基于采样的
图神经网络高效训练

罗意

2021-11-19

目录

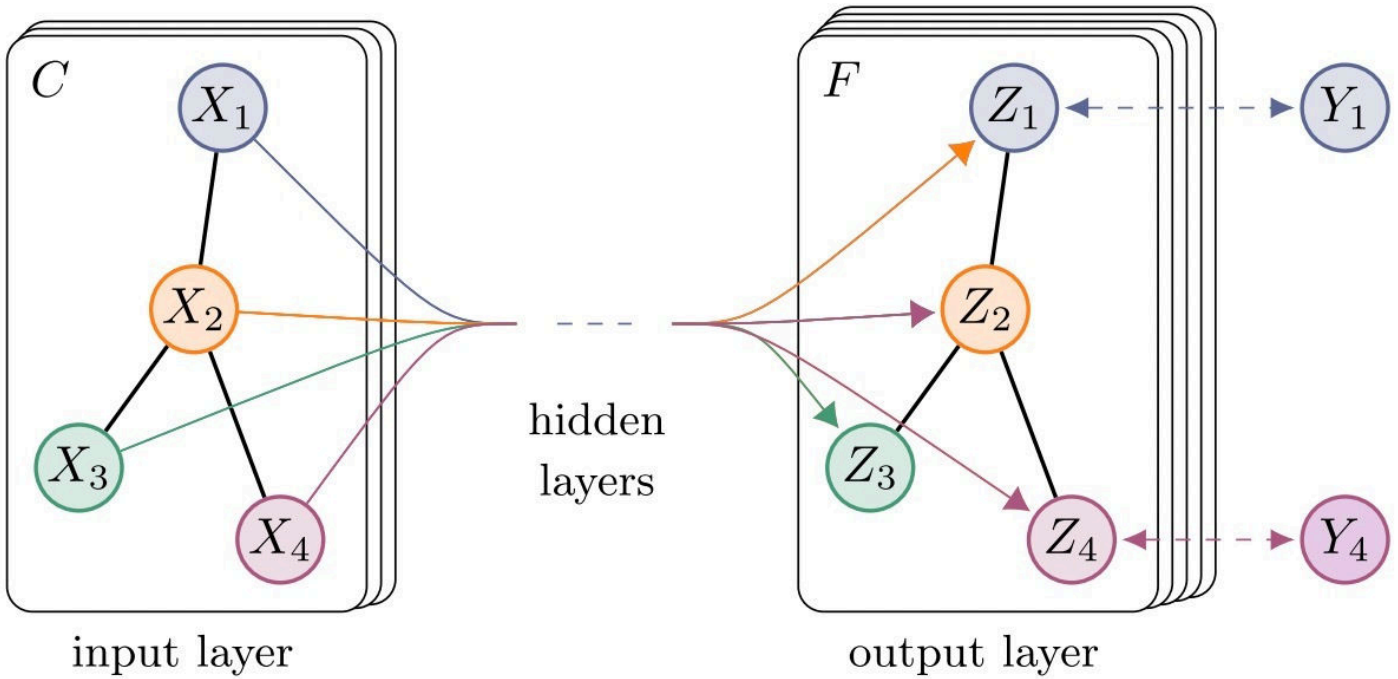
CONTENTS

1

CONTENTS

图神经网络GNN

◆ 图神经网络与节点分类



◆ 图神经网络的一般方式

1. 聚合节点邻域特征
2. 变换

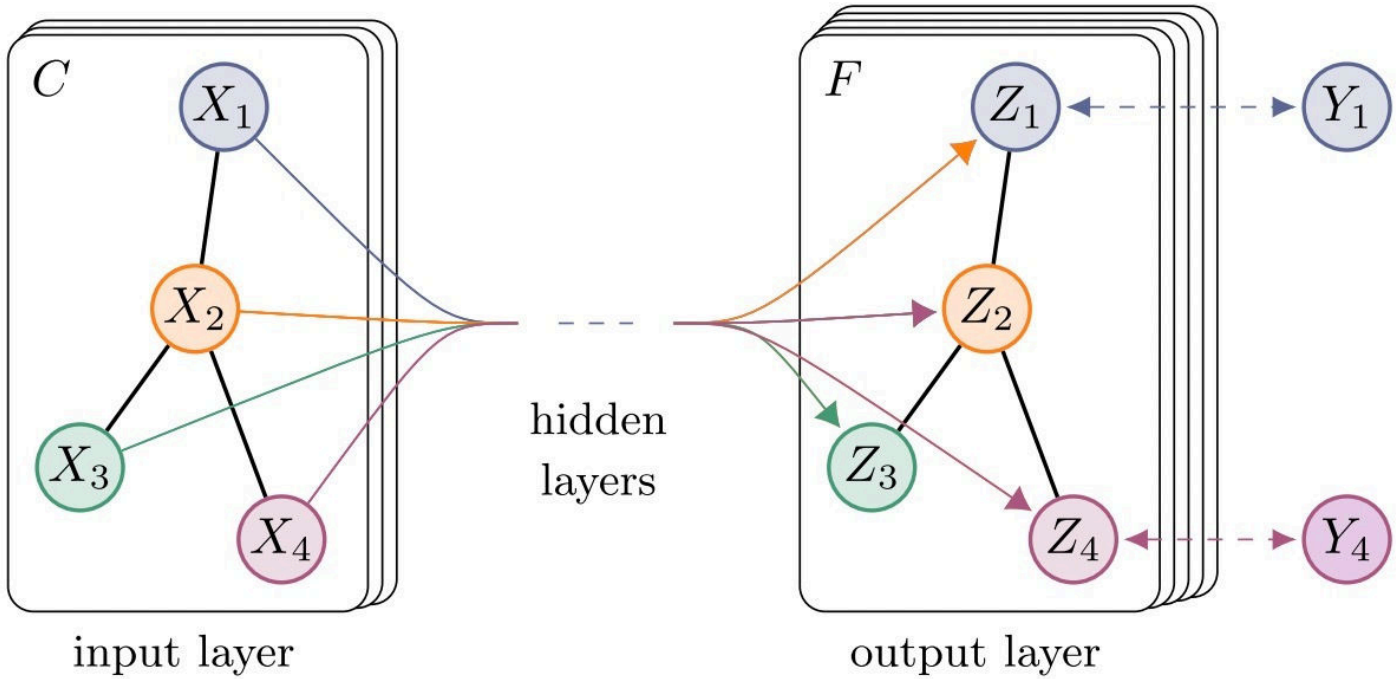
T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," ICLR, 2017.

数据集趋势

发布时间	数据集	节点数	连接数	特征维度	标签种类
2008	Cora	2708	5429	1433	7
	Citeseer	3327	4732	3703	6
	Pubmed	19717	44338	500	3
2017	Reddit	232965	11606919	602	41
2020	Yelp	716847	6977410	300	100
	ogbn-products	2449029	61859140	100	47
	ogbn-papers100M	111059956	1615685872	-	172

Weihua Hu, et al, "Open Graph Benchmark: Datasets for Machine Learning on Graphs," NeurIPS 2020

GNN大规模应用中的问题



- 多层卷积时计算量指数级增长引发的**难扩展**问题
- 全量训练（full-batch）收敛慢产生的**低效**问题

采样方法

◆ 采样

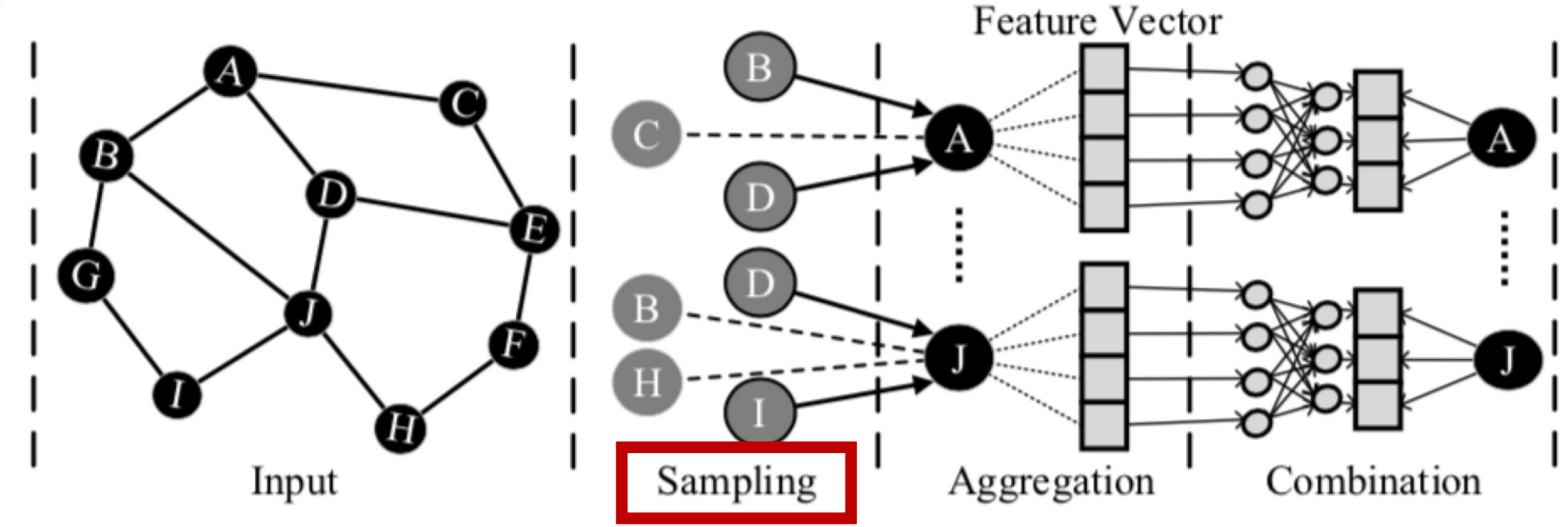
在聚合前仅选择部分节点作为全部节点的估计，以有限的精度损失，提升GNN的扩展性和学习效率。

1. 扩展性：限制多层卷积涉及的节点规模，减小计算量
2. 学习效率：支持小批量（mini-batch）训练，加快参数更新

◆ 对采样方法的要求

1. 提出采样策略
2. 证明以该策略采样，估计量是无偏的
3. 减小以该策略采样时估计量的方差

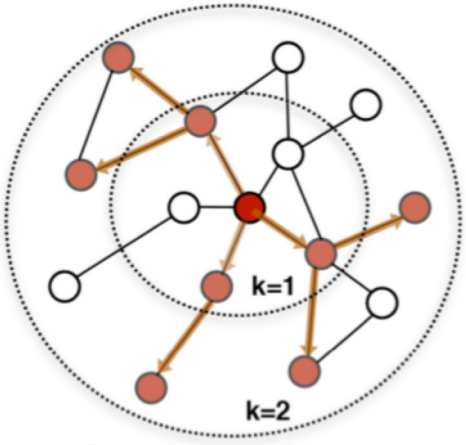
在大规模图中基于采样训练GNN

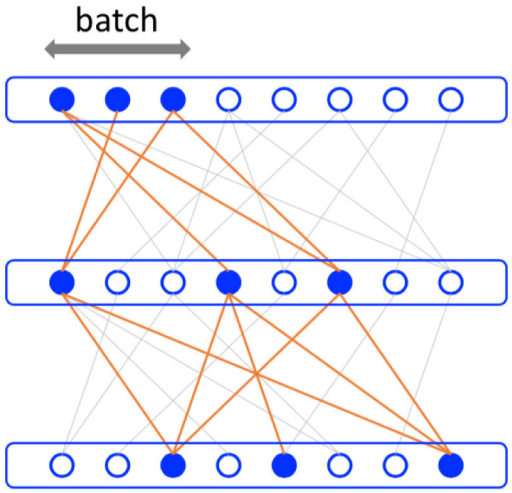


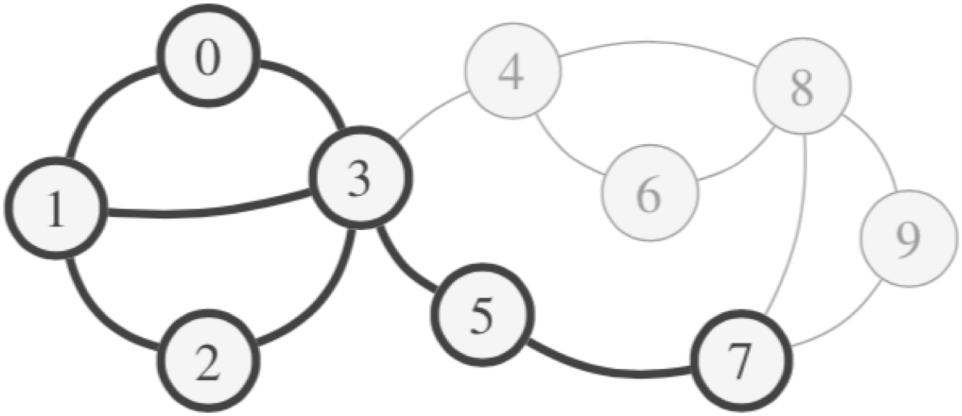
- **基于采样训练GNN**
  1. 采样部分节点作为整体的估计
  2. 聚合节点邻域特征
  3. 变换

Xin Liu, et al, "Sampling methods for efficient training of graph convolutional networks: a survey," IEEE CAA J. Autom. Sinica (2022)

基于采样训练GNN的方法分类

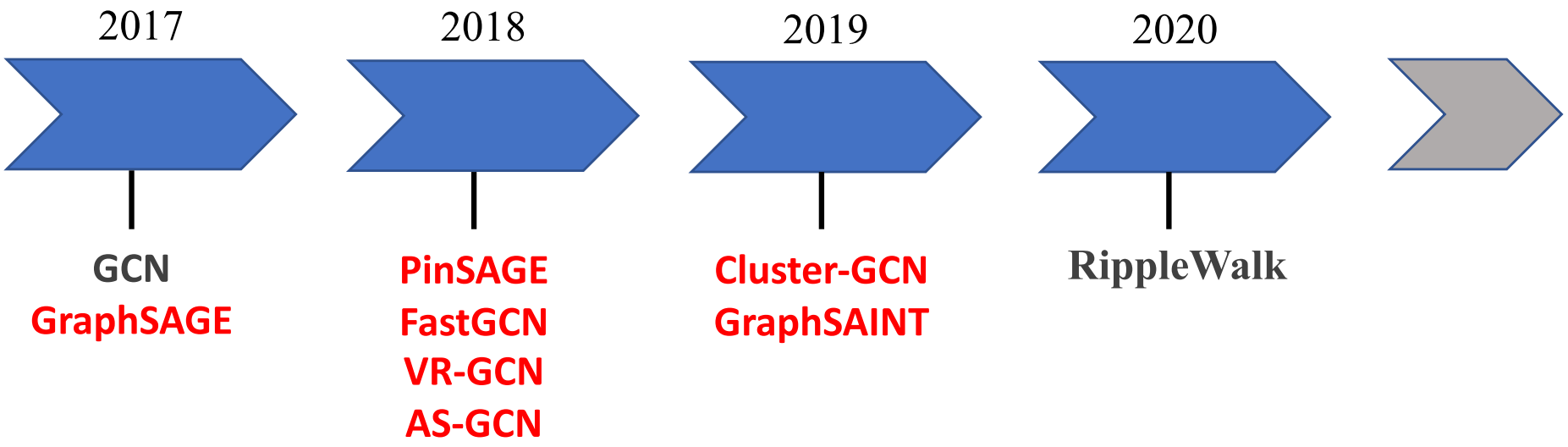






- 采样邻域节点的方法（node-wise）
- 分层采样节点的方法（layer-wise）
- 基于子图样本的方法（subgraph-based）

基于采样训练GNN方法的发展



时间	类型	方法	无偏性	有效性	解决问题	产生问题
2017	Node	GraphSAGE	否	否	扩展、效率	邻域爆炸
2018	Node	PinSAGE	否	否	邻域爆炸	启发式
2018	Layer	FastGCN	否	是	邻域爆炸	稀疏连接
2018	Node	VR-GCN	是	是	邻域爆炸	高内存
2018	Layer	AS-GCN	是	是	稀疏连接	采样复杂
2019	Subgraph	Cluster-GCN	否	否	邻域爆炸	引入偏差
2019	Subgraph	GraphSAINT	是	是	偏差、方差	-

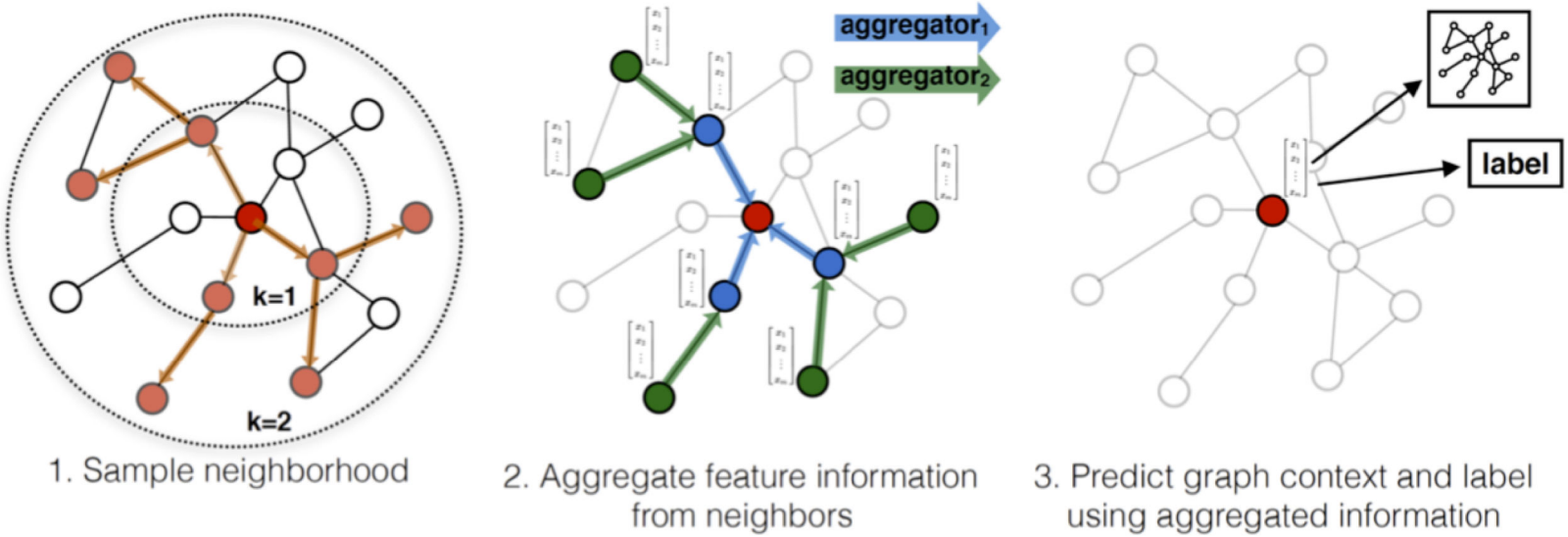
2

CONTENTS

采样邻域节点的方法

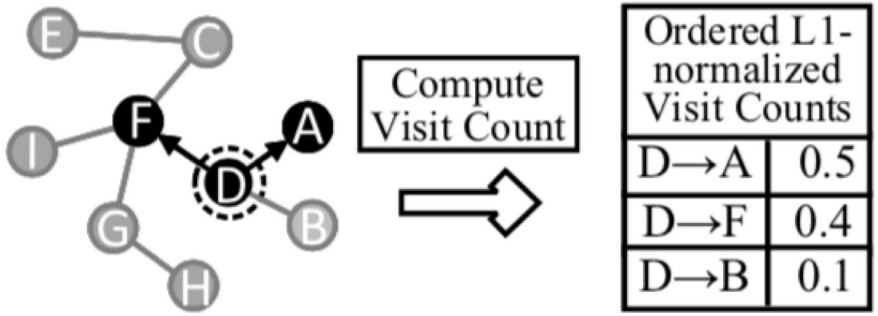
GraphSAGE/PinSAGE

- **GraphSAGE：邻域爆炸**



1. Sample neighborhood　2. Aggregate feature information from neighbors　3. Predict graph context and label using aggregated information

- **PinSAGE：寻求相关性更强的1-hop邻居**



Ordered L1-normalized Visit Counts	
D→A	0.5
D→F	0.4
D→B	0.1

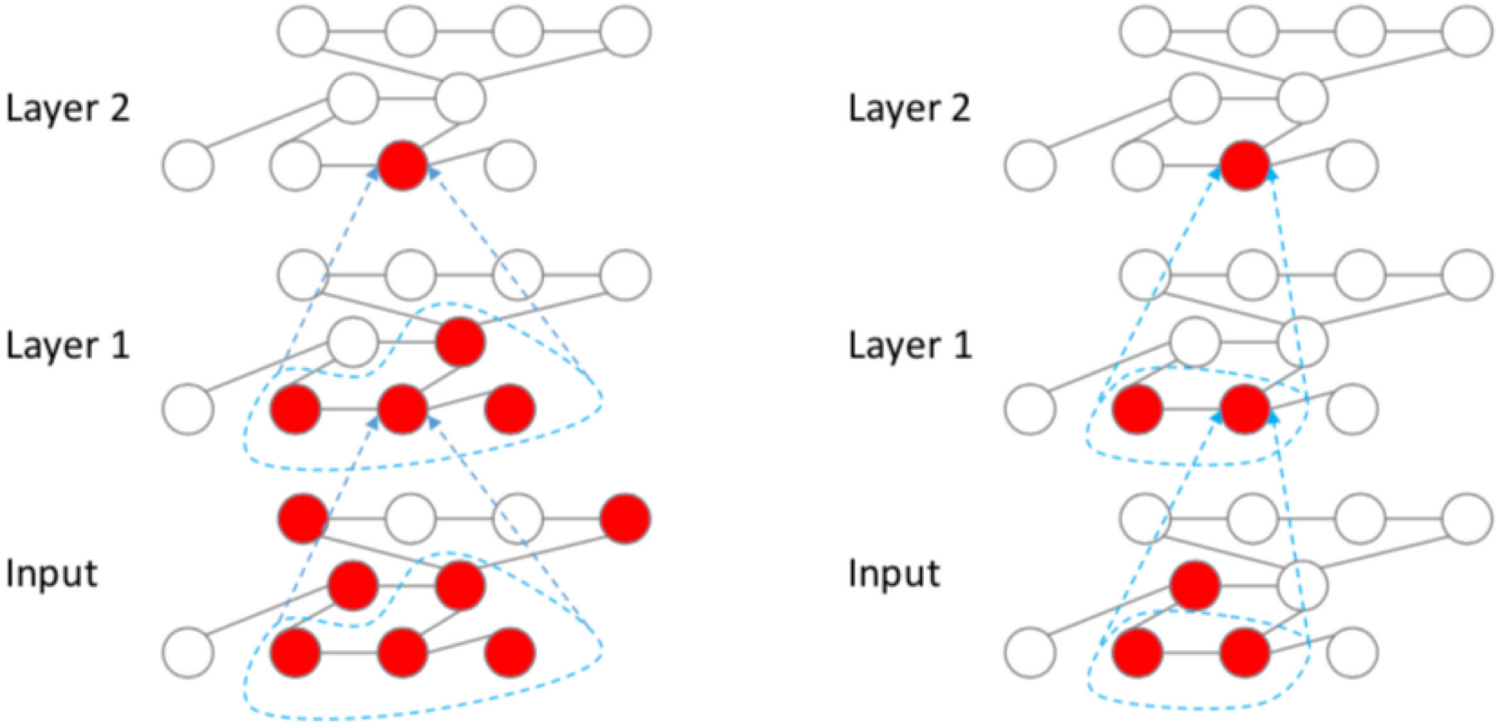
W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," NIPS, 2017
R. Ying, R. He, et al, "Graph convolutional neural networks for web-scale recommender systems," SIGKDD, 2018

VR-GCN

◆ 采样方法

感受野内每次只采样自己和1个邻居，邻域线性扩大

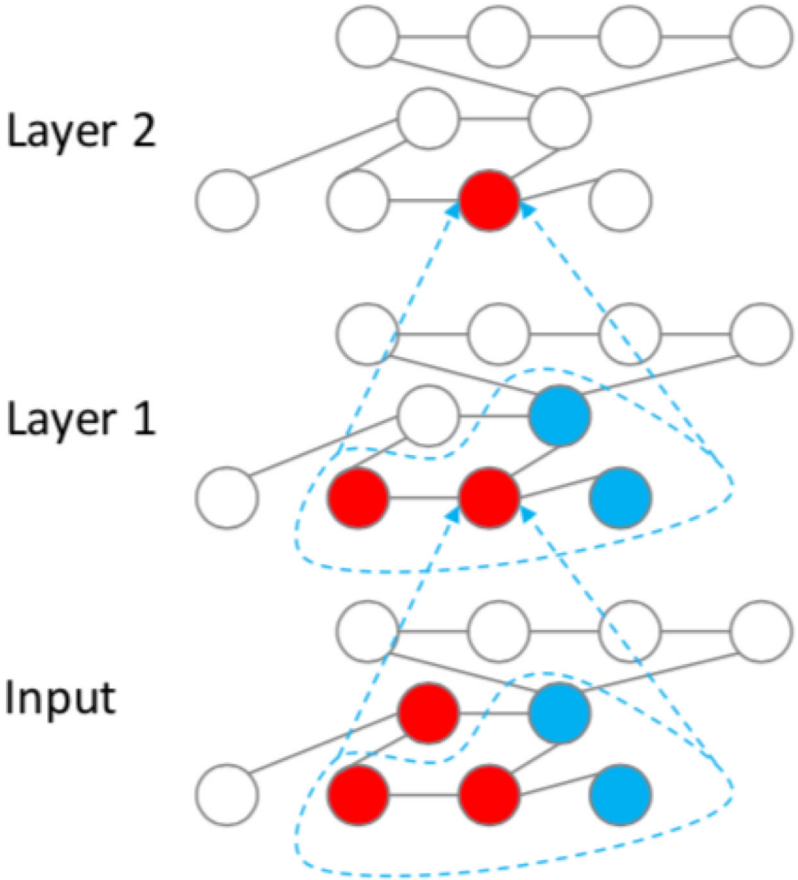
（对比GraphSAGE采样25&10）

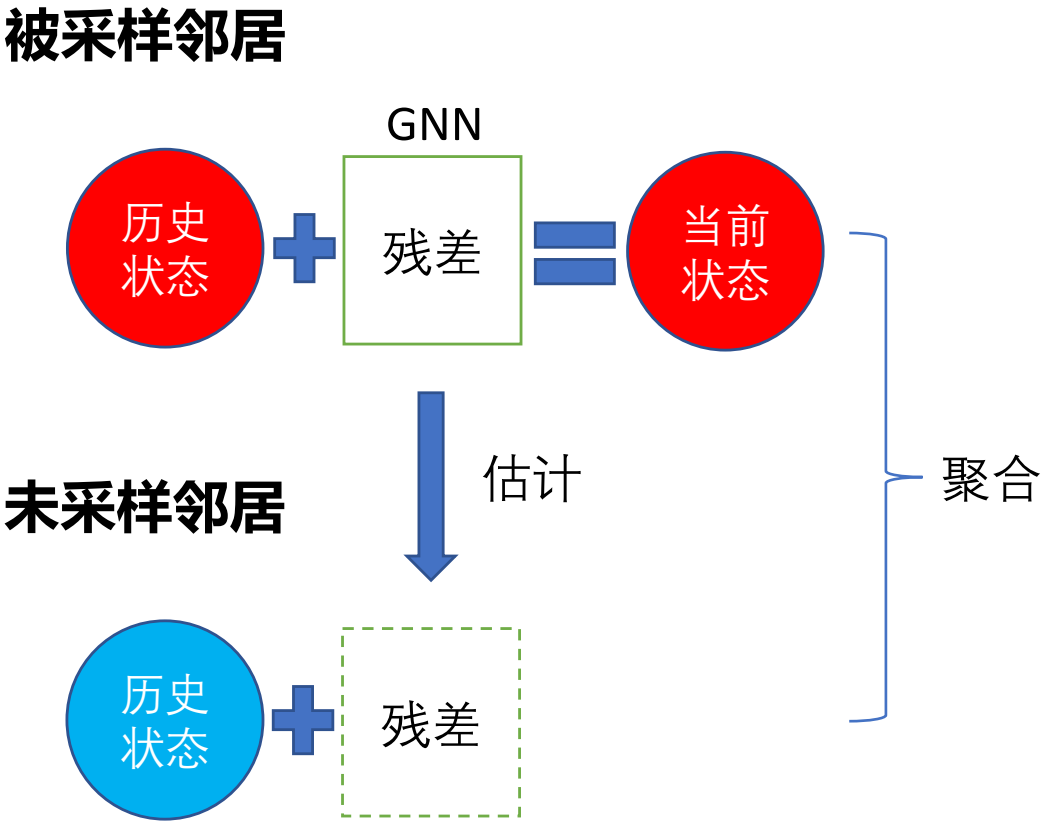


J. Chen, J. Zhu, and L. Song, "Stochastic training of graph convolutional networks with variance reduction," ICML, 2018

VR-GCN

◆ **方差控制**


Layer 2
Layer 1
Input


被采样邻居
GNN
历史状态
残差
当前状态
估计
聚合
未采样邻居
历史状态
残差

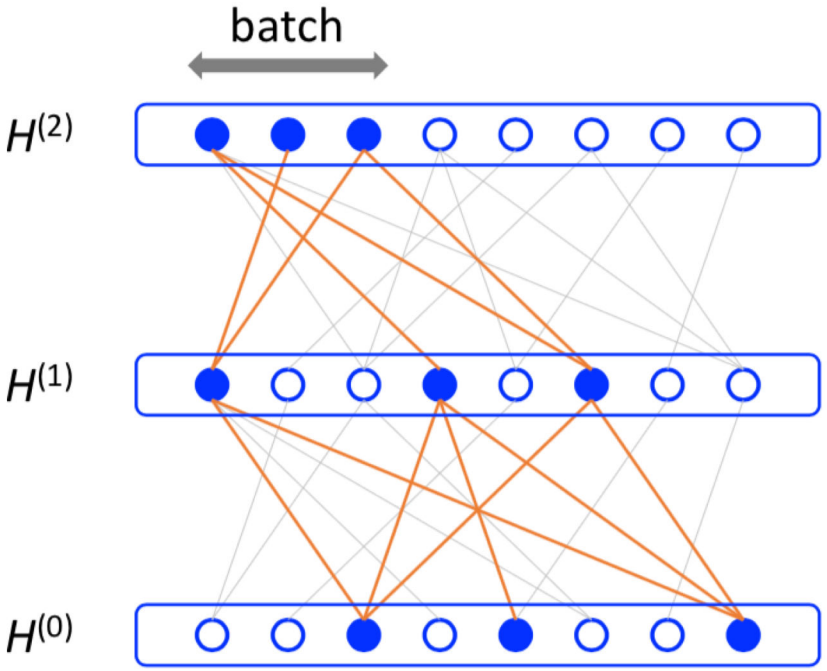
3

CONTENTS

分层采样节点的方法

FastGCN

◆ 采样方法



估计量

$$h^{(l+1)}(v) = \sigma(E_p[\hat{A}(v,u)h^{(l)}(u)W^{(l)}])$$

引入重要性采样

$$h^{(l+1)}(v) = \sigma(E_q[\hat{A}(v,u)h^{(l)}(u)W^{(l)} \cdot \frac{p(u)}{q(u)}])$$

理论最优的q

$$q^* \propto \sqrt{E_v[\hat{A}(v,u)^2]} \cdot |h^{(l)}(u)W^{(l)}| \cdot p(u)$$

简言之：邻居越多越重要

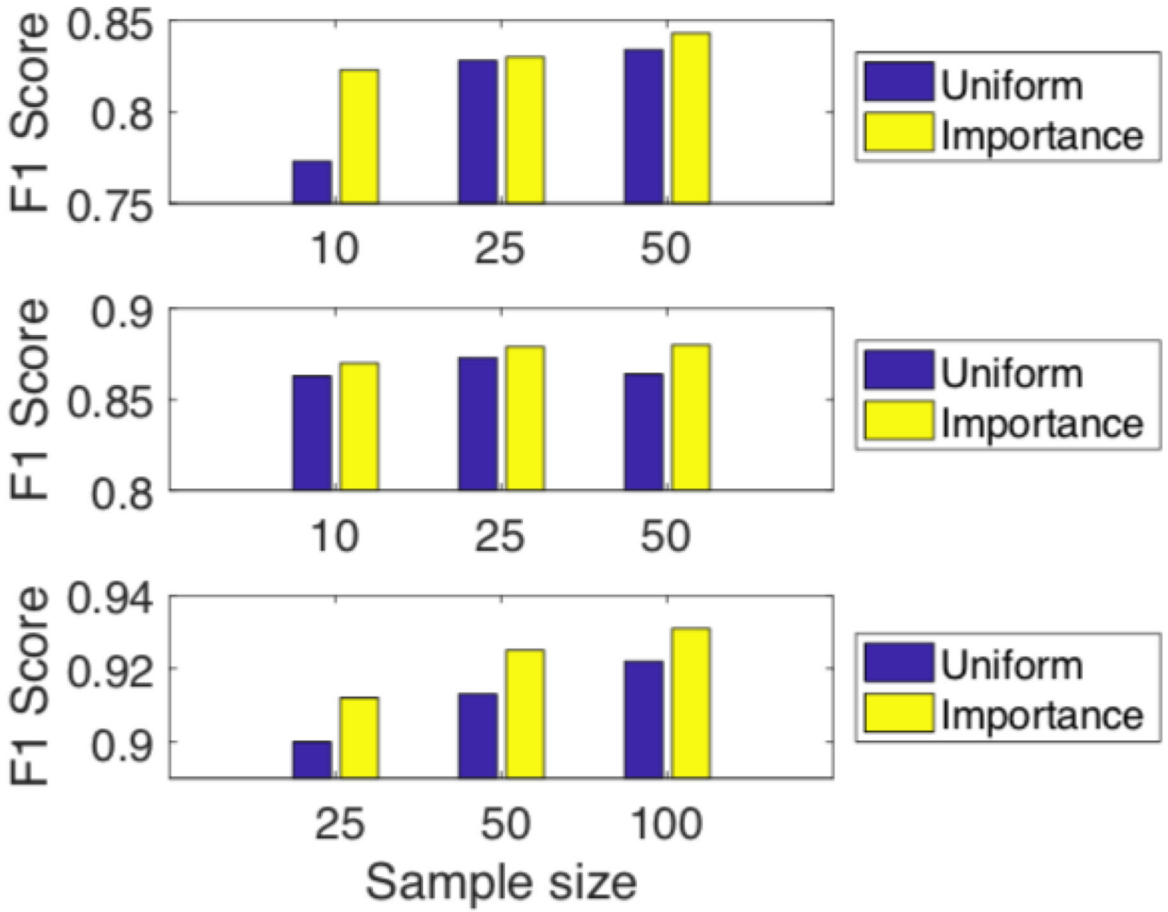
$$\hat{q} \propto \sqrt{E_v[\hat{A}(v,u)^2]} = ||\hat{A}(:,u)||^2$$

J. Chen, T. Ma, and C. Xiao, "Fastgcn: Fast learning with graph convolutional networks via importance sampling," ICLR, 2018

重要性采样的重要性

- **重要性采样降方差后，效果比均匀采样更好**

数据集为Cora, Pubmed, Reddit



AS-GCN

◆ 按条件采样节点

解决稀疏连接问题

◆ 方差控制

最优采样分布 $$q^*(u_j) = \frac{p(u_j|v_i)|h^{(l)}(u_j)|}{\sum_{j=1}^{N} p(u_j|v_i)|h^{(l)}(u_j)|}$$

方差 $$\mathrm{Var}_q(\hat{\mu}_q(v_i)) = \frac{1}{n}\mathbb{E}_{q(u_j)}[\frac{(p(u_j|v_i)|h^{(l)}(u_j)| - \mu_q(v_i)q(u_j))^2}{q^2(u_j)}].$$

引入可学习的线性变换W，假装 $h(u_i) = Wx_i$

一举两得：

1. 可用于指导采样
2. 方差项作为正则项进行优化

W. Huang, T. Zhang, Y. Rong, and J. Huang, "Adaptive sampling towards fast graph representation learning," Advances in Neural Information Processing Systems, 2018

4

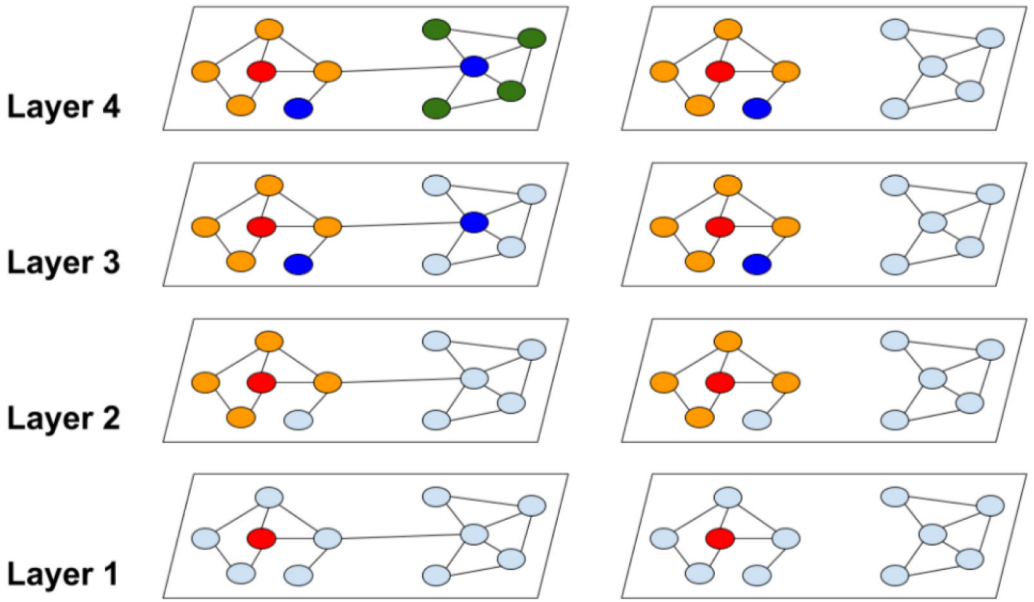
CONTENTS

基于子图样本的方法

Cluster-GCN

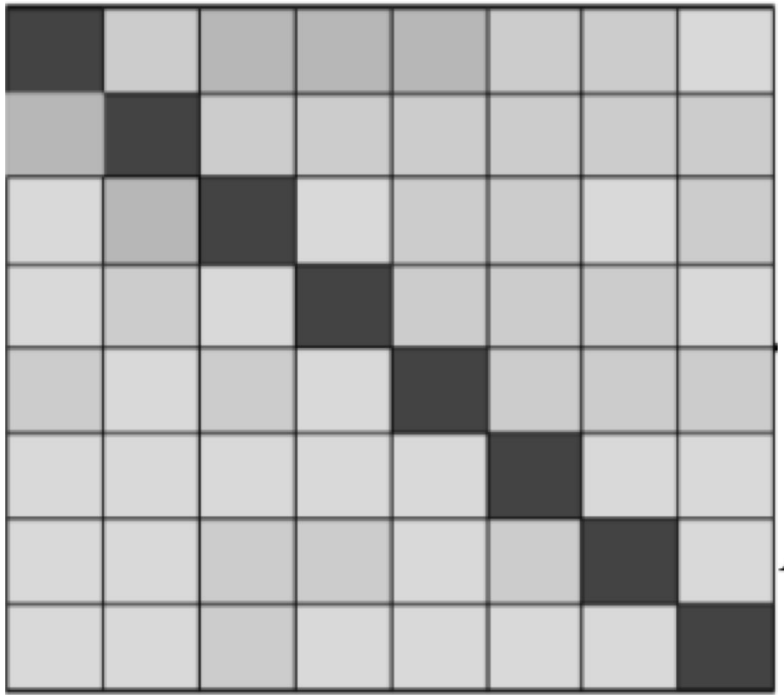
◆ Motivation

1. 图网络中存在按社群聚集现象
2. 一批被采样的节点间连接数越多，其表示的学习应越充分
3. 所以跨社群的采样得不偿失



W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh,
"Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks," SIGKDD, 2019

Cluster-GCN



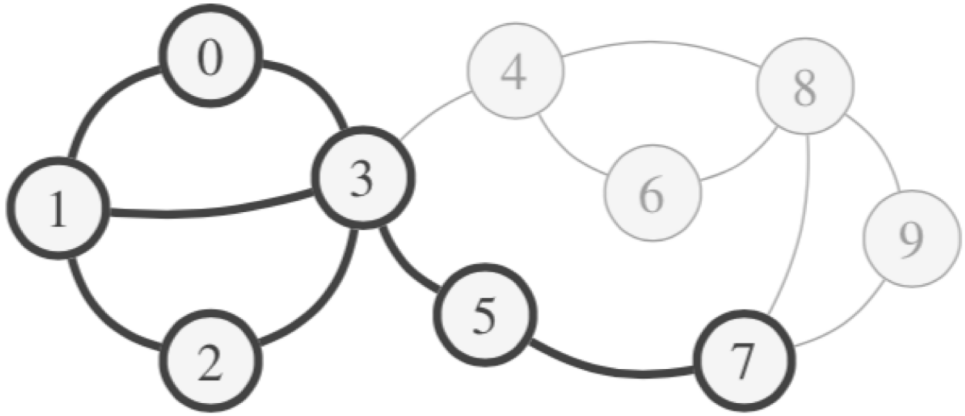
$$A = \bar{A} + \Delta = \begin{bmatrix} A_{11} & \cdots & A_{1c} \\ \vdots & \ddots & \vdots \\ A_{c1} & \cdots & A_{cc} \end{bmatrix}$$

$$\bar{A} = \begin{bmatrix} A_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{cc} \end{bmatrix}, \Delta = \begin{bmatrix} 0 & \cdots & A_{1c} \\ \vdots & \ddots & \vdots \\ A_{c1} & \cdots & 0 \end{bmatrix}$$

- **步骤**

1. 使用社群发现算法划分图为c个cluster
2. 每个batch采样若干个cluster，拼成子图（包括cluster间的连接）
3. 在子图上使用GCN

GraphSAINT



- **三个采样器（无偏、有效）**

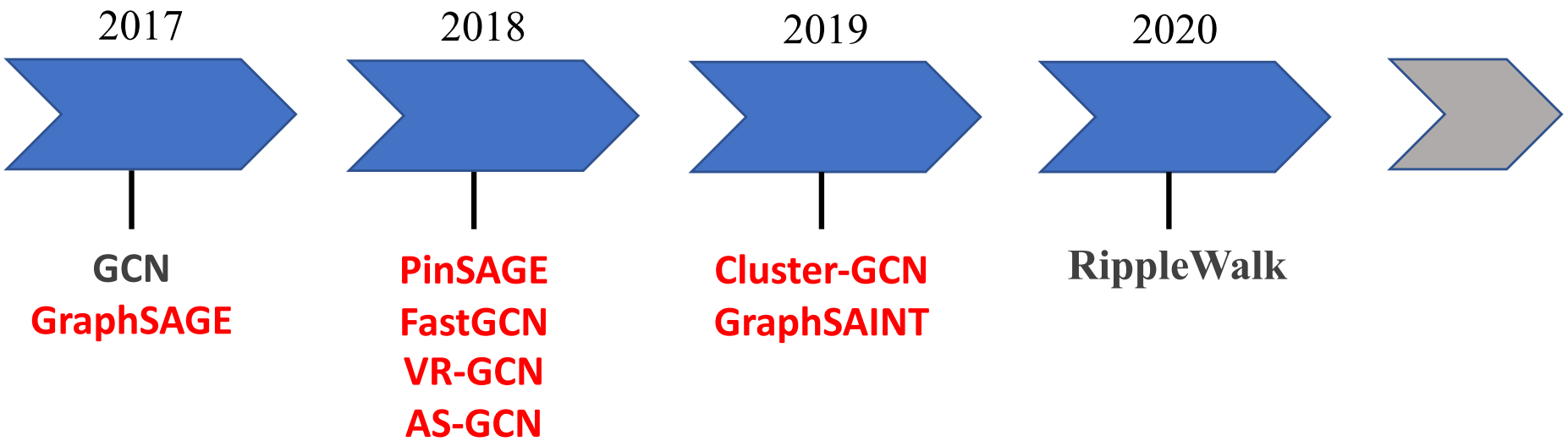
1. 以邻接矩阵所在列的2范数（FastGCN）为概率采样节点，组成子图
2. 以两端点度的倒数和为概率采样边，组成子图
3. 随机游走，路径形成子图

H. Zeng, H. Zhou, A. Srivastava, et al, "GraphSAINT: Graph sampling based in- ductive learning method," ICLR, 2020.

谢谢

敬请批评指正！

基于采样训练GNN方法的发展



时间	类型	方法	无偏性	有效性	解决问题	产生问题
2017	Node	GraphSAGE	否	否	扩展、效率	邻域爆炸
2018	Node	PinSAGE	否	否	邻域爆炸	启发式
2018	Layer	FastGCN	否	是	邻域爆炸	稀疏连接
2018	Node	VR-GCN	是	是	邻域爆炸	高内存
2018	Layer	AS-GCN	是	是	稀疏连接	采样复杂
2019	Subgraph	Cluster-GCN	否	否	邻域爆炸	引入偏差
2019	Subgraph	GraphSAINT	是	是	偏差、方差	-