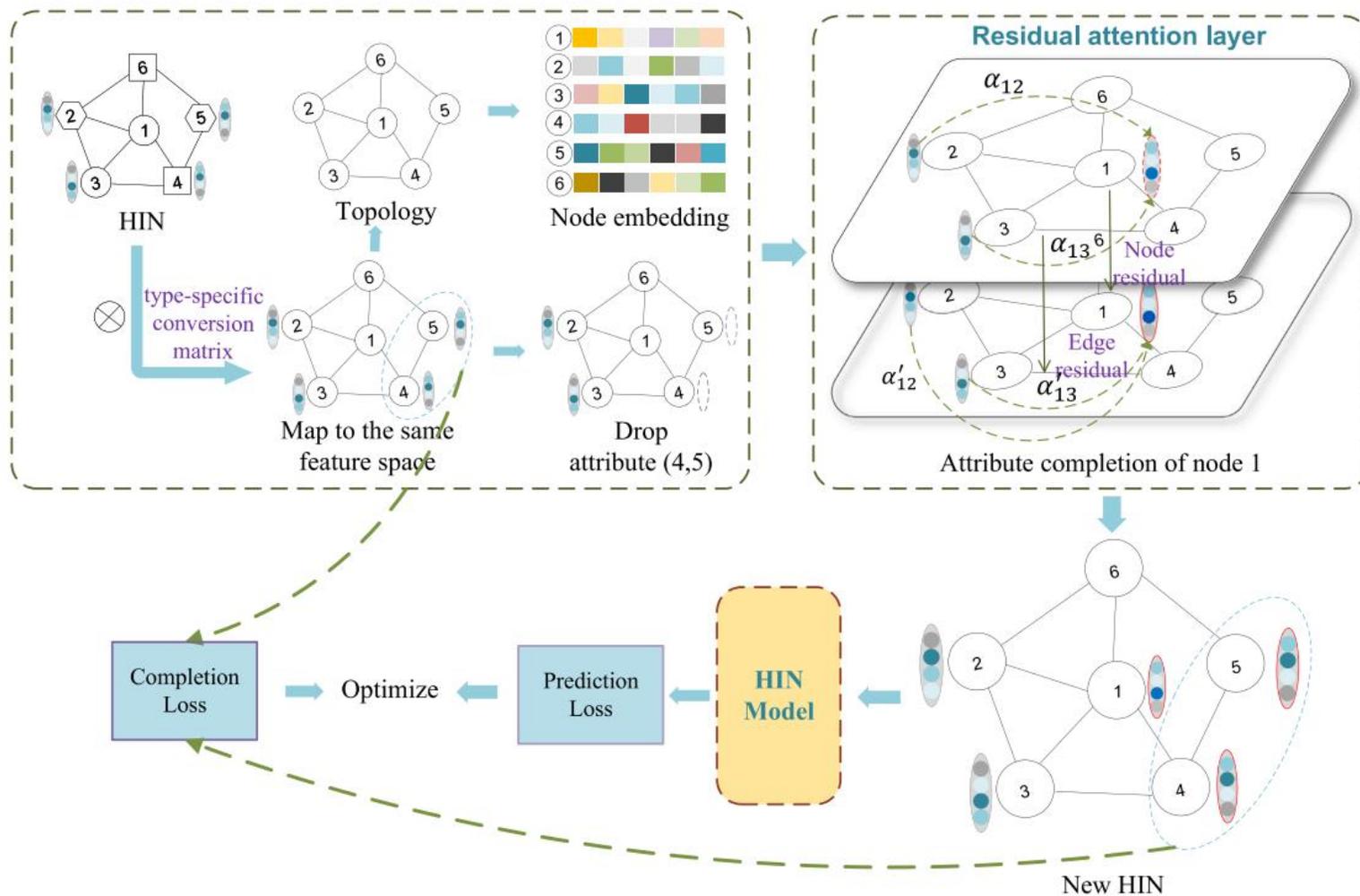
- 随机游走算法
- skip-gram模型

残差注意力机制

- 主动学习节点邻居的重要度排序
- 解决**过度平滑**和**梯度消失**

重构损失

- 评判模型效果
- 增强鲁棒性

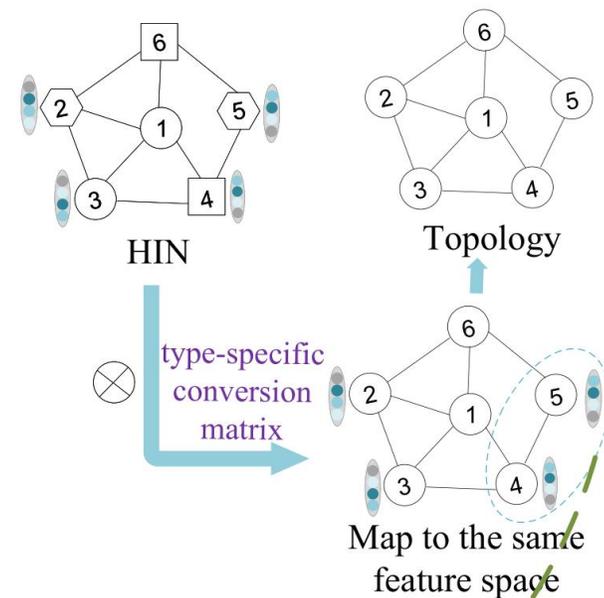


属性处理预训练通用框架

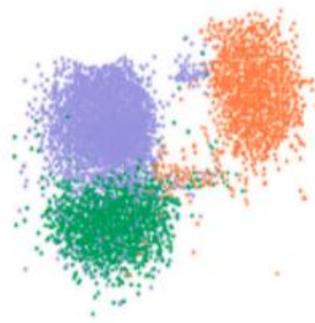
- 现有方法存在问题
 - 聚合来自相邻节点的信息时存在大量**不同类型的节点**
 - 如何整合不同的特征空间
- AC-HEN的解决方法
 - 设计类型特定的转换矩阵 M
 - 将各类节点的初始特征映射到统一特征空间

$$h'_i = M_i \cdot h_i$$

- M_i 针对不同节点类型单独设计，确保**异构特征对齐**



(a) Original graph



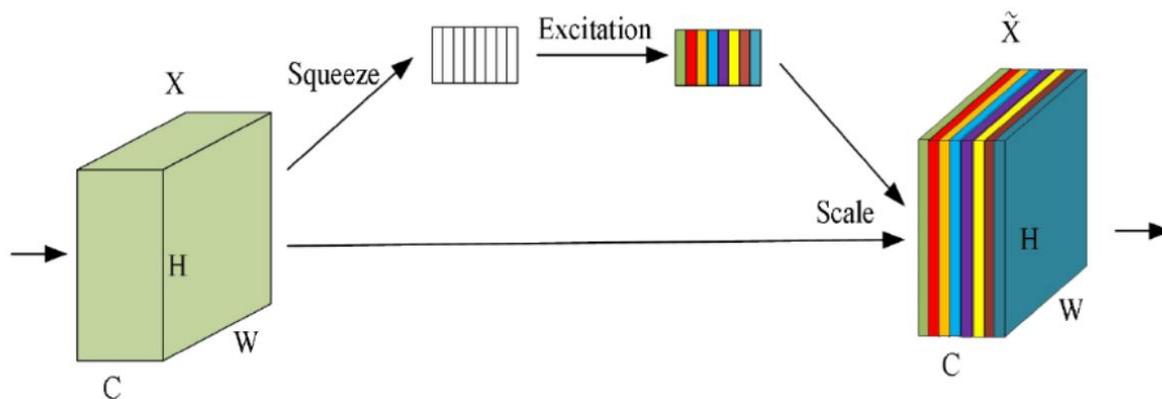
(b) MAGNN-AC



(c) RA-MAGNN

问题与动机

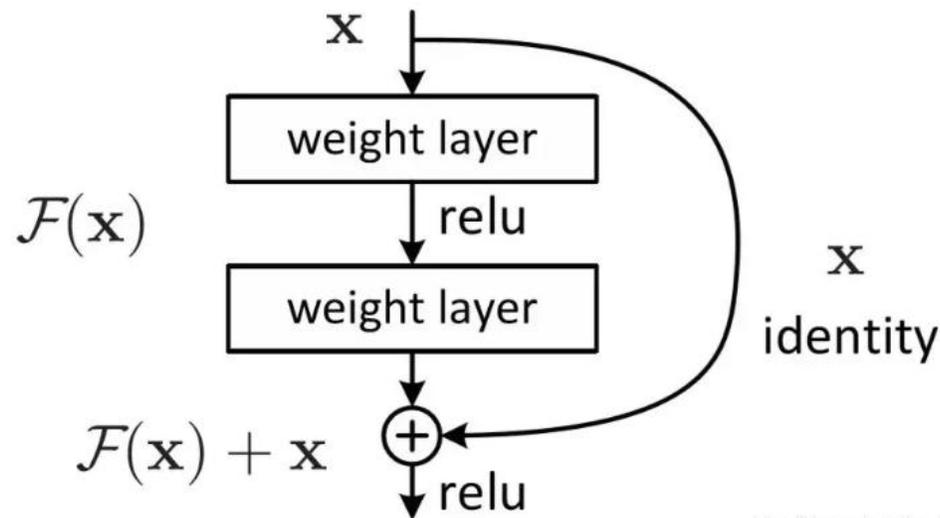
- 在多数GNN网络中存在过度平滑和梯度消失
 - 过度平滑：随着网络层数的增加，节点嵌入向量变得**越来越相似**
 - 梯度消失：当反向传播时，梯度逐层传递的过程中，逐渐变得非常小，甚至**接近零**
- RA-HGNN的解决方法
 - 使用残差注意力机制（残差学习+注意力机制）
 - 引入**残差连接**，帮助信息在网络中更容易流动，避免梯度消失和信息丢失
 - 在每一层或每个块的输入特征上应用注意力机制，生成加权的特征图



• 残差注意力机制

– 残差连接

- 输出表述为：**输入**和**输入的一个非线性变换**的线性叠加
- 帮助信息在网络中更容易流动，避免梯度消失和信息丢失



– RA-HGNN

- 节点残差：保留节点自身的原始信息，缓解深层网络训练中的梯度消失问题
- 边缘残差：保留历史层的注意力权重信息，增强对复杂关系的建模能力

$$h_i^{(m+1)} = \sigma \left[\sum_{j \in N_i^+} a_{ij}^{(m+1)} W^{(m+1)} (h'_j)^m + W_{res}^{(m+1)} (h'_i)^m \right]$$

$$h_i^{(m+1)} = \sigma \left(\left\|_{k=1}^K \hat{h}_{ik}^{(m+1)} + W_{res(k)}^{(m+1)} (h'_i)^m \right\| \right)$$

• 数据资源

– 数据集:

数据集	节点类型1	节点类型2	节点类型3	节点类型4
DBLP	论文*14328	作者*4057	期刊/会议*20	术语*8798
ACM	论文*4019	作者*7167	主题*60	-
IMDB	电影*780	演员*5841	导演*2269	-

Datasets.

Dataset	Node	Edge	Attribute
ACM ^a	paper(P):4,019 author(A):7,167 subject(S):60	paper-paper:9,615 paper-author:13,407 paper-subject:4,019	paper:raw author:handcrafted subject:handcrafted
DBLP ^b	author(A):4,057 paper(P):14,328 term(T):7,723 venue(V):20	author-paper:19,645 paper-term:85,810 paper-venue:14,328	author:handcrafted paper:raw term:handcrafted venue:handcrafted
IMDB ^c	movie(M):4,278 director(D):2,081 actor(A):5,257	movie-director:4,278 movie-actor:12,828	movie:raw director:handcrafted actor:handcrafted

• 评价指标

- 节点分类准确率
- 节点聚类准确率

• 对比方法

- GCN (2016)
- GCN (2017)
- GAT (2018)
- HAN (2019)
- MAGNN (2020)
- AC-HEN (2022)
- HetReGAT-FC (2023)

对比实验

- 实验设置
 - 使用K-means对训练后得到的节点进行聚类
- 实验结果
 - RA-HGNN算法的属性补全性能优越
 - 评价指标: NMI、ARI

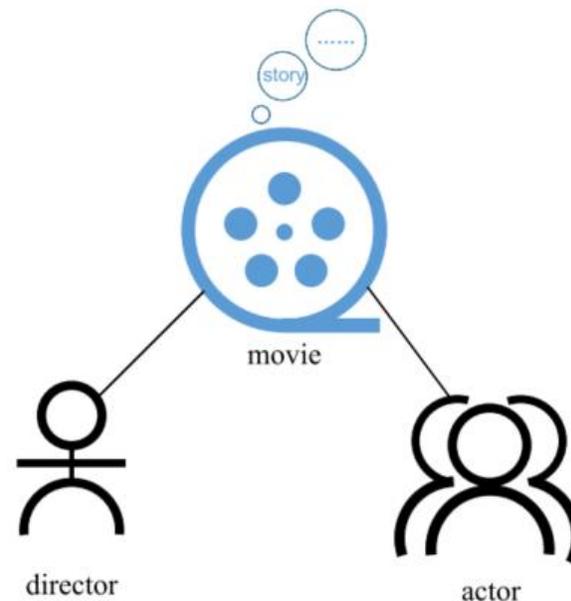


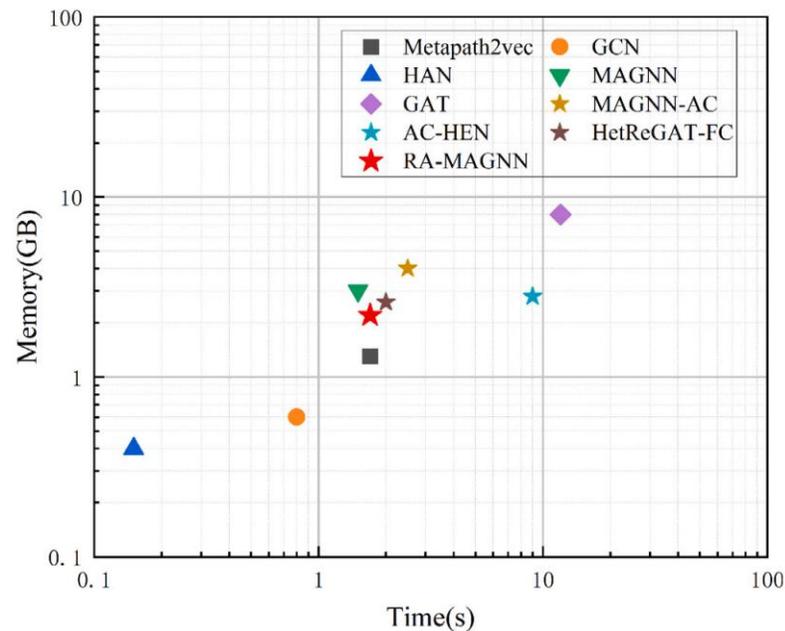
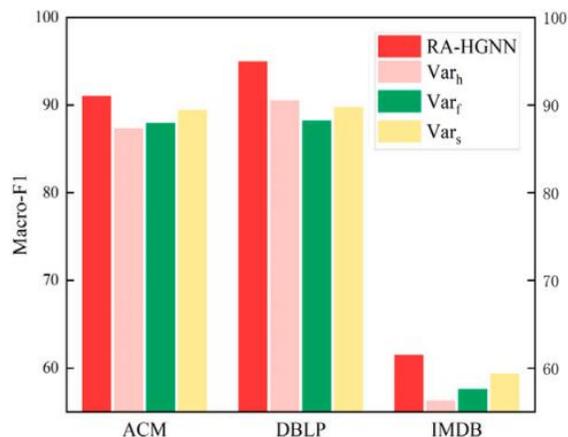
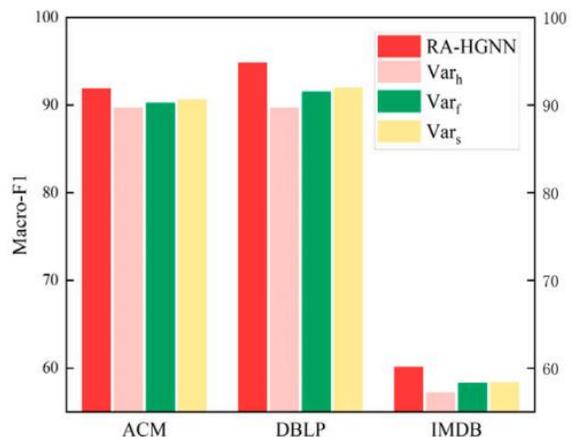
Table 4
Quantitative results (%) on the node clustering task.

Datasets	Metrics	Metapath2vec	GCN	HAN	GAT	MAGNN	MAGNN-AC	AC-HEN	HetReGAT-FC	RA-MAGNN
ACM	NMI	21.22	51.40	61.56	57.29	62.51	64.93	64.82	64.97	65.21
	ARI	21.00	53.01	64.39	60.43	67.23	68.52	68.84	69.05	69.88
IMDB	NMI	1.20	5.45	10.87	8.45	12.87	14.29	14.22	14.07	14.31
	ARI	1.70	4.40	10.01	7.46	11.98	14.09	14.01	14.10	14.12
DBLP	NMI	74.30	75.01	79.12	71.50	77.64	79.49	79.51	79.35	79.59
	ARI	78.50	80.49	84.76	77.26	82.60	84.83	84.51	84.73	84.92

实验设置

- Var_h : 没有对初始输入数据进行特征空间进行统一的维度映射
- Var_r : 没有使用残差连接, 而是替换为传统的HGNN
- Var_s : 没有使用多个注意力头的互补机制, 使用传统的加权聚合

实验结果



首次总结

- 算法贡献

- 将异构图中的节点映射到相同的特征空间中进行节点拓扑嵌入
- 使用**残差注意力机制**减小过度平滑和梯度消失
- 作为通用的端到端框架，可以与其他网络结合
- 复杂度能在资源占用和时间消耗上取得**均衡结果**

- 算法不足

- 特征统一转换矩阵需人工设计，无法自学习，人工耗时长
- 在属性缺失度较高的数据集上补全准确率较低



北京林业大学
景观规划设计学院



特点总结与未来展望

• 特点总结

算法	AC-HEN	RA-HGNN
优势	1. 探索高阶邻居节点对目标节点属性补全的不同影响 2. 考虑特征空间中的特征聚合和网络结构空间中的一阶和高阶邻居	1. 使用 残差注意力机制 减小过度平滑和梯度消失 2. 通用的端到端框架，可以与其他网络结合
劣势	1. 节点属性不能完全缺失，否则无法计算相似度 2. 复杂度较高	特征统一转换矩阵需人工设计，无法自学习

• 未来展望

- 引入**进化算法**来优化初始图结构
- 探究更加复杂的异质图、超图等图结构的属性补全
- 从静态异质图演化至**动态异质图**（引入时间维度，图结构和属性动态变化）

- [1] Zhao Z, Liu Z, Wang Y, et al. RA-HGNN: Attribute completion of heterogeneous graph neural networks based on residual attention mechanism[J]. *Expert Systems with Applications*, 2024, 243: 122945.
- [2] Wang K, Yu Y, Huang C, et al. Heterogeneous graph neural network for attribute completion[J]. *Knowledge-Based Systems*, 2022, 251: 109171..

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！



Table 3
Results (%) on ACM, IMDB and DBLP datasets on the node classification task.

Datasets	Metrics	Training	Without attribute completion					With attribute completion			
			Metapath2vec	GCN	HAN	GAT	MAGNN	MAGNN-AC	AC-HEN	HetReGAT-FC	RA-MAGNN
ACM	Macro-F1	1%	35.23	66.83	86.95	88.58	85.58	84.27	85.11	85.37	88.93
		5%	42.37	72.45	88.65	89.20	86.12	85.47	86.72	86.51	89.25
		10%	44.29	70.79	89.39	89.29	86.60	86.23	86.95	87.69	89.51
		20%	69.95	70.41	89.40	89.59	88.01	88.44	88.15	88.25	89.55
		40%	71.15	70.82	89.79	89.77	89.42	89.93	89.33	89.51	90.69
		60%	71.47	69.67	89.51	89.72	90.39	90.67	90.66	90.42	91.52
		80%	72.18	67.23	90.63	89.42	90.79	91.08	90.82	90.93	91.96
	Micro-F1	1%	63.18	71.74	87.37	88.60	85.95	84.82	85.05	85.33	88.96
		5%	68.88	74.95	88.62	89.10	86.24	85.52	86.73	87.21	89.26
		10%	70.29	71.40	89.32	89.19	86.67	86.19	87.52	87.39	89.34
		20%	72.12	74.02	89.22	89.47	88.08	88.42	88.25	88.26	89.41
		40%	73.17	74.57	89.64	89.65	89.48	89.95	89.51	89.63	90.70
		60%	73.65	74.10	89.33	89.60	90.42	90.66	90.46	90.77	91.55
		80%	74.14	72.69	90.54	89.29	90.80	91.05	90.83	90.92	91.98
IMDB	Macro-F1	1%	35.23	39.72	52.49	49.52	50.78	49.75	49.88	50.57	52.63
		5%	42.37	42.95	56.16	53.08	54.28	53.60	54.49	54.41	56.29
		10%	44.29	43.70	57.02	53.61	56.39	55.25	55.35	56.81	57.23
		20%	46.42	44.75	50.00	54.81	58.11	58.17	58.33	58.19	58.26
		40%	47.70	45.26	52.71	55.09	59.39	59.26	59.41	59.32	59.46
		60%	48.25	46.72	54.24	55.71	59.97	59.45	59.82	59.91	59.95
		80%	48.73	47.13	54.38	55.40	60.02	60.08	60.08	60.11	60.19
	Micro-F1	1%	39.55	44.01	54.38	51.32	51.62	50.47	51.36	51.87	54.69
		5%	44.33	46.41	56.74	53.73	54.46	53.77	54.22	55.35	56.90
		10%	46.15	47.02	57.35	54.14	56.53	55.48	55.67	55.41	57.42
		20%	48.08	47.44	55.73	55.02	58.16	57.27	58.14	58.17	58.27
		40%	49.55	47.62	57.97	55.29	59.46	59.18	59.33	59.31	59.49
		60%	50.06	48.49	58.32	55.91	60.05	59.58	59.57	60.08	59.98
		80%	50.68	48.73	58.51	55.67	60.15	60.13	60.11	60.05	60.24
DBLP	Macro-F1	1%	88.76	86.99	89.37	32.68	92.45	92.69	92.16	92.38	92.73
		5%	90.49	89.03	90.83	57.20	92.44	93.10	93.29	93.36	93.70
		10%	91.09	89.53	91.24	64.57	92.44	93.18	92.98	93.34	93.97
		20%	91.50	90.06	92.24	66.92	93.13	93.21	93.26	93.71	94.20
		40%	92.55	90.37	92.40	73.23	93.23	93.35	93.55	93.83	94.35
		60%	93.25	90.57	92.80	77.17	93.57	93.38	93.75	94.02	94.38
		80%	93.48	90.74	93.08	78.20	94.10	94.02	94.22	94.19	94.63
	Micro-F1	1%	89.91	87.55	90.12	48.74	93.11	93.25	92.54	93.21	93.29
		5%	91.19	89.58	91.49	70.79	93.02	93.51	93.22	93.01	94.14
		10%	91.74	90.02	91.88	75.90	93.02	93.55	92.37	94.02	94.39
		20%	92.14	90.53	93.11	76.98	93.61	93.60	94.14	94.33	94.60
		40%	93.09	90.83	93.30	79.61	93.68	93.37	93.84	94.01	94.73
		60%	93.76	91.01	93.70	81.62	93.99	93.75	94.06	94.53	94.75
		80%	93.94	91.15	93.99	82.22	94.47	94.17	94.68	94.52	94.98