

# 图神经网络导论

## 图结构学习

授课教师：周晟

浙江大学 软件学院

2023.12



# 课程内容

- ① 研究背景
- ② 基于度量学习的图结构学习
- ③ 基于优化的图结构学习
- ④ 基于生成模型的图结构学习
- ⑤ 无监督 GSL

## ① 研究背景

② 基于度量学习的图结构学习

③ 基于优化的图结构学习

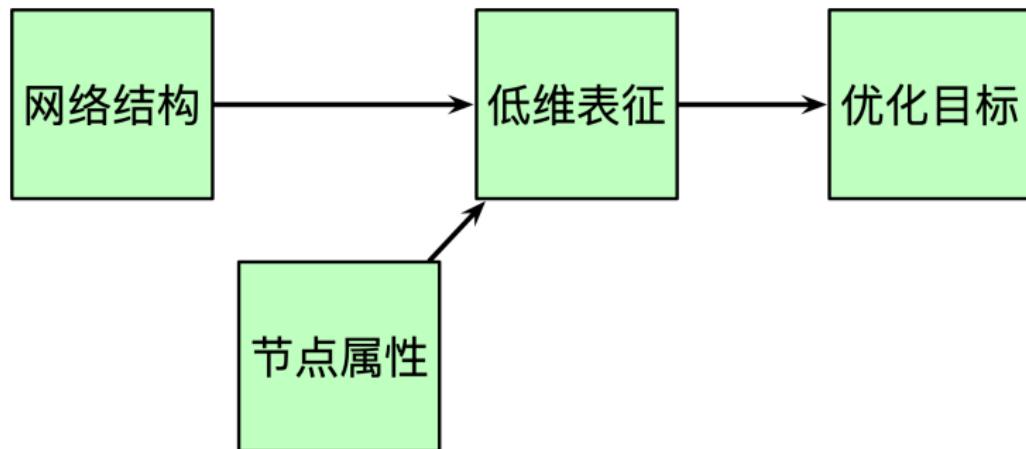
④ 基于生成模型的图结构学习

⑤ 无监督 GSL

# 研究背景

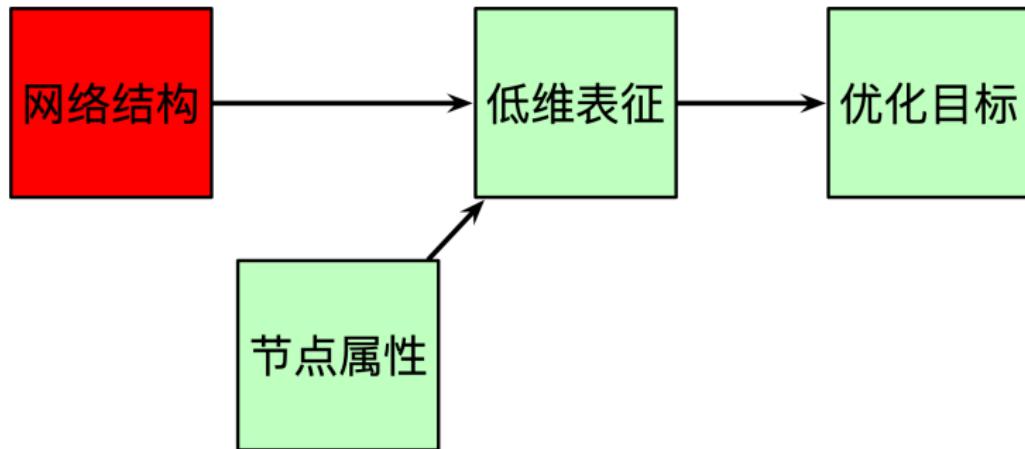
## 图神经网络的原理

通过将节点的信息沿着网络结构传递 (Message Passing)，捕获节点间的关系和节点的特征，学习更好的图 (节点，边，图) 表征。



图神经网络基本框架

# 研究背景



网络结构受损下的图神经网络

## 问题与挑战

实际场景中，网络结构往往面临受损或缺失的问题，使得依赖网络结构信息的图神经网络效果严重下降。

# 图结构缺失问题

- ① 数据天然的稀疏性（社交网络关注量有限）
- ② 数据采集、存储过程中丢失
- ③ 人为隐藏关系（犯罪）
- ④ 数据集本身没有关系型结构



# 图结构冗余问题

- ① 数据采集、存储过程中错误
- ② 电商网络中，刷单行为（诈骗高发）
- ③ 人为注入，“浑水摸鱼”



# 图结构学习的意义

图结构学习对于图神经网络的理论和应用均有重要意义：

- ① 理解数据间的真实关系
- ② 预测图结构演化趋势
- ③ 发掘数据中异常关系

本次课内容：

- ① 图结构学习方法
  - ① 度量学习：将网络中的边表示成由图神经网络学习的节点表征的函数
  - ② 直接优化：将结构作为参数，利用图的性质进行优化
  - ③ 生成模型：假设图结构是生成的，利用生成模型进行优化
- ② 图结构生成方法 (Optional)

# 图结构学习分类

## 按照图结构来源分类

- ① 度量学习：将网络中的边表示成由图神经网络学习的节点表征的函数
- ② 直接优化：将结构作为参数，利用图的性质进行优化
- ③ 生成模型：假设图结构是按照某种规律生成的，利用生成模型进行优化

## 按照训练流程分类

- ① 两阶段方法：先学习结构，再输入 GNN 用于下游任务。
- ② 同时优化方法：结构学习模块和 GNN 模块同时优化。
- ③ 迭代优化方法：图结构和 GNN 交替更新。

# 图结构的特性：同质性 (homophily)

## 同质性 (homophily)

同质性 (homophily)，指一个图网络中相邻节点相似的程度。在节点分类任务中，同质性通常被定义为相邻节点属于同一类别的概率。

可以证明，在同质性为 1 的极端情况下，GNN 可以达到100%的准确率。因此，**如何产生更具有同质性的图结构**是图结构学习中值得思考的问题。

# 图统计量

除了节点级的局部特征，大规模真实图数据往往具备特殊的宏观性质。

常用如下四种统计量来描述一个图的全局性质：

- ① 度分布, Degree distribution:  $P(k)$
- ② 聚类系数, Clustering coefficient:  $C$
- ③ 连通区域, Connected components:  $s$
- ④ 路径长度, Path length:  $h$

# 图统计量

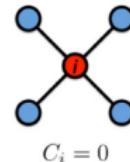
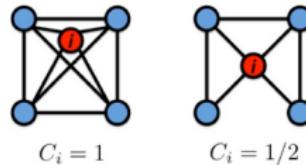
- 度分布, Degree distribution  $P(k)$  是指: 度为  $K$  的节点数量的分布,  $N_k =$  度为  $k$  的节点数量,

$$P(k) = N_k/N$$

- 聚类系数, Clustering coefficient

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

- $e_i$  是  $i$  节点的邻居中互相连边的数量
- $k_i$  是  $i$  节点的度

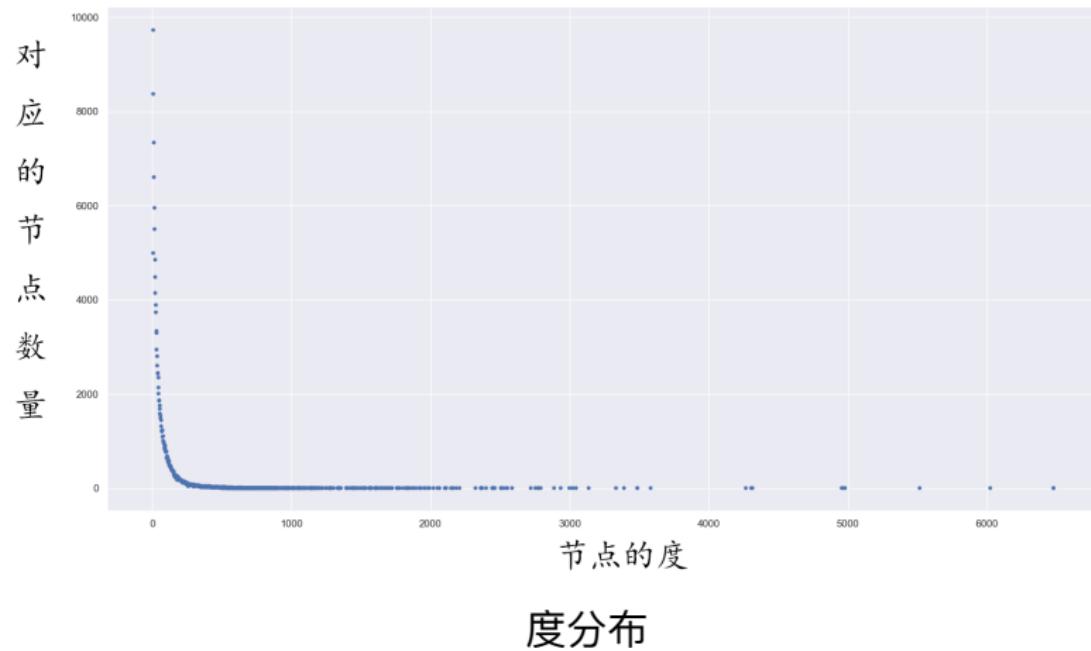


$$C_i \in [0,1]$$

# 图统计量

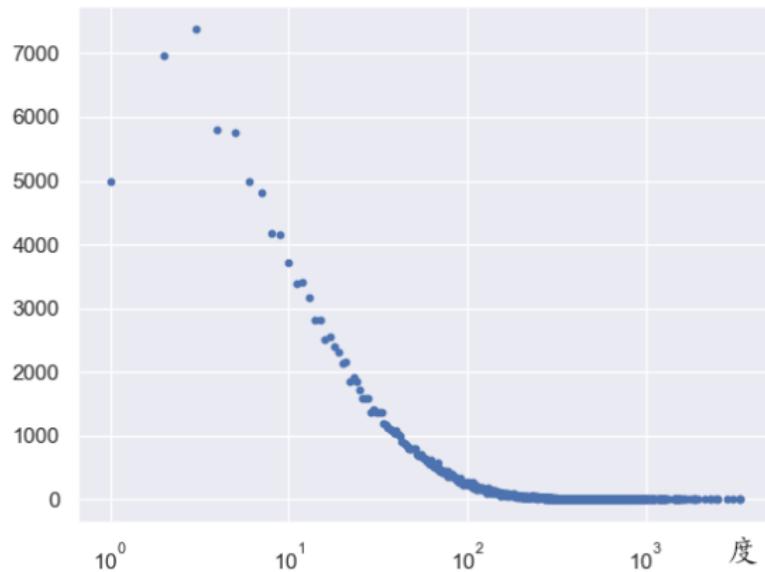
- 连通区域，Connected components，指图中联通区域按大小排序，其不同大小连通区域数量的分布
- 路径长度，Path length，指图中所有存在的最短路径的长度分布
- 以 Twitter 的社交网络图数据集为例，该数据集的邻接矩阵大小为  $157575 \times 157575$

# 图统计量-以 Twitter 数据集为例



# 图统计量-以 Twitter 数据集为例

点的数量



横轴指数化的度分布

# 随机图的度分布

$$p(k) = \binom{N-1}{K} P^k (1-P)^{N-1-k}$$

- 一个点的度为  $k$  的概率 (有  $k$  个点与之相连), 等价于除它本身之外的  $N-1$  个点选  $k$  个和它相连, 剩下  $N-1-k$  和它不连的概率。 $P$  是两点之间存在一条边的概率,  
 $P = \frac{E}{N*N}$ 。典型二项分布。
- 二项分布可以由泊松分布近似:

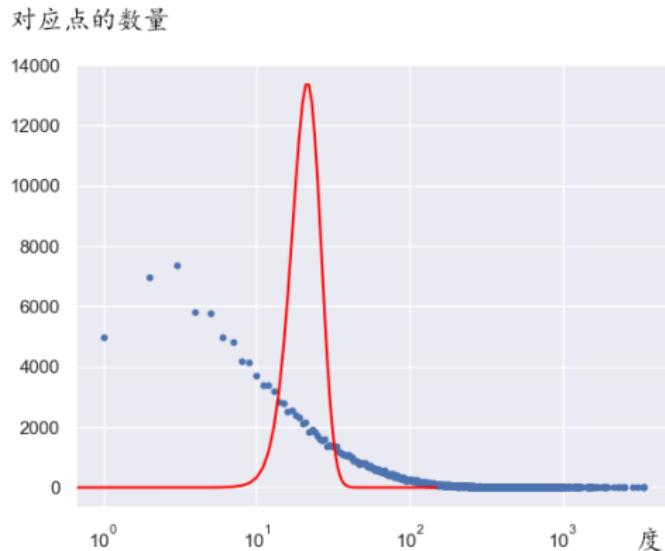
$$p(k) = \binom{N-1}{K} P^k (1-P)^{N-1-k} \approx \frac{\bar{k}^k}{k!} e^{-\bar{k}}$$

其中  $\bar{k} = p(N-1)$ , 是度的均值

- 好的图结构学习结果应保留图的这种天然特性

# 随机图的度分布

- 以 Twitter 数据集为例，其中共有 157575 点 3530653 边，把真实分布和随机分布（红色）画在一起



① 研究背景

② 基于度量学习的图结构学习

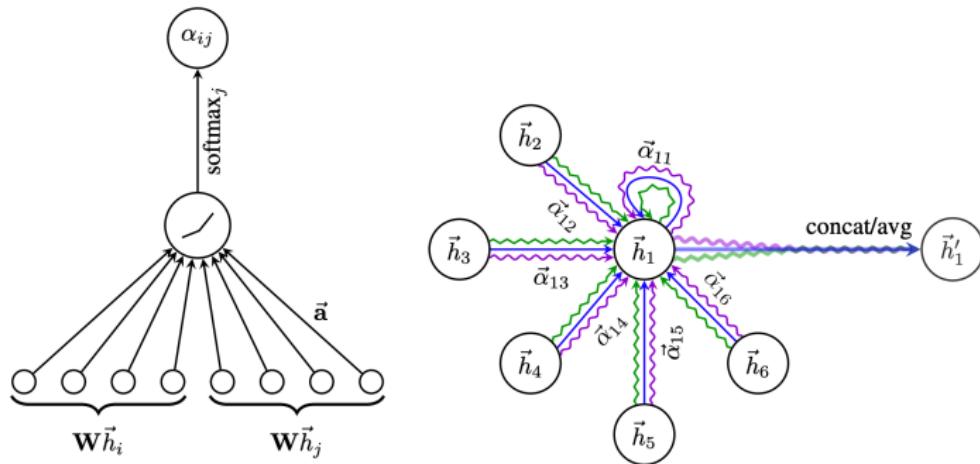
③ 基于优化的图结构学习

④ 基于生成模型的图结构学习

⑤ 无监督 GSL

# GAT 与图结构学习

$$\alpha_{ij} = \frac{\exp(\text{LeakReLU}(\vec{a}^T [W\vec{h}_i || W\vec{h}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i || W\vec{h}_k]))}$$

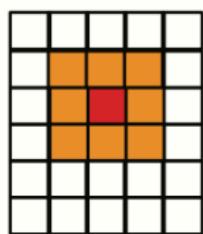


GAT 可看作一种特殊的图结构学习方法。

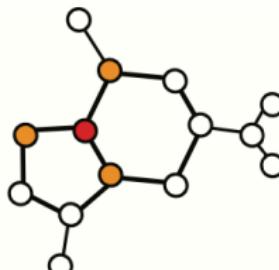
# Adaptive Graph Convolutional Neural Networks(AGCN)<sup>1</sup>

## 研究动机

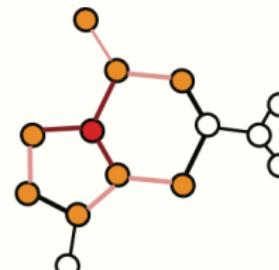
现有的图神经网络采用消息传递机制，但仅能在已观测的边上传递信息。由于网络稀疏性问题，实际可传递有效信息的节点并没有被连接。



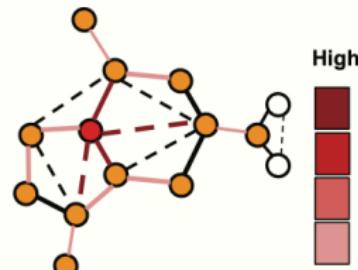
(1) CNN



(2) graphconv



(3) GCN



(4) AGCN

AGCN 与 CNN、图卷积和 GCN 的对比

# Adaptive Graph Convolutional Neural Networks

GCN 模型更新方式：

$$H^{(l+1)} = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

一个新的图结构即对应一个新的**拉普拉斯矩阵**。

## AGCN 对图结构学习的设想

- 图结构可参数化学习
- 参数量少（拉普拉斯矩阵参数化需要  $O(N^2)$  的空间）
- 可被应用于不同的图结构（可优化更多任务，如点云分类）

解决方案：基于核（kernel）的方法

## AGCN 的图结构学习策略

使用广义 Mahalanobis 距离实现基于核的距离度量

$$\mathbb{D}(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)}$$

- $x_i$  是图卷积网络某一层中节点  $i$  的特征。
- $M = W_d W_d^T$ ,  $W_d \in \mathbb{R}^{d \times d}$  是可学习的 kernel。

将参数量从  $O(N^2)$  降至  $O(d^2)$ !

# Adaptive Graph Convolutional Neural Networks

给定距离度量，使用 Gaussian Kernel 归一化得到稠密的邻接矩阵  $\tilde{A}$ 。

$$\tilde{A}_{ij} \leftrightarrow \mathbb{G}_{x_i, x_j} = \exp(-\mathbb{D}(x_i, x_j)/(2\sigma^2))$$

将其作为残差结构与原图结构叠加，得到更新后的拉普拉斯矩阵并用于卷积图神经网络：

$$\hat{L} = L + \alpha L_{res} = L + \alpha(I - \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2})$$

为什么要加上原来的图结构？

总结与归纳：

## 优点

- ① 模型简单，启发了很多后续的图结构学习工作
- ② 参数量小，便于优化

## 缺点

- ① 结构仅通过线性变换学习
- ② 学习的拉普拉斯矩阵过于稠密
- ③ 图神经网络没有给结构学习带来信息
- ④ 模型缺乏明确的结构学习优化目标

# Learning to Drop: Robust Graph Neural Network via Topological Denoising

## 研究动机

网络中观察到的边往往存在与任务无关的边，通过这些节点传递信息往往给节点表征学习带来噪声。以节点分类任务为例，标签一致的节点才能提供有效的分类信息。

		Ratio of positive edges removed										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
ratio of negative edges removed	0	60.1	60.1	55.7	55.2	54.8	54.8	54.2	53.8	53.6	53.5	53.5
	0.1	64.2	63.7	61.4	60.0	59.6	59.4	59.3	58.7	58.7	58.6	57.4
	0.2	69.6	68.2	66.5	66.4	66.1	65.4	63.6	63.8	62.6	62.1	61.2
	0.3	72.8	72.3	71.5	70.5	70.2	69.0	68.3	67.7	68.9	67.6	66.8
	0.4	79.3	76.9	74.5	73.5	73.5	72.9	72.6	71.8	71.2	70.3	69.5
	0.5	80.4	79.2	78.0	76.6	75.6	75.3	75.1	74.3	73.7	73.6	72.3
	0.6	83.6	82.4	81.3	80.6	80.3	78.6	78.1	77.3	76.8	75.0	74.1
	0.7	83.9	82.6	81.6	81.5	81.0	80.1	79.5	78.2	78.1	77.7	76.5
	0.8	85.5	83.8	83.5	82.8	81.1	80.7	80.7	79.9	79.6	79.9	79.4
	0.9	86.3	86.1	84.8	83.6	83.6	82.6	82.4	81.8	81.3	81.1	81.0

# PTDNet: 图的稀疏化

图的稀疏化，即将 dropout 的思想用于图结构以加强模型鲁棒性，从图中删除一部分边。

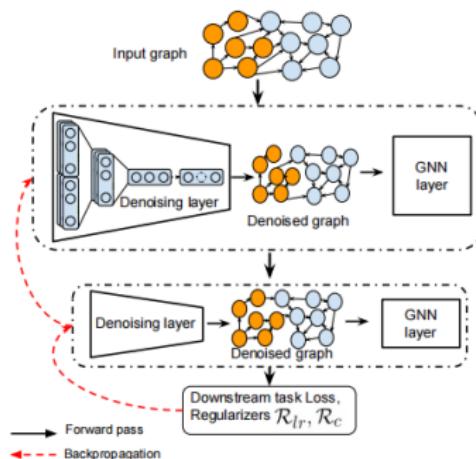
具体可分为：随机去边、根据预先定义的规则、参数化三种。

## 图的稀疏化和图结构学习

- ① 图的稀疏化可以视作一种图结构学习，但仅在初始存在的边上进行，即可以去除冗余的边，但不能增加新边（后者往往需要耗费更多空间）。
- ② 该不同之处通常表现在得到新的结构后与原结构乘还是加。

# PTDNet

PTDNet (parameterized topological denoising network), **参数化图结构降噪网络**, 使这一过程受任务损失监督, 并联合优化 GNN 网络和降噪网络。

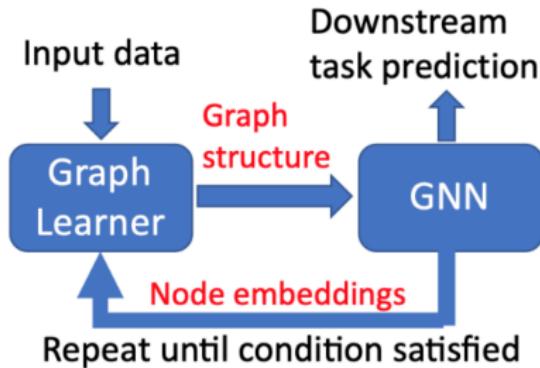


每层 GCN 均配备一个降噪网络, 使用上一层表征产生新结构给该卷积层使用。(整体架构和 GAT 很类似)

# Iterative Deep Graph Learning for Graph Neural Networks: Better and Robust Node Embeddings<sup>2</sup>

## 研究动机

图结构与图神经网络学习可以相互帮助：好的图结构可以学习好的图神经网络，好的图神经网络结果可以帮助建模更准确的网络结构。



<sup>2</sup>[Chen et al., 2020]

# 从图结构学习到图神经网络

- 图结构沿用 AGCN 的思想：融合原始网络结构与学习过程中的网络结构

$$\tilde{\mathbf{A}}^{(t)} = \lambda \mathbf{L}^{(0)} + (1 - \lambda) \{ \eta \mathbf{f}(\mathbf{A}^{(t)}) + (1 - \eta) \mathbf{f}(\mathbf{A}^{(1)}) \}$$

- 图神经网络沿用经典的 GCN：

$$\mathbf{Z} = \text{ReLU} \left( \mathbf{MP}(\mathbf{X}, \tilde{\mathbf{A}}) \mathbf{W}_1 \right), \hat{\mathbf{y}} = \sigma \left( \mathbf{MP}(\mathbf{Z}, \tilde{\mathbf{A}}) \mathbf{W}_2 \right)$$

$$\mathcal{L}_{\text{pred}} = \ell(\hat{\mathbf{y}}, \mathbf{y})$$

# 根据图神经网络学习图结构

## 基本假设

具有相似的低维表征的节点对，有更大的概率生成边。

基于节点表征的边概率计算：

- 第一层：基于向量的相似度计算

$$S_{ij} = \cos(\vec{w}_p \odot \vec{z}_i, \vec{w}_p \odot \vec{z}_j)$$

- 第二层：基于多头向量的相似度计算

$$S_{ij} = \frac{1}{m} \sum_{p=1}^m \cos(\vec{w}_p \odot \vec{z}_i, \vec{w}_p \odot \vec{z}_j)$$

# 根据图神经网络学习图结构

使用节点表征相似度预测图结构的缺陷：

- ① 计算复杂度高
- ② 网络过于稠密

IDGL 的解决方案：

- 第三层：Anchor-based 度量学习：

$$a_{ik}^p = \cos(\mathbf{w}_p \odot \mathbf{v}_i, \mathbf{w}_p \odot \mathbf{u}_k), \quad a_{ik} = \frac{1}{m} \sum_{p=1}^m a_{ik}^p$$

- 第四层：图稀疏化。只选择每个节点最近的 K 个节点，或相似度大于阈值  $\epsilon$  的邻居节点。

# 图结构质量控制

## IDGL 迭代优化的缺陷

图结构学习模块缺少监督信号，容易陷入局部平凡解 (Trival Solution)

好的图结构应具有如下性质：

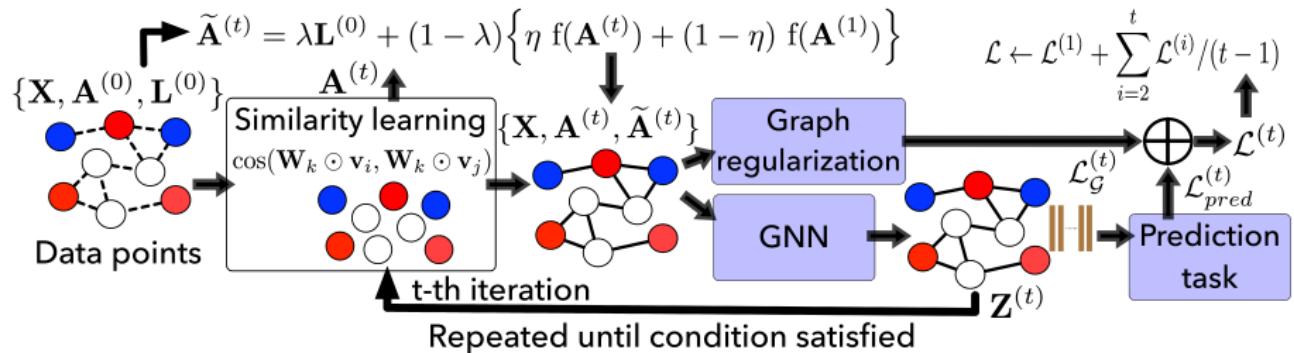
- ① 平滑性：临近的节点的属性相似度高

$$\Omega(\mathbf{A}, \mathbf{X}) = \frac{1}{2n^2} \sum_{i,j} A_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \frac{1}{n^2} \text{tr} (\mathbf{X}^T \mathbf{L} \mathbf{X})$$

- ② 连通性：网络中的节点尽可能构成联通图
- ③ 稀疏性：网络中的边总体稀疏

$$f(\mathbf{A}) = \frac{-\beta}{n} \mathbf{1}^T \log(\mathbf{A} \mathbf{1}) + \frac{\gamma}{n^2} \|\mathbf{A}\|_F^2$$

# IDGL 模型



IDGL 方法示意图

## 优点

- ① 提出了统一的迭代学习框架
- ② 明确了结构学习的优化目标
- ③ 使用多种手段兼顾效率与精度

# 循环优化图神经网络

一般形式：

- ① 使用已有结构学习节点表征  $z$

$$\mathbf{z} = GNN(X, A)$$

- ② 利用节点表征相似度学习图结构：

$$\tilde{\mathbf{A}}_{ij} = \phi(\mathbf{z}_i, \mathbf{z}_j)$$

- ③ 融合输入结构和学习图结构

$$\mathbf{A}^* = g(\mathbf{A}, \tilde{\mathbf{A}})$$

迭代循环，直至收敛。

# 基于节点相似度的图结构生成

## ① Gaussian Kernel

$$\begin{aligned}\phi(\mathbf{z}_i, \mathbf{z}_j) &= \sqrt{(\mathbf{z}_i - \mathbf{z}_j)^\top \mathbf{M} (\mathbf{z}_i - \mathbf{z}_j)} \\ \tilde{\mathbf{A}}_{ij} &= \exp\left(-\frac{\phi(\mathbf{z}_i, \mathbf{z}_j)}{2\sigma^2}\right)\end{aligned}$$

## ② 向量内积

$$\tilde{\mathbf{A}} = \sigma(\mathbf{Z}\mathbf{Z}^\top)$$

## ③ Cosine 相似度

$$\phi(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i \mathbf{z}_j^\top}{\|\mathbf{z}_i\|_2 \|\mathbf{z}_j\|_2}.$$

- ① 研究背景
- ② 基于度量学习的图结构学习
- ③ 基于优化的图结构学习
- ④ 基于生成模型的图结构学习
- ⑤ 无监督 GSL

# 基于优化的图结构学习

## 研究动机

图结构可以看作节点之间是否有边的参数的集合，可以直接对图结构进行参数化学习。

## 困难与挑战

- ① 优化目标如何确定？
- ② 参数量大 ( $O(N^2)$ )
- ③ 离散优化，难以训练

# Graph Structure Learning for Robust Graph Neural Networks<sup>3</sup>

真实图结构的性质 (properties) :

- ① 低秩
- ② 稀疏
- ③ 邻接的节点特征相近

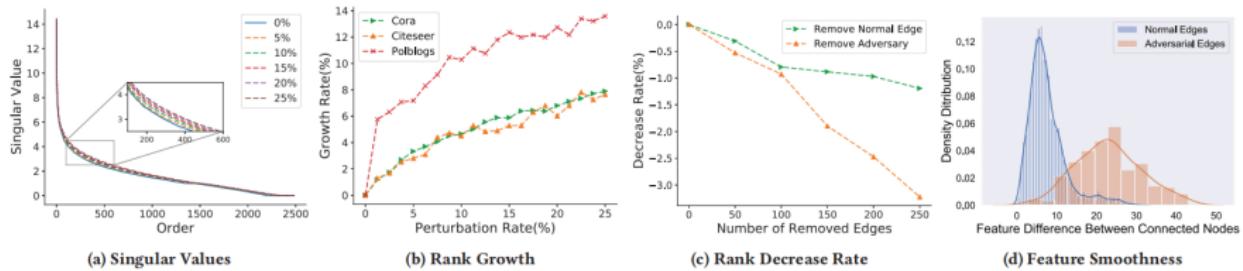


Figure 1: An illustrative example on the property changes of the adjacency matrix by adversarial attacks

<sup>3</sup>[Jin et al., 2020]

# Pro-GNN: 性质

1. 如何利用这些性质引导图结构的学习?

ProGNN 采取了直接优化邻接矩阵  $S$  的方法, 将  $S$  视作一个  $n * n$  的参数矩阵, 这样这些性质可以表达为有关  $S$  的一系列损失函数。

## 图性质的损失函数

① 低秩和稀疏:

$$\mathcal{L}_0 = \|A - S\|_F^2 + \alpha\|S\|_1 + \beta\|S\|_*, s.t., S = S^T$$

② 特征平滑:

$$\mathcal{L}_s = \text{tr}(X^T \hat{L} X) = \frac{1}{2} \sum_{i,j=1}^N S_{ij} \left( \frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2, s.t., S = S^T$$

# Pro-GNN: 优化

## 2. 如何联合学习图结构和图神经网络?

最终损失:

$$\begin{aligned} \arg \min_{S \in \mathcal{S}, \theta} \mathcal{L} &= \mathcal{L}_0 + \lambda \mathcal{L}_f + \gamma \mathcal{L}_{GNN} \\ &= \|A - S\|_F^2 + \alpha \|S\|_1 + \beta \|S\|_* + \gamma \mathcal{L}_{GNN}(\theta, S, X, \mathcal{Y}_L) \\ &\quad + \lambda \text{tr}(X^T \hat{L} X), \text{s.t. } S = S^T \end{aligned}$$

由于一起求解较为困难，使用**交替优化** (Alternative Optimization) 迭代求解。

### 交替优化

交替优化经常在 GSL 方法中出现，用来解决双层优化等较困难的优化问题。其做法是固定一个，优化另一个，交替进行。

# Pro-GNN: 优化

$$\begin{aligned} \arg \min_{S \in \mathcal{S}, \theta} \mathcal{L} = & \|A - S\|_F^2 + \alpha \|S\|_1 + \beta \|S\|_* + \gamma \mathcal{L}_{GNN}(\theta, S, X, \mathcal{Y}_L) \\ & + \lambda \text{tr}(X^T \hat{L} X), \text{s.t. } S = S^T \end{aligned}$$

优化 GNN:

$$\theta \leftarrow \eta' \frac{\partial \mathcal{L}_{GNN}(\theta, S, X, \mathcal{Y}_L)}{\partial \theta}$$

优化图结构  $S$ : 核范数和  $l_1$  范数不可导, 需使用 Proximal Optimization 数学工具<sup>4</sup>

---

<sup>4</sup>[Jin et al., 2020]

# Learning Discrete Structures for Graph Neural Networks<sup>5</sup>

基于双层优化的思路：

- 外层优化：在验证集上学习离散结构
- 内层优化：在训练集上学习 GCN 的参数

$$\min_A F(w_A, A) \text{ such that } w_A \in \arg \min_w L(w, A)$$

$$L(w, A) = \sum_{v \in V_{train}} l(f_w(X, A)_v, y_v) + \Omega(w)$$

$$F(w_A, A) = \sum_{v \in V_{val}} l(f_{w_A}(X, A)_v, y_v)$$

其中内层优化使用训练集中的点，外层优化使用验证集中的节点。

---

<sup>5</sup>[Franceschi et al., 2019]

# LDS: Reparameterization

A 是离散变量，无法直接求导，LDS 假设 A 从多个独立的伯努利分布中生成： $A \sim \text{Ber}(\theta)$ ,  $\theta \in (0, 1)^{n \times n}$   
优化问题变为：

$$\min_{\theta} \mathbb{E}_{A \sim \text{Ber}(\theta)} [F(w_{\theta}, A)]$$

$$\text{such that } w_{\theta} = \arg \min_w \mathbb{E}_{A \sim \text{Ber}(\theta)} [L(w, A)]$$

# LDS: GCN 内层优化

内层问题:

$$w_\theta = \arg \min_w \mathbb{E}_{A \sim \text{Ber}(\theta)} [L(w, A)]$$

$$\mathbb{E}_{A \sim \text{Ber}(\theta)} [L(w, A)] = \sum_{A \in \mathcal{H}_N} P_\theta(A) L(w, A)$$

上式不可能直接求出，按照 SGD 的思想迭代近似，每次采样一个结构梯度下降：

$$w_{\theta,t+1} = w_{\theta,t} - \gamma \nabla L(w_{\theta,t}, A_t), A_t \sim \text{Ber}(\theta)$$

# LDS：图结构外层优化

外层问题：

$$\min_{\theta} \mathbb{E}_{A \sim \text{Ber}(\theta)} [F(w_{\theta}, A)]$$

求解对  $\theta$  的梯度，将对期望的梯度近似为梯度的期望：

$$\begin{aligned}\nabla_{\theta} \mathbb{E}_{A \sim \text{Ber}(\theta)} [F(w_{\theta,T}, A)] &\approx \mathbb{E}_{A \sim \text{Ber}(\theta)} [\nabla_A F(w_{\theta,T}, A)] \\ &= \mathbb{E}_A [\partial_w F(w_{\theta,T}, A) \nabla_A w_{\theta,T} + \partial_A F(w_{\theta,T}, A)]\end{aligned}$$

同样使用 SGD 迭代近似最优解：

$$\theta' = \theta - \eta (\partial_w F(w_{\theta,T}, A) \nabla_A w_{\theta,T} + \partial_A F(w_{\theta,T}, A)), A \sim \text{Ber}(\theta)$$

# LDS：交替优化

进行  $T$  次内层迭代：

$$w_{\theta,t+1} = w_{\theta,t} - \gamma \nabla L(w_{\theta,t}, A_t), A_t \sim \text{Ber}(\theta)$$

进行一次外层迭代：

$$\theta' = \theta - \eta (\partial_w F(w_{\theta,T}, A) \nabla_A w_{\theta,T} + \partial_A F(w_{\theta,T}, A)), A \sim \text{Ber}(\theta)$$

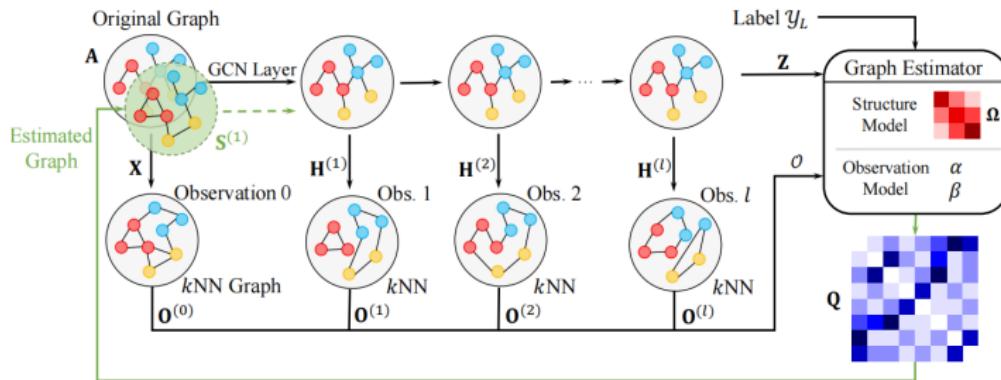
其中外层梯度计算中第一项  $\nabla_A w_{\theta,T}$  可在内层迭代中回溯。

- 1 研究背景
- 2 基于度量学习的图结构学习
- 3 基于优化的图结构学习
- 4 基于生成模型的图结构学习
- 5 无监督 GSL

# Graph Structure Estimation Neural Networks<sup>6</sup>

## 研究动机

现有的图结构学习方法仅考虑图数据的判别特性（节点之间是否有边），对网络的生成过程缺乏可解释性。



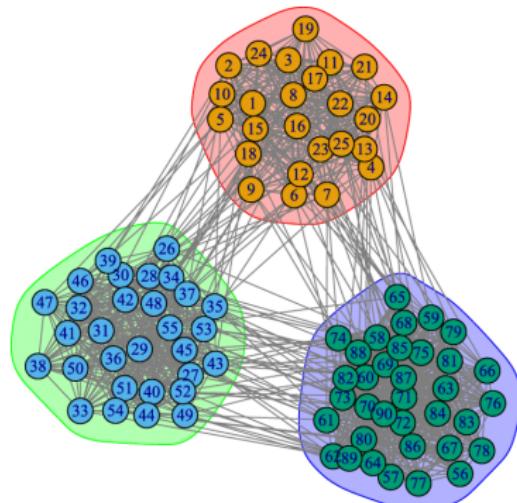
GEN 的模型示意图

<sup>6</sup>[Wang et al., 2021]

# SBM 模型简介

## Stochastic Block Model(SBM)

Stochastic Block Model(SBM) 是一种经典的网络生成模型。它假设网络中边的生成是由两个节点的社区分布决定，即属于同一个社区的节点有更大的概率生成边，不同社区的节点生成边的概率更小。



# GEN 的图结构建模

在 SBM 模型的假设下，网络中存在边的概率只与两个节点所属的社区有关。给定参数矩阵  $\Omega \in \mathcal{S}^{|C|*|C|}$ ，观测到网络的概率为：

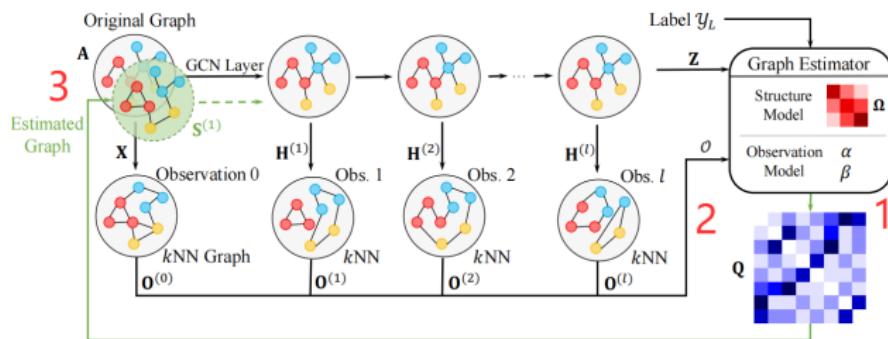
$$P(G|\Omega, Z, \mathcal{Y}_L) = \prod_{i < j} \Omega_{c_i c_j}^{G_{ij}} (1 - \Omega_{c_i c_j})^{1-G_{ij}}$$

$c_i$  表示节点  $i$  的类别。当其属于训练集时，使用真实标签；当属于无标签节点时，使用预测标签。

$$c_i = \begin{cases} y_i & \text{if } v_i \in \mathcal{V}_L \\ z_i & \text{otherwise} \end{cases}$$

GEN 的基本思路:

- ① 假设存在一个图的**最优结构**, 从一个**SBM 模型**中生成。
- ② 假设该最优结构存在多个**观测**, 从节点不同阶段的表示中各**自生成 KNN 图**作为观测。
- ③ **交替优化**, 在训练好的 GCN 基础上使用**EM 算法**求解最优结构, 将其作为新的结构重新训练。



GEN 的模型示意图

# GEN 的图结构建模

## 模型假设

给定最优的图结构  $G$ , 真实观测的图结构由最优图结构随机采样后生成。

定义最优图结构下有边和没有边, 在观测图结构下存在边的概率为  $\alpha$ ,  $\beta$ , 观测图结构的概率为:

$$P(O \mid G, \alpha, \beta) = \prod_{i < j} [\alpha^{E_{ij}} (1 - \alpha)^{M - E_{ij}}]^{G_{ij}} \times [\beta^{E_{ij}} (1 - \beta)^{M - E_{ij}}]^{1 - G_{ij}}$$

其中  $E_{ij}$  为给定的一组  $M$  个观测图结构中  $ij$  之间出现边的次数

# GEN 的模型优化

GEN 模型需要优化的参数有最优图结构  $G$  和图神经网络参数  $\Theta$ 。然而参数优化存在两大类困难：

- ① 两种参数高度耦合（迭代求解）
- ② 最优图结构  $Q$  为离散变量（转化为伯努利分布参数）

优化 GNN:

$$\theta \leftarrow \eta' \frac{\partial \mathcal{L}_{GNN}(\theta, S, X, \mathcal{Y}_L)}{\partial \theta}$$

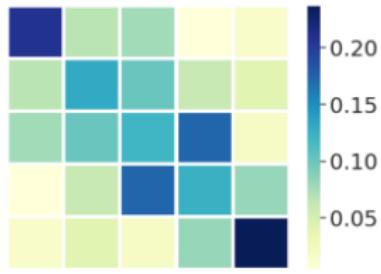
利用原图结构和属性图作为输入优化图结构  $S$ :

$$S \leftarrow EM(O^{(0)}, O^{(1)}, O^{(2)}, \dots, O^{(l)}, O^{(l+1)})$$

式中  $O^{(i)}$  为不同阶段的表征产生的 KNN 图。

# GEN: 同质性

在 toy dataset 上验证了: GEN 可以增加相同类别节点之间的边,减少连接不同类别的边。  
体现了 GEN 可以学习具有同质性的图结构的能力。



(a) Original graph



(b) Estimated graph

GEN 产生结构的同质性

- ① 研究背景
- ② 基于度量学习的图结构学习
- ③ 基于优化的图结构学习
- ④ 基于生成模型的图结构学习
- ⑤ 无监督 GSL

# Unsupervised GSL

以上的方法都基于有监督情况，结构学习依赖监督信息

## Drawbacks

- ① 不平衡。标签有限时有标签节点周围边受到监督信息引导，但距离有标签节点距离远的边难以学习。
- ② 不实用。相较于监督情况，无监督情况更广泛。
- ③ 不通用。对于每个任务都要重新学习结构，不能学习一个通用的结构。

近期出现的 Unsupervised GSL 工作大多利用了对比学习的思想，关键在于如何设计不同的 view，如何融入结构学习。

# Reliable Representations Make A Stronger Defender: Unsupervised Structure Refinement for Robust GNN<sub>[Li et al., 2022]</sub>

聚焦基于距离度量的 GSL 方法，认为监督情况下得到的表征容易受扰动，本身质量不佳，不能产生好的结构。

提出了一个 two-stage 的无监督方法 STABLE，分为两个阶段。

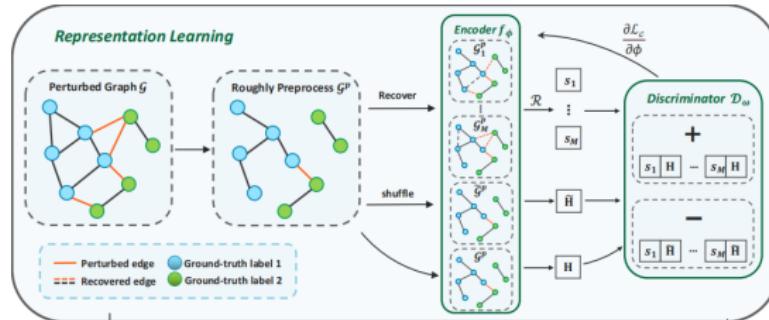
Representation Learning：用**对比学习**得到 reliable 的表征

Graph Refining：根据该表征的相似度计算结构。

# Representation Learning

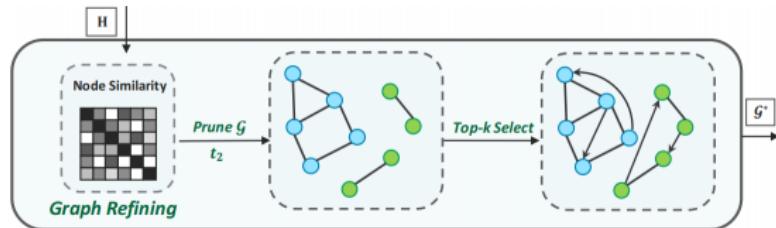
如何设计不同的 view，使得学到的表征可以对抗噪声？  
对于一个含噪声的图  $G$ ：

- ① 采取一个简单预处理，删除一部分边，得到噪声程度低的图  $G^p$
- ② 从  $G^p$  中随机恢复被删除的边，构成  $M$  个噪声程度较高的图  $G_1^p, G_2^p, \dots, G_M^p$
- ③ 要求从  $G^p$  和  $G_1^p, G_2^p, \dots, G_M^p$  得到的表征接近，同时 shuffle  $G^p$  得到  $\tilde{G}^p$  作为负样本，要求  $\tilde{G}^p$  和  $G_1^p, G_2^p, \dots, G_M^p$  得到的表征远离



# Graph Refining

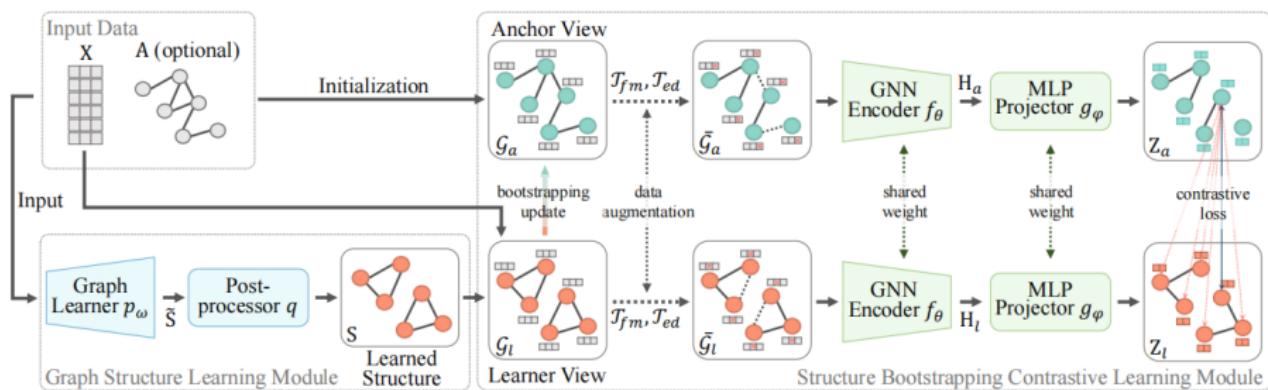
在获得 reliable 的表征  $H$  后，使用常规的相似度计算和后处理得到最终的结构：



该结构与任务无关的，从无监督方法获得，可以直接用于多个下游任务。

# Towards Unsupervised Deep Graph Structure Learning<sup>7</sup>

可学习的 GSL 模块产生 learner view，原结构作为 anchor view，最大化两个 view 之间的互信息（MI）。



<sup>7</sup>[Liu et al., 2022]

涵盖了之前 GSL 工作中的多种产生图的方法：

1. 直接将整个图作为参数：

$$S = g(\Omega), \Omega \in R^{n \times n}$$

2. Learnable encoder+metric：

$$S = metric(encoder_w(X, A))$$

再经过后处理操作（对称化、稀疏化等）得到 learner view

$$G_l = (S, X)$$

# Contrastive Learning

得到 learner view  $G_l(A, S)$  后和预先指定的 anchor view (文中设置为原图  $G_a = (A, X)$ ) 进行对比学习。

使用共享的 GCN+MLP 得到表征：

$$Z_l = g_\phi(f_\theta(\overline{G}_l)), Z_a = g_\phi(f_\theta(\overline{G}_a))$$

对比学习最大化互信息：

$$\begin{aligned} \mathcal{L} &= \frac{1}{2n} \sum_{i=1}^n [\ell(z_{l,i}, z_{a,i}) + \ell(z_{a,i}, z_{l,i})] \\ \ell(z_{l,i}, z_{a,i}) &= \log \frac{e^{\text{sim}(z_{l,i}, z_{a,i})/t}}{\sum_{k=1}^n e^{\text{sim}(z_{l,i}, z_{a,k})/t}} \end{aligned}$$

# 更新 Anchor view

一直使用原图作为 anchor view 会带来原结构中的误差，文中每隔若干轮次更新，和当前学习的结构加权：

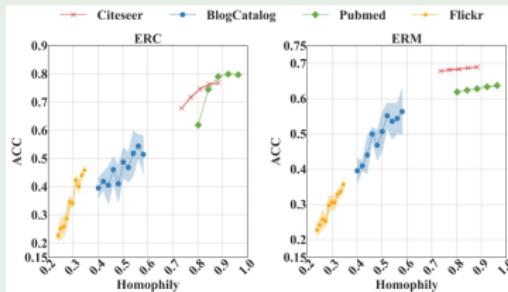
$$A_a = \tau A_a + (1 - \tau)S$$

# Homophily-enhanced Structure Learning for Graph Clustering<sup>8</sup>

## GSL+Graph Clustering

### 研究动机

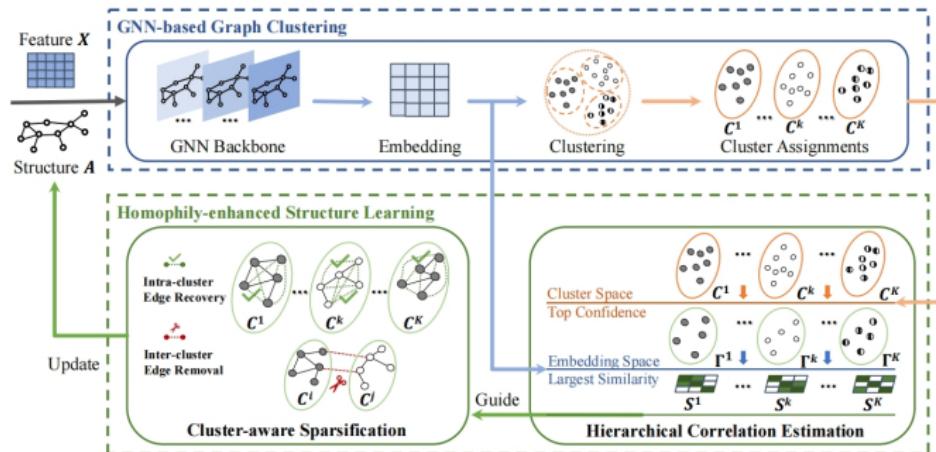
- ① 现有 GSL 方法大多是有监督的，而无监督方法未考虑图聚类问题的特性。
- ② 预实验发现提高图结构的 homophily 可以有效提高聚类效果。



<sup>8</sup>[Gu et al., 2023]

# HoLe: 模型架构

HoLe 采用类似 IDGL 的迭代更新机制，聚类和结构学习交替进行。



HoLe 方法示意图

每一个迭代中，HoLe 根据聚类结果，以同质性为目标更新结构。

# HoLe: Homophily-enhanced Structure Learning

- ① 筛选每个类中的高置信度节点:

$$\Gamma^k = \left\{ v_i \mid c_i = k, \text{Rank}_{\{v_i \mid c_i = k\}}(q_{ik}) \leq \gamma * |\{v_i \mid c_i = k\}| \right\}$$

- ② 计算隐空间内的相似度:  $S_{ij}^k = Z_i^k Z_j^{k^T} \in \mathbb{R}^{|\Gamma^k| \times |\Gamma^k|}$

- ③ 筛选相近同质边:

$$\mathcal{E}_{rc} = \bigcup_k \left\{ (v_i, v_j) \mid \text{Rank}(S_{ij}^k) \leq \xi * |\mathcal{E}| * \frac{N_k}{N}, \text{ and } c_i = c_j = k \right\}$$

- ④ 筛选不相似异质边:

$$\mathcal{E}_{rm} = \left\{ (v_i, v_j) \mid \text{Rank}(S_{ij}) \geq (1 - \eta) * |\mathcal{E}|, (v_i, v_j) \in \mathcal{E}, c_i \neq c_j \right\}$$

- ⑤  $A^* = A - A_{\mathcal{E}_{rm}} + A_{\mathcal{E}_{rc}}$

# PROSE: Graph Structure Learning via Progressive Strategy<sup>9</sup>

## 研究动机

现有的图结构学习方法同时学习图结构的不同部分，但一张图中各部分重要性不同，应当先学习重要节点之间的连接。如社交网络中先学习名人之间的连接。

## 困难与挑战

- ① 如何得到图结构中不同区域的重要性排序？
- ② 如何根据重要性依次学习结构？

<sup>9</sup>[Wang et al., 2023]

# PROSE：问题形式化

$$P_{\phi}(A^* | G) \propto \sum_{(A_m, \dots, A_1) \sim \prod_{i=1}^m P_{\phi_i^{pss}}(A_i | A_{>i}, \mathcal{G})} \prod_{i=1}^m P_{\phi_i^{psr}}(A_i^* | A_{<i}^*, A_i)$$

PROSE 基于应当按照重要性次序学习不同子结构的思想，将问题形式化为两个模块。

- Progressive Structure Splitting(PSS):

$$(A_m, \dots, A_1) \sim \prod_{i=1}^m P_{\phi_i^{pss}}(A_i | A_{>i}, \mathcal{G})$$

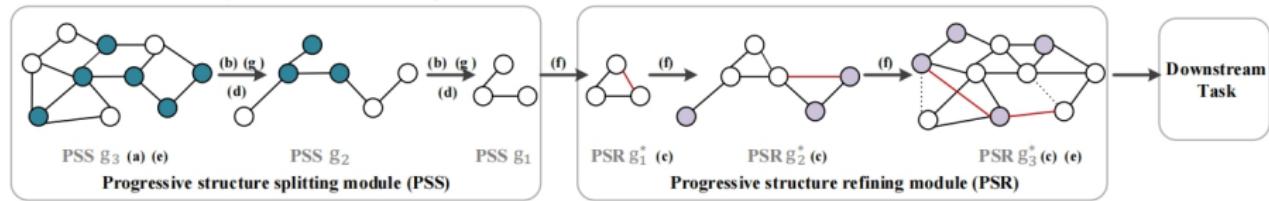
自顶向下划分原结构为重要性由小到大的层级

- Progressive Structure Refining(PSR):

$$\prod_{i=1}^m P_{\phi_i^{psr}}(A_i^* | A_{<i}^*, A_i)$$

自底向上依次学习结构，重要的子结构优先学习。

# PROSE: 模型



PROSE 方法示意图

- 按照渐进策略，使用 Graph Pooling 划分原始结构，再自底向上学习结构
- 事实上形成了一种 encoder-decoder 架构

# OpenGSL: A Comprehensive Benchmark for Graph Structure Learning<sup>10</sup>

## GSL 领域存在的问题

- ① 不同文章采用的实验设置（数据划分、处理等）不同
- ② 对学习的图结构缺少观察
- ③ 效率问题被忽视

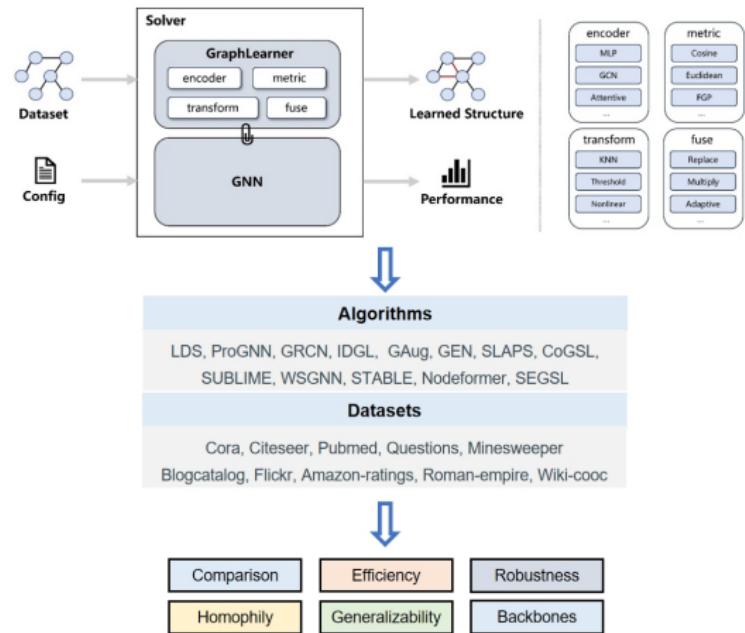
OpenGSL 旨在建立一个详细的 benchmark 来解决上述问题。

---

<sup>10</sup>[Zhou et al., 2023]

# OpenGSL: 介绍

- ① OpenGSL 是针对图结构学习的简单易用的开源包。
- ② 囊括了前沿的 GSL 方法，在统一设置下实现公平对比。
- ③ 对各 GSL 方法进行了多维度的分析。



Model	Cora	Citeseer	Pubmed	Questions	Minesweeper
GCN	$81.95 \pm 0.62$	$71.34 \pm 0.48$	$78.98 \pm 0.35$	<b><math>75.80 \pm 0.51</math></b>	<b><math>78.28 \pm 0.44</math></b>
LDS	<b><math>84.13 \pm 0.52</math></b>	<b><math>75.16 \pm 0.43</math></b>	—	—	—
ProGNN	$80.27 \pm 0.48$	$71.35 \pm 0.42$	$79.39 \pm 0.29$	—	$51.43 \pm 2.22$
IDGL	<b><math>84.19 \pm 0.61</math></b>	<b><math>73.26 \pm 0.53</math></b>	<b><math>82.78 \pm 0.44</math></b>	$50.00 \pm 0.00$	$50.00 \pm 0.00$
GRCN	<b><math>84.61 \pm 0.34</math></b>	$72.34 \pm 0.73$	$79.30 \pm 0.34$	<b><math>74.50 \pm 0.84</math></b>	$72.57 \pm 0.49$
GAug	$83.43 \pm 0.53$	$72.79 \pm 0.86$	$78.73 \pm 0.77$	—	<b><math>77.93 \pm 0.64</math></b>
SLAPS	$72.29 \pm 1.01$	$70.00 \pm 1.29$	$70.96 \pm 0.99$	—	$50.89 \pm 1.72$
WSGNN	$83.66 \pm 0.30$	$71.15 \pm 1.01$	$79.78 \pm 0.35$	—	$67.91 \pm 3.11$
Nodeformer	$78.81 \pm 1.21$	$70.39 \pm 2.04$	$78.38 \pm 1.94$	<b><math>72.61 \pm 2.29</math></b>	$77.29 \pm 1.71$
GEN	$81.66 \pm 0.91$	<b><math>73.21 \pm 0.62</math></b>	$78.49 \pm 3.98$	—	<b><math>79.56 \pm 1.09</math></b>
CoGSL	$81.46 \pm 0.88$	$72.94 \pm 0.71$	$78.38 \pm 0.41$	—	—
SEGSL	$81.04 \pm 1.07$	$71.57 \pm 0.40$	$79.26 \pm 0.67$	—	—
SUBLIME	$83.33 \pm 0.73$	$72.44 \pm 0.89$	<b><math>80.56 \pm 1.32</math></b>	$67.21 \pm 0.99$	$49.93 \pm 1.36$
STABLE	$83.25 \pm 0.86$	$70.99 \pm 1.19$	<b><math>81.46 \pm 0.78</math></b>	—	$70.78 \pm 0.27$

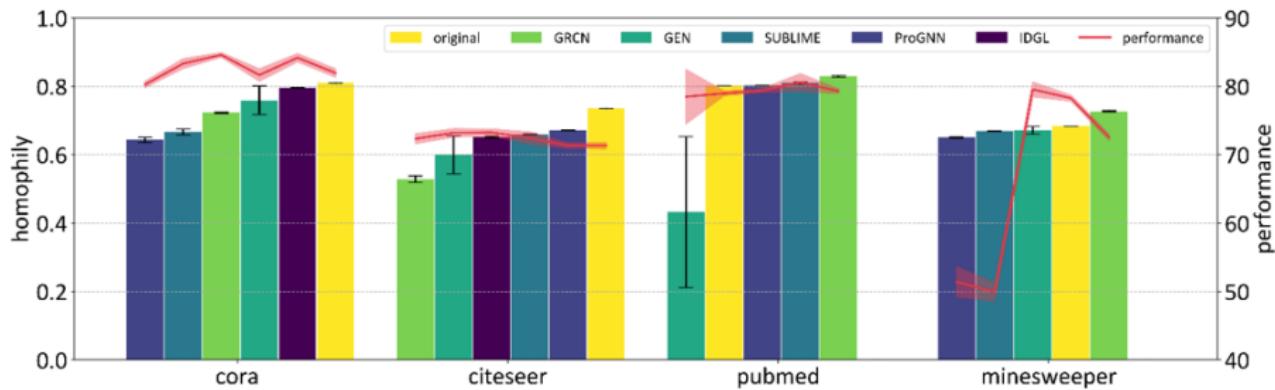
## 同质图上的分类表现

对于同质图，大部分 GSL 方法有提升作用；但是在类别不平衡的数据集上失效。

Model	BlogCatalog	Flickr	Amazon-ratings	Roman-empire	Wiki-cooc
GCN	$76.12 \pm 0.42$	$61.60 \pm 0.49$	$45.24 \pm 0.29$	<b><math>70.41 \pm 0.47</math></b>	<b><math>92.03 \pm 0.19</math></b>
LDS	$77.10 \pm 0.27$	—	—	—	—
ProGNN	$73.38 \pm 0.30$	$52.88 \pm 0.76$	—	$56.21 \pm 0.58$	$89.07 \pm 5.59$
IDGL	$89.68 \pm 0.24$	<b><math>86.03 \pm 0.25</math></b>	$45.87 \pm 0.58$	$47.10 \pm 0.65$	$90.18 \pm 0.27$
GRCN	$76.08 \pm 0.27$	$59.31 \pm 0.46$	<b><math>50.06 \pm 0.38</math></b>	$44.41 \pm 0.41$	$90.59 \pm 0.37$
GAug	$76.92 \pm 0.34$	$61.98 \pm 0.67$	<b><math>48.42 \pm 0.39</math></b>	$52.74 \pm 0.48$	<b><math>91.30 \pm 0.23</math></b>
SLAPS	<b><math>91.73 \pm 0.40</math></b>	$83.92 \pm 0.63$	$40.97 \pm 0.45$	<b><math>65.35 \pm 0.45</math></b>	$89.09 \pm 0.54$
WSGNN	<b><math>92.30 \pm 0.32</math></b>	<b><math>89.90 \pm 0.19</math></b>	$42.36 \pm 1.03$	$57.33 \pm 0.69$	$90.10 \pm 0.28$
Nodeformer	$44.53 \pm 22.62$	$67.14 \pm 6.77$	$41.33 \pm 1.25$	$56.54 \pm 3.73$	$54.83 \pm 4.43$
GEN	$90.48 \pm 0.99$	$84.84 \pm 0.81$	<b><math>49.17 \pm 0.68</math></b>	—	<b><math>91.15 \pm 0.49</math></b>
CoGSL	$83.96 \pm 0.54$	$75.10 \pm 0.47$	$40.82 \pm 0.13$	$46.52 \pm 0.48$	—
SeGSL	$75.03 \pm 0.28$	$60.59 \pm 0.54$	—	—	—
SUBLIME	<b><math>95.29 \pm 0.26</math></b>	<b><math>88.74 \pm 0.29</math></b>	$44.49 \pm 0.30$	<b><math>63.93 \pm 0.27</math></b>	$76.10 \pm 1.12$
STABLE	$71.84 \pm 0.56$	$51.36 \pm 1.24$	$48.36 \pm 0.21$	$41.00 \pm 1.18$	$80.46 \pm 2.44$

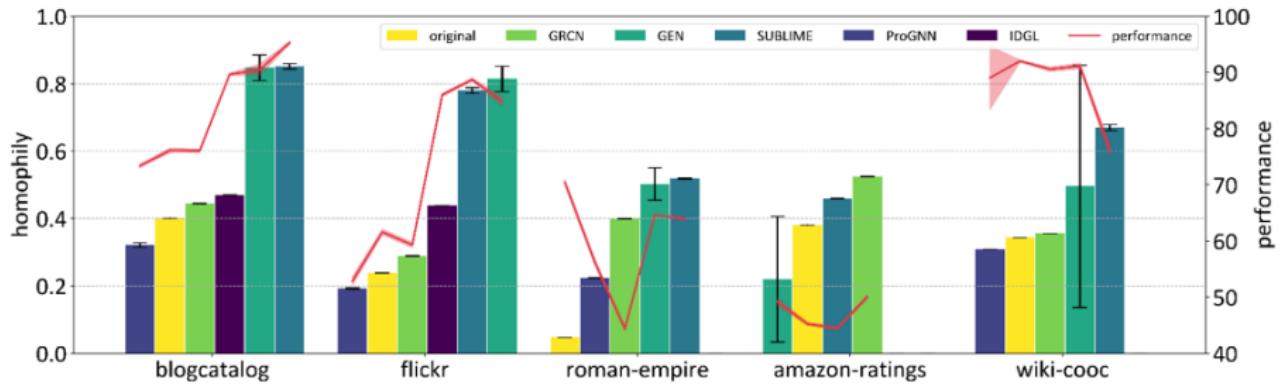
## 异质图上的分类表现

对于异质图，只有少数 GSL 方法在特定数据集上起作用。



同质图分类表现和学到结构的关系

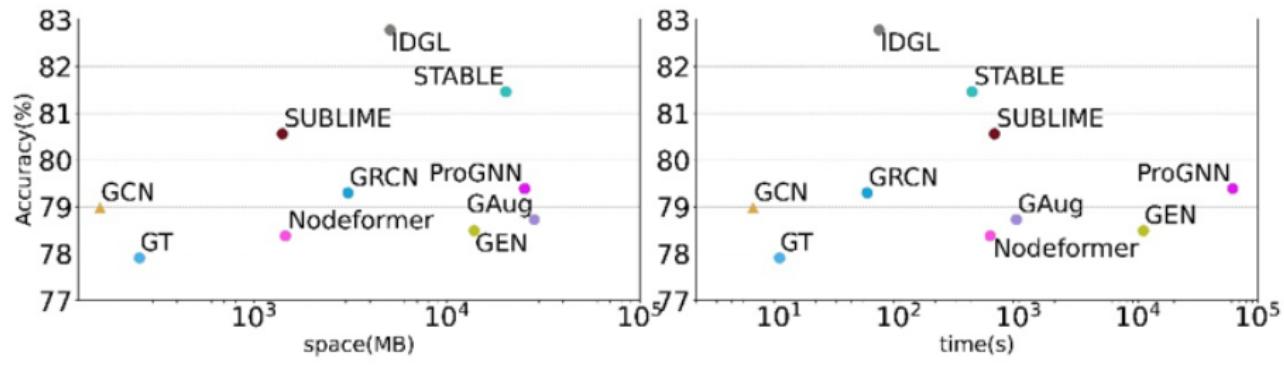
在异质图上，GSL 很容易提升结构的同质程度；在同质图上，学到结构的同质性很难进一步提升。



异质图分类表现和学到结构的关系

在大部分数据集上，分类表现和学到结构的同质性**无关**。

# OpenGSL：效率分析



PubMed 上的各 GSL 方法效率

大部分 GSL 方法的时间和空间消耗大。

# OpenGSL：未来研究

基于以上观察，有以下几个潜在的研究方向：

- ① 设计可以适应不同数据集（类别不平衡图、异质图）的 GSL 方法。
- ② 重新思考同质性的必要性，深入探索 GSL 的作用机制。
- ③ 发展可拓展的 GSL 方法以用于大规模数据集。



Paper :



Library :



# 课程总结

- ① 研究背景
- ② 基于度量学习的图结构学习
- ③ 基于优化的图结构学习
- ④ 基于生成模型的图结构学习
- ⑤ 无监督 GSL

# References I

-  Chen, Y., Wu, L., and Zaki, M. (2020). Iterative deep graph learning for graph neural networks: Better and robust node embeddings. *Advances in neural information processing systems*, 33:19314–19326.
-  Franceschi, L., Niepert, M., Pontil, M., and He, X. (2019). Learning discrete structures for graph neural networks. In *International conference on machine learning*, pages 1972–1982. PMLR.

## References II

-  Gu, M., Yang, G., Zhou, S., Ma, N., Chen, J., Tan, Q., Liu, M., and Bu, J. (2023).  
Homophily-enhanced structure learning for graph clustering.  
In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 577–586.
-  Jin, W., Ma, Y., Liu, X., Tang, X., Wang, S., and Tang, J. (2020).  
Graph structure learning for robust graph neural networks.  
In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 66–74.

# References III

-  Li, K., Liu, Y., Ao, X., Chi, J., Feng, J., Yang, H., and He, Q. (2022). Reliable representations make a stronger defender: Unsupervised structure refinement for robust gnn. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 925–935.
-  Li, R., Wang, S., Zhu, F., and Huang, J. (2018). Adaptive graph convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

# References IV

-  Liu, Y., Zheng, Y., Zhang, D., Chen, H., Peng, H., and Pan, S. (2022).  
Towards unsupervised deep graph structure learning.  
In *Proceedings of the ACM Web Conference 2022*, pages 1392–1403.
-  Wang, H., Fu, Y., Yu, T., Hu, L., Jiang, W., and Pu, S. (2023).  
Prose: Graph structure learning via progressive strategy.  
In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2337–2348.
-  Wang, R., Mou, S., Wang, X., Xiao, W., Ju, Q., Shi, C., and Xie, X. (2021).  
Graph structure estimation neural networks.  
In *Proceedings of the Web Conference 2021*, pages 342–353.

# References V

-  Zhou, Z., Zhou, S., Mao, B., Zhou, X., Chen, J., Tan, Q., Zha, D., Wang, C., Feng, Y., and Chen, C. (2023).  
OpenGSL: A comprehensive benchmark for graph structure learning.  
*Advances in neural information processing systems.*