

图卷积神经网络

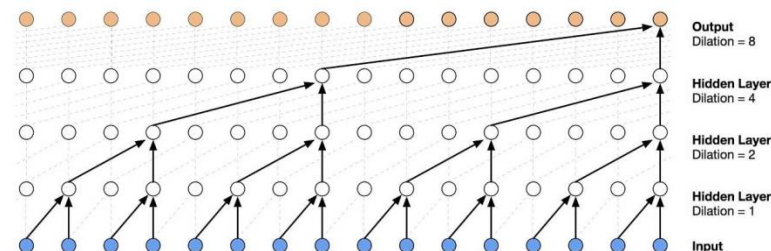
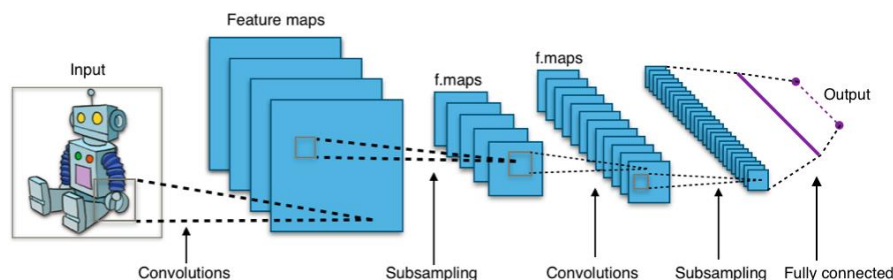
沈华伟

中国科学院计算技术研究所

2019年5月7日

» 卷积神经网络

- 卷积神经网络(CNN)在图像、视频、语音、文本等分布在欧式空间中的数据上取得了很大的成功
 - 图像分类、对象检测、人机对话、机器翻译.....



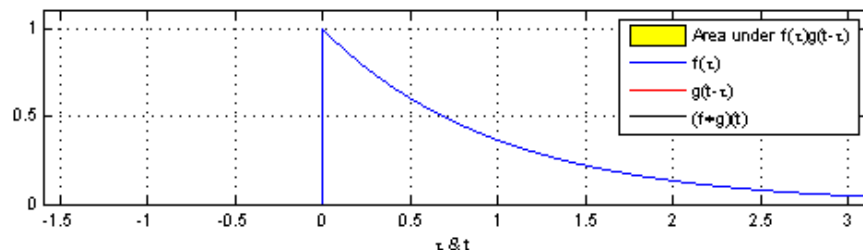
- 一个有趣的问题: 如何把卷积神经网络迁移到图数据或者网络数据这类非欧数据上?

卷积

■ 卷积是一种关于两个函数(f 和 g) 的数学运算

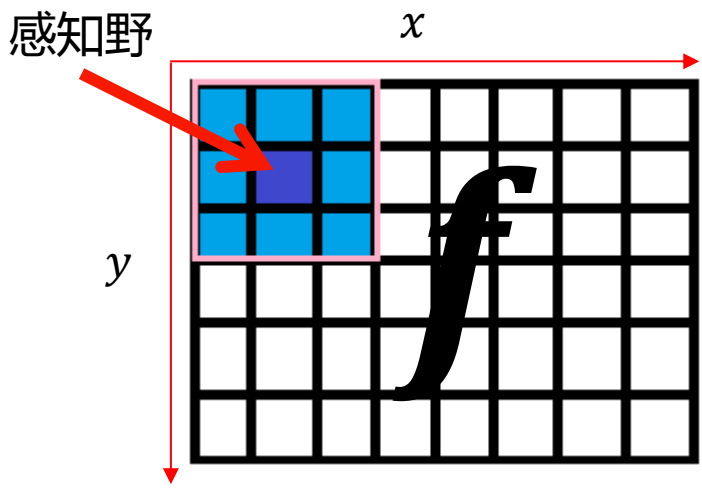
□ 一维（连续）时域信号的卷积

$$h(t) = (f * g)(t) \stackrel{\text{def}}{=} \int f(t)g(t - \tau) d\tau$$



□ 二维（离散）空域信号的卷积

$$h(x, y) = (f * g)(x, y) \stackrel{\text{def}}{=} \sum_{m, n} f(x - m, y - n)g(m, n)$$



$g =$

$g(1,1)$	$g(0,1)$	$g(-1,1)$
$g(1,0)$	$g(0,0)$	$g(-1,0)$
$g(1,-1)$	$g(0,-1)$	$g(-1,-1)$

» 理解卷积

■ 卷积的意义

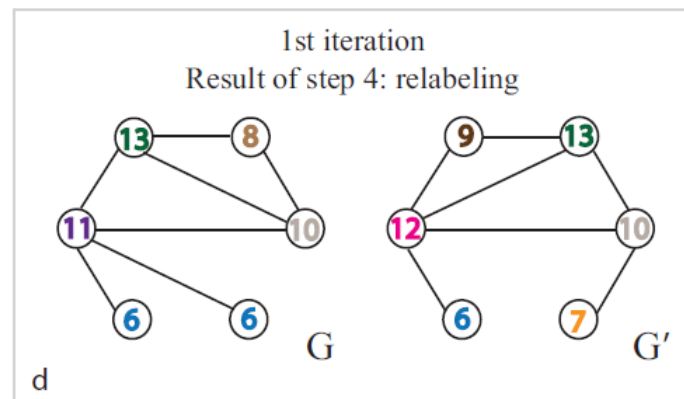
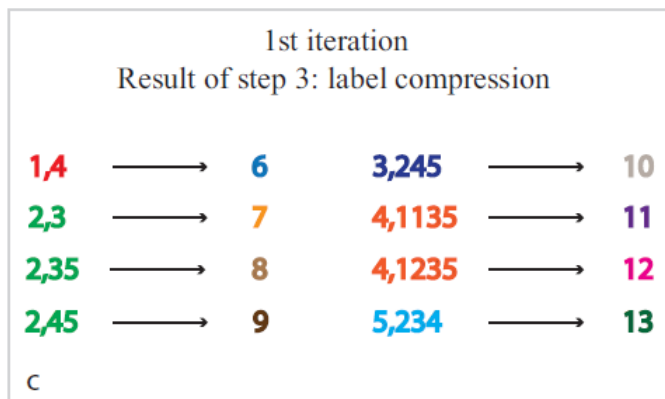
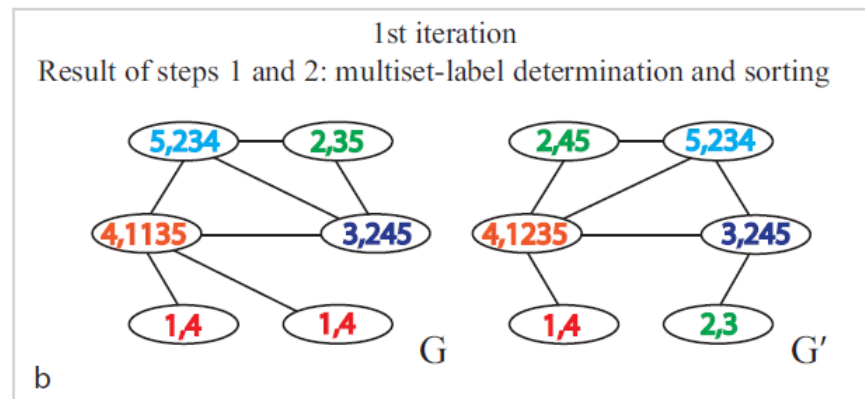
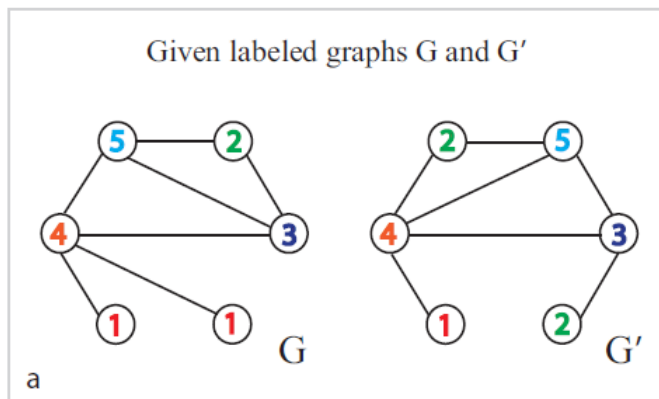
- 直观上看，卷积是函数 f 的加权平均，权重由卷积核 g 确定，因此卷积也被解释为采用模板 g 对函数 f 进行局部模板匹配

■ 卷积的性质

- 卷积后得到的函数一般比原函数更平滑
- 卷积具有平移不变性

理解图卷积神经网络

■ Weisfeiler-Lehman isomorphism Test

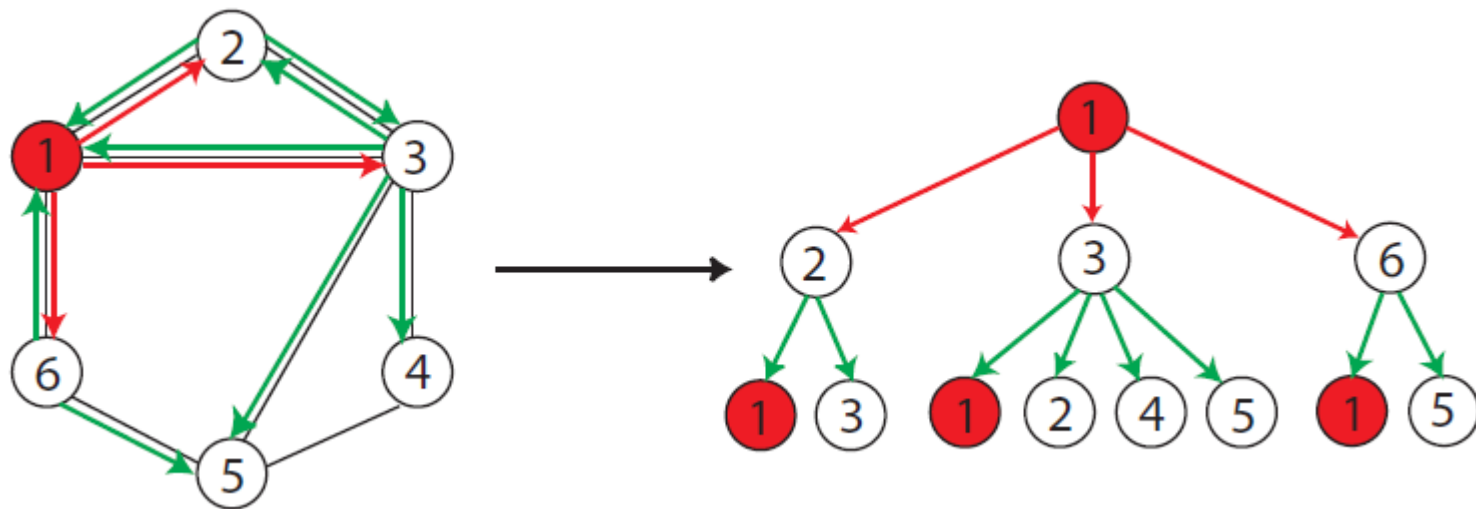


$$\phi_{WLsubtree}^{(1)}(G) = (2, 1, 1, 1, 1, 2, 0, 1, 0, 1, 1, 0, 1)$$

$$\phi_{WLsubtree}^{(1)}(G') = (1, 2, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1)$$

理解图卷积神经网络

■ Subtree kernel



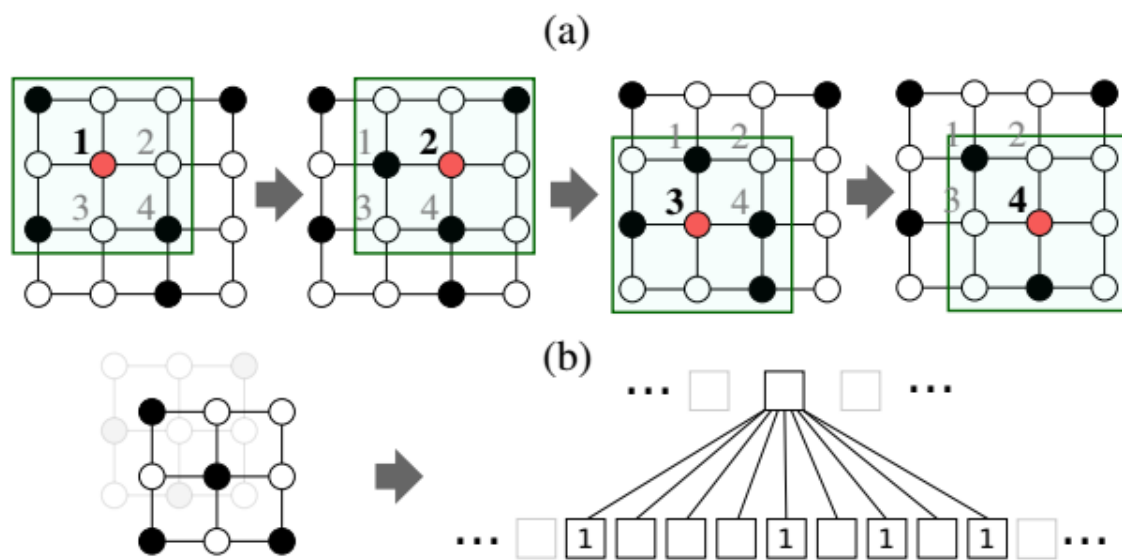
■ 单射聚合函数

- 将结构和节点特征映射为子图的表达

图卷积神经网络

■ 设计图卷积神经网络所面临的挑战

- 图数据不满足平移不变性，如何在图上定义卷积算子？
- 图数据规模大，如何实现参数共享从而降低参数个数和计算复杂性？



① 确定邻域



② 邻域定序

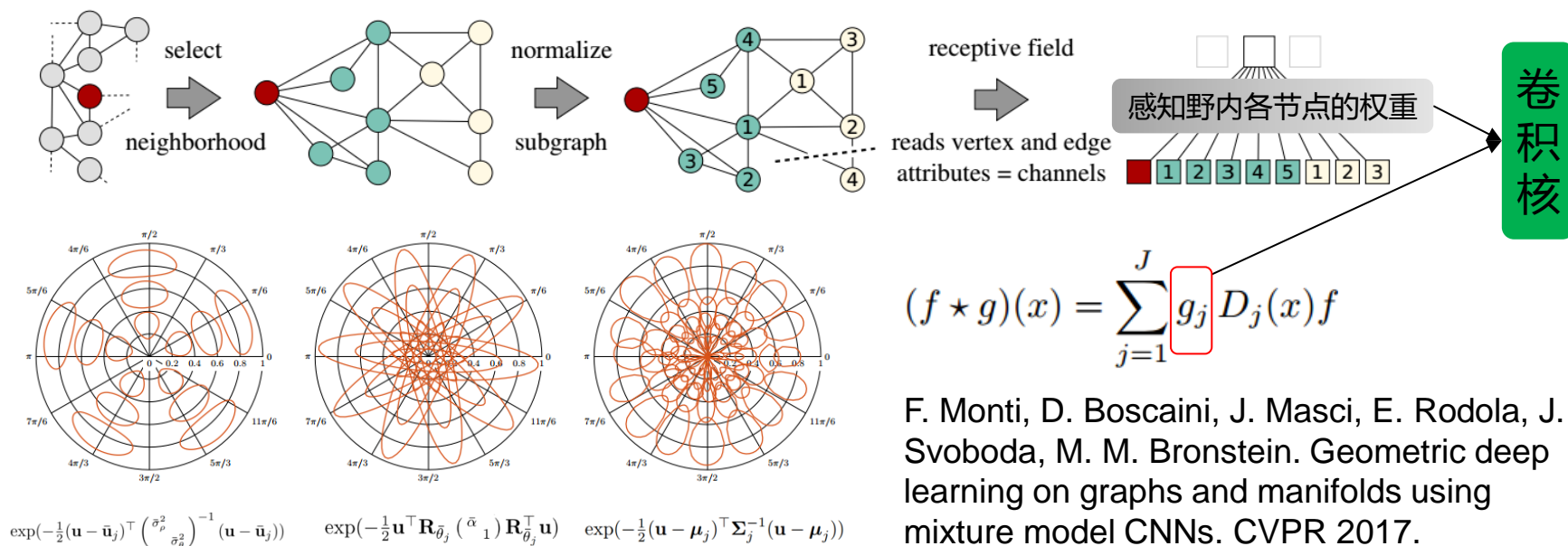


③ 参数共享

图卷积神经网络的现有方法

■ 空间域方法：Spatial Methods

- 类比的方式：类比CNN定义每个节点的**感知野**（同等数量的临近节点），在**空间域直接定义卷积核**实现参数共享
- 核函数方法：在大小不同的领域上定义多个核函数，然后将**卷积核定义为核函数的系数**实现参数共享

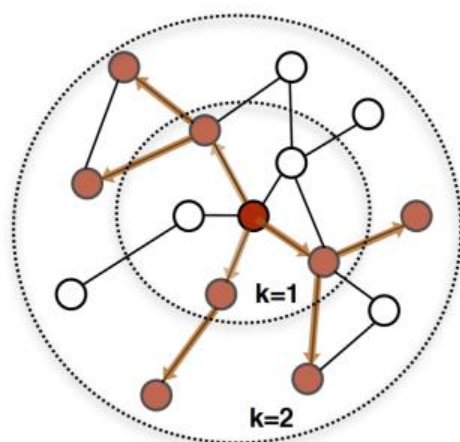


图卷积神经网络的现有方法

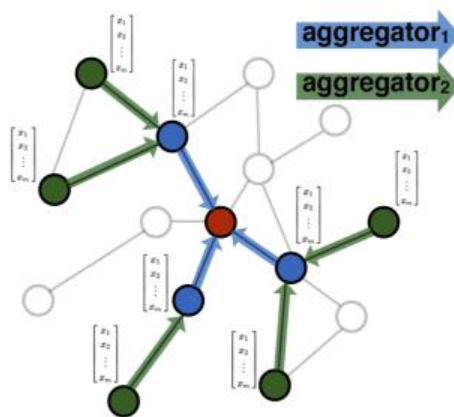
■ 空间域方法：Spatial Methods

□ GraphSAGE：邻居采样+聚合

□ GAT：采用注意力机制将聚合操作中的权重进行参数化

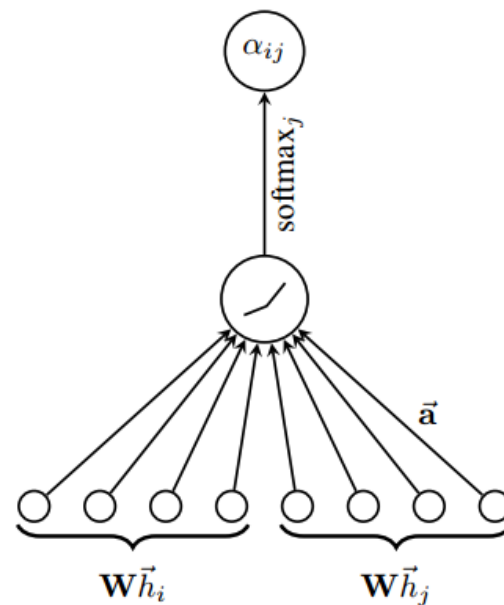


1. Sample neighborhood



2. Aggregate feature information from neighbors

GraphSAGE: Inductive Learning

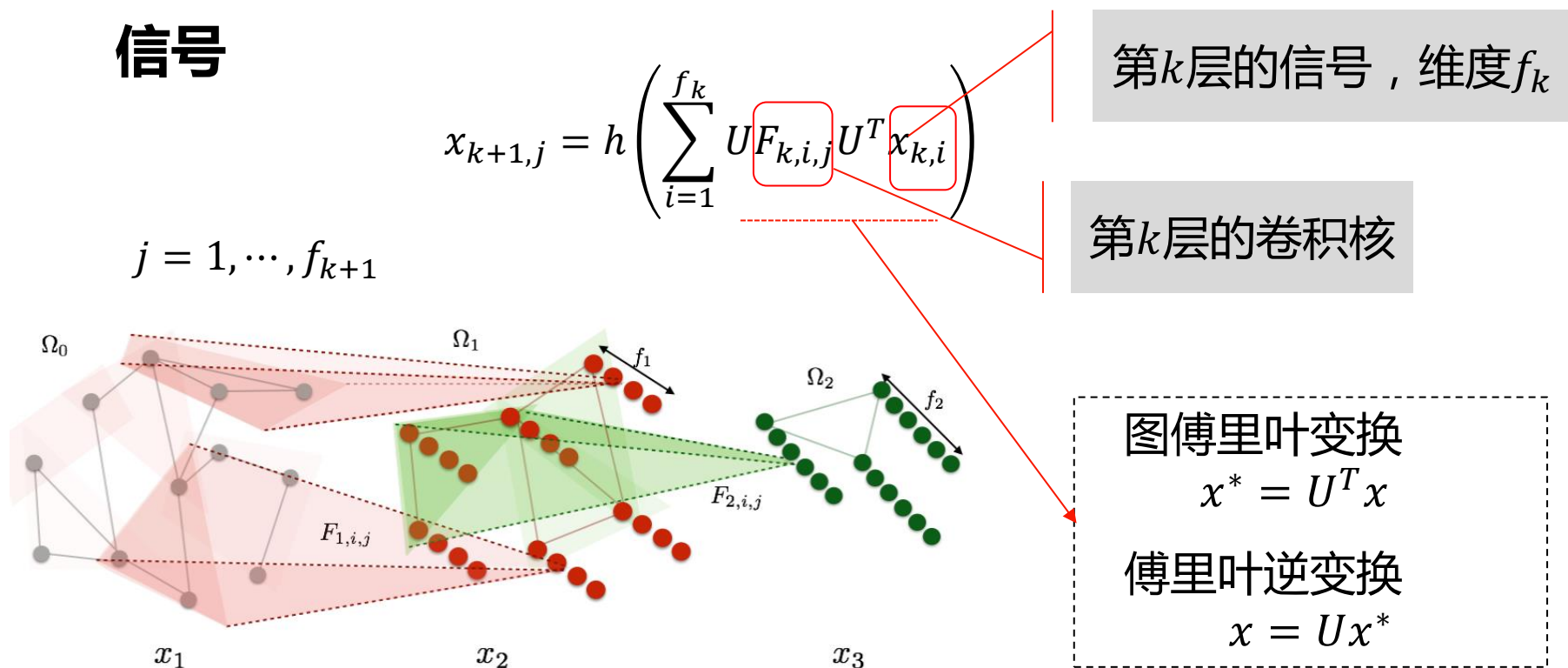


GAT

图卷积神经网络的现有方法

■ 谱域方法：Spectral Graph CNN

- 将图上的信号通过傅里叶变换映射到谱域，在谱域定义卷积核实现参数共享，然后通过傅里叶逆变换得到卷积后的信号



» 谱方法的不足和改进

■ 谱方法的不足

- 依赖于特征分解，**计算效率低**
- 卷积操作的**局部性差**

■ 改进方法：**卷积核参数化——多项式近似**

$$\text{diag}(g_1, g_2, \dots, g_n) = g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k \quad \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

- 卷积核的自由参数从 n 降低到了 K
- 计算复杂性从 $O(n^3)$ 降低到 $O(m * K)$ ， m 表示网络中的边数

基于小波变换的图卷积神经网络

Graph Wavelet Neural Network

(ICLR 2019)

» 基于小波变换的图卷积神经网络

- Graph Wavelet Neural Network (ICLR 2019)
 - 采用图小波变换代替图傅里叶变换实现信号从节点域到谱域的映射，在谱域定义卷积核实现参数共享，进而通过图小波逆变换得到卷积后的信号

Graph Fourier transform

$$x^* = U^T x$$

Inverse Fourier transform

$$x = U x^*$$

Graph Wavelet transform

$$x^* = \psi_s^{-1} x$$

Inverse Wavelet transform

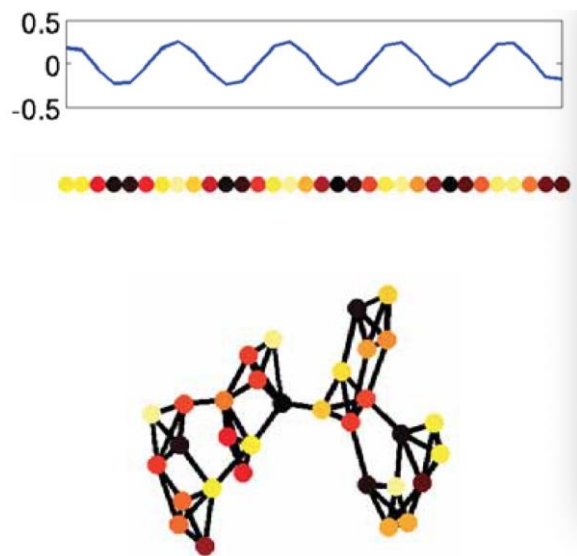
$$x = \psi_s x^*$$

注： U 是拉普拉斯矩阵的单位正交特征向量构成的矩阵
 ψ_s 是图小波变换的基

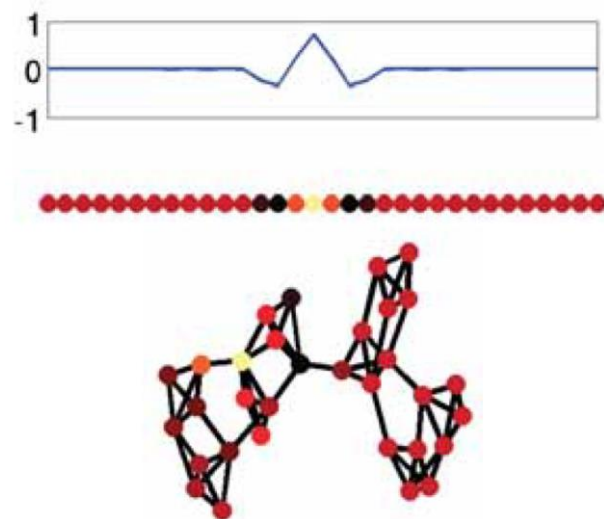
基于小波变换的图卷积神经网络

■ 小波变换的优点

- 傅里叶变换的基是全域信号、小波变换的基是局域信号(稀疏)
- 傅里叶变换计算效率低、小波变换**计算效率高**
- 基于小波变换的图卷积**局部性好**



傅里叶基： U



小波基： $\psi_s = Ue^{\lambda s}U^T$

基于小波变换的图卷积神经网络

基于小波变换的图卷积

$$x *_{\mathcal{G}} y = U((U^{\top} y) \odot (U^{\top} x)),$$

$$x *_{\mathcal{G}} y = \psi_s((\psi_s^{-1} y) \odot (\psi_s^{-1} x))$$

换基

图小波卷积神经网络

$$x_{k+1,j} = h\left(\sum_{i=1}^p U F_{k,i,j} U^{\top} x_{k,i}\right) \rightarrow x_{k+1,j} = h\left(\sum_{i=1}^p \psi_s F_{k,i,j} \psi_s^{-1} x_{k,i}\right)$$

$$j = 1, \dots, q$$

参数复杂性： $O(n * p * q)$

基于小波变换的图卷积神经网络

■ 创新点

- 传统的方法：每一组特征变换单独拥有一个卷积核，卷积和特征变换是耦合的
- 我们提出：将特征变换和图卷积解耦

■ 大大降低了参数复杂性

- 传统的谱方法： $O(n * p * q)$
- 多项式近似的谱方法： $O(K * p * q)$
- 图小波卷积神经网络： $O(n + p * q)$

对于大规模网络，
参数复杂度有数个
数量级的降低

» 基于小波变换的图卷积神经网络

■ 应用：图上的半监督分类

- 部分节点有标签，根据节点特征和网络结构对没有标签的节点打上标签

■ 传统方法

- 标签传播
- 监督学习+图正则

$$\mathcal{L} = \mathcal{L}_0 + \lambda \mathcal{L}_{\text{reg}}$$

$$\mathcal{L}_{\text{reg}} = \sum_{ij} A_{ij} \|f(X_i) - f(X_j)\|^2$$

■ 图卷积神经网络

- 完全监督

$$\mathcal{L} = f(X, A)$$

- 前向进行特征扩散、
后向进行标签传播

基于小波变换的图卷积神经网络

数据集

Dataset	Nodes	Edges	Classes	Features	Label Rate
Citeseer	3,327	4,732	6	3,703	0.036
Cora	2,708	5,429	7	1,433	0.052
Pubmed	19,717	44,338	3	500	0.003

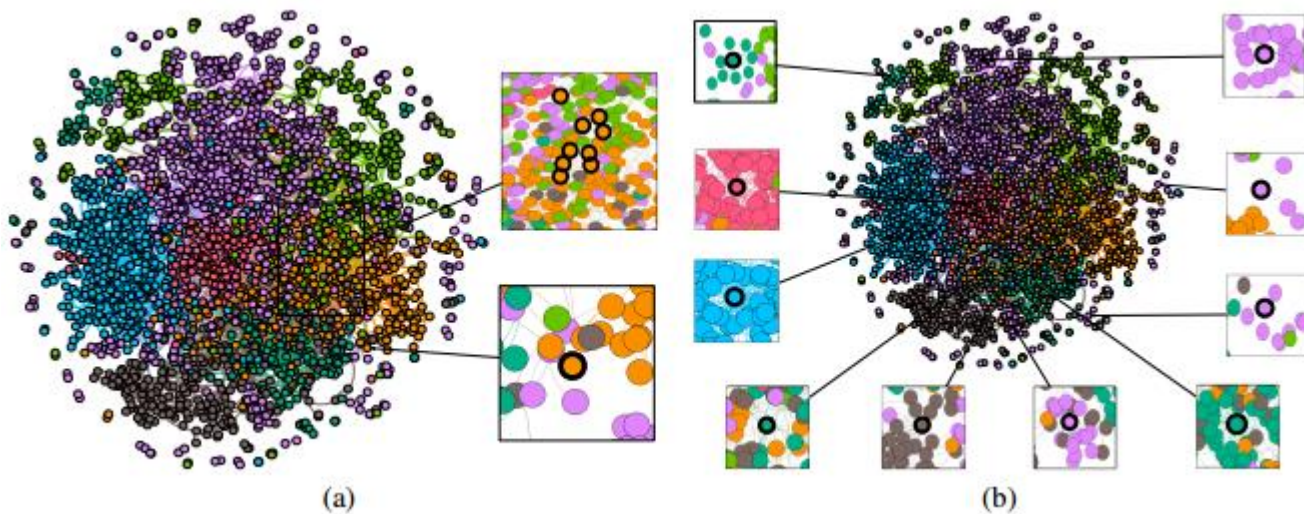
结果

Method	Cora	Citeseer	Pubmed
MLP	55.1%	46.5%	71.4%
ManiReg	59.5%	60.1%	70.7%
SemiEmb	59.0%	59.6%	71.7%
LP	68.0%	45.3%	63.0%
DeepWalk	67.2%	43.2%	65.3%
ICA	75.1%	69.1%	73.9%
Planetoid	75.7%	64.7%	77.2%
Spectral CNN	73.3%	58.9%	73.9%
ChebyNet	81.2%	69.8%	74.4%
GCN	81.5%	70.3%	79.0%
MoNet	81.7±0.5%	—	78.8±0.3%
GWNN	82.8%	71.7%	79.1%

基于小波变换的图卷积神经网络

■ 基的可解释性

- 以节点 i 为中心的小波体现了在 i 的视角下其他节点在空间域中的临近性
- 选择合适的基相当于在空间域选择合适的核函数



» 小结

■ 图卷积神经网络

- 卷积神经网络在非欧数据（例如：图）上的拓展应用
- 特别适合网络化数据的处理，实现数据在时间、空间、关联约束下的简约计算

■ 发展趋势

- 参数共享和计算效率问题是未来要解决的问题
- 高维短时数据的建模和预测：交通、股市、社交媒体等

面向图半监督学习的热核图卷积神经网络

**Graph Convolutional Networks using Heat
Kernel for Semi-supervised Learning
(IJCAI 2019)**

» 从空间方法的角度审视谱方法

- 联系：谱方法是空间方法的**特例**



$$(f \star g)(x) = \sum_{j=1}^J g_j D_j(x) f$$

核函数：
刻画节点相似度
或距离

- 区别：

- 谱方法定义核函数时需要**显式**的空间映射

$$D_i = u_i u_i^T, \quad \text{这里 } u_i \text{ 定义了显式的空间映射}$$

- 空间方法直接定义核函数，不需要显式的空间映射

» 谱方法回顾：空间方法视角

■ Spectral CNN

$$y = U g_{\theta} U^T x = (\theta_1 \boxed{u_1 u_1^T} + \theta_2 \boxed{u_2 u_2^T} + \cdots + \theta_n \boxed{u_n u_n^T}) x$$

■ ChebyNet

$$y = (\theta_0 I + \theta_1 L + \theta_2 L^2 + \cdots + \theta_{K-1} L^{K-1}) x$$

■ GCN

$$y = \theta(I - L)x$$

思考这样一个问题：

GCN为什么在半监督学习任务上能够优于ChebyNet？

图信号处理：滤波

■ 信号关于网络的平滑程度

$$x^T L x = \sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{d_u}} - \frac{x_v}{\sqrt{d_v}} \right)^2$$

x 表示图上的信号， A 是网络的邻接矩阵， E 是网络的边集， L 是拉普拉斯矩阵， d_u 表示节点 u 的度

■ 基础滤波器

- $u_i u_i^T$ ($1 \leq i \leq n$) 构成一组基础滤波器， $\lambda_i = u_i^T L u_i$
- 对于一个信号 x ， $u_i u_i^T$ 只允许频率为 λ_i 的信号分量通过

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n,$$

$$u_i u_i^T x = \alpha_i u_i$$

组合滤波器：高通 vs. 低通

■ 组合滤波器

- 由基础滤波器的线性组合构成

$$\theta_1 u_1 u_1^T + \theta_2 u_2 u_2^T + \cdots + \theta_n u_n u_n^T$$

- L^k 是组合滤波器的一个例子，对应的系数为： $\{\lambda_i^k\}_{i=1}^n$
- L^k 赋予高频基础滤波器更大的权重，是一种高通滤波器

■ GCN只考虑 $k = 0$ 和 $k = 1$ 两阶，具有更好的平滑性

- 部分解释了GCN在图半监督学习任务上比ChebyNet好

思考方向：构造抑制高频信号的组合滤波器，来刻画平滑性

» GraphHeat : 热核图卷积神经网络

■ 低通组合滤波器

$\{e^{-skL}\}$, s 是尺度参数, k 是阶数

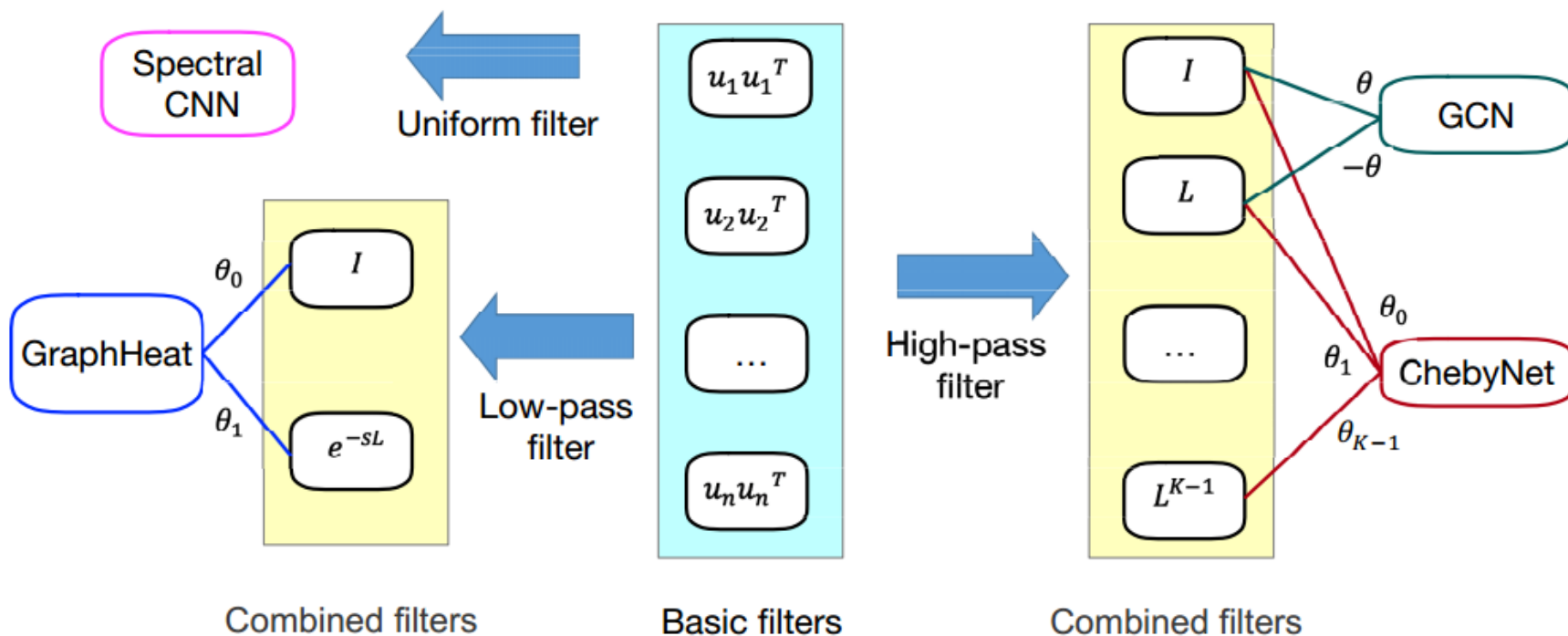
- e^{-sL} 是图的热核 (Heat Kernel) , 该核函数通过热扩散定义节点间的相似度

$$e^{-sL} = Ue^{-s\Lambda}U^T, \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

- $u_i u_i^T$ ($1 \leq i \leq n$) 基础滤波器的系数为 $e^{-s\lambda_i}$, 从而抑制了高频信号

GraphHeat : 热核图卷积神经网络

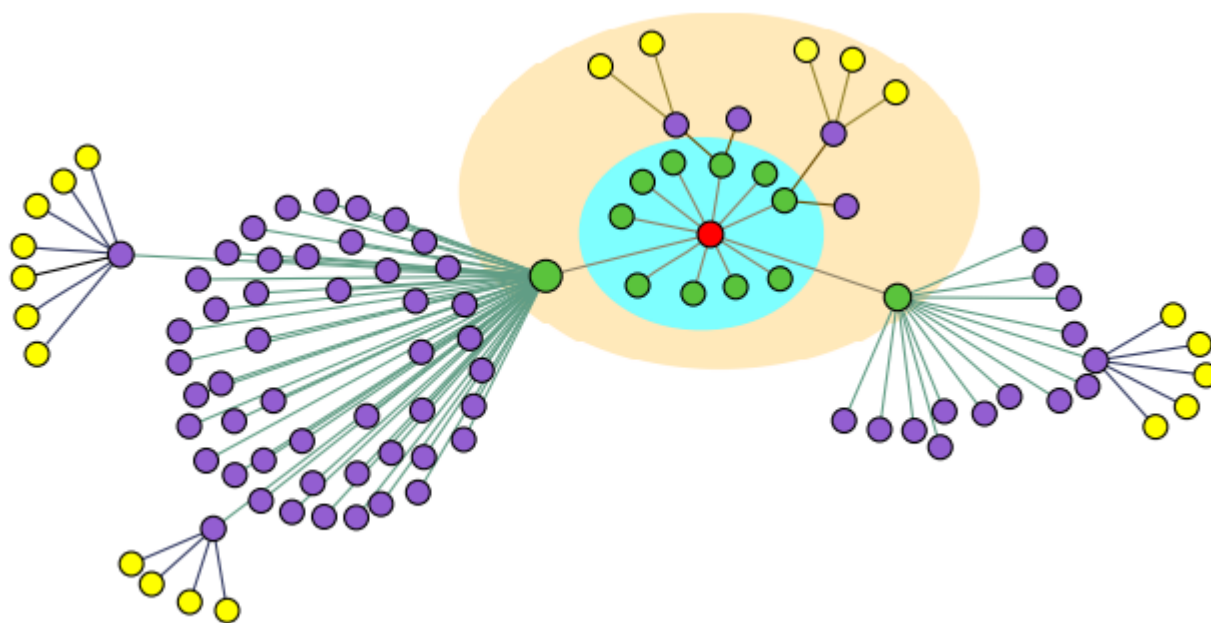
方法对比



邻域对比

■ GCN和GraphHeat的邻域对比

- GCN：按照最短路径定义临近(颜色区分)
- GraphHeat：按照热扩散距离定义临近(圆圈标记)



» GraphHeat : 热核图卷积神经网络

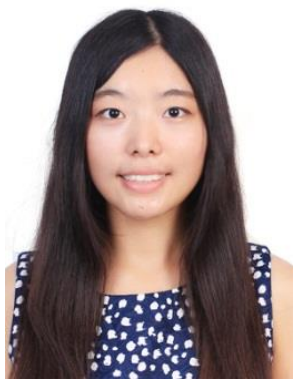
■ 在节点分类任务上的结果：

Method	Cora	Citeseer	Pubmed
MLP	55.1%	46.5%	71.4%
ManiReg	59.5%	60.1%	70.7%
SemiEmb	59.0%	59.6%	71.7%
LP	68.0%	45.3%	63.0%
DeepWalk	67.2%	43.2%	65.3%
ICA	75.1%	69.1%	73.9%
Planetoid	75.7%	64.7%	77.2%
ChebyNet	81.2%	69.8%	74.4%
GCN	81.5%	70.3%	79.0%
MoNet	81.7±0.5%	—	78.8±0.3%
GAT	83.0±0.7%	72.5±0.7%	79.0±0.3%
GraphHeat	83.7%	72.5%	80.5%

» 几点思考

- 应用场景还比较少，数据集也不多
 - 节点级
 - 半监督学习：node classification
 - 链路预测、推荐等
 - 网络级：graph classification
 - 信号级：应用较少
- 规模能否上的去
 - PinSAGE：GraphSAGE在Pinterest真实场景上的应用
- 图神经网络（GNN）到底解决了什么问题？
 - 和Network Embedding孰优孰劣？

» 致谢



徐冰冰



曹琦



岑科廷



仇韞琦



程学旗