

doi:10.3969/j.issn.1672-4348.2023.01.012

自适应频率和动态节点嵌入的图卷积网络

陈林凯¹, 毛国君^{1,2}

1. 福建工程学院 计算机科学与数学学院, 福建 福州 350118;
2. 福建省大数据挖掘与应用重点实验室, 福建 福州 350118)

摘要: 图卷积网络由于能够直接处理图结构数据的优点而受到广泛研究。当前的多数图卷积网络是基于图信号的平滑性(低频信息), 且不能根据各节点适合的接受域生成对应的节点嵌入, 随着网络层数的增加, 易出现图卷积网络特有的过平滑问题而导致性能下降。为此, 提出了基于自适应频率和动态节点嵌入的图卷积网络模型(adaptive frequency and dynamic node embedding based graph convolutional network, FDGCN)。FDGCN 模型能够自适应聚合不同频率的信息, 同时利用每层网络的输出, 平衡每个节点来自全局和局部领域的信息, 动态地调节节点嵌入。通过在 4 个公共数据集上进行实验, 对比了 6 个现有模型, 证明了 FDGCN 模型的有效性。

关键词: 图神经网络; 图卷积神经网络; 过平滑; 节点分类; 频率自适应

中图分类号: TP18

文献标志码: A

文章编号: 1672-4348(2023)01-0078-06

An adaptive frequency and dynamic node embedding based graph convolutional network

CHEN Linkai¹, MAO Guojun^{1,2}

- (1. School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China;
2. Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou 350118, China)

Abstract: Graph convolutional networks have been extensively studied due to their advantages of being able to directly handle graph-structured data. Most of the current graph convolutional networks are based on the smoothness of the graph signal (low frequency information) and cannot generate corresponding node embedding according to the suitable acceptance domain of each node. However, as the number of network layers increases, the problem of over-smoothing unique to graph convolutional networks is prone to occur, resulting in performance degradation. Therefore, an adaptive frequency and dynamic node embedding based graph convolutional network (FDGCN) was proposed. FDGCN model is capable of adaptively aggregating information at different frequencies; meanwhile, it dynamically adjusts node embedding by using the output of each network layer to balance the information from the global and local domains of each node. Experiments were conducted on four public datasets comparing six existing models to demonstrate the effectiveness of the FDGCN model.

Keywords: graph neural networks; graph convolutional neural networks; over-smoothing; node classification; frequency adaptation

近年来,学术界对图卷积网络的研究取得了一定的进展,但是在实践中发现,堆叠多层的图卷积网络的性能比少层的性能明显下降,文献[1]

表明重复堆叠许多层可能导致无法区分来自不同类别的节点的表示,从而使模型出现过平滑现象。图卷积网络(graph convolutional network, GCN)的

收稿日期: 2022-10-15

基金项目:国家重点研发项目(2019YFD0900805)

第一作者简介:陈林凯(1996—),男,广东揭阳人,硕士研究生,研究方向:图神经网络。

过平滑问题使得目前较为成功的 GCN 基本都是浅层结构,即在 2~4 层就几乎收敛到某个固定状态。但是,仅使用 2~4 层卷积层的图卷积网络模型很难完成大型网络的处理任务。因此,如何缓解图卷积网络的过平滑问题、改善图卷积网络学习能力受到广泛关注。

2020 年,Chen 等^[2]将残差卷积、恒等映射融入 GCN,提出了基于初始残差和恒等映射的图卷积网络 (graph convolutional network via initial residual and identity mapping, GCNII) 模型。2021 年,Bo 等^[3]的实验结果表明低频信息和高频信息都有助于学习节点的表示。Xu 等^[4]提出了跳越知识网络 (jumping knowledge networks, JKNet) 模型,证明了跳跃链接对于提高图卷积神经网络性能的作用。DropEdge^[5]在每个训练阶段之前随机删除一定比例的边,以缓解过平滑问题。

文献[6]表明,信号的平滑性即低频信息,是 GCN 成功的关键。多数图神经网络模型在学习过程中保留节点特征的共性(低频信息),这种机制适用于协调性网络,即同类别节点倾向于相互连接的图数据。然而,仅利用低频信息不可避免会忽略节点特征信息之间的差异。此外,每个节点适合的接受域,即节点能聚合到其他节点信息的范围,是不一样的,不合适的节点接受域将导致模型性能下降^[7]。

为了缓解图学习中产生的过平滑问题,本研究引入自适应频率和动态节点嵌入机制,使得模型可以合理利用图信号的低频信息和高频信息;同时根据每个节点适合的接受域,动态地平衡每个节点来自全局和局部领域的信息,生成更具区别性的节点嵌入。

1 FDGCN 模型设计

1.1 自适应频率聚合函数

图卷积网络模型可以看成特殊形式的滤波器,根据聚合函数的不同,可以分为低通滤波器、高通滤波器等。若给定输入节点特征矩阵 $\mathbf{Z} = \{z_1, z_2, \dots, z_N\} \in \mathbf{R}^{N \times P}$, N 为节点个数, P 为节点特征的维度,则特征矩阵经过 FDGCN 的聚合函数的节点嵌入更新如公式(1):

$$z_i^k = a_{ij}^L(F_L \cdot \mathbf{Z})_i^{k-1} + a_{ij}^H(F_H \cdot \mathbf{Z})_i^{k-1} = \mu z_i^0 + \sum_{j \in V_i} \frac{a_{ij}^L - a_{ij}^H}{\sqrt{d_i d_j}} z_j^{k-1} =$$

$$\mu z_i^0 + \sum_{j \in V_i} \frac{a_{ij}^G}{\sqrt{d_i d_j}} z_j^{k-1} \quad (1)$$

其中, z_i^k 表示第 k 层节点 i 的节点嵌入, V_i 为节点 i 的一阶邻居节点集, d_i 为节点 i 的度, a_{ij}^L 和 a_{ij}^H 分别表示节点 i 聚合节点 j 的低频和低频信息比例系数, $\mu \in [0, 1]$ 是一个超参数。 a_{ij}^G 数学表达式如公式(2):

$$a_{ij}^G = \tanh(\mathbf{g}^T [z_i^{k-1} \parallel z_j^{k-1}]) \quad (2)$$

式中, \mathbf{g} 是可学习的映射向量, $[\cdot \parallel \cdot]$ 表示拼接操作, $\tanh(\cdot)$ 是激活函数。 $a_{ij}^G \in [-1, 1]$ 是一个门控系数,用于控制低频信息和高频信息的比例。 F_H 代表高通滤波器,用于提取高频信息, F_L 低通滤波器用于提取低频信号,如公式(3)、公式(4)所示:

$$F_L = \mu \mathbf{I} + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} = (\mu + 1) \mathbf{I} - \mathbf{L} \quad (3)$$

$$F_H = \mu \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} = (\mu - 1) \mathbf{I} + \mathbf{L} \quad (4)$$

\mathbf{A} 、 \mathbf{D} 和 \mathbf{I} 分别为邻接矩阵、度矩阵及单位矩阵,拉普拉斯矩阵 $\mathbf{L} = \mathbf{D} - \mathbf{A}$ 。此外, $z_i^0 = \text{MLP}(z_i) \in \mathbf{R}^{1 \times p}$, MLP (multilayer perceptron) 为多层感知机,即将原始特征先进行线性变换,传播过程中不再进行特征变换。

1.2 动态节点嵌入调节机制

节点的接受域随着网络层数的增加而增大,但每个节点的最佳接受域都不一样。为此,本研究在进行特征传播后引入动态节点嵌入机制。即利用一个可学习的映射向量来生成比例系数,该映射向量通过 Xavier 初始化方法进行初始化操作并由所有节点共享。其中随着模型训练的迭代更新为公式(5), $l = [\gamma \log \hat{y} + (1 - \gamma) \log(1 - \hat{y})]$ 表示含有可学习映射向量 $\mathbf{t} (\mathbf{t} \in \mathbf{R}^{P \times 1})$ 的交叉熵损失函数, lr 表示学习率。

$$\mathbf{t} = \mathbf{t} - \text{lr} \frac{\partial l}{\partial \mathbf{t}} \quad (5)$$

1.3 FDGCN 模型

图 1 为 FDGCN 的结构示意图。输入邻接矩阵 \mathbf{A} 、节点特征矩阵 \mathbf{Z} 。首先将节点特征矩阵 \mathbf{Z} 经过多层感知机进行特征变换,压缩特征维度生成原始节点嵌入 \mathbf{Z}^0 再根据输入的邻接矩阵 \mathbf{A} 得到每个节点的邻居集合进行特征传播,传播过程中每次聚合保留一定比例的原始节点嵌入 $\mu \mathbf{Z}^0$,生成 K 层节点嵌入 $\mathbf{Z}^k (k=1, \dots, K)$ 。然后将 $K+1$ 层的节点嵌入(包含原始节点嵌入 \mathbf{Z}^0)与映射向量 \mathbf{t} 运算生成系数矩阵 $\mathbf{S} = (\mathbf{S}_0, \dots, \mathbf{S}_K)$ 。根据系

数矩阵 S 与 $K+1$ 层节点嵌入的运算,生成每个节点最佳的节点嵌入。最后再经过多层感知机和

softmax 激活函数输出 Z_{out} , 根据输出 Z_{out} 进行节点分类等下游任务。

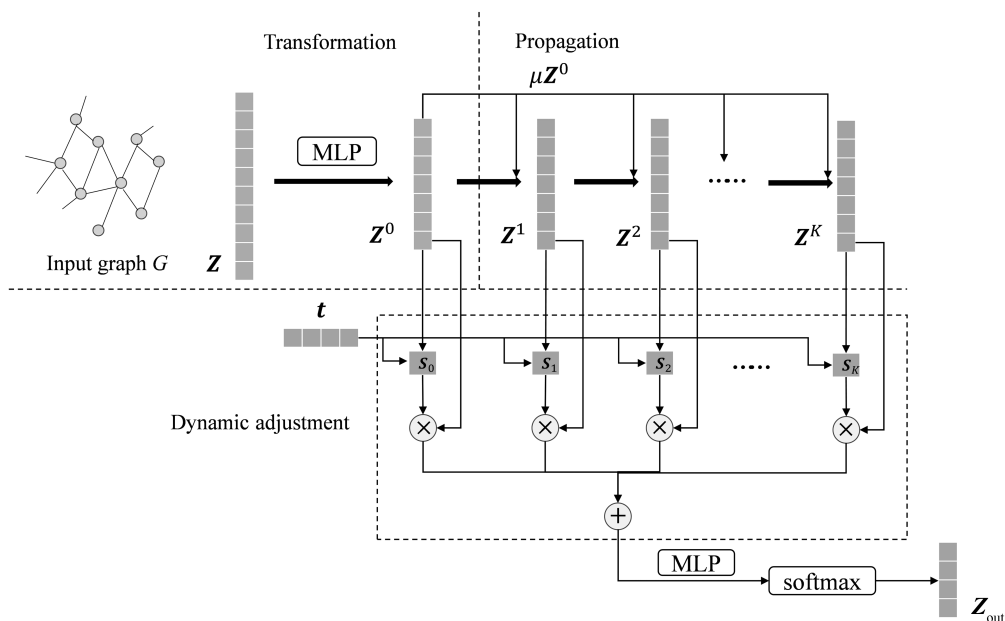


图 1 FDGCN 网络结构示意图

Fig.1 Schematic diagram of FDGCN network structure

算法 1 给出了利用 FDGCN 网络对图的节点进行概率分类的具体流程。

算法 1.class in FDGCN

输入:邻接矩阵 A (图 G), 特征矩阵 Z , 网络总层数 K , 超参数 μ

输出: 节点类别概率矩阵 Z_{out}

1. $Z^0 = \text{MLP}(Z)$, $\in \mathbb{R}^{N \times P}$;
2. for $k = 1$ to K ;
3. for $i = 1$ to N ;
4. $z_i^k = \mu z_i^0 + \sum_{j \in V_i} \frac{a_{ij}^G}{\sqrt{d_i d_j}} z_j^{k-1}$;
5. end for
6. end for
7. $Z' = \text{stack}(Z^0, \dots, Z^K)$, $\in \mathbb{R}^{N \times (K+1) \times P}$;
8. $S' = \sigma(Z't)$, $\in \mathbb{R}^{N \times (K+1) \times 1}$;
9. $S = (S')^T$, $\in \mathbb{R}^{N \times 1 \times (K+1)}$;
10. $\hat{Z} = \text{squeeze}(SZ')$, $\in \mathbb{R}^{N \times P}$;
11. $Z_{out} = \text{softmax}(\text{MLP}(\hat{Z}))$;
12. Classfying all nodes in G according to Z_{out} .

其中, $\text{stack}(\cdot)$, $\text{squeeze}(\cdot)$ 操作为数据维度进行重新排列, 以便在计算过程中进行维度匹配; $\sigma(\cdot)$ 为激活函数; $t \in \mathbb{R}^{P \times 1}$ 为可训练的映射向量,

通过映射向量 t 与不同的网络层的节点特征嵌入 $Z^k (k=0, \dots, K)$ 进行运算生成系数矩阵 S , 系数矩阵 S 用于调节来自不同网络层的每个节点嵌入表示的比例, 动态地衡量每个节点在不同接受域范围下生成的节点特征嵌入的重要性。

2 实验分析

2.1 实验环境配置

实验在 NVIDIA RTX 3060Ti 服务器上进行, Python 版本为 3.7, PyTorch 版本为 1.7, CUDA 版本为 11.0。

2.2 数据集

本研究使用 3 个标准的引文网络数据集 Cora、Citeseer 和 Pubmed^[8] 以及 1 个来自维基百科的网络数据集 Squirrel^[9] 来进行半监督节点分类, 其中 Cora、Citeseer 和 Pubmed 为协调性数据集, 而 Squirrel 为非协调性数据集。实验对所有数据集均按照无向图处理。

引文网络数据集 Cora、Citeseer 和 Pubmed 中的节点对应于文档, 边代表了两篇论文之间的引文关系(无向的), 节点特征是论文的词袋向量, 标签(类别)是论文的领域。在每个网络中, 使用每个类别 20 个标记节点进行训练, 500 个节点用于验证, 1 000 个节点用于测试。

维基百科的网络数据集 Squirrel 的节点代表网页,边表示两个页面之间的超链接,节点特征是页面中的一些关键信息名词,标签对应于页面的流量级别。在数据集 Squirrel 中,使用 20%作为验证集,20%为测试集,训练集比例为 60%。

表 1 包含 Cora、Citeseer、Pubmed 和 Squirrel 数据集的协调性、节点数、边数等相关信息,协调性的数值越高,数据集同类节点趋向相连的特性越显著。在本研究的实验中,Cora、Citeseer 和 Pubmed 数据集学习率 lr 设为 0.01,剪输出系数设为 dropout 0.6,原始特征比例 μ 分别设为 0.2、0.2、0.3,权重衰减 β 设为 $1e-3$;对于 Squirrel 数据集学习率 lr 设为 0.01,剪输出系数 dropout 设为 0.5,原始特征比例 μ 设为 0.3,权重衰减 β 设为 $5e-5$ 。

表 1 实验数据集
Tab.1 Experimental dataset

数据集	协调性	节点数/个	边数/条	类别/个	特征维度
Cora	0.771	2 708	5 429	7	1 433
Citeseer	0.671	3 327	4 732	6	3 703
Pubmed	0.686	19 717	44 338	3	500
Squirrel	0.018	5 201	217 073	3	2 089

2.3 模型对比实验

为验证模型的有效性,将 FDGCN 与目前现有的模型作对比,进行节点分类精度实验。对比模型包括基于谱域的图神经网络模型,如 SGC (simplifying graph convolutional networks)^[10]、切比雪夫网络 (ChebNet)^[11],基于空域的图神经网络模型,如图同构网络 (graph isomorphism network, GIN)^[12]、图注意力网络 (graph attention network, GAT)^[13]、APPNP (approximate personalized propagation of neural predictions)^[14]、GraphSAGE (sample and aggregate)^[15]。表 2 为不同模型在不同数据集上的对比实验结果 (实验结果采用 30 次平均值)。

在 Cora、Citeseer、Pubmed 和 Squirrel 数据集上,FDGCN 模型的实验精度分别达到 84.7%、72.7%、79.4%、65.3%。实验结果表明,与对比的模型相比,FDGCN 模型在协调性网络和非协调性网络上都具有更高的性能,获得的精度更高,分类效果更好。而且,相对于在协调性网络上的提升,

FDGCN 模型在非协调性网络上获得更为显著的效果。

表 2 在不同数据集下各模型的节点分类精度
Tab.2 Node classification accuracy of each model under different datasets

算法	节点分类精度/%			
	Cora	Citeseer	Pubmed	Squirrel
SGC	81.00	71.90	78.90	55.20
ChebNet	81.20	69.80	74.40	62.90
GIN	77.60	66.10	77.00	61.20
GAT	83.00	72.50	79.00	51.20
APPNP	83.70	72.10	79.20	55.70
GraphSAGE	82.30	71.20	78.50	53.30
FDGCN	84.7±0.6	72.7±0.5	79.4±0.4	65.3±0.6

训练时长也是评价模型优劣的重要指标之一。在相同的实验参数下,每个对比模型训练 300 个 epoch,图 2 展示了在 4 个数据集上每个对比模型的平均训练时长。从图 2 可以看出:在不同数据集上 SGC 和 GraphSAGE 模型的训练效率优于其他模型;对比 GAT、APPNP 模型,FDGCN 模型的训练时长更少;在不同数据集上,FDGCN 与 ChebNet、GIN 模型的训练效率各有高低。

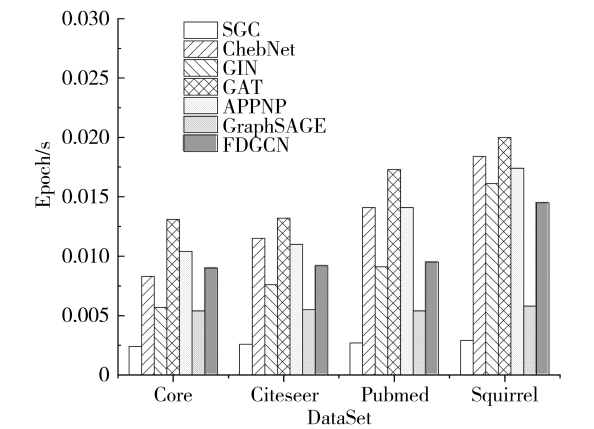


图 2 平均训练时长
Fig.2 Average training duration

SGC 是通过移除 GCN 层之间的非线性,达到减少参数量简化 GCN 的目的;而 GraphSAGE 是对邻居节点进行采样,再聚合采样邻居的节点信息更新节点表示。SGC 和 GraphSAGE 获得较高训练效率,但模型精度会有所不足。与其他模型

相比,FDGCN 在训练效率相差不大的情况下,取得了较大的精度提升。

2.4 自适应频率和动态节点嵌入的有效性分析

为验证技术融合的有效性,根据本研究使用的自适应频率(简称 F)、动态节点嵌入(简称 D)技术,在不同的网络层数 k 下,对使用不同技术的 FDGCN 模型在 Cora、Citeseer、Pubmed 及 Squirrel 数据集上进行实验,其节点分类精度结果如表 3~表 6 所示,实验结果采用 30 次平均值。

从表 3~表 6 可以看出,各个模型在随着网络层数增加时,不会出现性能急剧下降的情况;仅使用自适应频率技术的模型在 Cora、Citeseer 和 Pubmed 数据集上性能较弱;仅使用动态节点嵌入技术的模型在 Squirrel 数据集上性能较弱;融合了自适应频率和动态节点嵌入技术的模型在 4 个数据集上均获得最高的精度。

表 3 在 Cora 数据集上的节点分类精度

Tab.3 Node classification accuracy on Cora dataset

F	D	节点分类精度 /%			
		$k=2$	$k=4$	$k=6$	$k=8$
✓		82.1	83.5	83.8	83.8
	✓	81.7	84.0	84.2	84.3
✓	✓	81.3	84.0	84.3	84.7

注:加✓代表 FDGCN 模型中保留 F 或 D 技术。

表 4 在 Citeseer 数据集上的节点分类精度

Tab.4 Node classification accuracy on Citeseer dataset

F	D	节点分类精度 /%			
		$k=2$	$k=4$	$k=6$	$k=8$
✓		69.7	71.6	71.4	71.5
	✓	71.9	72.5	72.2	72.0
✓	✓	71.8	72.7	72.3	72.2

注:加✓代表 FDGCN 模型中保留 F 或 D 技术。

表 5 在 Pubmed 数据集上的节点分类精度

Tab.5 Node classification accuracy on Pubmed dataset

F	D	节点分类精度 /%			
		$k=2$	$k=4$	$k=6$	$k=8$
✓		78.4	78.8	78.6	78.8
	✓	78.3	79.0	79.2	79.1
✓	✓	78.1	79.0	79.2	79.4

注:加✓代表 FDGCN 模型中保留 F 或 D 技术。

表 6 在 Squirrel 数据集上的节点分类精度

Tab.6 Node classification accuracy on Squirrel dataset

F	D	节点分类精度 /%			
		$k=2$	$k=4$	$k=6$	$k=8$
✓		64.7	64.5	64.2	63.7
	✓	51.7	52.3	53.3	52.1
✓	✓	64.0	65.0	65.3	65.1

注:加✓代表 FDGCN 模型中保留 F 或 D 技术。

为了更直观地展示使用自适应频率的有效性,验证 FDGCN 是否可以学习不同的系数来适应不同的网络,将 FDGCN 最后一层中的系数 a_{ij}^G 提取并进行可视化操作。图 3 展示了在不同数据集上系数 a_{ij}^G 的变化,在 3 个引文网络数据集 Cora、Citeseer 和 Pubmed 上,系数 a_{ij}^G 的数值集中在 0.8~1.0 之间。而在维基百科的网络数据集 Squirrel 上,系数 a_{ij}^G 的数值多为负值。即在协调性数据集 Cora、Citeseer 和 Pubmed 上,低频信息起到主要作用,模型着重学习节点之间的共性。而在非协调性数据集 Squirrel 上,高频信息起到主要作用,模型着重学习节点之间的差异。

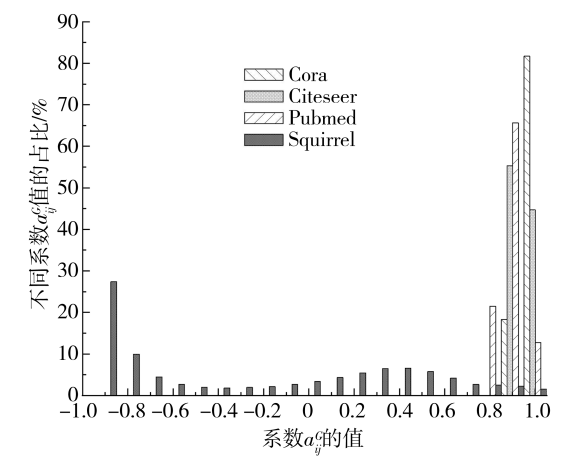


图 3 不同数据集上 a_{ij}^G 系数的变化

Fig.3 Variation of coefficient on a_{ij}^G different datasets

3 结语

随着对图学习的研究不断深入,过平滑问题在一定程度上影响着图神经网络的发展。本研究提出的网络模型 FDGCN 可以自适应地聚合不同频率信息,再利用动态节点嵌入机制平衡来自每个节点的局部和全局邻域的信息,生成更具鉴别

性的节点嵌入。实验结果明,FDGCN 模型在 Cora、Citeseer、Pubmed 及 Squirrel 数据集上的精度获得不同程度提升。FDGCN 模型为缓解图卷

积神经网络中的过平滑问题提供了一个有效途径。

参考文献:

- [1] LI Q M, HAN Z C, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 3538–3545.
- [2] CHEN M, WEI Z W, HUANG Z F, et al. Simple and deep graph convolutional networks[EB/OL]. 2020; arXiv:2007.02133 [cs.LG]. <https://arxiv.org/abs/2007.02133>.
- [3] BO D Y, WANG X, SHI C, et al. Beyond low-frequency information in graph convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(5): 3950–3957.
- [4] XU K, LI C, TIAN Y, et al. Representation learning on graphs with jumping knowledge networks[C]// International Conference on Machine Learning. New York: ACM, 2018: 5453–5462.
- [5] RONG Y, HUANG W, XU T, et al. DropEdge: towards deep graph convolutional networks on node classification[C]// International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net, 2019.
- [6] XU B B, SHEN H W, CAO Q, et al. Graph convolutional networks using heat kernel for semi-supervised learning[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 1928–1934.
- [7] LIU M, GAO H Y, JI S W. Towards deeper graph neural networks[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 338–348.
- [8] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3): 93.
- [9] TANG J, SUN J M, WANG C, et al. Social influence analysis in large-scale networks[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD’09. New York: ACM Press, 2009: 807–816.
- [10] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks[C]// International Conference on Machine Learning. California, 2019: 6861–6871.
- [11] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]// NIPS’16: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 3844–3852.
- [12] XU K, HU W, LESKOVEC J, et al. How powerful are graph neural networks? [C]// International Conference on Learning Representations. New Orleans, 2019.
- [13] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[C]// International Conference on Learning Representations, Vancouver, 2018.
- [14] GASTEIGER J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: graph neural networks meet personalized PageRank[C]// International Conference on Learning Representations, 2018.
- [15] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[EB/OL]. 2017; arXiv:1706.02216[cs.SI]. <https://arxiv.org/abs/1706.02216>.

(责任编辑:方素华)