面向隐私保护的图节点遗忘学习方法研究

姓名：许语轩

学号：B21060202

南京邮电大学
Nanjing University of Posts and Telecommunications

目录

CONTENTS

南京郵電大學 Nanjing University of Posts and Telecommunications

PART ONE

南京郵電大學 Nanjing University of Posts and Telecommunications

01

研究背景

Research Background

研究背景

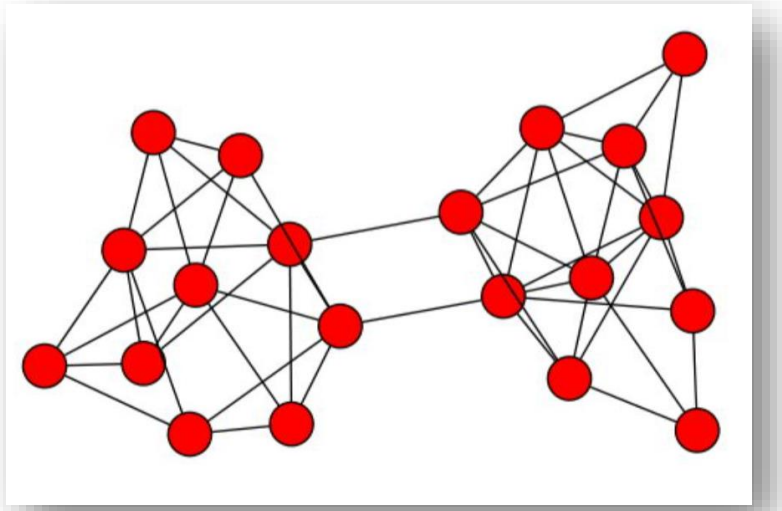
南京邮电大学 Nanjing University of Posts and Telecommunications

“被遗忘权”的提出

大数据时代，用户隐私保护需求增长，相关法律提出了“被遗忘权”，它赋予数据主体从存储数据的实体中删除数据的权利。

图神经网络的广泛应用

图神经网络（GNN）在多领域广泛应用，但图数据含大量隐私信息。在此背景下，如何在不影响 GNN 模型性能的同时，实现图节点遗忘学习以保护隐私，成为亟待解决的重要问题。



意义：通过实现图节点遗忘学习方法，不仅能够增强图神经网络的隐私保护能力，还能为数据处理提供更加灵活和高效的解决方案

研究内容——图遗忘学习

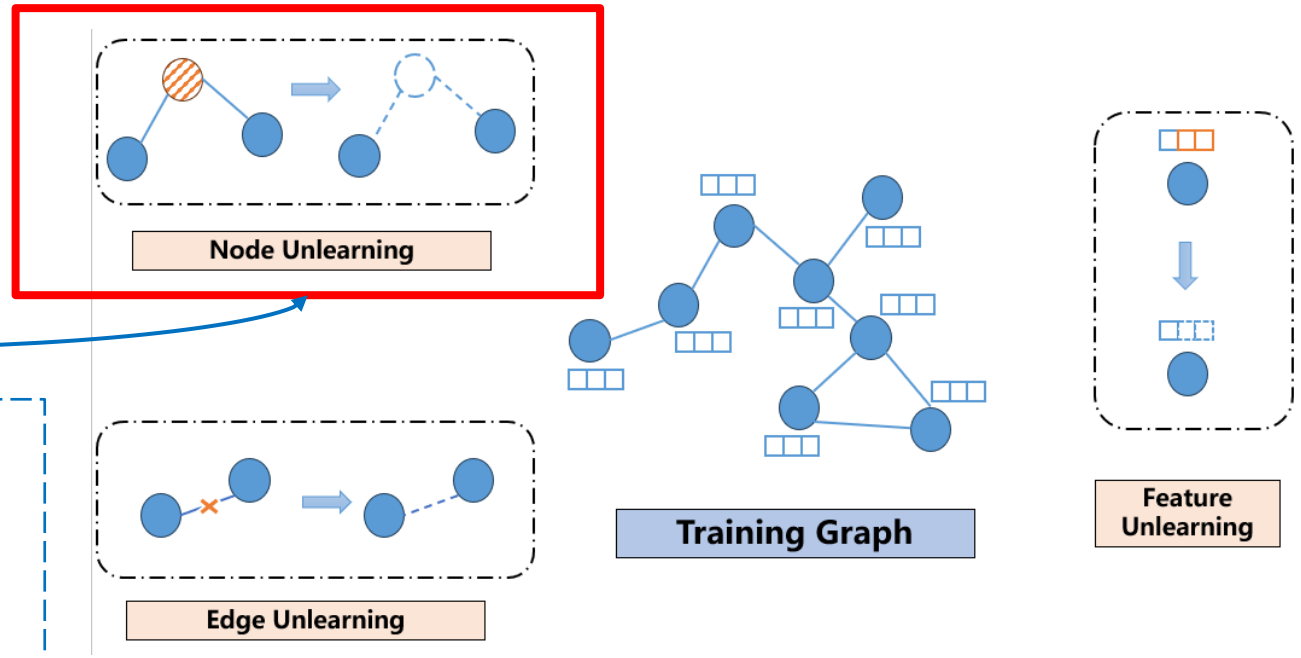
遗忘学习 从模型中删除已学习信息的过程，要求模型能够在不完全重新训练的情况下，删除或遗忘特定数据点的影响。

图遗忘学习 图遗忘学习是针对图神经网络（GNN）的遗忘学习。由于 GNN 在处理图结构数据时，节点和边之间存在丰富的交互信息，使得图遗忘学习不仅需要从模型中删除节点的特征信息，还需要考虑该节点与其他节点之间的连接关系及其对图结构的影响。

遗忘场景

- 节点遗忘（Node Unlearning）
- 边遗忘（Edge Unlearning）
- 特征遗忘（Feature Unlearning）

其中，图节点遗忘指从图中删除一个或多个节点及其相关的特征信息，这不仅意味着从数据中删除该节点的特征数据，还涉及到该节点在图中的位置、作用以及与其他节点的关系的丧失。



研究现状

南京邮电大学 Nanjing University of Posts and Telecommunications

图遗忘算法的优缺点比较

算法	模型	优点	缺点
基于划分	GraphEraser GUIDE GraphRevoker	依靠分区实现一定程度的图遗忘	效果很大程度依赖图划分质量、后续处理策略，如分区质量、聚合器选择等对最终性能影响大
基于影响函数	GIF CGU CEU	借助严谨数学公式，能快速、准确计算删除特定数据后模型参数变化	参数调整方式简单，处理复杂参数变化情况能力不足，易导致模型丢失关键特征
基于学习	GNNDelete MEGU GCU	努力在数据遗忘和模型推理能力间找平衡，设计特定损失函数优化训练进程	只适用于简单模型，模型层数增加时运行效率大幅降低，不能契合模型持续训练时的遗忘需求
基于投影	Projecter	通过对权重进行投影操作，能在一定程度上完成图遗忘任务	目前研究较少，投影策略效果、适用性以及在复杂图中的表现需更多研究验证
基于结构	UtU	直接对图的结构进行修改，无需重新大规模训练模型，无需实施复杂优化过程	容易让图原本的含义与特征大幅改变，需在达成图遗忘效果的同时最大程度留存图既有结构信息

PART TWO

南京郵電大學 Nanjing University of Posts and Telecommunications

02

研究方法

Research Methods

基于结构熵的图节点遗忘学习方法（SEGU）

南京邮电大学 Nanjing University of Posts and Telecommunications

问题形式化

聚焦于半监督节点分类任务，考虑图G=(V,E,X)，其中包含n个节点、m条边和特征矩阵X等要素。接收图节点遗忘学习请求后，输出非遗忘实体的预测结果且受遗忘实体的影响最小。

SEGU (Structural Entropy-based Graph Unlearning)
一种图遗忘学习互进化范式，旨在同时发展图遗忘学习的预测和遗忘能力，确保在统一训练框架中进行互补优化，以满足图数据场景下的隐私保护和模型性能需求。

模型架构

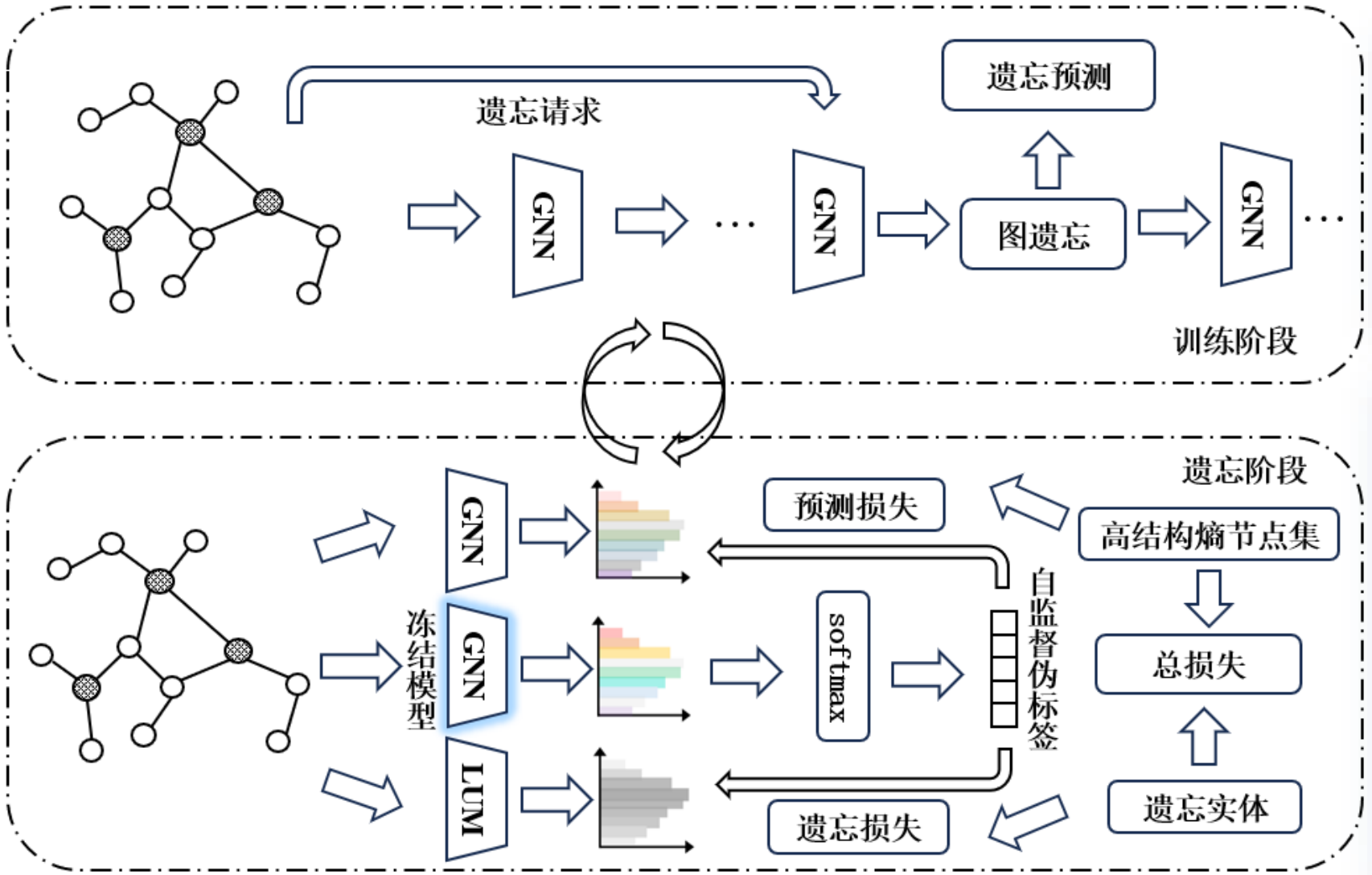
- **双模块设计**：SEGU包含**预测模块和遗忘模块**。预测模块负责保持对非遗忘实体的预测性能，而遗忘模块则负责消除遗忘实体的影响。
- **自适应高影响力邻居选择**：通过计算节点的结构熵，挑选出和目标节点紧密关联且有结构多样性的邻居节点，优化预测模块的损失函数。
- **拓扑感知的遗忘传播**：利用图拓扑结构和预测模块的自监督信息，实现了预测模块和遗忘模块的相互优化，提升了最终预测的准确性。

优化目标

- **预测模块**：基于原始训练模型，通过焦点损失（Focal Loss）保持对非遗忘实体的预测性能，同时通过KL散度损失（KL Loss）消除遗忘实体的影响。
- **遗忘模块**：通过反向焦点损失（Reverse Focal Loss）增强对遗忘实体的遗忘能力，同时利用KL损失和预测模块的输出，确保对非遗忘实体的预测性能。

最终，SEGU 通过公式 $L = Lp + \kappa Lu$实现基于互进化的遗忘学习优化

模型框架



训练阶段

- 图数据输入 GNN 处理
- 接收遗忘请求，进入遗忘阶段

遗忘阶段

- **预测模块**：原始模型初始化，依据冻结模型推理，交互消除影响，聚焦高结构熵节点
- **线性遗忘模块（LUM）**：强化遗忘能力、借助线性映射操作并结合自监督伪标签生成针对遗忘实体的预测结果

基于结构熵的邻居选择

算法步骤

- **步骤 1：计算节点的结构熵**

  对于图中的每个节点 v_i，计算其结构熵 $H(v_i)$。

$$H(v_i) = -\sum \frac{g_{v_i}}{2m} \log\left(\frac{vol_{v_i}}{vol_{p(v_i)}}\right)$$

- **步骤2：确定结构熵阈值**

  计算所有节点结构熵的平均值，将其作为初步的结构熵阈值。

$$\bar{H} = \frac{1}{|\mathcal{V}|} \sum_{v_i \in V} H(v_i)$$

- **步骤3：筛选高结构熵节点**

  选择结构熵$H(v_i) > \bar{H}$的节点，将这些节点组成候选邻居节点集合C。

- **步骤4：结合k-跳邻居结构筛选**

  设目标节点集合为T，对于k-跳邻居子图$\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$，从候选邻居节点集合$C$中筛选出满足以下条件的节点：①节点在$k$-跳邻居子图 $\mathcal{V}_k$中；②节点不属于目标节点集合T。将满足条件的节点组成最终的邻居节点集合 $\mathcal{N}_{final}$。

拓扑感知的遗忘传播

- 基于预测模块和非遗忘实体提出拓扑感知的遗忘传播策略。
- 考虑预测模块的拓扑结构和自监督信息 L，有效融合预测和遗忘模块。
- 遵循同质性假设：图中相连节点具有相似标签，通过标签传播促使标签分布平滑。

$$\mathbf{Y}(\hat{\mathbf{Y}}, \mathbf{E}(\mathbf{L})) := \mathbf{Y}_u = \hat{\mathbf{Y}}_u, \mathbf{Y}_v = G(\hat{\mathbf{Y}}_v + G(\mathbf{E}_v))$$

——预测计算

$$\mathbf{E}(\mathbf{L}) := \mathbf{E}_u^{(0)} = \vec{0}, \mathbf{E}_v^{(0)} = \mathbf{L} - \hat{\mathbf{Y}}_v, \forall u \in \mathcal{V}_L, \forall v \in \mathcal{V}_U$$

——误差校正矩阵初始化

$$G(\mathbf{T}) := \mathbf{T}_i^{(l)} = \alpha \mathbf{T}_i^{(0)} + (1-\alpha) \sum_{j \in \mathcal{N}_i^{(1)}} \frac{1}{\sqrt{\tilde{d}_i \tilde{d}_j}} \mathbf{T}_j^{(l-1)}$$

——标签传播函数

$$\mathbf{Y}^* := \mathbf{Y}^*(\hat{\mathbf{Y}}^*, \mathbf{E}(\hat{\mathbf{Y}}))$$

——为非遗忘实体生成最终预测

$$\hat{\mathbf{P}} = \text{Encoder}(\mathbf{A}^*, \mathbf{X}^*, \mathbf{W}^*)$$

$$\hat{\mathbf{P}}^* = \mathbf{W}_u \hat{\mathbf{P}}$$

——预测模块推理

$$\hat{\mathbf{Y}} = \text{Softmax}(\hat{\mathbf{P}})$$

$$\hat{\mathbf{Y}}^* = \text{Softmax}(\hat{\mathbf{P}}^*)$$

——转换为概率分布

损失函数与优化目标

南京邮电大学 Nanjing University of Posts and Telecommunications

预测模块损失

$$\mathcal{L}_p = \sum_{u \in \text{HSEN}} \mathcal{L}_{\text{Focal}}(\widehat{\mathbf{Y}}_u, \widetilde{\mathbf{Y}}_u) + \sum_{v \in \text{HSEN}} \mathcal{L}_{\text{KL}}(\widehat{\mathbf{Y}}^*_v, \widehat{\mathbf{Y}}_v)$$

通过基于冻结模型输出的焦点损失保持推理能力，结合遗忘模块输出的 KL 散度消除遗忘影响，将优化范围限定于高结构熵节点（HSEN）

遗忘模块损失

$$\mathcal{L}_u = -\sum_{u \in \text{遗忘实体}} \mathcal{L}_{\text{Focal}}(\widehat{\mathbf{Y}}^*_u, \widetilde{\mathbf{Y}}_u) + \sum_{v \in \text{遗忘实体}} \mathcal{L}_{\text{KL}}(\widehat{\mathbf{Y}}_v, \widehat{\mathbf{Y}}^*_v)$$

将反向焦点损失作用于遗忘实体来增强遗忘能力，结合预测模块输出的KL散度保障预测性能

整体优化目标

$$\mathcal{L} = \mathcal{L}_p + \kappa \mathcal{L}_u$$

通过调节参数κ达成预测与遗忘模块协同优化

PART THREE

南京郵電大學 Nanjing University of Posts and Telecommunications

03 研究成果

Research results

实验设置

南京郵電大學 Nanjing University of Posts and Telecommunications

01

数据集

Cora
CiteSeer
PubMed
Amazon Photo
Amazon Computers
Coauthor CS
Coauthor Physics

02

GNN模型

GCN
GAT
GIN
SGC
SAGE

03

基线模型

Retrain
Eraser-LPA
Eraser-KMeans
GIF
GNNDelete

04

评价指标

F1分数
遗忘时间

05

实验环境

在 Windows 10 系统下进行，CPU 为 AMD Ryzen 7 5700U with Radeon Graphics 。Python 版本为 3.7.9，Pytorch 版本为 1.13.1 。

性能比较——模型效用

不同GNN模型在不同数据集上，本文方法SEGU和使用重新训练Retrain、Eraser-LPA、Eraser-KMeans、GIF方法时的 F1 分数对比

	数据集	Retrain	Eraser-LPA	Eraser-KMeans	GIF	SEGU
GCN	Cora	0.847±0.003	0.676±0.004	0.493±0.006	0.822±0.007	**0.854±0.005**
	CiteSeer	0.744±0.004	0.450±0.006	0.332±0.006	0.693±0.006	**0.761±0.002**
	PubMed	**0.882±0.005**	0.718±0.010	0.482±0.003	0.854±0.006	0.862±0.003
	Photo	0.918±0.003	0.452±0.000	0.544±0.000	0.898±0.003	**0.919±0.004**
	Computers	0.853±0.001	0.382±0.000	0.404±0.000	0.832±0.003	**0.854±0.001**
	CS	**0.928±0.000**	0.750±0.023	0.812±0.012	0.914±0.002	0.911±0.001
	Physics	**0.961±0.000**	0.858±0.008	0.815±0.001	0.936±0.001	0.957±0.000
GAT	Cora	0.863±0.005	0.727±0.009	0.754±0.009	0.865±0.007	**0.874±0.001**
	CiteSeer	0.764±0.004	0.676±0.004	0.746±0.006	**0.766±0.007**	0.760±0.006
	PubMed	**0.863±0.001**	0.858±0.003	0.860±0.003	0.845±0.001	0.851±0.001
	Photo	0.896±0.001	0.816±0.001	0.807±0.000	0.883±0.003	**0.897±0.000**
	Computers	0.791±0.002	0.748±0.001	0.763±0.002	**0.806±0.003**	0.789±0.004
	CS	**0.918±0.003**	0.858±0.004	0.906±0.002	0.902±0.003	0.909±0.001
	Physics	**0.955±0.000**	0.921±0.004	0.925±0.001	0.922±0.001	0.954±0.001

实验结论

在多个数据集上，SEGU方法的F1 分数接近retrain甚至更高，这表明 SEGU 方法在非遗忘实体预测性能方面表现效果较好。

性能比较——遗忘效率

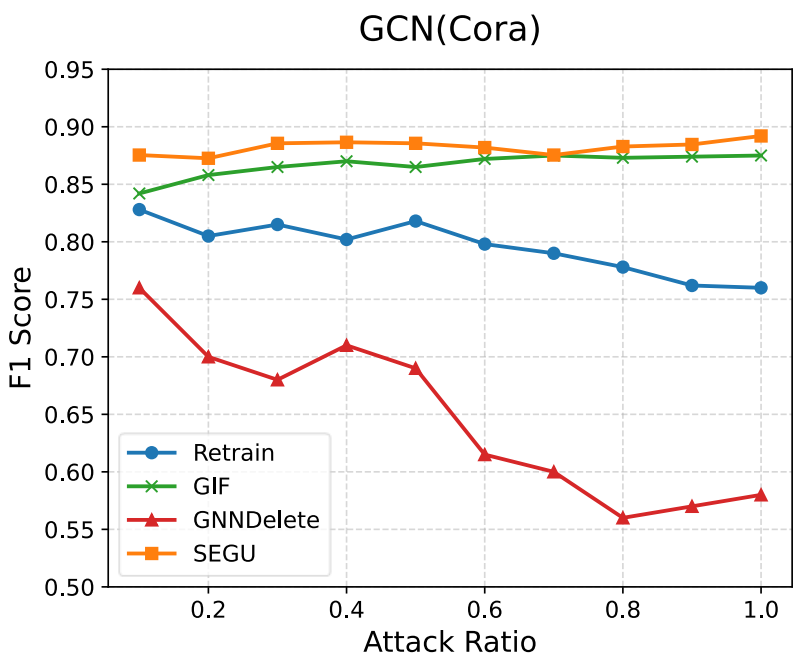
	GCN			GAT		
	Retrain	GIF	SEGU	Retrain	GIF	SEGU
Cora	3.6391	1.4882	0.4380	11.4470	4.8666	2.4534
CiteSeer	8.1513	2.8197	0.9111	32.5483	6.2719	6.6049
PubMed	40.1830	8.1757	2.2368	115.7410	34.9201	10.5557
Photo	23.5828	12.9564	3.2543	57.7359	60.3251	12.0476
Computers	58.3952	27.6307	6.5609	150.7310	131.0290	24.3414
CS	252.1700	29.9849	9.4565	1037.0300	81.2348	61.6028
Physics	847.5050	401.6030	20.1007	3827.5200	202.5920	143.3050

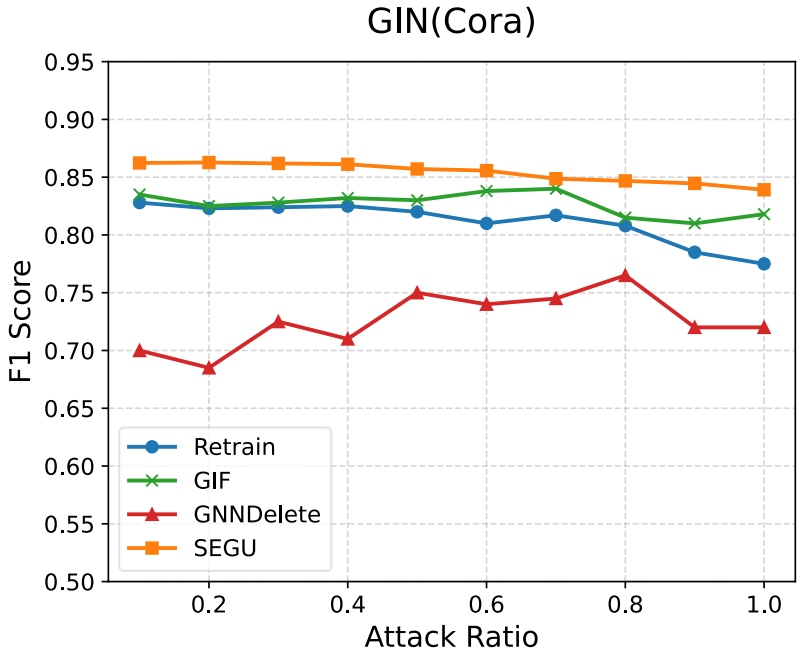
不同GNN模型在不同数据集上，本文方法 SEGU和使用重新训练Retrain 、 Eraser-LPA 、 Eraser-KMeans、 GIF方法时的遗忘时间对比

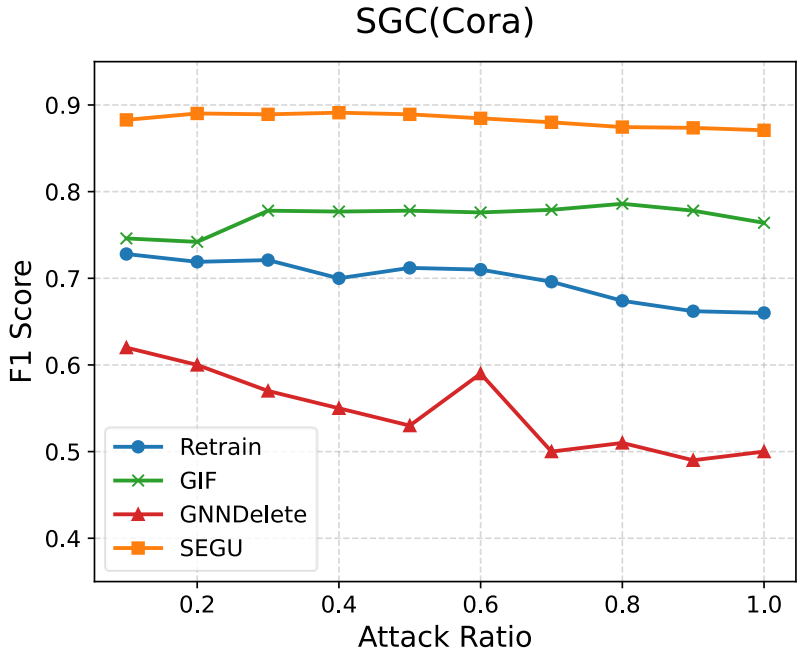
实验结论

SEGU方法在时间消耗上显著低于Retrain和GIF方法。这说明 SEGU更高效，能有效减少计算资源的消耗和训练时长。

遗忘效果







实验方法

为了验证遗忘效果，本文采用了边攻击，随机选择两个具有不同标签的节点作为添加噪声边的目标，这些边被视为遗忘实体。

实验结论

SEGU方法在应对噪声边所引发的负面效应方面呈现出优势，遗忘能力始终优于其他基线。

消融实验

南京邮电大学 Nanjing University of Posts and Telecommunications

实验目的

探索高结构熵节点选择（HSENS）及拓扑感知的遗忘传播（Topo. UP）在 SEGU 模型中的作用机制，明确对模型性能的影响

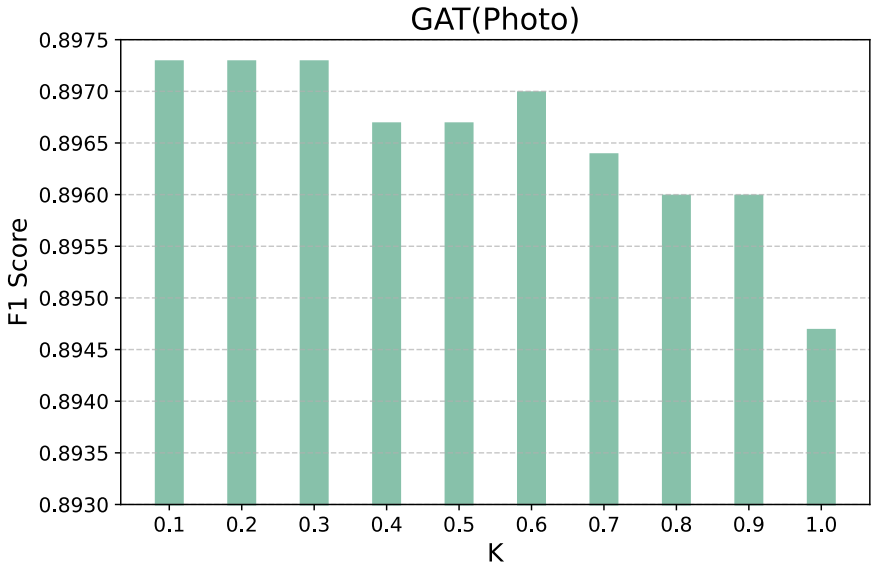
三种原型及变体的F1分数

模型	组件	Cora	CiteSeer
GCN	w/o HSENS	0.8530±0.0048	0.7597±0.0030
	w/o Topo. UP	0.8533±0.0027	0.7537±0.0015
	SEGU	0.8542±0.0036	0.7605±0.0037
GAT	w/o HSENS	0.8699±0.0009	0.7597±0.0012
	w/o Topo. UP	0.8634±0.0055	0.7712±0.0025
	SEGU	0.8736±0.0009	0.7597±0.0015

实验结论

HSENS 和 Topo. UP 在提升 SEGU 模型性能方面均起关键作用，从不同路径优化模型，使其在图节点遗忘学习任务中表现更优。

参数敏感性分析

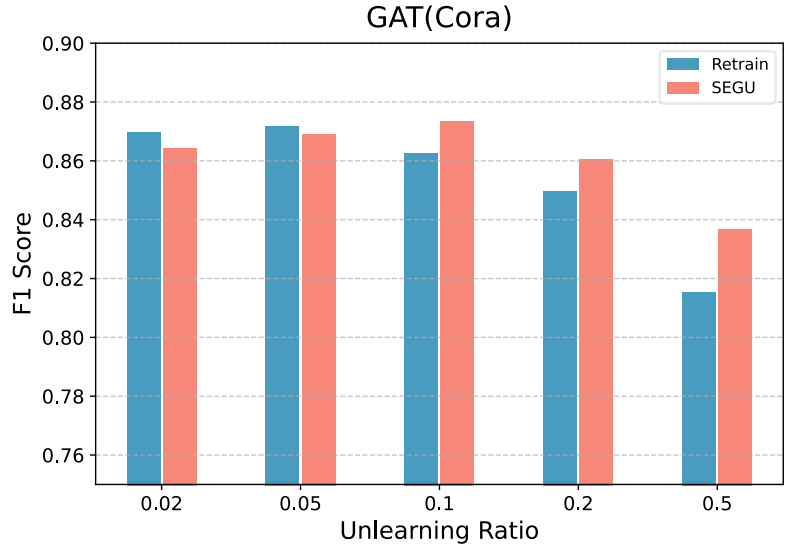


超参数κ

- SEGU 在节点任务中对非遗忘实体的预测性能随κ增加而降低或波动。
- 对κ值变化不敏感的模型，实际应用可灵活选值；对κ值变化敏感的模型，选值需谨慎，充分考虑κ变动对性能的影响，以保障模型性能。

不同遗忘比例下的性能分析

- 不同遗忘比例下重新训练和 SEGU 方法的 F1 分数对比
- SEGU可高效处理不同GNN模型的遗忘任务，在性能表现上与Retrain方法不相上下，甚至在高比例遗忘场景中实现超越。



PART FOUR

南京郵電大學 Nanjing University of Posts and Telecommunications

04

总结展望

Conclusion and Prospect

总结与展望

1 总结

本研究围绕图节点遗忘学习领域，针对现有方法在处理图结构信息和遗忘效率方面的不足，提出基于结构熵的图遗忘学习方法 SEGU。
在此基础上，构建预测模块与线性遗忘模块相互进化机制，同时提出基于结构熵的邻居选择方法和拓扑感知的遗忘传播策略，并设计包含焦点损失与 Kullback-Leibler 散度的复合损失函数。
最后通过在七个真实世界图数据集上实验，验证了 SEGU 在遗忘效率和模型效用方面的优越性。

2 展望

本研究需深入挖掘图的结构特征，探索更为先进的图表示学习技术，对邻居选择以及模块协同机制进行优化，以提高模型对于复杂图结构的适应能力。
随着隐私需求日益复杂，本研究需要把 SEGU 和新兴隐私保护技术进行深度融合，全面提升数据隐私保护的强度以及模型的安全性。
不同的应用场景对于图节点遗忘学习有着不一样的需求，后续需要仔细分析各个领域所有的独特需求以及约束条件，针对 SEGU 展开定制化的改进工作。

敬请各位老师批评指正！

姓名：许语轩

学号：B21060202

南京郵電大學
Nanjing University of Posts and Telecommunications