Diffusion Improves Graph Learning

NeurIPS 2019

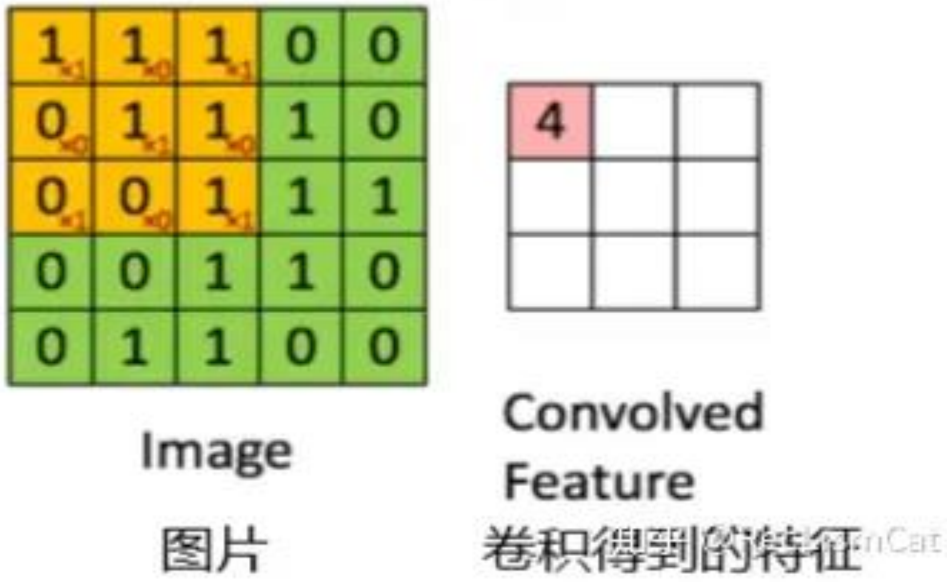
汇报人：庞媛媛

2024/03/24

目　录

CNN

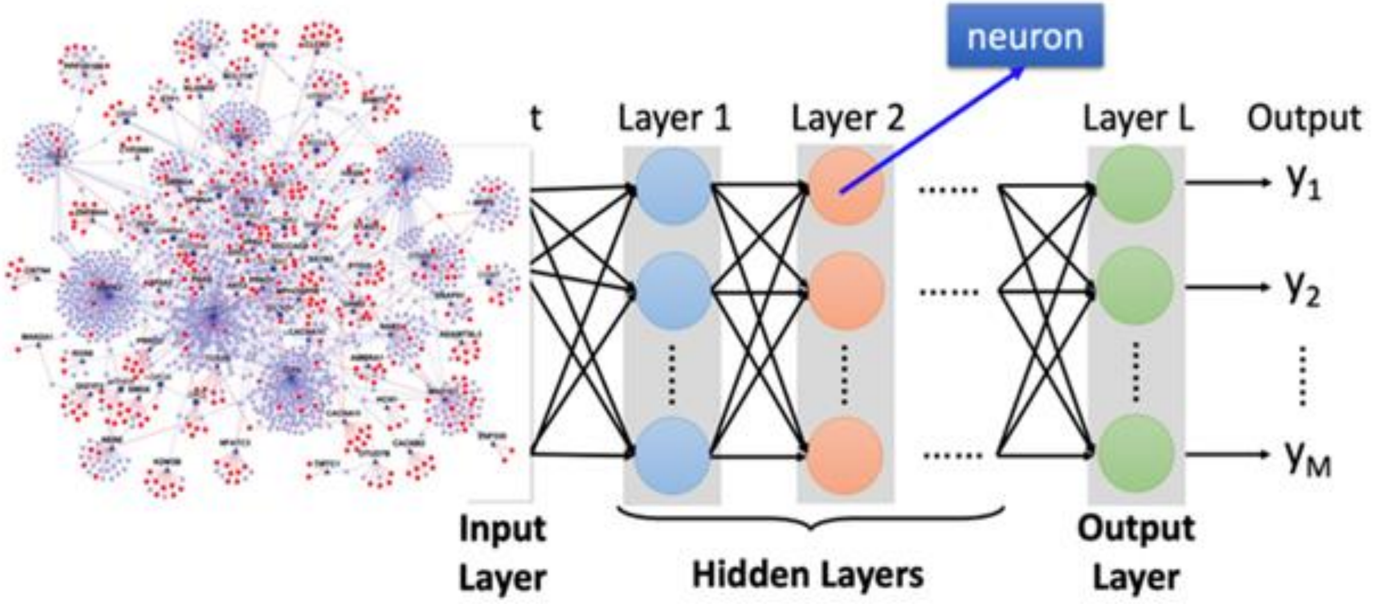
固定数量的邻域节点排序，与相同数量的卷积核参数相乘求和。



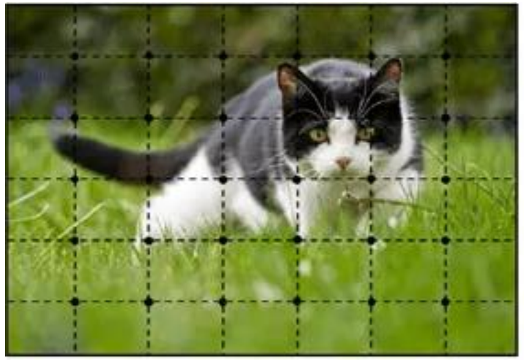
Step1：构建邻域。
找到固定数量的邻居节点。
对找到的邻居节点进行排序。
Step2：对邻域的点与卷积核参数做内积。

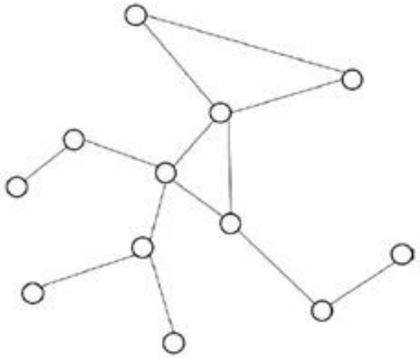
GNN

借助神经网络的能力如深度特征抽取等来处理图结构的数据，直观结构：

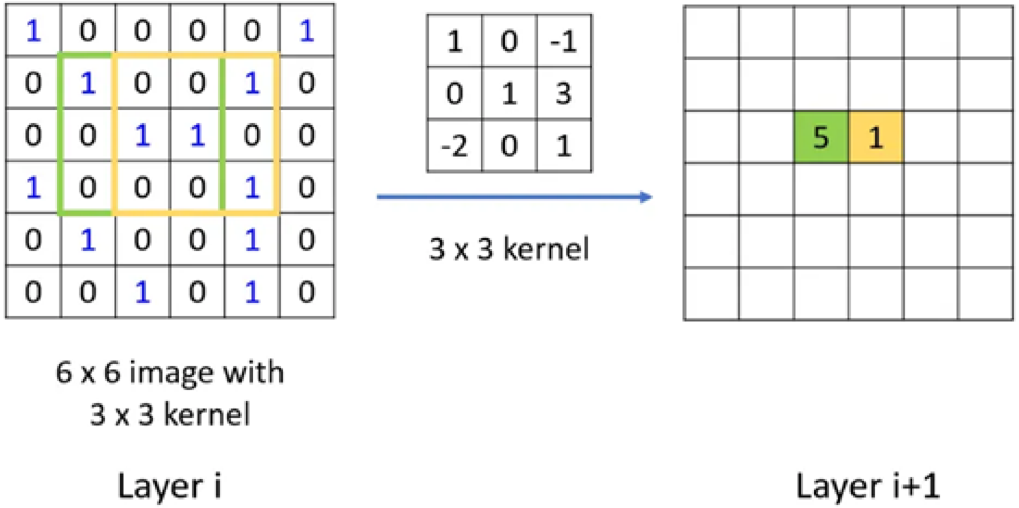


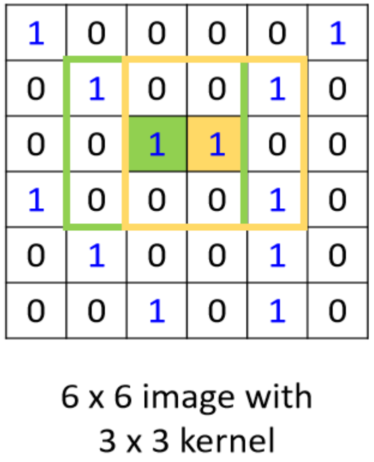
为什么需要？

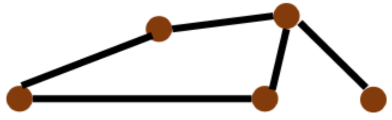




欧氏空间卷积&非欧式空间卷积





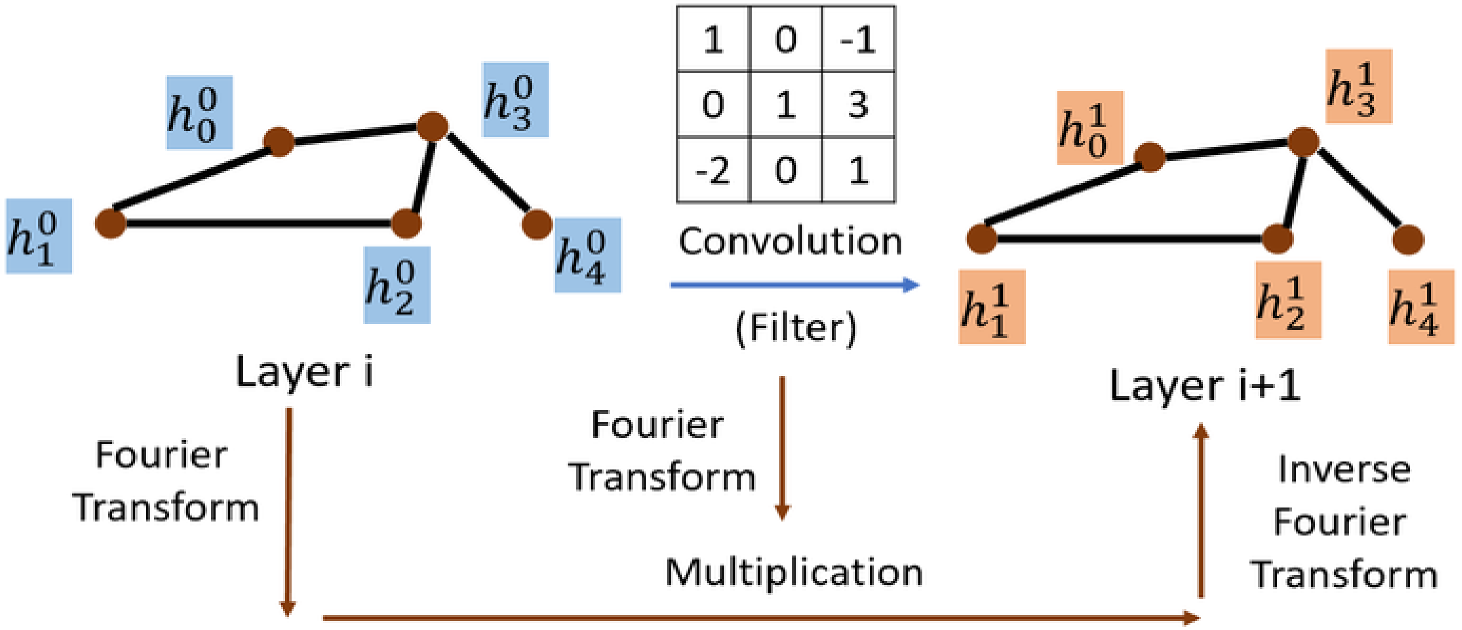


对于欧氏空间的图结构，通过卷积核滑动整个图区域。

对于非欧氏空间的图结构，无法找到固定大小的卷积核滑动整个图区域。

GCN

方案一：谱域GCN（图卷积运算~从图信号中去除噪声）



- **基本思路：** 先将空域输入信号和空域卷积核通过图傅里叶变换转换到谱域，然后在谱域中相乘，再通过反傅里叶变换转换回空域。
- **经典谱域GCN：** SCNN、ChebNet、GCN
- **缺点：** 不适用于有向图；假定图结构是固定的；复杂度问题

- **图傅里叶变换：**

拉普拉斯特征向量~基函数，拉普拉斯矩阵的特征值 ~ “频率”

① 图上的信号：$x = [x_1, x_2, \ldots, x_n]^\top \in \mathbb{R}^n$

② 空间域→谱域：$\hat{x} = U^\top x \in \mathbb{R}^n$

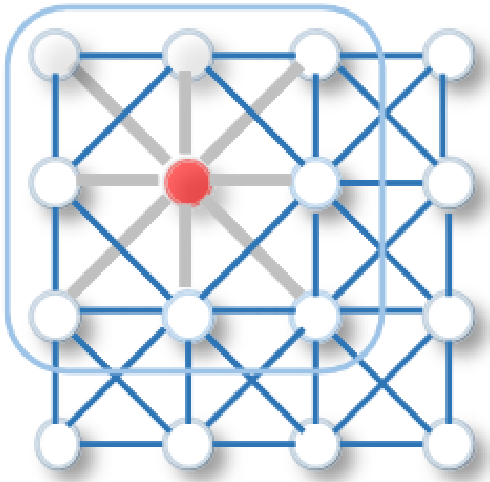
③ 谱域→空间域：$x = U\hat{x}$

$$L = U\Lambda U^{-1} = U\begin{bmatrix}\lambda_1 & & \\ & \cdots & \\ & & \lambda_n\end{bmatrix}U^{-1}, U = (\overrightarrow{u_1}, \overrightarrow{u_2}, \ldots, \overrightarrow{u_n})$$

GCN

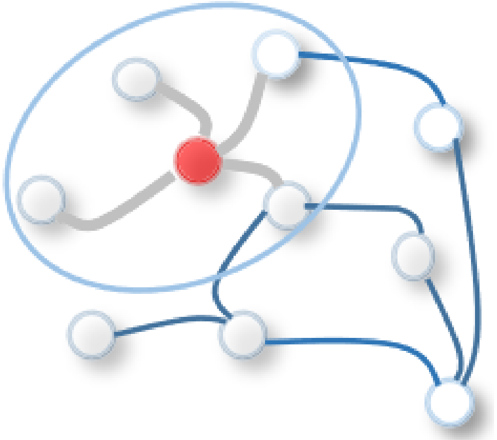
方案二：空域GCN（参考CNN对每个节点的邻居节点加权求和）

➢ 问题：图结构不存在固定的邻域结构
- 每一个节点的邻域大小不同且是变化的
- 同一领域内的节点不存在顺序性



(a)

(a)2D Convolution：每个像素是一个节点，红色节点的像素值与邻居节点的像素值加权平均。节点的**邻居是有序的，且大小相同且固定**。



(b)

(b) Graph Convolution：取红色节点及其邻域的节点特征的平均值。与图像数据不同，节点的**邻域是无序的，大小不同且是可变的**。

GCN

方案二：空域GCN

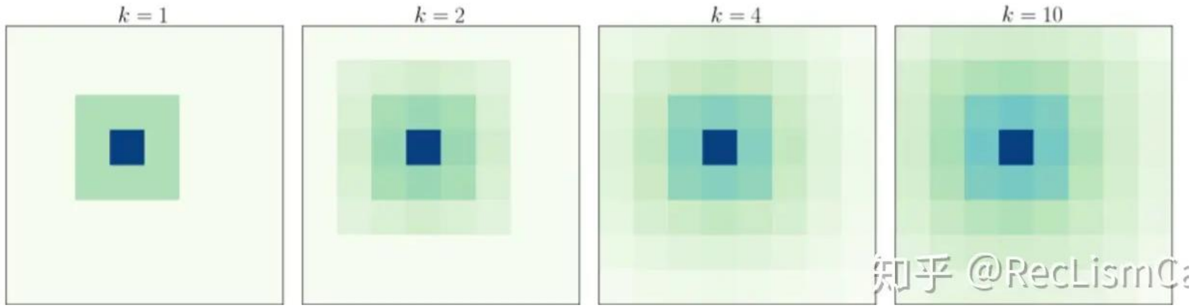
➢ 解决思路：①使用随机游走的方法，根据被选中的概率期望大小选择固定数量的邻居节点。②根据节点被选择的概率期望对邻域排序。

Step1：$P=D^{-1}S$（P：图上的随机游走转移矩阵；S：相似度矩阵（类似邻接矩阵）；D：度矩阵）

Step2：归一化P作为转移矩阵

Step3：定义多步的转移期望$Q^{(0)}=I, Q^{(1)}=I+P, \ldots Q^{(k)}=\sum_{i=0}^{k} P^k$。

每一项$Q_{ij}^{(k)}$表示k步内，节点i出发到节点j的期望访问数。



Step4：根据期望大小选择邻域

$\pi_i^{(k)}(c)$表示k步内由节点i出发的访问期望数第c大的节点，则选择的邻域顺序：$Q_{i\pi_i^{(k)}(1)} > Q_{i\pi_i^{(k)}(2)} > \cdots > Q_{i\pi_i^{(k)}(N)}$

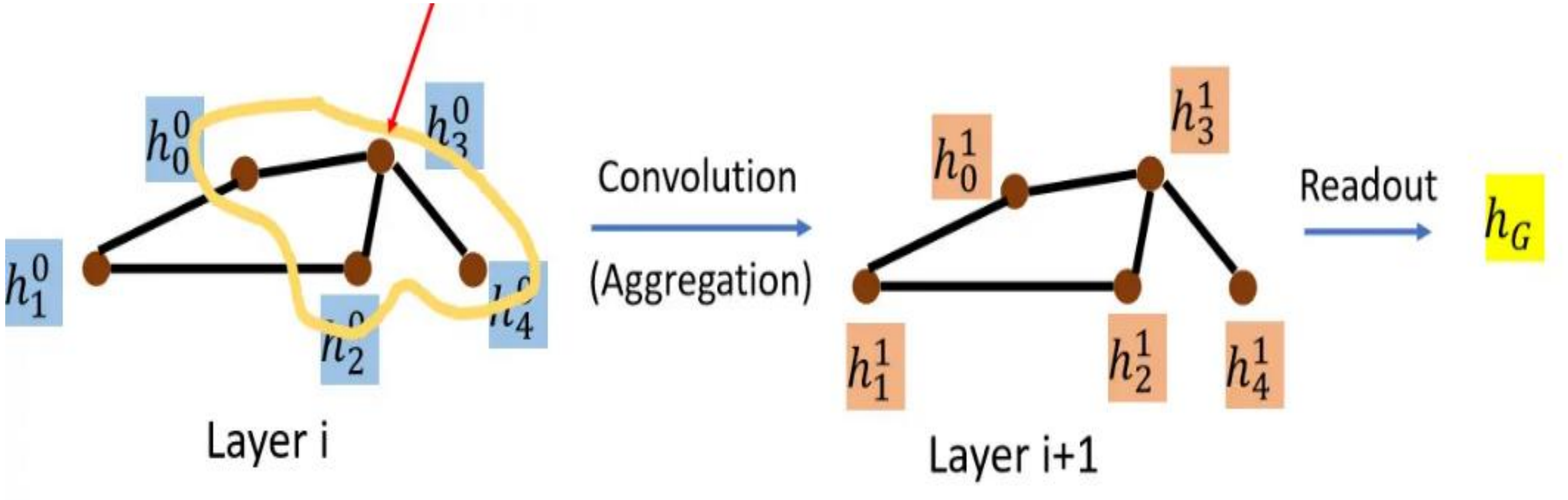
Step5：做1D卷积 $Conv_1(\mathbf{x}) = \begin{bmatrix} x_{\pi_1^{(k)}(1)} & \cdots & x_{\pi_1^{(k)}(p)} \\ x_{\pi_2^{(k)}(1)} & \cdots & x_{\pi_2^{(k)}(p)} \\ \vdots & \ddots & \vdots \\ x_{\pi_N^{(k)}(1)} & \cdots & x_{\pi_N^{(k)}(p)} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$

GCN

2个术语

Aggregate：用当前层的某节点邻居状态特征更新下一层该节点的状态特征。

Readout：把所有节点的状态特征集合起来代表整个图的状态特征。

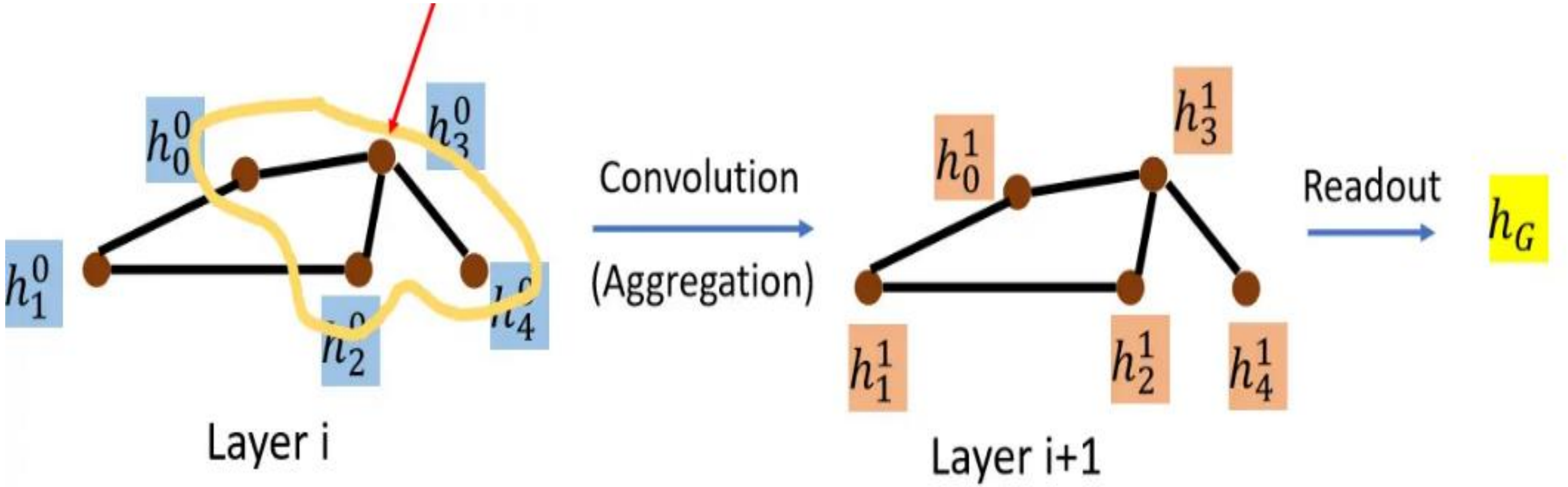


GCN

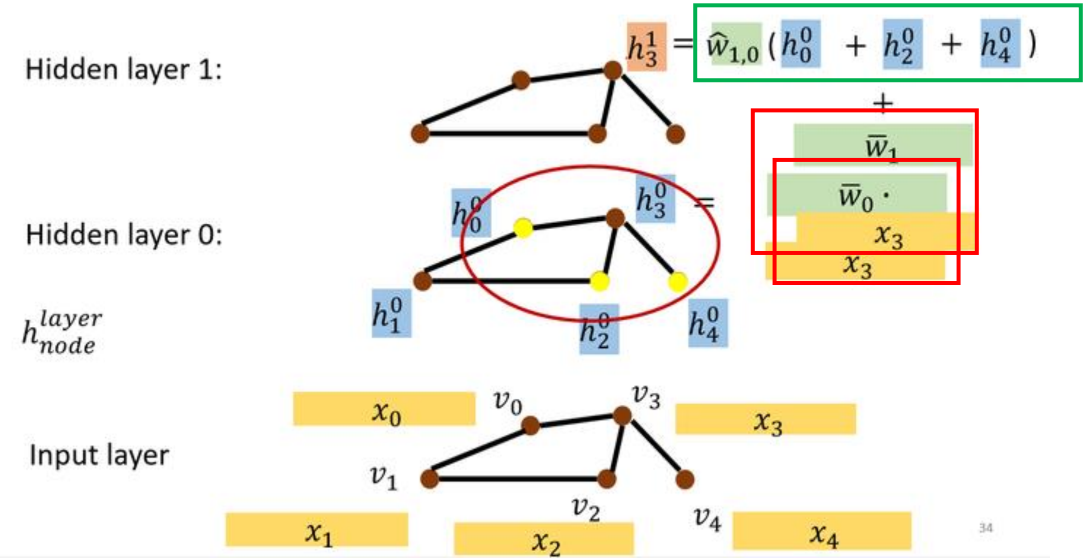
2个术语

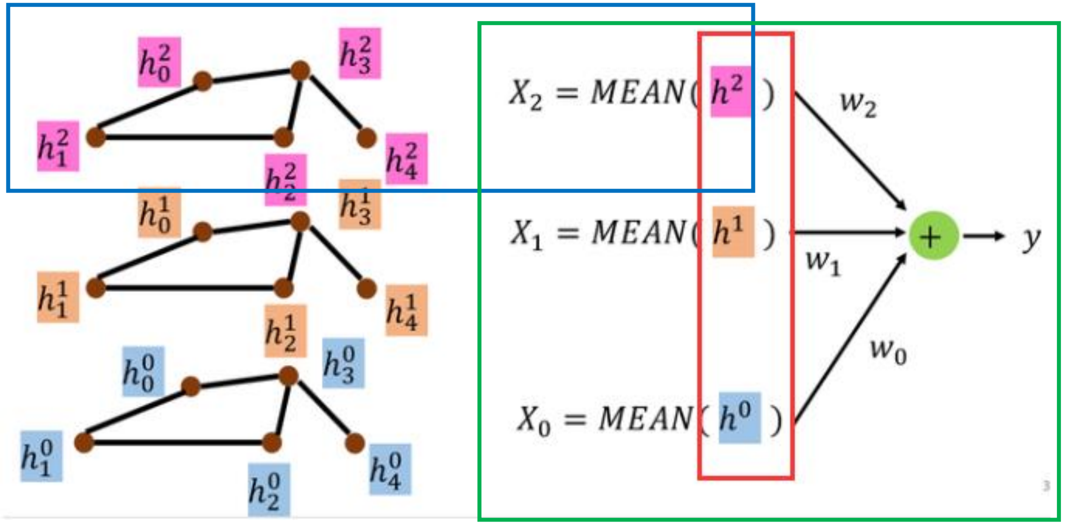
Aggregate：用当前层的某节点邻居状态特征更新下一层该节点的状态特征。

Readout：把所有节点的状态特征集合起来代表整个图的状态特征。



NN4G—经典空域GCN





方法：直接将节点的邻域信息相加来进行图卷积

Aggregate：$\mathbf{h}_v^{(k)} = f\left(\boxed{\mathbf{W}^{(k)^T}\mathbf{x}_v} + \boxed{\sum_{i=1}^{k-1}\sum_{u\in N(v)}\Theta^{(k)^T}\mathbf{h}_u^{(k-1)}}\right)$

矩阵形式$\mathbf{H}^{(k)} = f(\boxed{\mathbf{X}\mathbf{W}^{(k)}} + \boxed{\sum_{i=1}^{k-1}\mathbf{A}\mathbf{H}^{(k-1)}\Theta^{(k)}})$

Readout：通过对每一层的所有节点信息的取平均获得每一层图的representation，并对每一层的信息通过求和平均的方式进行整个图的representation更新，即：

$$\mathbf{h}_G = sum/mean/sum(\mathbf{h}_1^{(K)}, \mathbf{h}_2^{(K)}, \ldots, \mathbf{h}_n^{(K)})$$

GDC

GCN通常通过**直接（一跳）邻居**进行信息传递，捕获信息受限。

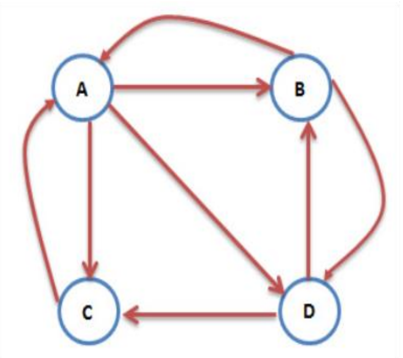
- 问题：如何克服？转向**更具表现力的邻居**
- 解决思路：空域GCN能在更深的层中利用高阶信息但将每层的信息限制在一跳邻域
  频域GCN能捕捉更复杂的图属性但表现经常劣于MPNN（消息传递神经网络，当时的图卷积网络）
  结合优势→受光谱启发的执行消息传递的新技术，聚合更大邻域的信息

图扩散卷积GDC

通过稀疏化广义的图扩散生成的新图构建的

Page Rank：利用网页简单的超链接来计算网页的分值，从而给网页进行排名的一种算法。
思想：估计悠闲的上网者分布在各个网页上的概率。顶点之间是随机游走的
模型：互联网~有向图，网页~节点

$$P(A) = P(C) + \frac{P(B)}{2}$$



Personalized PageRank：顶点之间偏向于一组起始顶点（个性化）

Heart kernel：局部扩散方法，相比PPR将更大的权重放在了距离种子节点（扩散起点）较近的节点

广义的图扩散

通过**扩散矩阵**定义广义图扩散：$\boldsymbol{S}=\sum_{k=0}^{\infty}\theta_k T^k$

无向图$\mathcal{G}=(\mathcal{V},\mathcal{E})$，$N=|\mathcal{V}|$表示节点数，$A\in\mathbb{R}^{N\times N}$表示邻接矩阵，$D$度矩阵。

- **广义过渡矩阵 $\boldsymbol{T}$：**

  随机游走过渡矩阵$T_{\mathrm{rw}}=AD^{-1}$：A中添加（加权）自循环来调整随机游走，即懒惰随机游走

  $$T_{\mathrm{sym}}=(w_{\mathrm{loop}}\boldsymbol{I}_N+D)^{-1/2}(w_{\mathrm{loop}}\boldsymbol{I}_N+A)(w_{\mathrm{loop}}\boldsymbol{I}_N+D)^{-1/2}$$

  对称过渡矩阵 $T_{\mathrm{sym}}=D^{-1/2}AD^{-1/2}$

- **加权系数$\boldsymbol{\theta_k}$：**

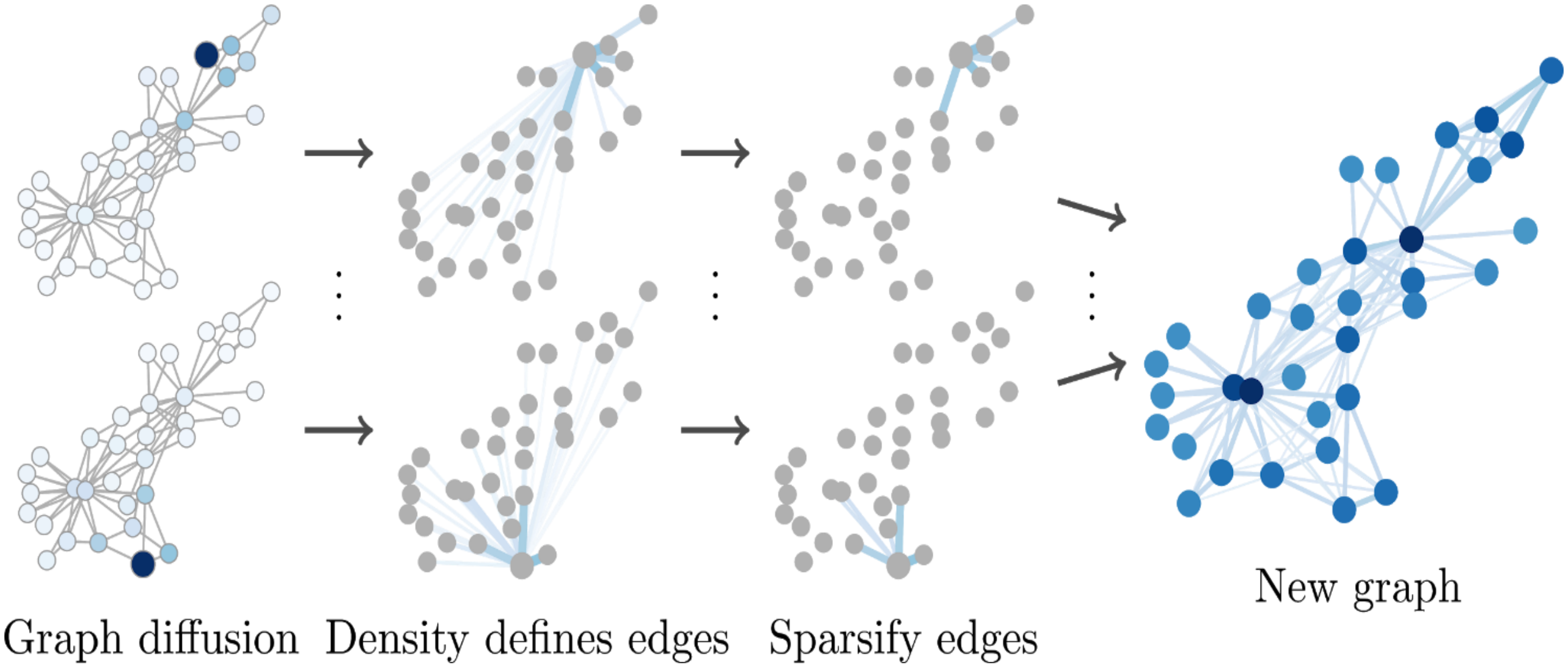
  PPR $T=T_{\mathrm{rw}}$ ， $\theta_k^{\mathrm{PPR}}=\alpha(1-\alpha)^k$ 远距传输概率$\alpha\in(0,1)$

  heart kernl $T=T_{\mathrm{rw}}$ ，$\theta_k^{\mathrm{HK}}=e^{-t}\frac{t^k}{k!}$ 扩散时间为 t

  Kipf & Welling 提出的近似图卷积，当$\theta_1=1\text{and}\theta_k=0\text{for}k\neq1$使用$T=\tilde{T}_{\mathrm{sym}}\text{with}w_{\mathrm{loop}}=1$

  其他方法：将优化稀疏类似于θ_k的方法作为其训练过程的一部分，但表现差于上述简单系数。

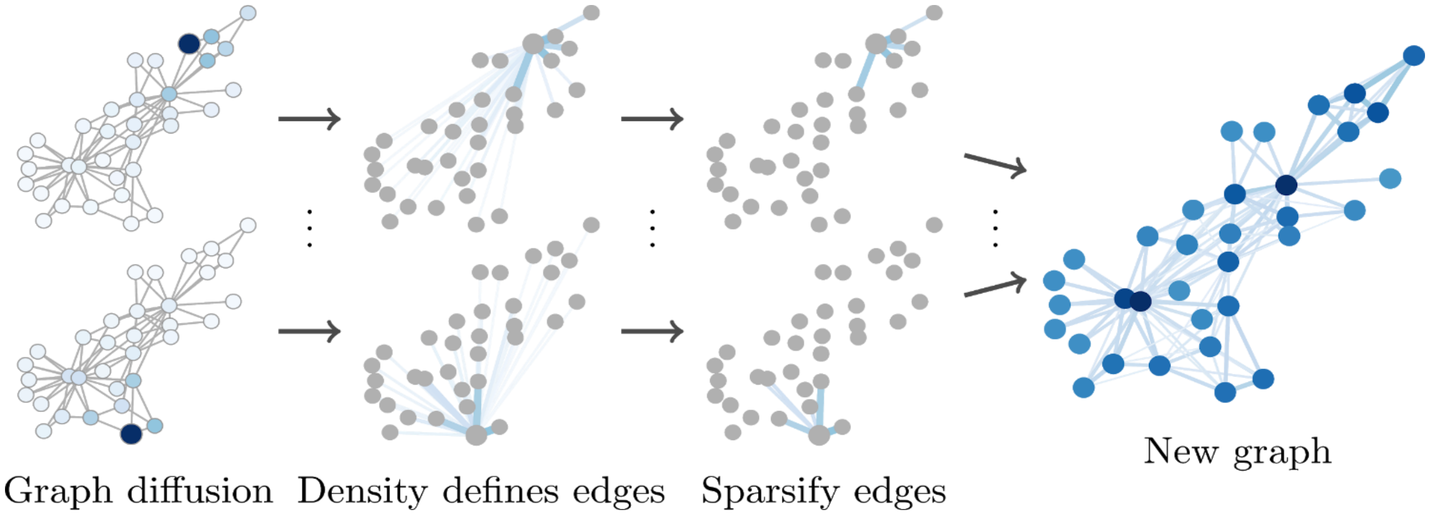
图扩散卷积



本质上：**GDC是用广义图扩散矩阵 S 的稀疏化版本$\tilde{S}$交换普通邻接矩阵A**

Step1：开始时所有注意力放在节点 v上，对原始图进行扩散，不断将一些注意力传递给邻居，将注意力分散开。
Step2：一段时间后，扩散后的密度定义了起始节点v的边。
Step3：通过对每个节点这样做，得到了一个矩阵S，它定义了一个新的连续加权图（稠密）。
Step4：移除所有权值较小的边。对每个节点这样做一次，得到一个新的稀疏的加权图和稀疏矩阵$\tilde{S}$。

图扩散卷积



几点思考：

- 图扩散平滑图上的邻域~图像的高斯滤波器去噪
- 稀疏化：前提“四/六度分离”导致的局部化；top-k或阈值
- 局限：基于同类假设，即 "物以类聚"。扩展到异亲关系（即 "异性相吸"）表现不佳。

GDC四步骤：

1.计算过渡矩阵 T：T_{rw} / T_{sym}

2.广义图扩散矩阵S：$S=\sum_{k=0}^{\infty}\theta_k T^k$

3. 截断小值稀疏化得到稀疏矩阵$\tilde{S}$

4.计算过渡矩阵$T_{\tilde{S}}$

实验设置、数据集、模型

- 过渡矩阵：带自循环的对称过渡矩阵$\widetilde{\boldsymbol{T}}_{\mathrm{sym}}=(\boldsymbol{I}_N+\boldsymbol{D})^{-\frac{1}{2}}(\boldsymbol{I}_N+\boldsymbol{A})(\boldsymbol{I}_N+\boldsymbol{D})^{-\frac{1}{2}}$
  列随机游走过渡矩阵$T_{rw}^{\tilde{\boldsymbol{S}}}=\tilde{\boldsymbol{S}}D_{\tilde{\boldsymbol{S}}}^{-1}$
  加权系数：heart kernl和 PPR
  扩散矩阵 S ：使用阈值或 top-k 进行稀疏化
- 6个数据集：引用图 CITESEER、CORA 和 PUBMED，共同作者图 COAUTHOR CS ，
  共同购买图 AMAZON COMPUTERS 和 AMAZON PHOTO
- 9个模型：

有监督的模型：图卷积网络（GCN）、图注意网络（GAT）、跳跃知识网络（JK）、图同构网络（GIN）、ARMA
无监督的模型：degree corrected随机块模型（DCSBM）、谱聚类、DeepWalk、Deep Graph Infomax（DGI）

实验结果

- 半监督的节点分类

Figure 3: Node classification accuracy of GNNs with and without GDC. GDC consistently improves accuracy across models and datasets. It is able to fix models whose accuracy otherwise breaks down.

- 无监督的聚类

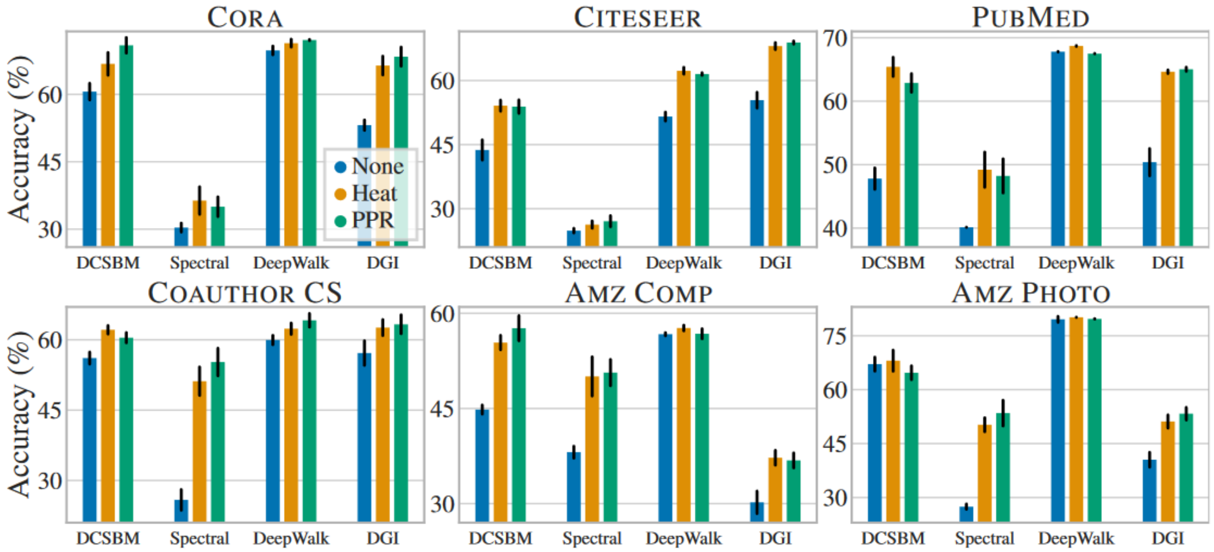


Figure 4: Clustering accuracy with and without GDC. GDC consistently improves the accuracy across a diverse set of models and datasets.

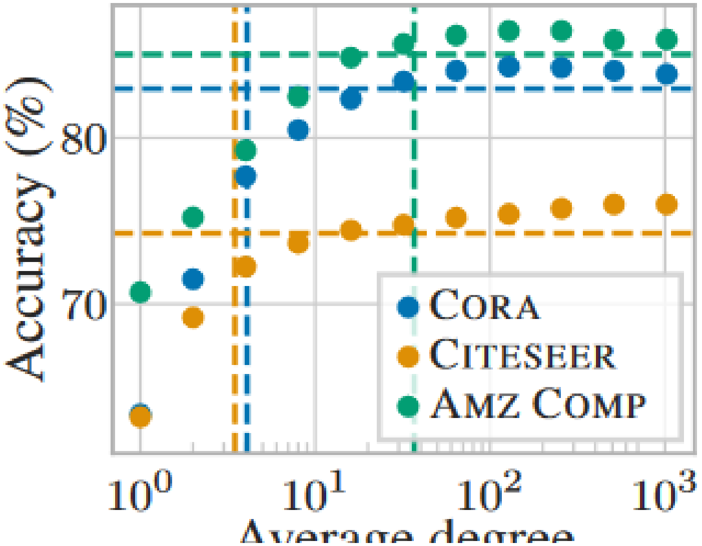
有 GDC 和没有 GDC 的 GNN 的节点分类准确率。在不同的模型和数据集上，GDC 都能持续提高准确率。它还能修复那些精度不佳的模型。

使用和不使用 GDC 时的聚类精度。在不同的模型和数据集中，GDC 始终能提高准确率。

5个问题

Q1：GDC 是否会增加图密度？

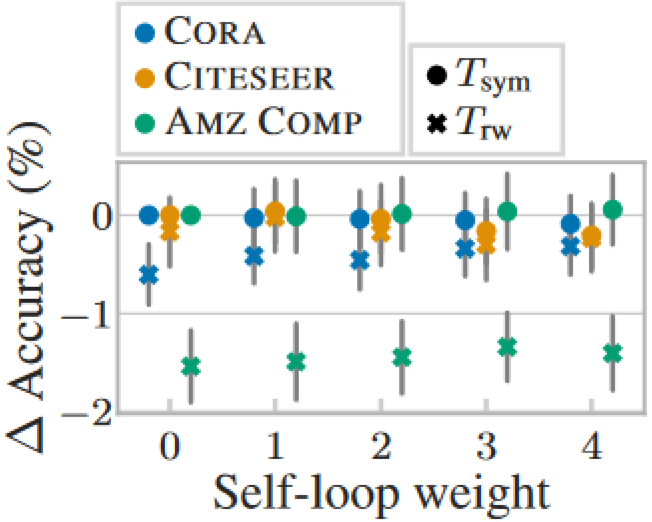
- GDC需要大致相同的平均度才能超越原始图的性能，与数据集及其平均度无关。
- 存在最佳稀疏度
- 稀疏化有利于计算，提高预测性能。



GCN+GDC 精确度（使用 PPR 和 top-k）

Q2：如何选择过渡矩阵 T？

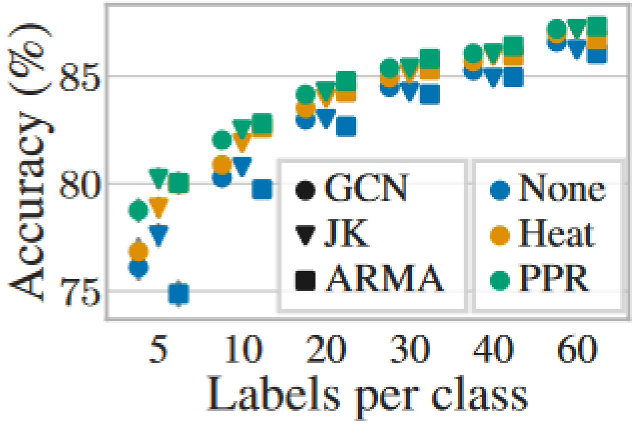
- T_{rw}性能更差
- 自循环没有明显影响



使用T_{sym}和T_{rw}对GCN+GDC 精度的差异（使用 PPR 和 top-k，百分点）

Q3：标签率对 GDC 有何影响？

- 标签率越稀疏，GDC 的改进幅度越大。

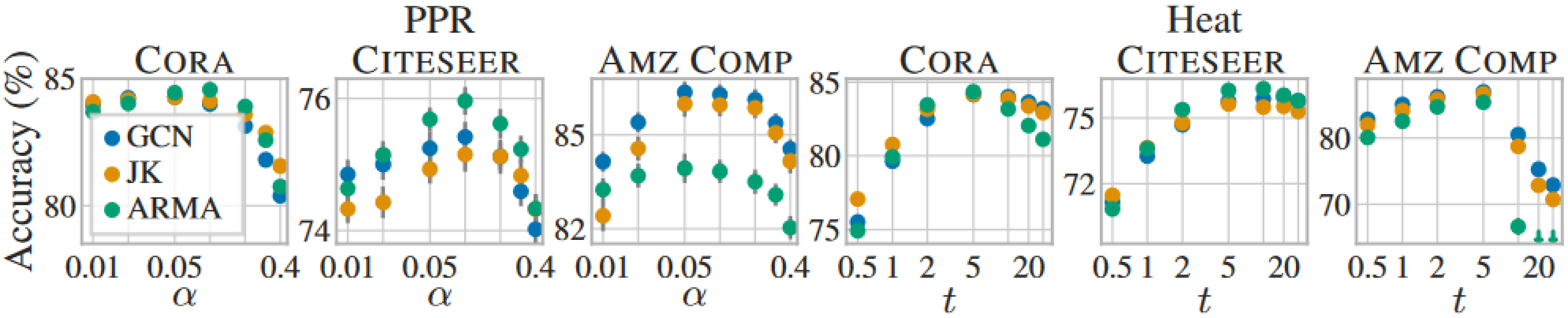


不同标签率下Cora的准确率

5个问题

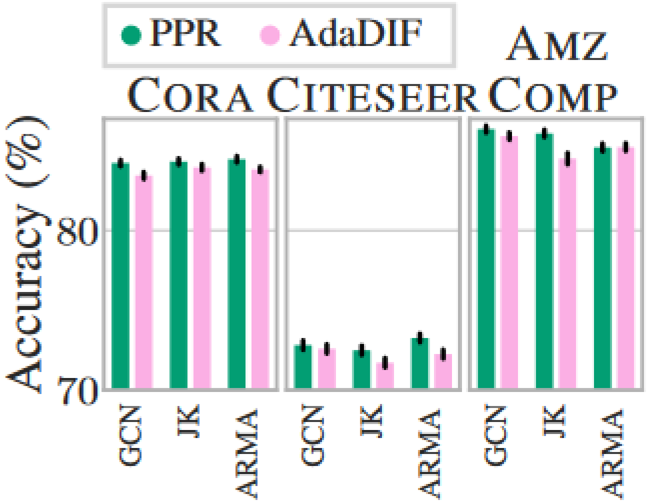
Q4：如何选择系数θ_k？

➢ 最佳超参数在很小范围内，不同数据集和模型的最佳值一致。



使用 GDC 并改变 PPR(α)和*heart kernel (t)* 的超参数所达到的精度

➢ 学习类似系数的模型中获取的θ_k，即使手工调整正则化，训练后也比不上原系数。

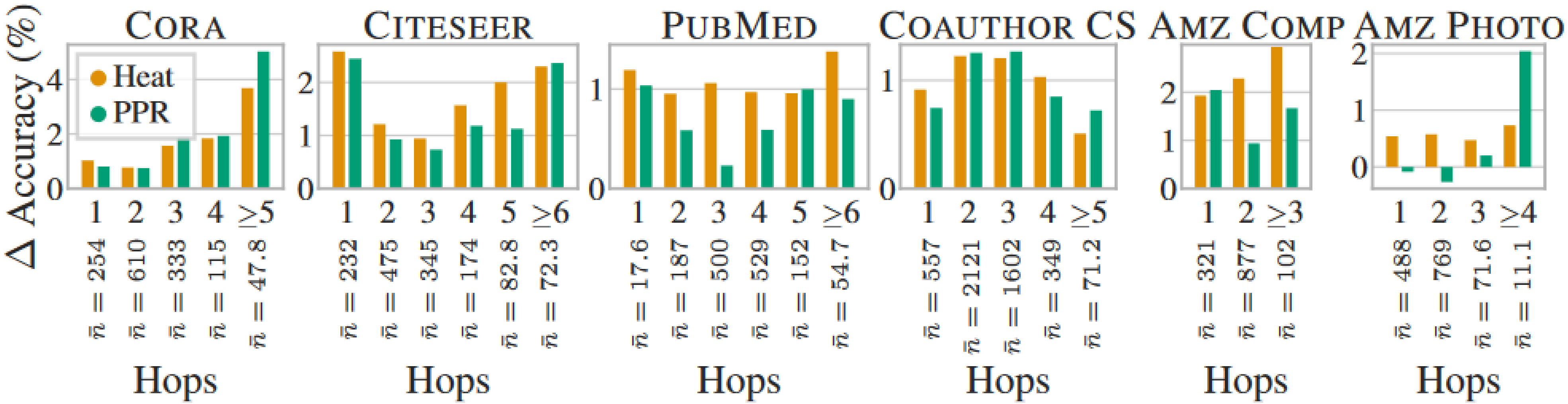


使用 PPR 定义的系数θ_k和 AdaDIF 学习的系数θ_k的 GDC 精度

5个问题

Q5: 哪些节点受益于 GDC？

➢ 距离较远的节点往往从 GDC 中获益更多



根据与训练集的距离（跳数），添加 GDC 后 GCN 精度的提高（百分点）

优点及扩展

- 提出基于稀疏广义图扩散的图扩散卷积（GDC）方法。
- GDC 是 GNN 中消息传递的一种更强大的空间局部扩展，够增强任何基于图的模型。
- 广泛而严谨的实验表明，GDC 在有监督和无监督任务中都能持续提高各种模型的准确性。
- 扩展包括：其他扩散系数θ_k，如 Fouss 等人[22]提出的方法所给出的扩散系数，以及更先进的随机游走和不是由过渡矩阵幂定义的算子。

谢 谢！