



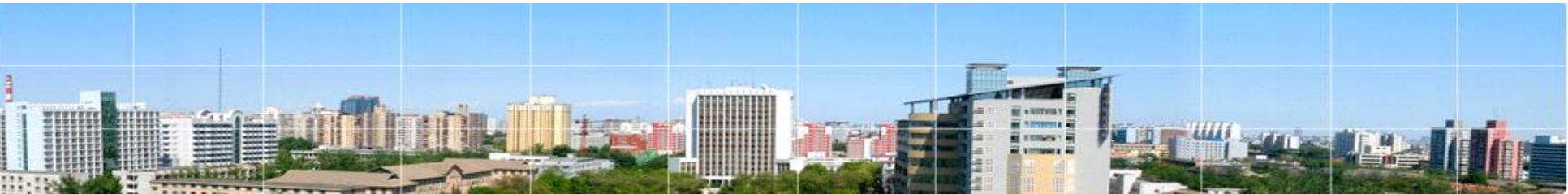
北京邮电大学
Beijing University of Posts and Telecommunications

第五章

图神经网络进阶（1）

石川 教授

数据科学与服务中心 计算机学院





- 课程介绍
- 讲课内容
 - 5.1 数据优化
 - 5.2 架构优化
 - 5.3 训练优化
- 重点难点
 - 掌握图数据优化技术的目的和技术
 - 理解图神经网络存在的问题，并掌握主要的解决方案
 - 掌握设计对比或生成式图自监督训练的方法



5.1 数据优化

- 5.1 数据优化
 - 5.1.1 图结构优化
 - ▣ 结构缩减 (Graph Reduction)
 - ▣ 结构增强 (Graph Augmentation)
 - ▣ 结构生成 (Graph Generation)
 - ▣ 结构学习 (Graph Structure Learning)
 - 5.1.2 图特征优化
 - ▣ 特征增强 (Feature Augmentation)
 - ▣ 特征选择 (Feature Selection)
 - ▣ 特征补全 (Feature Completion)
 - 5.1.3 图标签优化
 - ▣ 标签混合 (Label Mixup)
 - ▣ 伪标签 (Pseudo Labeling)
 - ▣ 主动学习 (Active Learning)



5.1.1 图结构优化概述

- 图结构：图数据中最核心的部分，描述节点之间关联信息
 - 结构缩减：减少图中的冗余节点和边，降低计算复杂度，提高可扩展性
 - 结构增强：通过低开销的方式丰富图结构信息，缓解过拟合
 - 结构生成：生成高质量且多样化的图样本
 - 结构学习：从图数据中发现和优化有价值的图结构，进一步提升图模型的表达能力

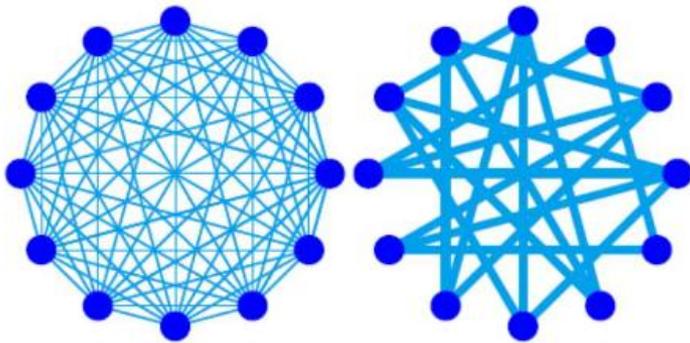


5.1.1 图结构优化-结构缩减

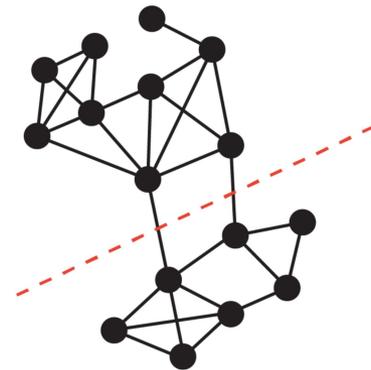
- 近年来，图数据集的规模和复杂性呈现出指数级增长。对于大规模网络，现有算法在可扩展性和效率方面面临着严峻挑战
- 结构缩减技术在保留关键信息的前提下，通过减少图数据集的规模，降低计算复杂度并提升模型的可扩展性。结构缩减方法可以分为三类：
 - 图稀疏化 (Graph Sparsification)
 - 图粗化 (Graph Coarsening)
 - 图压缩 (Graph Condensation)

5.1.1 图结构优化-结构缩减

- 图稀疏化通过移除原始图 G 中的部分边，生成一个简化图 $G_s = \{\mathcal{V}, \mathcal{E}_s\}$ ，其中 $\mathcal{E}_s \subseteq \mathcal{E}$ 。
- 通常简化图 G_s 需要保持原图 G 的某些关键性质，如：
 - 图切割权值 (Cut Value)
 - 图的切割 (Cut) 将图的节点分为两部分，可以表示为 $C = (\mathcal{V}', \mathcal{V} - \mathcal{V}')$
 - 图切割权值 $w(C)$ 是指跨越该切割的所有边的总权值



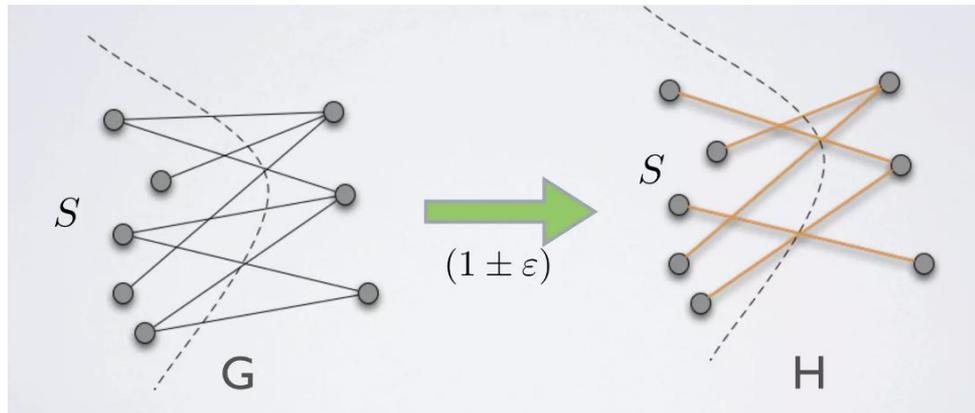
图稀疏化



图切割

5.1.1 图结构优化-结构缩减

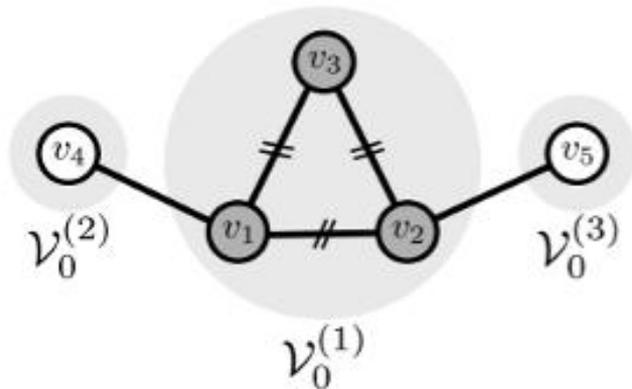
- 图稀疏化典型例子：割稀疏化 (Cut Sparsification)
 - 在保留图切割权值的前提下，减少图中的边，简化图结构
- ϵ -割稀疏化器 (ϵ -cut sparsifier)：一个简化图 G_S
 - 其图切割权值在所有切割的情况下与原始图 G 的图切割权值保持接近
$$(1 - \epsilon)w_G(C) \leq w_{G_S}(C) \leq (1 + \epsilon)w_G(C)$$
 - $\epsilon \in (0, 1)$ 表示允许的偏差
 - 割稀疏化广泛应用于解决图的连通性问题、最大流问题等





5.1.1 图结构优化-结构缩减

- **图粗化**：将一组紧密连接的节点合并成超级节点的方法来简化图结构。
 - 简化图 G_S 可以记为 $G_S = (\mathcal{V}_S, \mathcal{E}_S)$ ，其中节点集合 $|\mathcal{V}_S| < |\mathcal{V}|$
- **谱保持的图粗化方法**因为可以较好地保持原始图的重要结构信息受到了较多关注。受限谱相似性方法[1]是其中一个典型例子。



(a) Graph G



(b) Coarse graph G_c



5.1.1 图结构优化-结构缩减

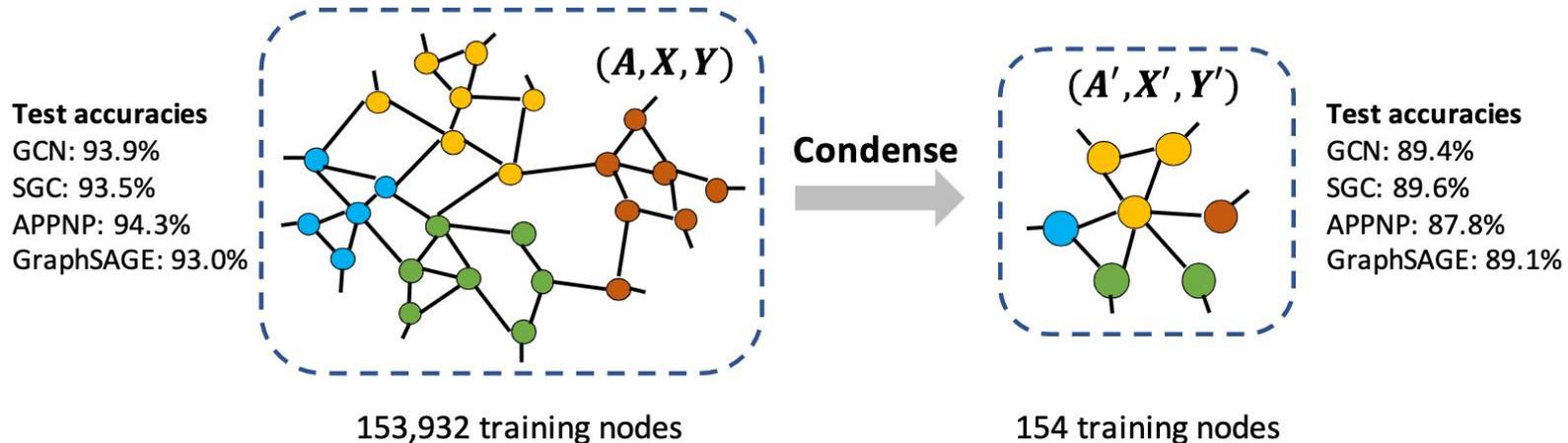
- 受限谱相似性[1] (Restricted Spectral Similarity) : 确保简化图能学习到原始图光谱特性的图粗化技术

$$(1 - \zeta_k) \lambda_k \leq \mathbf{u}_k^T \tilde{\mathbf{L}} \mathbf{u}_k \leq (1 + \zeta_k) \lambda_k$$

- λ_k 和 \mathbf{u}_k 分别表示原始图 G 的拉普拉斯矩阵 L 的第 k 个特征值和特征向量
- $\tilde{\mathbf{L}} \in \mathbb{R}^{N \times N}$ 是简化图 G_s 拉普拉斯矩阵的近似
- $\zeta_k \in (0, 1)$ 是误差容忍度
- 受限光谱相似性方法通过确保简化图的拉普拉斯矩阵特征值和特征向量在一定误差范围内接近原始图, 从而确保简化图能有效地学习和保留原始图的光谱特性

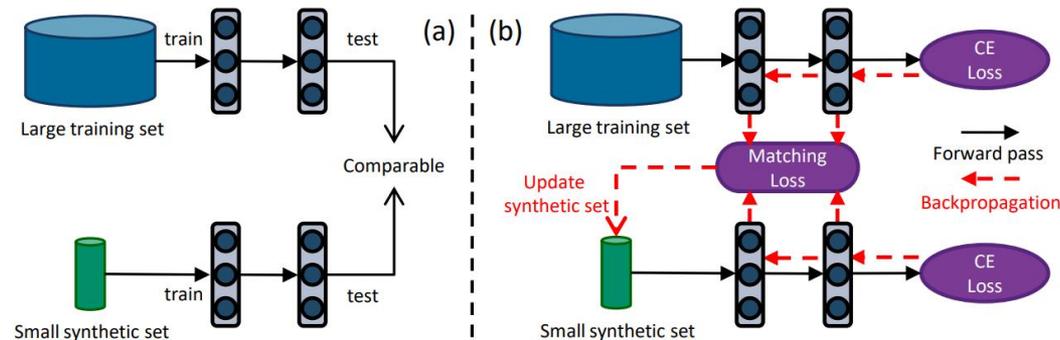
5.1.1 图结构优化-结构缩减

- 图压缩：合成一个新的更小的图来实现原始图的压缩
 - 图压缩的目的是生成一个包含较少节点和边的简化图 $G_s = \{\mathcal{V}_s, \mathcal{E}_s\}$ ，其中 $|\mathcal{V}_s| \ll |\mathcal{V}|$
- 在这种简化后的图上训练的模型能表现出与原始图上训练的模型相似的性能



5.1.1 图结构优化-结构缩减

- 经典的图压缩框架是利用**梯度匹配**的方法对齐原始图和简化图的梯度[1]



- 通过最小化梯度之间的距离来保证在简化图上训练模型时，模型的参数更新与原始图上训练时相似

$$Y_S = \nabla_{\theta} \mathcal{L}(f_{\theta}(A_S, X_S), Y_S),$$

$$Y = \nabla_{\theta} \mathcal{L}(f_{\theta}(A, X), Y),$$

$$Dis(Y_S, Y) = \frac{Y_S \cdot Y}{\|Y_S\| \|Y\|}$$

- f_{θ} : 图模型, θ : 模型参数, \mathcal{L} : 损失函数, Y_S : 在简化图 G_S 上计算得到的梯度, Y : 在原始图 G 上计算得到的梯度



5.1.1 图结构优化-结构增强

- 由于图数据的稀缺性和稀疏性，模型在训练过程中往往无法充分拟合数据的底层分布，容易陷入局部最优解
- 为了解决这一问题，结构增强方法以低成本的方式在不改变图的关键信息前提下对图的拓扑结构进行适当的扰动
- 结构增强方法主要分为：
 - 启发式增强方法
 - 丢弃、子图替换、图扩散
 - 自适应增强方法
 - 基于边的方法、基于子图的方法

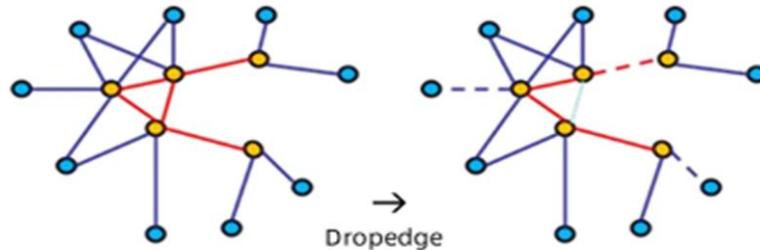


5.1.1 图结构优化-启发式增强方法

- 丢弃 (Dropping) : 启发式结构增强技术
 - 通过随机丢弃图中的边、节点和特征等改善模型的训练效果
- DropEdge[1]: 以固定概率随机丢弃图中的边
 - 在图的边集中选择 $p|\mathcal{E}|$ 条边进行丢弃, p 表示去边率, 经过丢弃后的邻接矩阵可以表示如下:

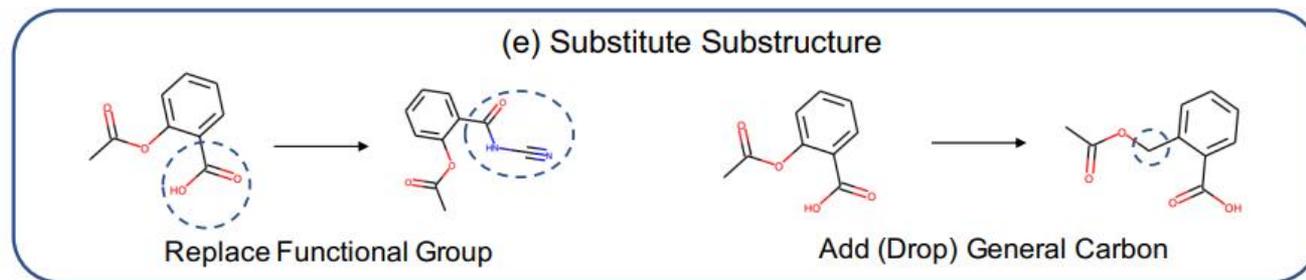
$$A_{\text{drop}} = A - A'$$

- A' : 随机选取的 $p|\mathcal{E}|$ 条边组成的邻接矩阵



5.1.1 图结构优化-启发式增强方法

- 子图替换:通过替换图中的特定子结构来实现图结构的增强
 - 弥补基于丢弃的方法仅关注到节点或边等最基本层次的信息,而忽视了更高层次的信息的不足
- MoCL[1]: 在生物医学领域用子图替换的方法进行图增强



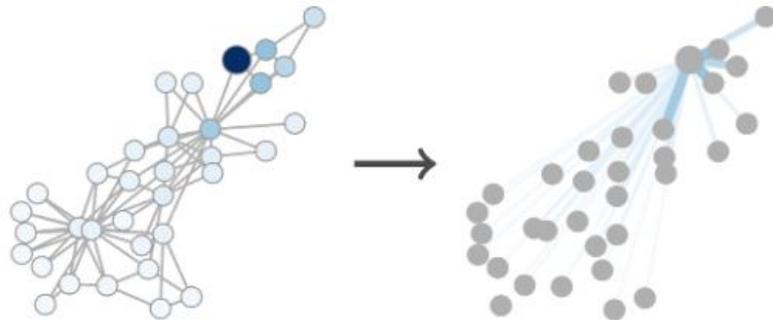
- 指出大多数丢弃的方法在增强过程中改变了分子图的语义,因此它通过注入领域知识来辅助增强过程
- 引入了生物电子等排体的概念来对分子中的特定有效子结构进行替换
 - ▣ 生物电子等排体是一类具有相似物理或化学性质的分子片段,替换后不会显著改变分子的整体性质

5.1.1 图结构优化-启发式增强方法

- **图扩散**：一种通过在不同距离上聚合邻居信息来丰富图的拓扑结构的方法

$$\tilde{A} = \sum_{k=0}^{\infty} \theta_k T^k$$

- \tilde{A} ：增强后的邻接矩阵； $T = AD^{-1}$ ：随机游走转移矩阵；
- θ_k ：权重系数，用于调整不同距离 k 的邻居节点在增强邻接矩阵中的影响程度
 - 例如在个性化PageRank中，随机游走转移矩阵系数为 $\theta_k = \alpha(1 - \alpha)^k$ ，随着步数 k 的增加，权重会迅速减小，这意味着越远的节点对最终PageRank值的贡献越小，从而强调了直接邻居的重要性





5.1.1 图结构优化-自适应增强方法

- 自适应增强方法在训练阶段基于具体任务优化自适应地进行结构增强
- 为了使可微损失函数指导边增强过程，**基于边的自适应增强**将边的权值视为可被优化的连续变量
- Pro-GNN[1]：引入特定的约束（如平滑性和稀疏性）来构建损失函数，从而生成用于优化图结构的梯度

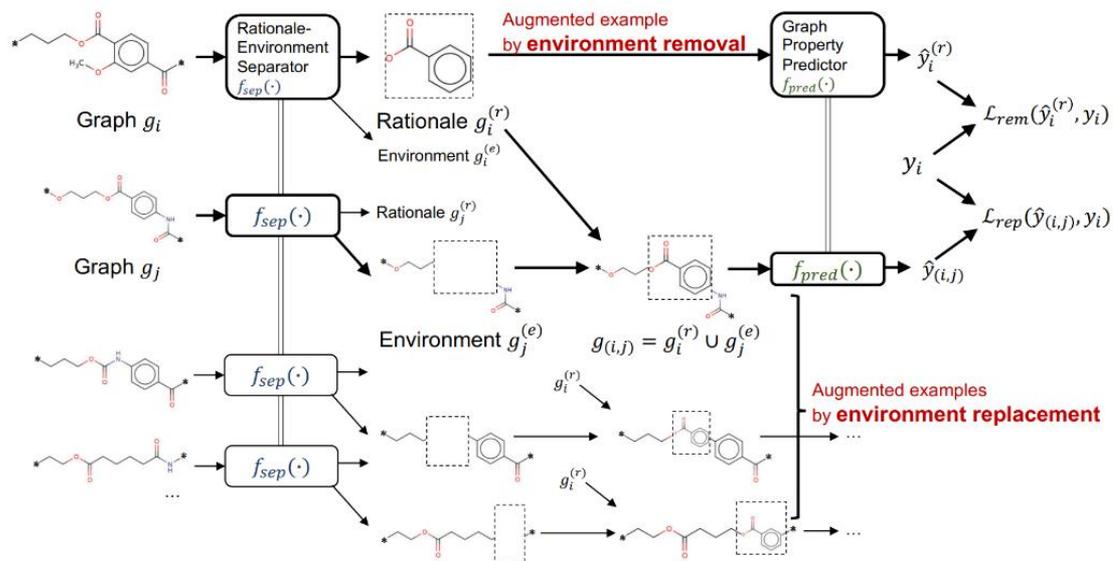
$$\mathcal{L} = \|\tilde{A} - A\|_F^2 + \eta\|\tilde{A}\|_1 + \beta\|\tilde{A}\|_* + \text{tr}(X^T \hat{L} X) + \gamma\mathcal{L}_{GNN}$$

- \hat{L} : 归一化的拉普拉斯矩阵
- $\|\tilde{A} - A\|_F^2$: 旨在让增强后的 \tilde{A} 接近原始邻接矩阵 A
- $\eta\|\tilde{A}\|_1$ 和 $\|\tilde{A}\|_*$: 确保图的稀疏性和低秩特性
- $\text{tr}(X^T \hat{L} X)$: 控制特征的平滑性
- γ 控制针对具体任务的神经网络损失函数 \mathcal{L}_{GNN} 的比重



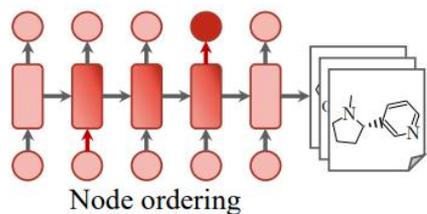
5.1.1 图结构优化-自适应增强方法

- 基于子图的自适应增强旨在找到最具代表性和信息量的子图
- GREA[1]: 一个典型的基于子图的增强过程
 - 利用核心子图与不同的环境子图的组合生成新的数据样本
 - 核心子图: 图结构中最能解释或支持模型预测的子结构
 - 环境子图: 在核心子图被识别并分离后, 图中剩余的部分
 - 通过一个MLP (Separator) 生成掩码向量指示哪些节点属于核心子图

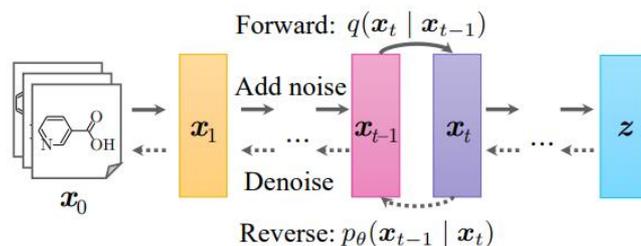


5.1.1 图结构优化-结构生成

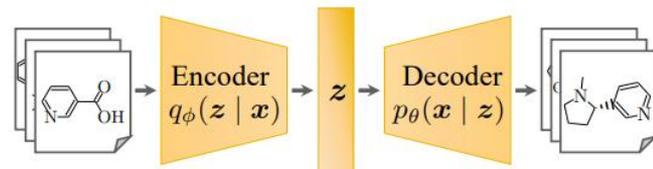
- 尽管图增强可以初步丰富拓扑信息，但不可避免地会引入噪声，从而影响模型性能。作为一种更高级的方法，结构生成旨在生成高质量和多样化的图样本
- 基于不同生成数据形式的代表性结构生成方法：
 - 节点序列 (Node Sequence) -- 自回归图生成
 - 邻接矩阵 (Adjacency Matrix) -- 扩散模型
 - 节点嵌入 (Node Embedding) -- 变分自编码器



(1) 自回归图生成



(2) 扩散模型



(3) 变分自编码器

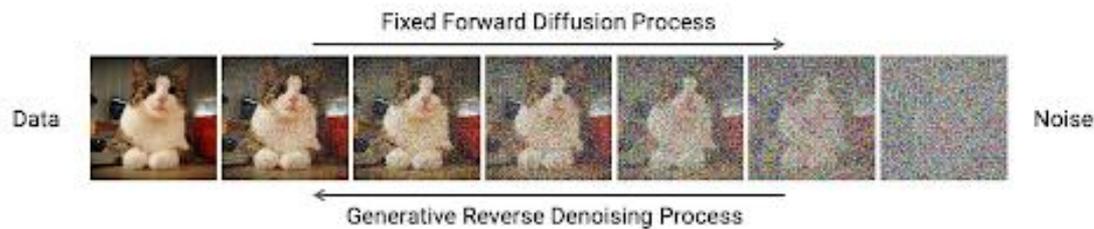


5.1.1 图结构优化-结构生成

- **节点序列生成：**将图简化为节点序列是结构生成的初步思路，催生了自回归图生成方法，旨在基于预先采样的节点顺序逐个生成图的节点
 - 巨大的生成空间 (n^2)；图的置换不变性带来的巨大输入空间 ($n!$)；复杂的依赖关系
- 为了解决这些挑战，GraphRNN[1]设计了三个步骤
 - 通过广度优先搜索（BFS）或深度优先搜索（DFS）遍历图来降低输入空间生成节点的序列；
 - 使用一个RNN根据当前已知节点预测下一个节点；
 - 使用另一个RNN模型预测新生成的节点与已知的节点之间的连接情况；

5.1.1 图结构优化-结构生成

- 邻接矩阵生成：一次性生成整个图的结构信息，尤其适用于小型图的生成
- EDGE[1]：基于离散扩散模型
 - 扩散模型最早用于连续空间数据的生成，如图像生成

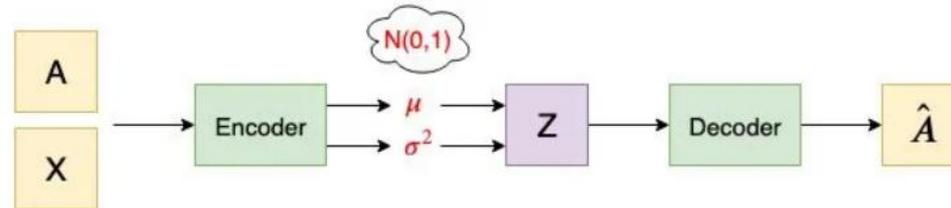


- EDGE方法将其扩展到离散空间，应用于图的生成
- 它的工作原理主要分为两个阶段
 - ▣ 第一阶段是正向扩散过程，逐步向图结构中加入噪声，直到数据接近于均匀噪声；
 - ▣ 第二阶段是反向去噪过程，从噪声数据开始，通过逐步去噪恢复出的图结构信息。另外，EDGE还为每个节点设计一个目标度向量，通过度匹配损失函数来确保生成的图结构符合预定的节点度分布



5.1.1 图结构优化-结构生成

- **节点嵌入生成**：生成图的邻接矩阵通常耗时较长，且无法扩展到大型图。通过生成节点嵌入进而得到结构
 - $A = H \cdot H^T$ ，通过生成一个较小的张量 $H \in \mathbb{R}^{N \times d}$ ，而不是大型邻接矩阵 $A \in \mathbb{R}^{N \times N}$ ，其中 $d \ll N$
- 变分图自编码器（VGAE） [1]



- 利用编码器学习潜在节点表示，VGAE 假设节点的潜在表示为服从高斯分布的随机变量，通过参数矩阵的变换输出其均值和方差，利用重参数技巧重新采样潜在表示
- 解码器从潜在表示中重构图的邻接矩阵（即预测节点之间是否存在边），这一步通常采用一个内积解码器，即通过计算节点潜在表示的内积来预测边的存在概率，在训练过程中，通过链接预测损失函数重建图结构



5.1.1 图结构优化-结构学习

- 现实世界中的图结构往往是噪声较多或不完整的，导致模型的学习效果下降，甚至产生错误的结果。这种图结构在社交网络、交通系统、生物网络等多个领域中普遍存在。图结构学习旨在从数据中发现和优化有价值的图结构，以增强图表示的学习。
- 根据是否考虑边的权重信息，现有的图结构学习方法大致可以分为两类：
 - 离散图结构学习 (Discrete Graph Structures Learning)
 - 加权图结构学习 (Weighted Graph Structures Learning)



5.1.1 图结构优化-结构学习

- **离散图结构学习**：将图结构视为随机变量，从概率邻接矩阵中进行采样
 - 核心在于利用概率模型捕捉图中节点之间的关系，并通过采样生成多样化的图结构，以反映节点连接关系的潜在不确定性和多样性
- **VGCNs [1]**：利用概率模型建模，并使用蒙特卡洛方法近似计算
 - 通过参数化的随机图模型纳入不确定的图信息
 - 蒙特卡洛方法是一类基于随机抽样的计算方法，常用于数值积分、概率分布的近似、优化问题等。它的核心思想是通过大量随机样本来估计一个期望值或者积分结果，这种方法特别适合高维复杂问题，尤其是解析解难以获得时

[1] Zhang Y, Pal S, Coates M, et al. Bayesian graph convolutional neural networks for semi-supervised classification[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 5829-5836.



5.1.1 图结构优化-结构学习

- VGCNs [1]通过已知的信息：（部分已知标签、特征和观察到的图结构）来推断节点或图的标签的后验概率：

$$p(Z|Y_{\mathcal{L}}, X, \mathcal{G}_{obs})$$

$$= \int p(Z|W, \mathcal{G}, X)p(W|Y_{\mathcal{L}}, X, \mathcal{G})p(\mathcal{G}|\lambda)p(\lambda|\mathcal{G}_{obs})dWd\mathcal{G}d\lambda$$

- $p(Z|W, \mathcal{G}, X)$ ：通过图卷积神经网络（GCN）来建模
- $p(W|Y_{\mathcal{L}}, X, \mathcal{G})$ ：在图神经网络上进行训练时，权重的更新过程
- $p(\mathcal{G}|\lambda)$ ：在给定参数 λ 的情况下，随机图的生成概率
- $p(\lambda|\mathcal{G}_{obs})$ ：在给定观察到的图的情况下，参数 λ 的后验概率，它表示我们从观察到的图中推断出随机图生成模型的参数



5.1.1 图结构优化-结构学习

- 蒙特卡洛近似：采用近似方法来计算上述积分

$$p(Z|Y_{\mathcal{L}}, X, \mathcal{G}_{\text{obs}}) \approx \frac{1}{V} \sum_v \frac{V}{N_G S} \sum_{i=1}^{N_G} \sum_{s=1}^S p(Z|W_{s,i,v}, \mathcal{G}_{i,v}, X)$$

- 蒙特卡罗近似通过采样的方法来估计标签 Z 的后验分布
 - 从图生成模型的参数分布中生成 V 个样本这些样本代表不同的模型参数
 - 对于每个模型参数，生成 N_G 个图样本 $\mathcal{G}_{i,v}$ 反映出在这些参数下可能的图结构
 - 对于每个图样本从贝叶斯图卷积神经网络中采样权重矩阵 $W_{s,i,v}$
 - 通过对所有样本的加权平均来估计积分



5.1.1 图结构优化-结构学习

- **加权图结构学习：**通过节点之间的相似性来得到图结构
 - 离散图结构学习的方法过于依赖已知的图结构和节点连接模式，缺乏对新节点的适应能力，在推理阶段面对未见节点时，往往无法有效进行归纳学习
 - 加权邻接矩阵能够编码更丰富的边的信息，这有利于后续的图表示学习
 - 学到的相似性度量函数以后可以应用于未见过的节点嵌入集来推断图结构，从而实现归纳图结构学习。
- **核心是设计节点对的相似性度量函数。常用的有：**
 - 基于注意力的方法
 - 基于核的方法



5.1.1 图结构优化-结构学习

- 一些代表性的基于注意力的相似性度量函数：
 - $S_{i,j} = \vec{v}_i^T \vec{v}_j$, 其中 $S \in \mathbb{R}^{n \times n}$, 计算任意一对节点嵌入之间的点积
 - $S_{i,j} = (\vec{v}_i \odot \vec{u})^T \vec{v}_j$, 其中 \odot 表示逐元素乘法, \vec{u} 是一个非负的可训练权重向量。通过引入可学习参数的修改点积, 表达能力更强
 - $S_{i,j} = \text{ReLU}(W\vec{v}_i)^T \text{ReLU}(W\vec{v}_j)$, 其中, $\text{ReLU}(x) = \max(0, x)$ 是一种修正线性单元。引入权重矩阵的度量方法进一步提高表达能力
 - $S_{i,j} = \text{softmax}((W_1\vec{v}_i)^T (W_2\vec{v}_j))$, 进一步对两个节点嵌入应用了不同的线性变换, 并引入了归一化操作



5.1.2 图特征优化概述

- 图特征是用于描述图中节点、边或整个图的属性和信息的特征表示。在本节中，将探讨如何从图特征的角度优化图数据。主要包括以下三部分内容：
 - 特征增强：通过扩展或修改节点特征，避免模型训练时的过拟合；
 - 特征选择：识别与标签高度相关的特征，避免维度灾难；
 - 特征补全：解决图数据中特征不完整的问题；



5.1.2 图特征优化-特征增强

- 特征增强通过对节点、边或图的原始特征进行扩展或修改，为图数据引入额外的、相关的信息，不仅能够缓解模型的过拟合，还可以提高模型的泛化性能，特别是在处理复杂图结构或稀疏图数据时表现尤为显著。
- 主要介绍两种特征增强方法：
 - 通用特征增强 (General Attribute Augmentation)
 - 位置编码 (Position Encoding)



5.1.2 图特征优化-特征增强

- **通用特征增强**一般用于特征预处理，旨在通过各种直观的调整来提高特征的效用和多样性
 - 对于节点本身有特征的情况
 - 特征损坏 (Feature Corruption) 通过向原始节点特征 x_i 中加入可控噪声 r_i 来产生增广数据，可以表示为 $\tilde{x}_i = x_i + r_i$;
 - 特征洗牌 (Feature Shuffling) 通过随机切换特征矩阵中的行和列改变原始节点特征数据的上下文信息，产生增广数据，可以形式化表示为 $\tilde{X} = P_r X P_c$ ，其中 P_r 和 P_c 分别是行排列矩阵和列排列矩阵，它们在每一行和每一列中恰好有一个元素为1, 其他位置均为0;
 - 特征掩码 (Feature Masking) 的核心操作是将节点特征矩阵中的一部分原始置零，通过与一个掩码矩阵 M 逐元素相乘来实现： $\tilde{X} = X \odot M$ ，如果节点 i 的特征向量的第 j 个元素被丢弃，则 $M_{i,j} = 0$;
 - 对于节点本身没有特征的情况
 - 通过将图自身的结构信息编码成节点特征，比如直接使用节点的度做为节点特征。更加复杂的方法包括利用随机游走算法来捕获结构信息，并仿照 word2vec 技术来生成节点特征等



5.1.2 图特征优化-特征增强

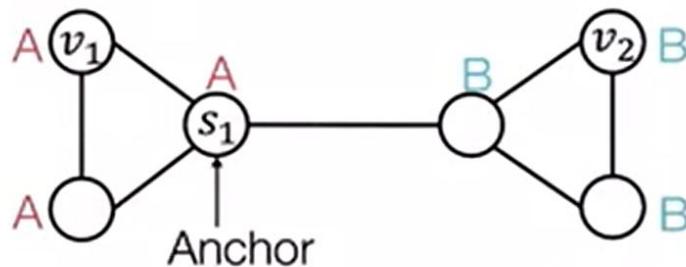
- **位置编码**：通过引入位置信息来增强节点特征
 - 绝对位置编码 (Absolute Position Encoding, APE)
 - 相对位置编码 (Relative Position Encoding, RPE)
- **绝对位置编码**目标是为每个节点分配一个位置表示，以指示其在整个图中的唯一位置。一种流行的方法是利用图拉普拉斯矩阵的特征向量，具体来说，通过对拉普拉斯矩阵进行特征分解，可以得到：

$$\Delta = I - D^{-1/2}AD^{-1/2} = U\Lambda U^T$$

- 其中， U 是特征向量矩阵， Λ 是特征值矩阵，一般来说，选择使用节点 k 个最小的非平凡特征向量，组成 $N \times k$ 的矩阵，每一行作为一个节点的位置编码

5.1.2 图特征优化-特征增强

- **相对位置编码**：通过将节点之间的距离作为位置编码来捕捉它们之间的关系信息。
 - 通过随机选择一个节点作为锚点节点，通过目标节点与锚点节点之间的相对距离来定位目标节点
 - 例如，使用图神经网络往往无法区分下图中的节点 v_1 和 v_2 ，因为两个节点虽然位置不同，但是所处位置图结构相同，而通过选择 s_1 作为锚点节点，把它作为坐标，便可以通过相对位置来区分 v_1 和 v_2



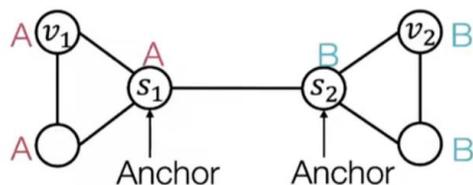
Relative Distances

	s_1
v_1	1
v_2	2

(a) 单个锚点

5.1.2 图特征优化-特征增强

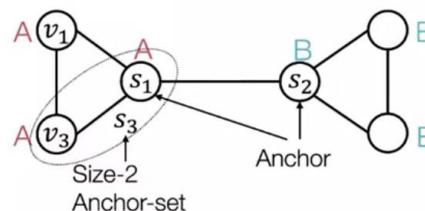
- 为了更精确捕捉节点位置信息，可以选择更多的锚点节点
- 锚点集[1]：通过锚点 s_1 和 s_2 无法区分节点 v_1 和 v_3 ，而通过将 s_1 和 v_3 组合为锚点集，把目标节点到锚点集中所有节点的最近距离作为相对距离，可以更为精确地表示节点的位置信息



Relative Distances

	s_1	s_2
v_1	1	2
v_2	2	1

(b) 多个锚点



Relative Distances

	s_1	s_2	s_3
v_1	1	2	1
v_3	1	2	0

(c) 锚点集

[1] You J, Ying R, Leskovec J. Position-aware graph neural networks[C]//International conference on machine learning. PMLR, 2019: 7134-7143.



5.1.2 图特征优化-特征选择

- 当机器学习算法中使用的数据特征维度过高，就会在高维特征空间中呈现出稀疏性，导致需要指数级增长的数据量去维持泛化性，使得模型训练的成本显著增加，这种现象被称为**维度灾难**。
- **特征选择**是一种缓解维度灾难的方法，旨在识别与标签高度相关的特征，并在模型训练过程中优先考虑这些特征。
- 在图学习中，常用的特征选择方法可以根据其与下游任务的关系分为两类：
 - 任务无关的特征选择 (Task-independent FS)
 - 任务特定的特征选择 (Task-specific FS)



5.1.2 图特征优化-特征选择

- **任务无关的特征选择**专注于选择能够适用于任何图神经网络模型或下游任务的特征。
- 主要围绕引入正则化目标进行特征选择。如：AsGNNS[1]将正则化方法引入GCN，通过 $\ell_{2,1}$ 范数来对每一个图卷积层的参数进行约束，优化目标可以表示为：

$$\min L_{gcn}(A, X; \Theta) + \sum_{k=1}^K \lambda_k \|\Theta^{(k)}\|_{2,1}$$

- 其中， $\|\Theta^{(k)}\|_{2,1} = \sum_{i=1}^N \sqrt{\sum_{j=1}^M |\theta_{ij}^{(k)}|^2}$ 表示参数矩阵每一行的二范数总和，这种约束确保了学习的参数矩阵 $\Theta^{(k)}$ 具备行稀疏性，从而能够自然地进行特征选择。

[1]Jiang B, Wang B, Luo B. Sparse norm regularized attribute selection for graph neural networks[J]. Pattern Recognition, 2023, 137: 109265.

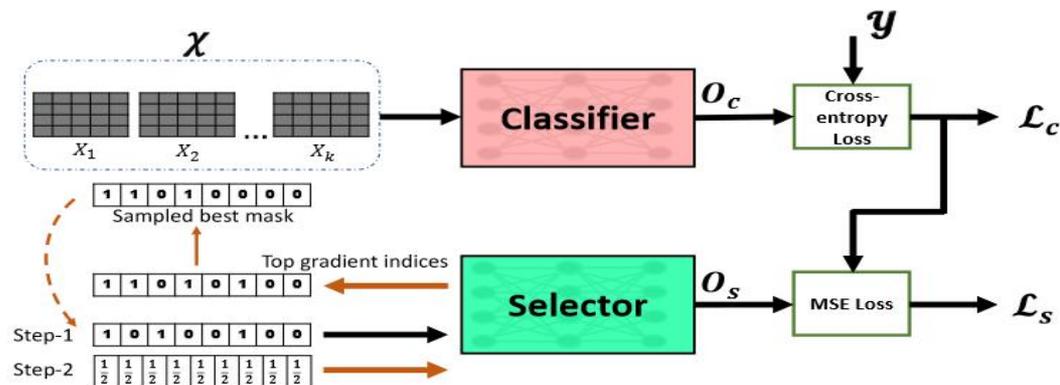


5.1.2 图特征优化-特征选择

- 任务特定的特征选择：考虑GNN具体任务
- Dual-Net GNN[1]:
 - 选择器模型：为分类器提供最佳输入子集，实现最佳性能
 - 两层的MLP: $f_s(\phi, m)$
 - ϕ : 选择器的参数，输出是一个标量值，表示输入掩码向量的预测性能
 - 分类器模型：基于输入节点特征的子集训练，预测节点标签
 - 两层的MLP: $f_c(\theta; \mathcal{X}, m)$
 - θ : 网络的参数， \mathcal{X} : 节点特征矩阵， m : 指示节点特征矩阵子集的掩码向量
 - 损失函数
 - 分类器：交叉熵损失函数
 - 选择器：均方误差损失函数，衡量选择器的预测性能与分类器真实的分类性能之间的差异

5.1.2 图特征优化-特征选择

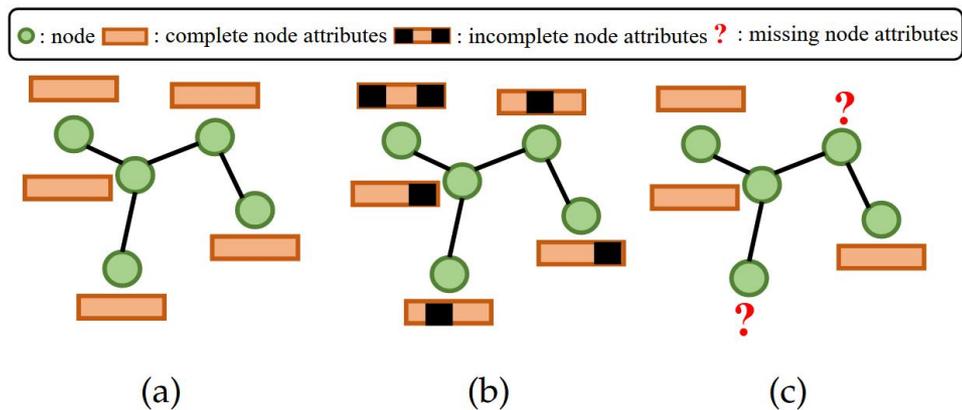
- 第一阶段：训练分类器和选择器
 - 分类器对不同的掩码组合产生稳定的不同损失
 - 选择器能够学会将这些掩码映射到分类器的分类性能
- 第二阶段：利用训练好的选择器为分类器生成可能的最优掩码



- 输入中有大梯度的成分对输出的贡献更大
- Dual-Net GNN识别具有TopP梯度的索引，并固定采样若干组合，计算分类器的验证损失
- 选择验证损失最小的掩码，在此掩码上对分类器和选择器进行一次训练
- 第三阶段：选择验证损失最小的输入掩码，仅对分类器继续进行训练，直到收敛

5.1.2 图特征优化-特征补全

- 大多数图神经网络假设图中的节点特征是完整的，但这一假设在实际应用中往往并不能成立
 - 数据收集过程中出现的机器或人为错误
 - 完全收集数据集在实际中成本很高
 - 许多用户由于隐私保护不愿提供完整的个人信息



- 特征补全：填补图中缺失的节点特征
 - 基于同质图的特征补全
 - 基于异质图的特征补全

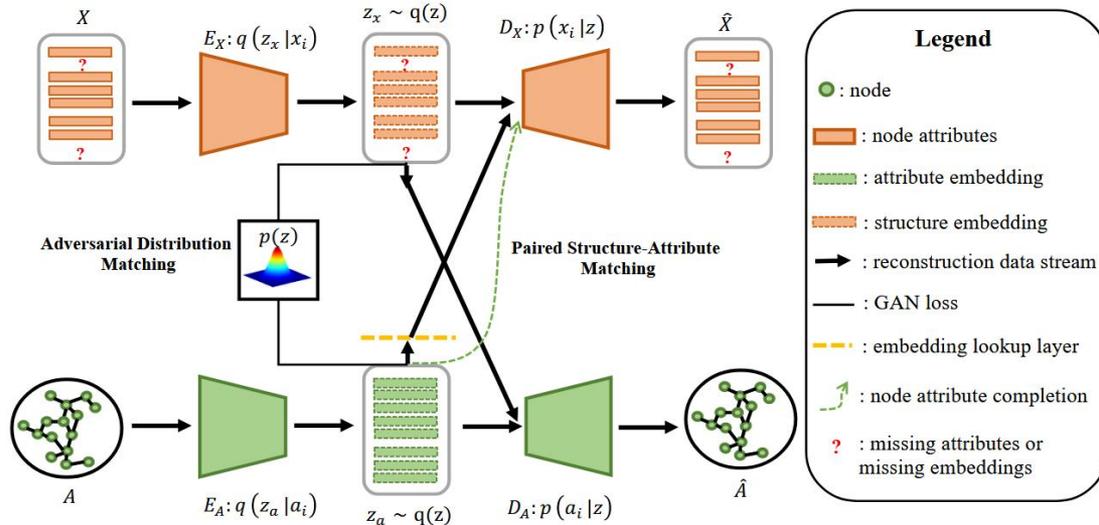


5.1.2 图特征优化-特征补全

- 大多数的同质图神经网络往往假设图具有完整的特征信息，没有针对特征缺失图进行设计，因此无法提供令人满意的学习效果。
 - 结构-属性转换器（SAT）[1]对图提出了共享潜在空间的假设，并开发了一种基于分布匹配的图神经网络，用于处理特征缺失的图。SAT采用了一种结构与属性解耦的方案，能够进行同质图节点特征补全任务
 - SAT认为学习属性缺失图的一种可能方法是以解耦的方式输入结构和属性，同时允许结构和属性的联合分布建模

5.1.2 图特征优化-特征补全

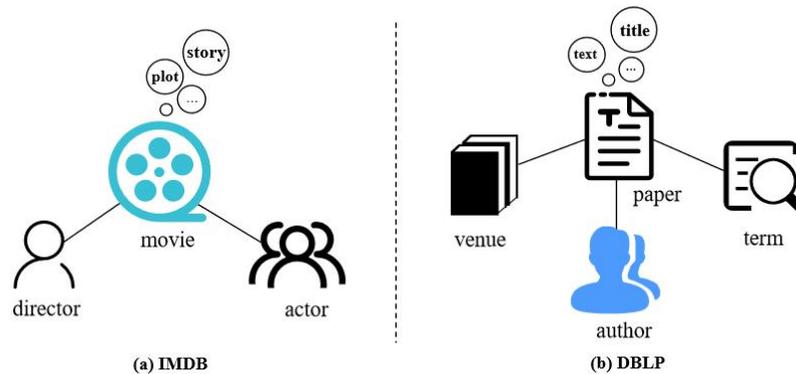
- SAT 的架构分为三步：



- SAT 使用结构编码器 (EA) 和属性编码器 (EX) 将输入的数据映射到一个共享的潜在空间中
- 在潜在空间中, SAT 通过对抗性分布匹配方法, 将编码得到的潜在表示与真实的先验分布进行对齐
- SAT通过结构解码器 (DA) 和属性解码器 (DX) 重构图结构和节点属性

5.1.2 图特征优化-特征补全

- 异质图特征缺失：异质图包含多种类型节点，各种节点的特征并不在同一特征空间；另外异质图存在特定类型节点整体特征缺失问题
 - 在 IMDB 数据集中，只有电影节点具有原始属性，而在 DBLP 数据集中，只有论文节点具有原始属性，其他类型的节点则没有属性

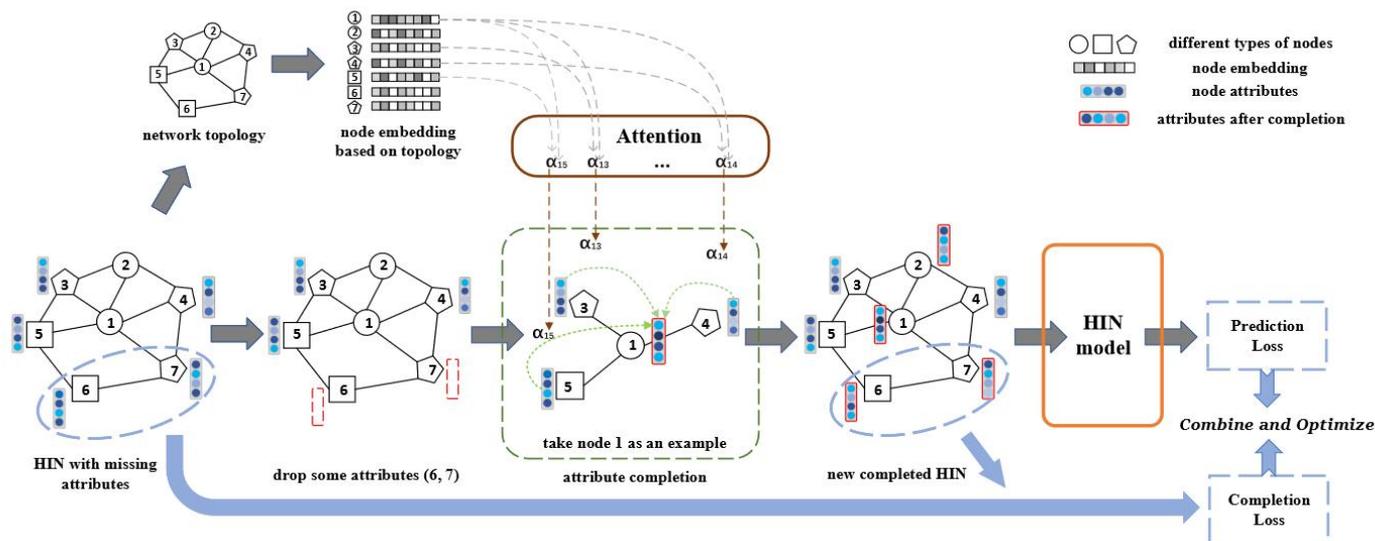


- HGNN-AC[1]：第一个异质图特征补全解决方案
 - 充分利用异质图的拓扑结构，结合不同类型节点和边之间的关联信息进行节点属性的恢复



5.1.2 图特征优化-特征补全

- HGNN-AC对于给定的缺失特征的异质图：
 - 使用现有的异质图嵌入方法从网络拓扑 A 计算节点嵌入
 - 随机丢弃一些属性并进行属性补全，在属性补全后，获得一个所有节点都有属性的新异质图
 - 属性补全过程本质上是直接相连邻居属性的加权聚合，权重由节点嵌入推导出的注意力机制决定。
 - 将新异质图输入到任意的异质信息网络（HIN）模型中
 - 损失函数
 - 补全损失：丢弃的属性与重构的属性之间存在一个补全损失
 - 模型预测损失





5.1.2 图标签优化概述

- 图标签：用于描述图中节点或边的特征或类别
- 图标签优化旨在增强现有图标签数据，并缓解与过拟合和噪声标签相关的问题
- 这些技术包括
 - 标签混合：通过混合来自不同实例的标签来创建新的训练示例
 - 伪标签：通过训练好的模型扩展标签集，为未标记的节点分配标签
 - 主动学习方法：在标记成本有限的情况下，从数据集中选择最有效的数据进行标记，以获得最佳的模型性能

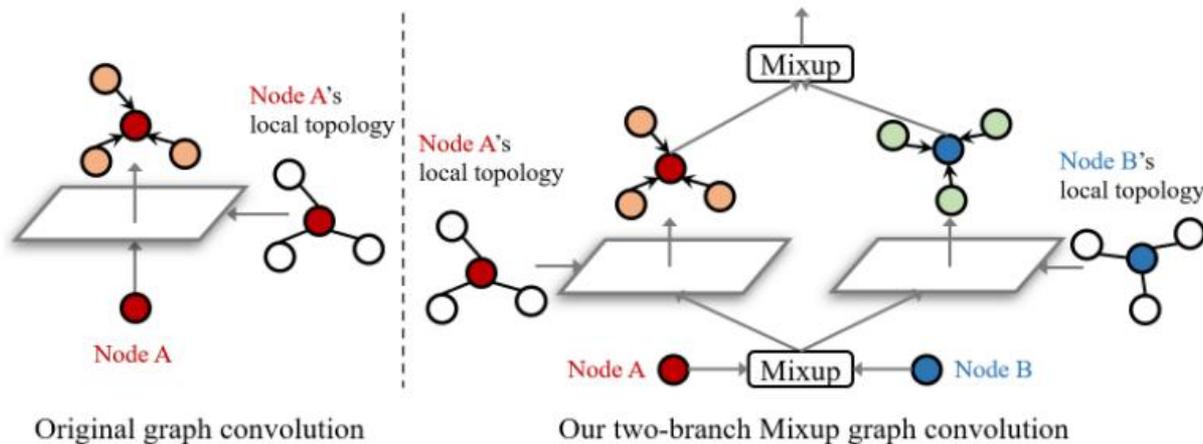


5.1.2 图标签优化-标签混合

- **标签混合**：基于标签的增强学习技术，通过将两个不同实例及其关联标签结合为一个新实例，从而扩展训练数据集
 - 创造了新的训练样本，增加了数据的多样性，对于提高模型的泛化能力至关重要
 - 模型能够更好地应对各种输入，提升其在未见数据上的表现，有效减少过拟合现象
- **标签混合主要包括**：
 - 节点级混合 (Node-level Mixup)
 - 图级混合 (Graph-level Mixup)

5.1.2 图标签优化-标签混合

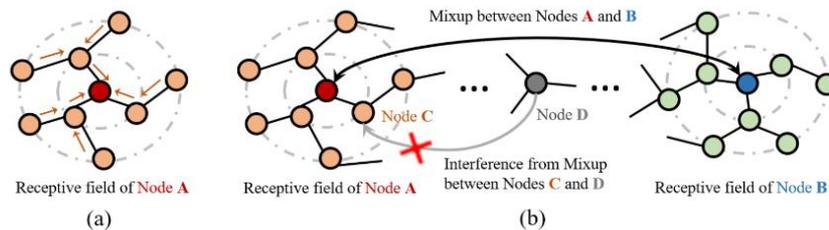
- 节点级混合：
 - 混合不同节点的标签，特别是针对节点分类任务
 - 图的拓扑结构复杂且不规则，进行节点混合时需要同时考虑节点的拓扑结构
 - 为了解决这个问题，经典的框架[1]提出了双分支图卷积方法：
 - 对节点特征和标签使用一样的过程进行混合
 - 左侧是传统的图卷积
 - 右侧是双分支图卷积：① 混合一对节点的属性；② 对于每个节点进行单独的图卷积 ③ 在传递到下一层之前进一步混合



[1] Y. Wang, W. Wang, Y. Liang, Y. Cai, and B. Hooi, "Mixup for node and graph classification," in Proceedings of the Web Conference 2021, 2021, pp. 3663–3674.

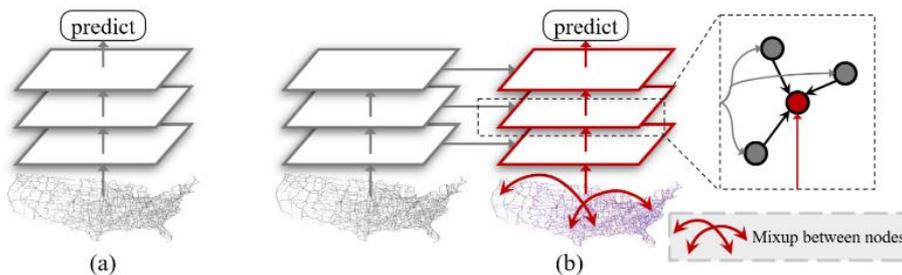
5.1.2 图标签优化-标签混合

- 除了如何考虑节点拓扑，还存在干扰问题：
 - 传统的混合方法：消息传递和节点混合同时进行，不同节点对的感受野有重叠，信息传递相互干扰，导致混合特征不准确



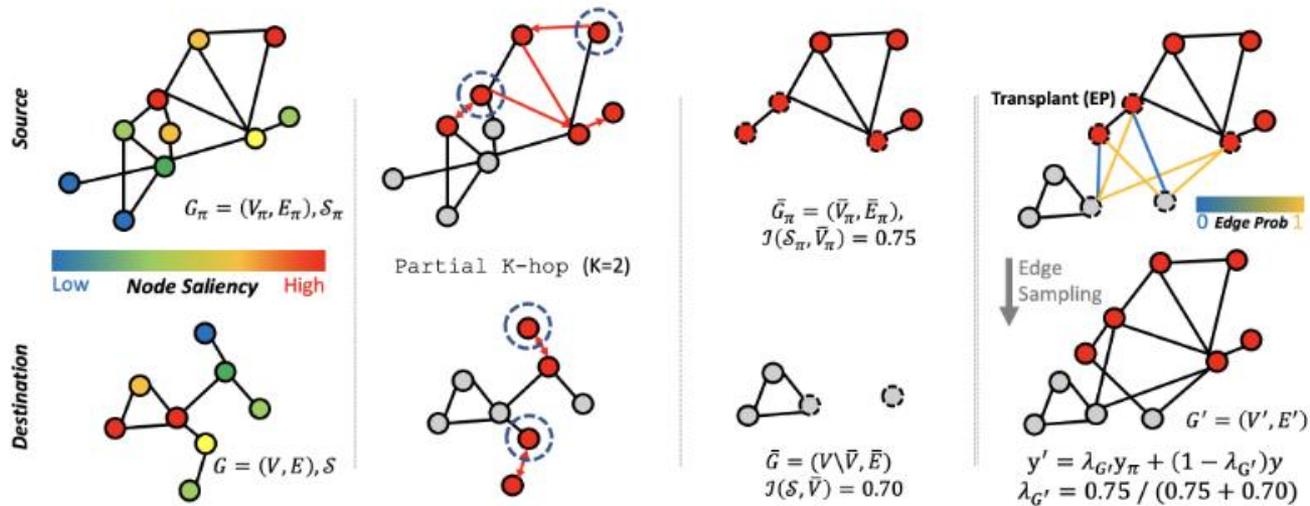
- 当我们混合节点A和节点B的特征时，如果同时对节点 C（橙色）和节点 D（灰色）进行混合，就会出现干扰的问题

- 双阶段混合方法：
 - 第一阶段：按照现有的 GNN 方式执行前向传播，但不进行混合
 - 第二阶段：使用双分支图卷积方法对配对的节点进行卷积操作，不直接使用邻居节点当前的特征，而是使用第一阶段得到的节点邻居表示。



5.1.2 图标签优化-标签混合

- **图级混合**: 直接混合图的标签, 往往涉及子图选择、图的移植、边链接预测、标签混合等多个步骤。
 - Transplant [1]: 一个图级混合方法



- 计算源图 G_π 和目标图 G 的节点重要性向量 S_π 和 S
- 从每个图中提取以重要节点 \hat{V}_π 和随机节点 \hat{V} 为锚的K-hop子图 \bar{G}_π 和 \bar{G}
- 计算源图子图和目标图子图的子图重要性
- 将源子图 \bar{G}_π 移植到剩余的目标子图 \bar{G} , 并根据子图重要性得到混合图的标签 y'



5.1.2 图标签优化-标签混合

- 重点介绍如何得到源图和目标图的节点重要性向量 \mathcal{S}_π 和 \mathcal{S} ，如何根据 \mathcal{S}_π 和 \mathcal{S} 定义子图重要性并自适应确定混合图的标签 y'
- 计算图的节点重要性向量：一个图的所有节点重要性构成向量 $\mathcal{S}_G = [s_1, \dots, s_{|V|}]^T$
 - 与之前提到的Dual-Net方法类似，Graph Transplant致力于简单快速地根据节点特征和梯度找到节点重要性，对于每个节点 v ，重要性 s_v 被定义为分类任务的损失函数关于该节点在第 l 层节点特征矩阵梯度值的 ℓ_2 范数，即

$$s_v = \left\| \left[\frac{\partial \mathcal{L}}{\partial X^{(l)}} \right]_{v,:} \right\|_2$$



5.1.2 图标签优化-标签混合

- 根据节点重要性自适应计算混合标签：
 - 定义子图重要性（源图子图 \bar{G}_π 以及目标图剩余部分 \bar{G} 的重要性）：子图中节点在原始完整图中的总重要性： $I(S, V) = \frac{\sum_{v \in V} s_v}{\|S\|_1}$
 - S : 节点重要性向量, V : 子图中的节点集合, s_v : 节点 v 的重要性
 - 自适应得到混合图的标签: $\lambda_{G'} y_\pi + (1 - \lambda_{G'}) y$, 其中 $\lambda_{G'} = \frac{I(S_{G\pi}, \bar{V}_\pi)}{I(S_{G\pi}, \bar{V}_\pi) + I(S_G, V \setminus \bar{V})}$
 - 其中 \bar{V}_π 是源图子图的节点集合, $V \setminus \bar{V}$ 是目标图剩余部分的节点集合



5.1.2 图标签优化-伪标签

- **伪标签**：将训练好的模型对未标记数据的预测结果作为“伪标签”，并与真实标记数据一起用于模型的后续训练
 - 图神经网络需要大量的标记数据用于模型的训练与验证，而标记数据往往非常昂贵
 - 用于在有大量未标记数据但只有少量标记数据的场景下训练深度学习模型



5.1.2 图标签优化-伪标签

- 通过协同训练和自训练方法来扩展训练集[1]
 - 协同训练：利用随机游走模型探索图的全局结构，弥补GCN仅能处理局部结构的不足

Algorithm 1 Expand the Label Set via ParWalks

- 1: $P := (L + \alpha\Lambda)^{-1}$
 - 2: **for** each class k **do**
 - 3: $p := \sum_{j \in \mathcal{S}_k} P_{:,j}$
 - 4: Find the top t vertices in p
 - 5: Add them to the training set with label k
 - 6: **end for**
-

- 自训练：

Algorithm 2 Expand the Label Set via Self-Training

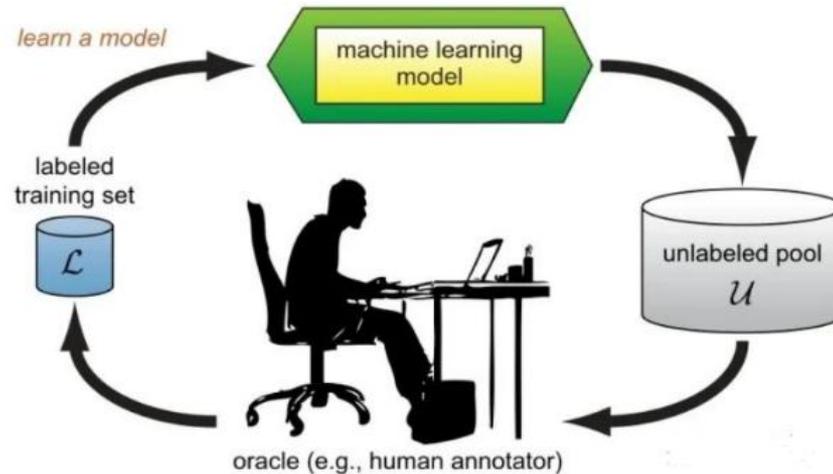
- 1: $Z := GCN(X) \in \mathbb{R}^{n \times F}$, the output of GCN
 - 2: **for** each class k **do**
 - 3: Find the top t vertices in $Z_{:,k}$
 - 4: Add them to the training set with label k
 - 5: **end for**
-

- 1: $P_{i,j}$ 表明顶点 i 与 j 归属同一类别的可能性
- 3: \mathcal{S}_k 是类别 k 数据集， $p \in \mathbb{R}^n$ 是置信向量，记录每一个节点属于类别 k 的置信度
- 4: 找到置信度top- t 顶点，加入类别 k 的训练集

- 1: 训练一个GCN模型，得到节点的预测结果 Z
- 3: 对于每个类别 k ，通过比较Softmax得分，选择最自信的预测实例
- 4: 将这些最自信的预测节点添加到对应类别的标签集中，并赋予它们类别 k 的标签

5.1.2 图标签优化-主动学习

- **主动学习**：在标记成本有限的情况下，从数据集中选择最有效的数据标记，以获得最佳的模型性能。
 - **主动图学习**：对于一个图 $G = (V, E)$ ，通过设计主动学习查询策略，从未标注节点集合 \mathcal{U} 中选择 B （标注预算）个节点给用于标注查询节点的预言机（oracle）标注，并将其加入标注节点集合 \mathcal{L} 中用于图嵌入训练



- 图上的主动学习往往假设节点之间存在关联，考虑节点之间的相互作用。这一思想避免了选择的节点相似且聚集在一起，导致信息的重复



5.1.2 图标签优化-主动学习

- FeatProp[1]: 一个基于聚类的通用图数据主动学习框架

Algorithm 1 Active Learning with Distance-based Clustering

Require: Node representation matrix X , graph structure matrix G and budget b

- 1: Compute a distance function $d_{X,G}(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ # for FeatProp: use Eqn. (7)
- 2: Perform clustering using $d_{X,G}$ with b centers # for FeatProp: use K-Medoids
- 3: Select s to be the centers
- 4: Obtain labels for $v \in s$ and train model \mathcal{M}

Ensure: Model \mathcal{M}

- 该框架主要包括两个步骤:

- **步骤一:** 使用节点特征表示 X 和图结构 G 计算距离矩阵 $d_{X,G}$, 对应节点特

征经过传播后的成对节点距离, 使用 L2 范数来表示:

$$d_{X,G}(v_i, v_j) = \|(S^K X)_i - (S^K X)_j\|_2,$$

- 其中 K 是传播总层数, 在传播过程中考虑了图结构
- **步骤二:** 在该距离矩阵上应用具有 b 个中心的聚类, 并从每个簇中选择离聚类中心最近的节点进行标注
- FeatProp应用 K-Medoids 聚类。K-Medoids 问题与 K-Means 相似, 但所选的中心必须是数据集中的真实样本节点。这对主动学习至关重要, 因为无法对 K-Means 产生的虚假聚类中心进行标记



5.2 架构优化

- 5.2 架构优化
 - 5.2.1 信息传递优化
 - 过平滑 (over-smoothing)
 - 过压缩 (over-squashing)
 - 表达能力受限 (limited expressive power)
 - 5.2.2 图采样优化
 - 重要性采样 (Importance Sampling)
 - 自适应采样 (Adaptive Sampling)
 - 5.2.3 图池化优化
 - 平铺池化 (Flat Pooling)
 - 层级池化 (Hierarchical Pooling)



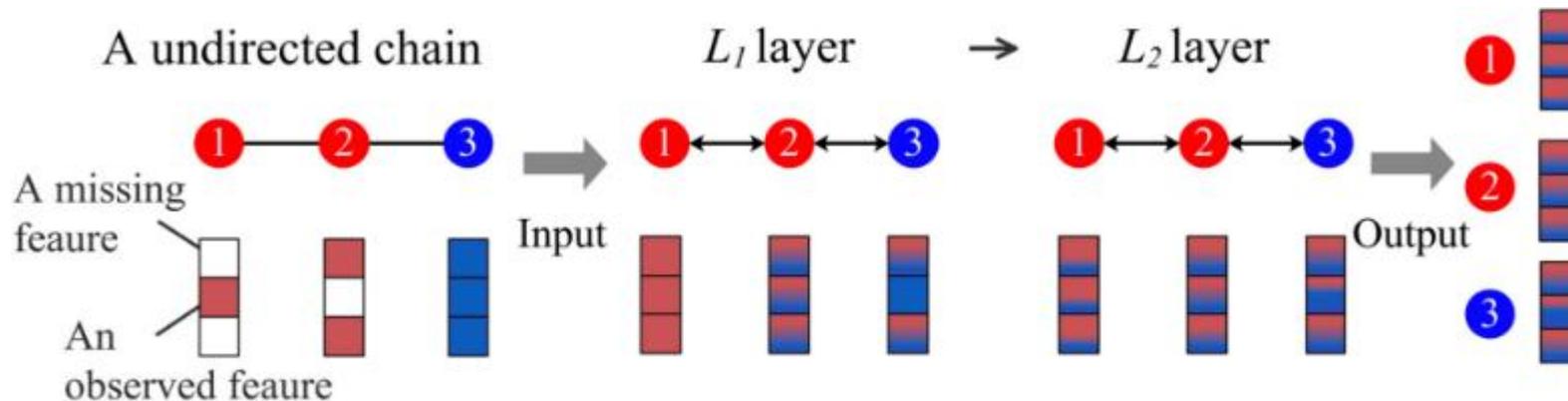
5.2.1 信息传递优化概述

- 基于消息传递的图神经网络在各种分类任务中表现出色，主要归功于其能够通过消息传递机制有效地聚合邻居节点的信息，从而捕捉复杂的图结构与节点特征之间的关系。然而，这种信息传递机制仍然存在一些局限性：
 - 过平滑 (over-smoothing)：模型的层数增加时，节点表示变得越来越相似，最终导致不同类别节点的特征无法区分
 - 过压缩 (over-squashing)：随着网络层数的增加，信息传递量呈指数级增长，但最终传递到目标节点的却是有限且固定大小的信息
 - 表达能力受限 (limited expressive power)：能够处理和区分图结构的能力上限为1-WL test算法。处理复杂图结构时，表达能力有限



5.2.1 信息传递优化-过平滑

- **定义 5.1 (过平滑):** 假设 G 是一个无向的连通图, $X^n \in \mathbb{R}^{v \times m}$ 表示在图 G 上定义的 N -层GNN的第 n 层隐藏特征。此外, 称 $\mu: \mathbb{R}^{v \times m} \rightarrow \mathbb{R}_{\geq 0}$ 为一个节点相似性度量, 如果它满足以下公理:
 - 1、 $\exists c \in \mathbb{R}^m$, 当对所有节点 $i \in V$, $X_i = c$ 时, 满足 $\mu(X) = 0$, $X \in \mathbb{R}^{v \times m}$
 - 2、 $\forall X, Y \in \mathbb{R}^{v \times m}$, $\mu(X + Y) \leq \mu(X) + \mu(Y)$
 - 然后, 过平滑被定义为相似性度量 μ 随着 GNN 层数 n 增加逐渐收敛到 0 的过程
 - 3、 $\mu(X^n) \leq C_1 e^{-C_2 n}$, $n = 0, \dots, N$, 其中 C_1 和 C_2 是正的常数





5.2.1 信息传递优化-过平滑

- **正则化**: 通过在训练过程中引入某种限制或约束, 抑制模型过度复杂化
 - 例如, 介绍过的DropEdge增加了训练过程中的随机性, 消息传播路径也随之改变, 有效阻止了信息在深层网络中过度传播
- **归一化**: 调整不同特征的尺度, 使其在模型训练时更具可比性
 - PairNorm[1] 通过在每一层的GNN之后对节点特征进行归一化处理, 确保节点对之间的距离保持恒定, 不会失去区分度, 同时保持稳定性, 可以形式化表示如下:

$$\hat{X}_i = X_i - \frac{1}{v} \sum_{j=1}^v X_j$$
$$X_i = \frac{s\hat{X}_i}{\sqrt{\frac{1}{v} \sum_{j=1}^v \|\hat{X}_j\|_2^2}}$$



5.2.1 信息传递优化-过平滑

- 改变图神经网络的信息传递过程

- 图耦合振荡器网络 GraphCON[1]:
- 通过改变图神经网络的消息传递过程来缓解深层图神经网络的过平滑问题
- GraphCON 通过将信息传播的机制转变为基于非线性振荡器的动态机制，可以形式化表示为：

$$Y^n = Y^{n-1} + \Delta t [\sigma(F_{\theta^n}(X^{n-1}, G)) - \gamma X^{n-1} - \alpha Y^{n-1}],$$
$$X^n = X^{n-1} + \Delta t Y^n,$$

- 其中， Y^n 表示节点特征的动态状态, 用于刻画特征随层数演化的速度。 Δt 表示时间，这一参数决定了特征更新的速度。

$F_{\theta^n}(X^{n-1}, G)$ 是图神经网络中的消息传递函数。 γ 和 α 控制特征消息传递的参数。它们影响特征更新的方式，能够调节节点特

[1]Rusch T K, Chamberlain B, Rowbottom J, et al. Graph-coupled oscillator networks[C]//International Conference on Machine Learning. PMLR, 2022: 18888-18909.



5.2.1 信息传递优化-过平滑

- 残差连接:
 - 在神经网络层中将输入与该层的输出相加，网络能够更好地保留原始输入的信息，从而防止特征在层与层之间的过平滑
 - 最初在图中引入残差连接 [1]

$$X^n = X^{n-1} + F_{\theta_n}(X^{n-1}, G)$$

- 引入可调节的缩放因子来控制残差学习中的信息流动 [2]:

$$X^n = \sigma \left[\left((1 - \alpha_n) \widehat{D}^{-\frac{1}{2}} \widehat{A} \widehat{D}^{-\frac{1}{2}} X^{n-1} + \alpha_n X^0 \right) \left((1 - \beta_n) I + \beta_n W^n \right) \right]$$

- α_n 是一个缩放因子，控制输入特征的贡献程度， β_n 是另一个缩放因子，在每一层中能够适应不同的特征传播方式

[1] Li G, Muller M, Thabet A, et al. Deepgens: Can gcnns go as deep as cnns?[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9267-9276.

[2] Chen M, Wei Z, Huang Z, et al. Simple and deep graph convolutional networks[C]//International conference on machine learning. PMLR, 2020: 1725-1735.

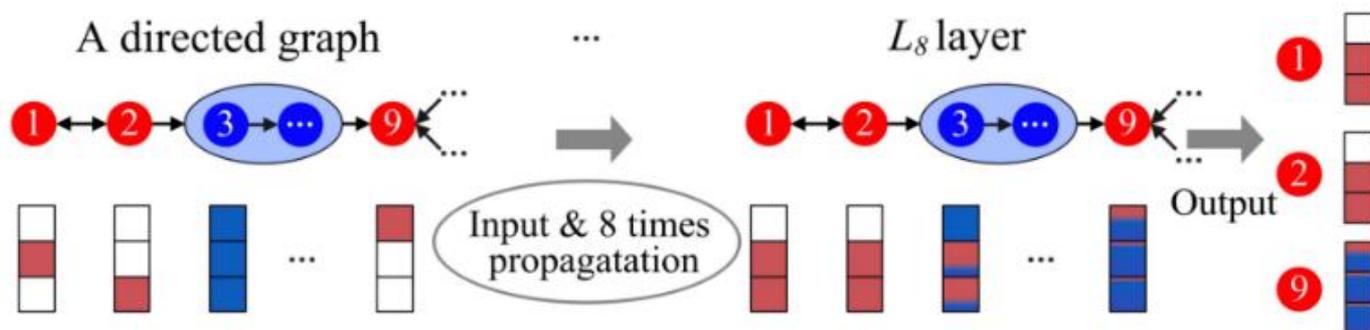


5.2.1 信息传递优化-过压缩

- **过压缩**: 尽管信息可能呈指数级增长, 但最终传递到目标节点的却是有限且固定大小的信息, 这个“瓶颈”被称为图神经网络中的过度压缩问题
- **定义 5.2 (过压缩分数)**: 经过 l 层消息传递后, 节点 i 和节点 s 之间的长距离依赖分数定义为:

$$\text{OSQ}(i, s) = \left\| \frac{\partial h_i^{(l)}}{\partial x_s} \right\|$$

- 其中 $\|\cdot\|$ 是矩阵的谱范数。





5.2.1 信息传递优化-过压缩

- **动力学启发的消息传递：基于分数邻接矩阵的消息传递**
 - 在经典的图神经网络中，消息传递主要集中在节点的局部邻域。为了解决这种局部邻域的消息传递导致的过压缩问题，最近开发了一些基于非局部动态的消息传递神经网络。
 - 分数邻接矩阵[1]：记作 \hat{A}^τ ，其中 $\tau \in \mathbb{R}$ 。
 - 基于分数邻接矩阵就可以实现非局部消息传递。具体而言，消息传递规则为 $\frac{\partial H}{\partial t} = -\hat{A}^\tau H$ ，其中 τ 决定了连接的密度和传播的范围。通过选择适当的 τ 值，可以使图的连接性变得更加密集，这种调整允许在不同的连接水平上传播节点特征信息，从而有效地将信息从一个节点传播到更远的节点



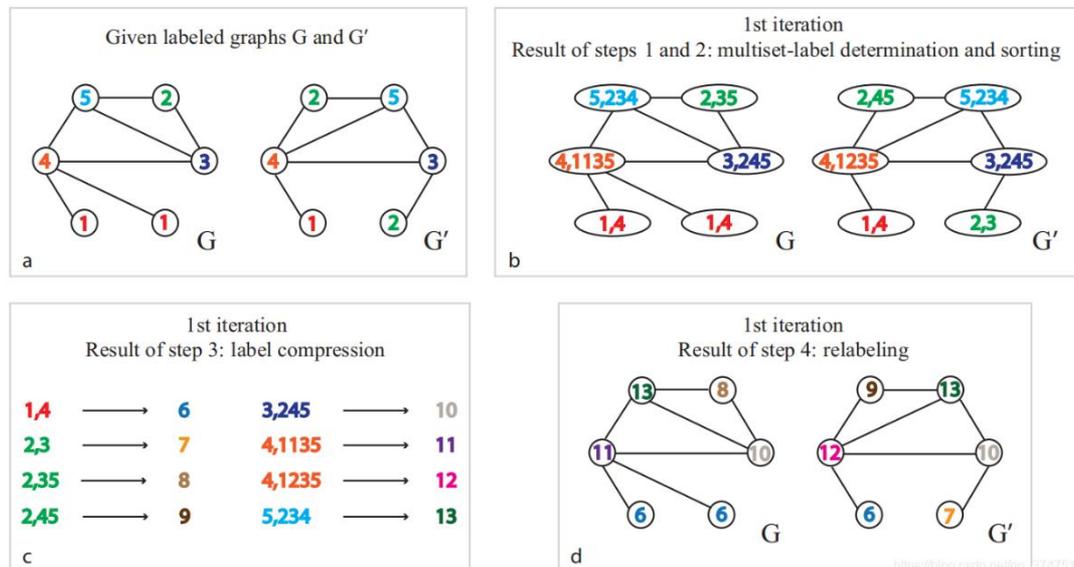
5.2.1 信息传递优化-过压缩

- **Graph transformer:** 为所有节点对分配注意力，而不论它们在原始图中是否相连，这种机制称为全局感受野
 - Graphormer[1]: Graphormer 是一种图神经网络模型，通过空间编码和边编码过程为图的邻接矩阵分配注意力。
 - 节点特征可以通过与度数相关的相对重要性进行增强， $h_i^{(0)} = x_i^{(0)} + \text{deg}^-(i) + \text{deg}^+(i)$ ，其中 $\text{deg}^-(i)$ 和 $\text{deg}^+(i) \in \mathbb{R}^{c_0}$ 是可学习的嵌入向量
 - 空间编码: 考虑一个函数 $\zeta(i, j): V \times V \rightarrow \mathbb{R}$ 用于测量节点之间的空间关系，如果节点相连，其值为最短路径距离，否则为-1。空间编码 $b_{\zeta(i, j)}$ 是一个根据 $\zeta(i, j)$ 索引的可学习向量，随索引增加递减，从而减少对距离较远邻居的关注
 - 边编码: $\forall i, j \in \mathcal{V}, c_{i, j} = \frac{1}{N} \sum_{n=1}^N x_{e_n} (w_n^E)^\top$ ， x_{e_n} 是节点之间最短路径的第n条边的特征， w_n^E 是对应的可学习权重嵌入



5.2.1 信息传递优化-表达能力受限

- 图同构测试方法：1-WL算法是著名的图同构测试方法
 - 工作原理：基于节点的“标签”传播：每个节点根据它的邻居来更新自己的标签，并迭代多次，直到标签不再变化。如果两个图在多轮迭代后每个节点的标签都相同，则认为这两个图在结构上是相同的（同构）。它是一种高效但有限的图同构检测方法
 - 基于1-WL的GNN的表达能力和1-WL是等效的
 - 下图是一次迭代过程





5.2.1 信息传递优化-表达能力受限

- 缺点：1-WL在许多场景中表现良好，但它并不能区分所有的图，表达能力有限。
 - 1-WL和基于1-WL的GNN难以检测和计数一些重要的图子结构，尤其是三角形，而三角形在社交网络分析、分子结构分析等许多应用中非常关键
- 解决方法：
 - 高阶图神经网络模型 (High-order GNNs)
 - 基于非等变操作的图神经网络 (Non-equivarant GNNs)



5.2.1 信息传递优化-表达能力受限

- 高阶图神经网络K-GNN[1]: 基于k-set WL算法
 - 具体来说, 考虑包含 k 个节点的子图, 从而使得算法更高效, 同时也能考虑图结构信息, k-set WL比1-WL更强大
 - k-set记为: $[V]_k = \{S \subseteq V \mid |S| = k\}$
 - k-set的邻居定义为只差一个节点的k-set, 即: $N_{V,k}(S) = \{J \in [V]_k \mid |J \cap S| = k - 1\}$

[1] Morris C, Ritzert M, Fey M, et al. Weisfeiler and leman go neural: Higher-order graph neural networks[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 4602-4609.



5.2.1 信息传递优化-表达能力受限

- K-GNN: 基于k-set WL算法

$$x_k^{(t)}(S) = \sigma \left(W_1^{(t)} x_k^{(t-1)}(S) + \sum_{U \in N_{V,k}(S)} W_2^{(t)} x_k^{(t-1)}(U) \right)$$

- $x_k^{(t)}(S)$ 表示第t层含有k个节点的子图的嵌入
- 由于k-set WL比1-WL更强大，因此k-GNN比消息传递神经网络更强大，并且通过适当的参数矩阵初始化，k-GNN 被证明可以与k-set WL一样强大



5.2.1 信息传递优化-表达能力受限

- 基于非等变操作的图神经网络：
 - 直接打破传统消息传递图神经网络的对称性，提高表达能力
 - 例如，RP-GNN[1]在计算图的嵌入时，通过对图上所有可能的排列取平均，确保了对称性，即结果不受节点顺序影响。形式上，关系池化通过任意函数 f 来获得图 G 的嵌入，具体如下：

$$f(A, X) = \frac{1}{n!} \sum_{\sigma \in S_n} f(\sigma \cdot A, \sigma \cdot X)$$

- 其中 σ 是在对称群中定义的置换。为了提高图神经网络的表达能力，为每个节点特征附加一个对置换敏感的标识符，模型变得非等变。由此衍生出的新神经网络架构定义如下：

$$f(A, X) = \frac{1}{n!} \sum_{\sigma \in S_n} f(A, [X, \sigma \cdot I_n]),$$

- 已经证明了这种网络架构在区分非同构图方面比原始图神经网络更强大

[1] Murphy R, Srinivasan B, Rao V, et al. Relational pooling for graph representations[C]//International Conference on Machine Learning. PMLR, 2019: 4663-4673.