

# 图卷积神经网络综述

徐冰冰<sup>1),2),3)</sup> 岑科廷<sup>1),2),3)</sup> 黄俊杰<sup>1),2),3)</sup> 沈华伟<sup>1),2)</sup> 程学旗<sup>1),2)</sup>

<sup>1)</sup>(中国科学院网络数据科学与技术重点实验室 北京 100190)

<sup>2)</sup>(中国科学院计算技术研究所 北京 100190)

<sup>3)</sup>(中国科学院大学 北京 100049)

**摘 要** 过去几年,卷积神经网络因其强大的建模能力引起广泛关注,在自然语言处理、图像识别等领域成功应用。然而,传统的卷积神经网络只能处理欧氏空间数据,而现实生活中的许多场景,如交通网络、社交网络、引用网络等,都是以图数据的形式存在。将卷积神经网络迁移到图数据分析处理中的核心在于图卷积算子的构建和图池化算子的构建。本文对图卷积神经网络进行综述,首先介绍了图卷积神经网络的背景并梳理了两类经典方法——谱方法和空间方法。针对图数据上平移不变性的缺失给图卷积算子的定义带来的困难,谱方法借助卷积定理在谱域定义图卷积,而空间方法通过在节点域定义节点相关性来实现图卷积。进而,本文介绍了图卷积神经网络的最新进展,这其中包括如何利用图卷积神经网络建模图上的复杂信息,如异质连接、高阶连接等,以及如何在大规模图上实现图卷积神经网络;此外,本文介绍了图卷积神经网络的相关应用,包括推荐系统领域、交通预测领域等;最后本文对图卷积神经网络的发展趋势进行了总结和展望。

**关键词** 图卷积神经网络;卷积;池化;非欧空间

**中图法分类号** TP18 **DOI号** 10.11897/SP.J.1016.2020.00755

## A Survey on Graph Convolutional Neural Network

XU Bing-Bing<sup>1),2),3)</sup> CEN Ke-Ting<sup>1),2),3)</sup> HUANG Jun-Jie<sup>1),2),3)</sup>

SHEN Hua-Wei<sup>1),2)</sup> CHENG Xue-Qi<sup>1),2)</sup>

<sup>1)</sup>(Key Laboratory of Network Data Science and Technology, Chinese Academy of Sciences, Beijing 100190)

<sup>2)</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

<sup>3)</sup>(University of Chinese Academy of Sciences, Beijing 100049)

**Abstract** In the past few years, convolutional neural network has attracted widespread attention due to its powerful modeling capabilities, and has achieved great improvement in areas such as natural language processing and computer vision. Traditional convolutional neural network can only process Euclidean data. However, many real-life scenarios, such as transportation networks, social networks and citation networks, are located in the form of graph data. The method used to process graph data previously is network embedding. Specifically, the researchers model the graph-based mission into a two-stage model. In the first stage, a fixed-length representation is learned for each node via capturing the proximity over nodes, this representation is fed into the second stage to solve downstream tasks, e. g., link prediction, node classification and graph classification. In recent years, the powerful modeling capabilities of convolutional neural network

收稿日期:2019-05-06;在线出版日期:2019-11-04。本课题得到国家自然科学基金项目(61425016,91746301)、北京智源人工智能研究院和王宽诚教育基金会资助。徐冰冰,博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为图神经网络、图上半监督学习。E-mail: xubingbing@ict.ac.cn。岑科廷(共同第一作者),博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为网络表示学习、图神经网络。E-mail: cenketing@ict.ac.cn。黄俊杰(共同第一作者),硕士研究生,主要研究方向为社会媒体计算、图神经网络。E-mail: huangjunjie17s@ict.ac.cn。沈华伟(通信作者),博士,研究员,中国计算机学会(CCF)高级会员,主要研究领域为社交网络分析和社会媒体计算。E-mail: shenhuawei@ict.ac.cn。程学旗,博士,研究员,中国计算机学会(CCF)会士,主要研究领域为大数据分析、网络科学、网络信息安全以及互联网搜索与数据挖掘。

and the ubiquity of graph data have inspired researchers to transfer convolutional neural network to graphs, which can solve the graph-based task via an end-to-end manner. The core of graph convolutional neural network is the construction of graph convolution operator and pooling operator. In this paper, we review the graph convolutional neural network. Firstly, the background of graph convolutional neural network and its classical methods are introduced, including spectral methods and spatial methods. The lack of translation invariance on the graph data makes it difficult to define the graph convolution operator. The spectral methods define the convolution in the spectral domain via the convolution theorem, while the spatial methods implement the graph convolution by defining the node correlation in the node domain. Then, the latest developments are introduced. Recent researchers focus on how to model the complicated information on graph via graph convolution neural network, e.g., heterogeneous connection and high-order connection. In addition, how to construct graph convolution neural network on large-scale network also attracts much attention. Moreover, we conclude the graph convolution neural network in many applications, including traffic prediction and recommender system. Finally, the developing trend of graph convolutional neural network is summarized and forecasted.

**Keywords** graph convolutional neural network; convolution; pooling; non-Euclidean space

## 1 引言

过去几年,卷积神经网络<sup>[1-2]</sup>快速发展,并借由其强大的建模能力引起广泛关注.相比传统方法,卷积神经网络的引入给图像处理<sup>[3]</sup>和自然语言处理等领域带来了很大的提升,如机器翻译<sup>[4]</sup>、图像识别<sup>[5]</sup>和语音识别<sup>[6]</sup>等.但是,传统的卷积神经网络只能处理欧氏空间数据(如图像、文本和语音),这些领域的数据具有平移不变性.平移不变性使得我们可以在输入数据空间定义全局共享的卷积核,从而定义卷积神经网络.以图像数据为例,一张图片可以表示为欧氏空间中一组规则散布的像素点,平移不变性则表示以任意像素点为中心,可以获取相同尺寸的局部结构.基于此,卷积神经网络通过学习在每个像素点共享的卷积核来建模局部连接,进而为图片学习到意义丰富的隐层表示.

尽管传统的卷积神经网络在文本和图像领域带来提升,但是它仅能处理欧氏空间数据.而同时,一种非欧空间数据:图数据,因其普遍存在性逐渐受到关注.图数据可以自然的表达实际生活中的数据结构,如交通网络、万维网和社交网络等.不同于图像和文本数据,图数据中每个节点的局部结构各异,这使得平移不变性不再满足<sup>[7]</sup>.平移不变性的缺失给在图数据上定义卷积神经网络提出了挑战.

近年来,由于图数据的普遍存在性,研究人员开

始关注如何在图上构造深度学习模型.借助于卷积神经网络对局部结构的建模能力及图上普遍存在的节点依赖关系,图卷积神经网络成为其中最活跃最重要的一支.近期陆续涌现出一些文章探索图上深度学习并对其进行综述<sup>[8-9]</sup>,但是针对其中最重要的分支,图卷积神经网络,其建模方法和应用方面的深入探讨和总结仍为空白.对此,我们在本篇文章中深入整理总结了图卷积神经网络的发展历程及其未来趋势.

图卷积神经网络的构建所面临的挑战主要来源于以下几个方面:

### (1) 图数据是非欧空间数据

图数据作为非欧空间数据,不满足平移不变性,即每个节点具有各异的局部结构.而传统卷积神经网络中的基本算子:卷积和池化,依赖于数据的平移不变性.此时如何在图数据上定义卷积和池化算子成为一个有挑战的工作.

### (2) 图数据具有多样的特性

实际生活中的多种应用都可以用图数据自然的表示,这使得图数据具有多样的特性,如社交网络中用户的有向连接,引文网络中作者和引文的异质连接,政治关系网络中的正负倾向带符号连接等.多样的图特性给图卷积神经网络的构建带来更多信息,但是多种特性的建模也要求图卷积神经网络的设计更加复杂精细,给图卷积神经网络带来新的挑战.

### (3) 图数据的规模很大

在大数据时代,实际应用中的图可能规模极大,含有百万甚至千万级别的节点,如推荐系统中的用户商品网络,社交网络中的用户网络.如何在时间和空间可接受范围内在大规模图上构建图卷积神经网络也是非常大的挑战.

### 1.1 现有方法及分类

图数据建模的历史由来已久.起初,研究人员关注统计分析的方法,这个时期没有机器学习模型的参与,如网页排序的常用算法 PageRank<sup>[10]</sup>、HITS<sup>[11]</sup>等.此外,研究人员也借用图谱理论的知识,如用拉普拉斯矩阵的特征值和特征向量做社区分析或者人群聚类<sup>[12]</sup>等.随着深度学习的崛起,研究人员开始考虑把深度学习的模型引入到图数据中,代表性的研究工作是网络嵌入(Network Embedding)<sup>[13]</sup>,即通过约束节点的邻近性为每个节点学习固定长度的表达,如 DeepWalk<sup>[14]</sup>、LINE<sup>[15]</sup>、node2vec<sup>[16]</sup>等.这一时期,在解决具体的应用问题时,研究人员通常将其建模为两阶段问题,以节点分类为例,第一阶段为每个节点学习统一长度的表达,第二阶段将节点表达作为输入,训练分类模型.近年来,研究人员对图数据建模的关注逐渐转移到如何将深度学习的模型迁移到图数据上,进行端到端的建模,而图卷积神经

网络则是其中最活跃的一支.

在建模图卷积神经网络时,研究人员关注如何在图上构建卷积算子. Bruna 等人<sup>[17]</sup>在 2013 年提出第一个图卷积神经网络,他们基于图谱理论从卷积定理出发,在谱空间定义图卷积.这一支后来发展为图卷积领域的谱方法.最初的谱方法具有时空复杂度较高的弊端, ChebNet<sup>[18]</sup>和 GCN<sup>[19]</sup>对谱方法中的卷积核进行参数化,大大降低了时空复杂度.这两个方法虽然被归为谱方法,但已经开始从空间角度定义节点的权重矩阵.在这两个方法的启发下,空间方法应用而生,开始考虑在节点域用注意力机制、序列化模型等建模节点间的权重.这一时期的图卷积神经网络在构建卷积算子过程中没有过多的考虑图特性,随着卷积算子的逐渐完善,人们开始考虑多样的图特性,如开始关注如何建模图上的高阶信息,并分别针对边上带特征的图、异质图等进行精细设计.此外,如何训练更高效的图卷积神经网络也受到广泛关注.研究人员开始试图训练更深层的图卷积神经网络,以增强模型的泛化能力.同时,模型到大规模图的可扩展性以及训练的速度也是图卷积神经网络中非常重点的研究方向.表 1 展示了图卷积神经网络的主要方法.

表 1 图卷积神经网络的主要方法

关注领域	方法	相关论文
关注卷积算子构建	基于卷积定理的图卷积神经网络	Spectral CNN <sup>[17]</sup> , ChebyNet <sup>[18]</sup> , GCN <sup>[19]</sup> , Henaff 等人 <sup>[20]</sup> , GWNN <sup>[21]</sup> , GraphHeat <sup>[22]</sup> , PPNP <sup>[23]</sup> , SPC <sup>[24]</sup>
	基于聚合函数的图卷积神经网络	MoNet <sup>[25]</sup> , MPNNs <sup>[26]</sup> , GNNs <sup>[27]</sup> , GAT <sup>[28]</sup> , GraphSAGE <sup>[29]</sup> , DCNN <sup>[30]</sup> , ConfGCN <sup>[31]</sup> , HGNN <sup>[32]</sup>
关注图上复杂信息建模	建模边上信息的图卷积神经网络	R-GCNs <sup>[33]</sup> , R-GAT <sup>[34]</sup> , SGCN <sup>[35]</sup> , DPGCNN <sup>[36]</sup> , EGAT <sup>[37]</sup> , ECC <sup>[38]</sup>
	建模高阶信息的图卷积神经网络	HA-GCN <sup>[39]</sup> , Motif-CNN <sup>[40]</sup> , HAN <sup>[41]</sup> , MCN <sup>[42]</sup> , Motif-Net <sup>[43]</sup>
关注训练过程优化	深层图卷积神经网络	Jumping Knowledge Network <sup>[44]</sup> , Co-Train <sup>[45]</sup>
	大规模图卷积神经网络	GraphSAGE <sup>[29]</sup> , PinSage <sup>[46]</sup> , Control Variate Based Algorithm(CV) <sup>[47]</sup> , FastGCN <sup>[48]</sup> , Adapt <sup>[49]</sup>

池化算子作为卷积神经网络的主要组成部分,作用是扩大感受野,降低参数.近期,也有研究开始关注图上池化算子的构建.图上池化算子主要用于图分类问题,目的是学习到图的层级结构.

### 1.2 任务

图数据建模所针对的应用场景非常广泛,这也使得图数据建模所处理的任务多样.我们将下游任务分为节点级别的任务和图级别的任务,节点级别的任务包括节点分类,链接预测等,如引文网络中的文章分类,推荐系统中用户对商品的偏好推断.图级别的任务包括图生成,图分类等,如药物网络生成,蛋白质网络中的蛋白质分类.

### 1.3 章节组织

本文第 2 节给定全文的符号定义;第 3 节回顾初期图卷积网络中卷积算子和池化算子的定义,其中卷积算子定义部分又包括了空间方法和谱方法;第 4 节主要关注图卷积神经网络的一些新进展,包括图特性建模和训练优化两个部分;第 5 节我们详细介绍了图卷积神经网络在现有各种典型应用下图的构造及卷积的定义;在第 6 节和第 7 节中,我们给出图卷积神经网络的未来发展展望,并总结全文.

## 2 符号定义

在这一节中,我们给出文中常见符号的定义.

表 2 展示了每种符号的代表含义.

表 2 符号定义	
符号	表示含义
$G$	图
$V$	节点集合
$E$	连边集合
$n$	节点个数
$I_n \in \mathbb{R}^{n \times n}$	单位阵
$A \in \mathbb{R}^{n \times n}$	邻接矩阵
$D \in \mathbb{R}^{n \times n}$	度矩阵(对角阵)
$L \in \mathbb{R}^{n \times n}$	拉普拉斯矩阵
$U \in \mathbb{R}^{n \times n}$	特征向量矩阵
$\Lambda \in \mathbb{R}^{n \times n}$	特征值矩阵(对角阵)
$u_i \in \mathbb{R}^n$	第 $i$ 个特征向量
$X \in \mathbb{R}^{n \times D}$	节点特征矩阵
$X_i \in \mathbb{R}^D$	第 $i$ 个节点的特征
$f \in \mathbb{R}^n$	信号

我们用  $G=\{V,E,A\}$  表示无向图,其中  $V$  表示节点集合,  $|V|=n$  表示图上共有  $n$  个节点,  $E$  表示边集合,  $A$  表示邻接矩阵,定义节点之间的相互连接,且在无向图中  $A_{i,j}=A_{j,i}$ .  $L=D-A$  表示图上的拉普拉斯矩阵,其中  $D$  是一个对角阵,  $D_{i,i}$  表示第  $i$  个节点的度且  $D_{i,i}=\sum_j A_{i,j}$ . 归一化后的拉普拉斯矩阵定义为  $L=I_n-D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ , 其中  $I_n \in \mathbb{R}^{n \times n}$  是单位阵. 因为  $L$  是一个实对称矩阵,对  $L$  做特征分解得到  $L=U\Lambda U^T$ . 其中  $U=\{u_i\}_{i=1}^n$  表示  $n$  个相互正交的特征向量,  $\Lambda=\text{diag}(\{\lambda_i\}_{i=1}^n)$  是一个对角阵,  $\lambda_i$  表示  $u_i$  对应的特征值.

我们用  $X$  表示图  $G$  上的节点特征,其中  $X \in \mathbb{R}^{n \times D}$ ,  $X_i \in \mathbb{R}^D$  是第  $i$  节点的特征,  $X_{ij}$  表示矩阵  $X$  的第  $i$  行第  $j$  列,即第  $i$  节点的第  $j$  个特征.  $f,x,y \in \mathbb{R}^n$  表示图上的信号,其中  $f_i$  表示节点  $i$  在信号  $f$  上的取值.

3 图卷积神经网络

图卷积神经网络主要包括卷积算子和池化算子的构建,其中卷积算子的目的是刻画节点的局部结构,而池化算子的目的是学到网络的层级化表示,降低参数. 在解决节点级别的任务时,研究人员更关注如何给每个节点学到更好的表达,此时池化算子并不必要,因此前期大量的工作仅关注图上卷积算子的构建,而池化算子通常用在图级别的任务上. 在这一章节中,我们将详细介绍图上卷积算子和池化算子的构建.

3.1 图卷积算子的构建

本节介绍关注卷积算子构建的图卷积神经网络

方法. 现有的图卷积神经网络分为谱方法和空间方法两类,谱方法利用图上卷积定理从谱域定义图卷积,而空间方法从节点域出发,通过定义聚合函数来聚合每个中心节点和其邻近节点.

3.1.1 图卷积神经网络谱方法

图上平移不变性的缺失,给在节点域定义卷积神经网络带来困难. 谱方法利用卷积定理从谱域定义图卷积. 我们首先给出卷积定理的背景知识.

(1) 图信号处理

卷积定理:信号卷积的傅立叶变换等价于信号傅立叶变换的乘积<sup>[7]</sup>:

$$F(f \times g) = F(f) \cdot F(g) \tag{1}$$

其中,  $f, g$  表示两个原始信号,  $F(f)$  表示  $f$  的傅立叶变换,  $\cdot$  表示乘积算子,  $*$  表示卷积算子. 对式(1)两端做傅立叶逆变换,可以得到

$$f * g = F^{-1}(F(f) \cdot F(g)) \tag{2}$$

其中,  $F^{-1}(f)$  表示信号  $f$  的傅立叶逆变换.

利用卷积定理,我们可以对谱空间的信号做乘法,再利用傅里叶逆变换将信号转换到原空间来实现图卷积,从而避免了因图数据不满足平移不变性而造成的卷积定义困难问题. 图上傅立叶变换依赖于图上的拉普拉斯矩阵,在下文中,我们将给出图上傅立叶变换的定义.

图上傅立叶变换的定义依赖于拉普拉斯矩阵的特征向量,以特征向量作为谱空间下的一组基底,图上信号  $x$  的傅立叶变换为

$$\hat{x} = U^T x \tag{3}$$

其中,  $x$  指信号在节点域的原始表示,  $\hat{x}$  指信号  $x$  变换到谱域后的表示,  $U^T$  表示特征向量矩阵的转置,用于做傅立叶变换. 信号  $x$  的傅立叶逆变换为

$$x = U\hat{x} \tag{4}$$

利用图上傅立叶变换和逆变换,我们可以基于卷积定理实现图卷积算子

$$x_G^* y = U((U^T x) \odot (U^T y)) \tag{5}$$

其中,  $_G^*$  表示图卷积算子,  $x, y$  表示图上节点域的信号,  $\odot$  指哈达玛乘法,表示两个向量的对应元素相乘. 我们用一个对角阵  $g_\theta$  代替向量  $U^T y$ ,那么哈达玛乘法可以转化成矩阵乘法. 将卷积核  $g_\theta$  作用在信号上,图卷积可以表示成如下形式

$$U g_\theta U^T x \tag{6}$$

卷积定理提供了通过傅立叶变换在图上定义卷积的方式. 基于此卷积算子的定义,国内外陆续涌现出一些图卷积神经网络.

(2) 基于卷积定理的图卷积神经网络

谱卷积神经网络(Spectral CNN)<sup>[17]</sup>是最早提

出在图上构建卷积神经网络的方法,该方法利用卷积定理在每一层定义图卷积算子,在损失函数指导下通过梯度反向回传学习卷积核,并堆叠多层组成神经网络.谱卷积神经网络第  $m$  层的结构如下

$$\mathbf{X}_j^{m+1} = h\left(\mathbf{U} \sum_{i=1}^p F_{i,j}^m \mathbf{U}^\top \mathbf{X}_i^m\right), j=1, \dots, q \quad (7)$$

其中,  $p, q$  分别是输入特征和输出特征的维度,  $\mathbf{X}_i^m \in \mathbb{R}^n$  表示图上节点在第  $m$  层的第  $i$  个输入特征,  $F_{i,j}^m$  表示谱空间下卷积核,  $h$  表示非线性激活函数.在谱卷积神经网络中,这样一层结构将特征从  $p$  维转化到  $q$  维,且基于卷积定理通过学习卷积核实现了图卷积.

谱卷积神经网络将卷积核作用在谱空间的输入信号上,并利用卷积定理实现图卷积,以完成节点之间的信息聚合,然后将非线性激活函数作用在聚合结果上,并堆叠多层形成神经网络.该模型不满足局部性,使得谱卷积神经网络的局部性没有保证,即产生信息聚合的节点并不一定是邻近节点.

建模图卷积神经网络的初衷是为了利用图结构刻画邻近节点的信息聚合,而上文介绍的谱卷积神经网络并不满足局部性,Henaff 等人<sup>[20]</sup>提出用带有平滑性约束的插值卷积核,这种方法降低参数个数且实现了图卷积神经网络的局部化.最近,小波神经网络(GWNN)<sup>[21]</sup>提出用小波变换代替傅立叶变换实现卷积定理.

小波神经网络指出,与傅立叶变换相似,小波变换也定义了一种将信号从节点域变换到谱域的方法<sup>[50]</sup>.这里用  $\Psi_s = \{\psi_{s1}, \psi_{s2}, \dots, \psi_{sn}\}$  表示小波变换的基底,其中  $\psi_{si}$  表示从第  $i$  个节点出发的能量扩散,刻画了第  $i$  个节点的局部结构.小波基底的定义依赖于拉普拉斯矩阵的特征向量,即  $\Psi_s = \mathbf{U} \mathbf{G}_s \mathbf{U}^\top$ , 其中  $\mathbf{G}_s = \text{diag}(\{g_s(\lambda_i)\}_{i=1}^n)$ , 对角线元素由  $g$  函数作用到特征值上得到.不同的  $g$  函数赋予小波基底不同的性质,在小波神经网络中,作者使用热核函数,即  $g_s(\lambda_i) = e^{s\lambda_i}$ .

以  $\Psi_s$  为谱空间的基底,图上小波逆变换的变换矩阵为  $\Psi_s^{-1} = \mathbf{U} \mathbf{G}_{-s} \mathbf{U}^\top$ , 其中  $\mathbf{G}_{-s}$  表示将上述  $g$  函数替换为  $g_{-s}(\lambda_i) = e^{-s\lambda_i}$ .

和傅立叶变换相比,小波变换的基底具有几个很好的性质:(1)小波变换的基底可以通过切比雪夫多项式近似得到,避免拉普拉斯矩阵特征分解的高昂代价;(2)小波变换的基底具有局部性;(3)小波基底的局部性使得小波变换矩阵非常稀疏,这大大降低了  $\Psi_s^{-1}x$  的计算复杂度,使计算过程更加高

效;(4)热核函数中的超参数  $s$  用以表示热量扩散的范围,通过调节超参数  $s$  可以灵活的适应于不同任务场景.图 1 指在不同  $s$  下黄色节点对应的小波变换基底.图 1(a)表示在  $s$  较小时,以黄色节点为中心进行的热量扩散,图 1(b)表示当  $s$  增大时,热量扩散范围也变大;此外,图 1(a)也反应了小波基底的局部性和稀疏性.

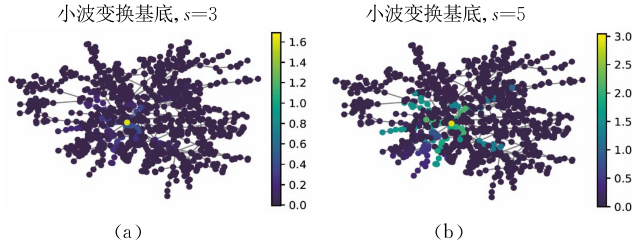


图 1 不同  $s$  下的小波变换基底<sup>[21]</sup> ((a)当  $s$  较小时,以黄色节点为中心的热量扩散范围较小;(b)当  $s$  增大时,热量扩散范围增大)

利用图上小波变换代替傅立叶变换,小波神经网络的第  $m$  层结构定义如下:

$$\mathbf{X}_j^{m+1} = h\left(\Psi_s \sum_{i=1}^p F_{i,j}^m \Psi_s^{-1} \mathbf{X}_i^m\right), j=1, \dots, q \quad (8)$$

与谱卷积神经网络相比,小波神经网络用小波变换代替傅立叶变换,即用  $\Psi$  和  $\Psi_s^{-1}$  替代了  $\mathbf{U}$  和  $\mathbf{U}$ .在这样一组小波基底,图卷积神经网络满足了局部性,且由于小波基底的可加速计算以及稀疏性,图卷积神经网络的计算复杂度也大大降低.

除小波神经网络外,还有一些工作致力于实现图卷积神经网络的局部性和加速计算,但不同于小波神经网络更换基底的方式,这些工作通过参数化卷积核实现局部性,同时降低参数复杂度和计算复杂度.下文我们将给出这类工作的具体介绍.

在式(6)中,  $\mathbf{g}_\theta$  是需要学的卷积核,在谱卷积神经网络中,  $\mathbf{g}_\theta$  是对角阵的形式,且有  $n$  个需要学的参数.切比雪夫网络(ChebyNet)<sup>[18]</sup>对卷积核  $\mathbf{g}_\theta$  进行参数化

$$\mathbf{g}_\theta = \sum_{k=0}^{K-1} \theta_k T_k(\hat{\Lambda}) \quad (9)$$

其中,  $\theta_k$  是需要学的系数,  $\hat{\Lambda} = \frac{2\Lambda}{\lambda_{\max}} - \mathbf{I}_n$ . 切比雪夫多项式是通过递归得到,递归表达式为

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x) \quad (10)$$

其中,  $T_0(x) = 1, T_1(x) = x$ .

令  $\hat{L} = \frac{2L}{\lambda_{\max}} - \mathbf{I}_n$ , 切比雪夫网络第  $m$  层的结构定义如下:

$$\begin{aligned} \mathbf{X}_j^{m+1} &= h\left(\mathbf{U} \sum_{i=1}^p \left(\sum_{k=0}^{K-1} \theta_k T_k(\hat{\mathbf{A}})\right) \mathbf{U}^\top \mathbf{X}_i^m\right) \\ &= h\left(\sum_{i=1}^p \sum_{k=0}^{K-1} \theta_k T_k(\hat{\mathbf{L}}) \mathbf{X}_i^m\right), j=1, \dots, q \quad (11) \end{aligned}$$

切比雪夫网络利用特征值矩阵的多项式参数化卷积核,实现谱卷积神经网络,且巧妙的利用  $\mathbf{L}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$  引入拉普拉斯矩阵,从而避免了拉普拉斯矩阵的特征分解,同时参数复杂度从  $O(n \times p \times q)$  下降到  $O(K \times p \times q)$ . 此外,在拉普拉斯矩阵中,当且仅当节点  $i, j$  满足  $K$  跳可达时,  $\mathbf{L}_{i,j}^K \neq 0$ , 这一性质使得当  $K$  较小时,切比雪夫网络具有局部性。

为了使图卷积神经网络在图上半监督学习领域发挥作用, Kipf 等人<sup>[19]</sup>对切比雪夫网络简化并提出一阶图卷积神经网络, Kipf 等人<sup>[19]</sup>令  $K=2$  且  $\lambda_{\max}=2$ , 则式(11)可以写成

$$\mathbf{X}_j^{m+1} = h\left(\sum_{i=1}^p (\theta_0 - \theta_1 (\mathbf{L} - \mathbf{I}_n)) \mathbf{X}_i^m\right), j=1, \dots, q \quad (12)$$

在图上半监督学习场景下,带标签的数据非常少,为了避免模型过拟合, Kipf 等人约束  $\theta = \theta_0 = -\theta_1$  来降低模型参数,并对权重矩阵做归一化处理,最终得到如下的一阶图卷积神经网络

$$\mathbf{X}_j^{m+1} = h\left(\sum_{i=1}^p \theta \hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}_i^m\right), j=1, \dots, q \quad (13)$$

其中,  $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$ , 且  $\hat{\mathbf{D}}_{ii} = \sum_j \hat{\mathbf{A}}_{i,j}$ .

图热核网络(GraphHeat)<sup>[22]</sup>从滤波器的角度对以上谱方法进行分析,指出谱卷积神经网络是非参滤波器,而切比雪夫网络和一阶图卷积神经网络都是高通滤波器,但这与图半监督学习的任务中的平滑性先验不一致,基于此,图热核网络利用热核函数参数化卷积核,进而实现低通滤波器。图热核网络的思想如图2所示。

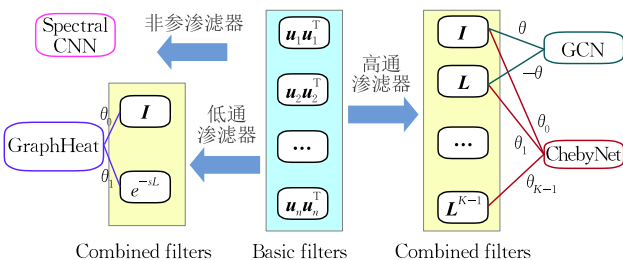


图2 图热核网络与原有谱方法对比<sup>[22]</sup>

基于个性化 PageRank 的图卷积神经网络(PPNP)<sup>[23]</sup>和简明一阶图卷积神经网络(SGC)<sup>[24]</sup>则是对一阶图卷积神经网络方法进行分析,并提出了一些简化和变体。PPNP 从深层网络的搭建出发,

指出随着模型层数加深,网络拟合能力增强,但是在—阶图卷积神经网络中会引起节点表达过于平滑,进而导致节点不可区分的问题。基于此,PPNP 解耦维度变换和特征传播,并引入个性化 PageRank,对输入数据先完成较少层数的维度变换,然后基于个性化 PageRank 进行特征传播,特征传播过程不进行参数学习,因此可以用在半监督学习任务中。PPNP 的结构如图3所示。

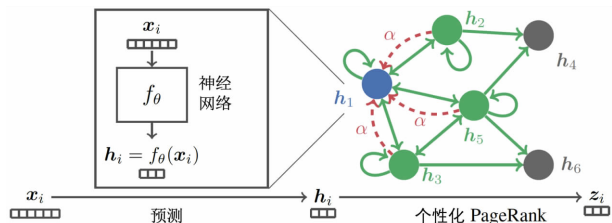


图3 基于个性化 PageRank 的图卷积神经网络结构<sup>[23]</sup>

简明一阶图卷积神经网络(SGC)指出,非线性变换在一阶图卷积神经网络中无足轻重,使得 GCN 发挥作用的是每一层的特征传播机制。基于此,SGC 抛弃层之间的非线性变换,将多个层的特征传播融合到一个层内,在完成特征传播后,SGC 对样本做一次维度变换。此模型等价于不含非线性变换的多层一阶图卷积神经网络。

切比雪夫网络和一阶图卷积神经网络着眼于参数化卷积核,图热核网络着眼于低通滤波器,以上方法虽然是从谱空间出发,但其最终形式已经包含定义节点相关性的聚合函数,从空间方法角度看,切比雪夫网络以拉普拉斯矩阵多项式作为聚合函数,一阶卷积神经网络以  $\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}}$  作为聚合函数,图热核网络以  $e^{-sL}$  作为聚合函数,其输出结果表示每个节点在该聚合函数下由自身和邻近节点加权得到的新表达。而基于个性化 PageRank 的图卷积神经网络(PPNP)和简明一阶图卷积神经网络(SGC)虽然是从谱方法一阶图卷积神经网络出发,但其已经是通过聚合函数分析节点特征传播。以上方法可以看作是谱方法和空间方法的桥梁。

### 3.1.2 图卷积神经网络空间方法

上述方法都是从卷积定理出发在谱域定义图卷积,空间方法旨在从节点域出发,通过定义聚合函数来聚合每个中心节点和其邻近节点。上文中的切比雪夫网络和一阶图卷积网络可以看作以拉普拉斯矩阵或其变体作为聚合函数。在此启发下,近期出现一些工作通过注意力机制或递归神经网络等直接从节点域学习聚合函数,此外,也有一些工作从空间角度定义了图卷积神经网络的通用框架并解释图卷积神



经网络的内部机制。

### (1) 通用框架

通用框架的定义指出图卷积网络的核心问题, 同时给已有工作提供一个对比分析的平台。近期出现两篇文章旨在定义图卷积网络的通用框架, 其中混合卷积网络(MoNet)<sup>[25]</sup>着眼于图上平移不变性的缺失, 通过定义映射函数将每个节点的局部结构映射为相同尺寸的向量, 进而在映射后的结果上学习共享的卷积核; 而消息传播网络(MPNNs)<sup>[26]</sup>立足于节点之间的信息传播聚合, 通过定义聚合函数的通用形式提出框架。

平移不变性的缺失给图卷积神经网络的定义带来困难。混合卷积网络在图上定义坐标系, 并将节点之间的关系表示为新坐标系下的一个低维向量。同时, 混合卷积网络定义一簇权重函数, 权重函数作用在以一个节点为中心的所有邻近节点上, 其输入为节点间的关系表示(一个低维向量), 输出为一个标量值。通过这簇权重函数, 混合卷积网络为每个节点获得相同尺寸的向量表示:

$$D_j(x)f = \sum_{y \in N(x)} w_j(\mathbf{u}(x, y))f(y), \quad j=1, \dots, J \quad (14)$$

其中,  $N(x)$  表示  $x$  的邻近节点集合,  $f(y)$  表示节点  $y$  在信号  $f$  上的取值,  $\mathbf{u}(x, y)$  指坐标系  $\mathbf{u}$  下节点, 关系的低维向量表示,  $w_j$  表示第  $j$  个权重函数,  $J$  表示权重函数的个数。这步操作使得每个节点都得到一个  $J$  维的表示, 且这个表示融合了节点的局部结构信息。而混合卷积模型正是在这个  $J$  维表示上定义共享卷积核,

$$(f_g^*)(x) = \sum_{j=1}^J g(j)D_j(x)f \quad (15)$$

其中,  $\{g(j)\}_{j=1}^J$  指卷积核。

不同于混合卷积网络, 消息传播网络指出图卷积的核心在于定义节点之间的聚合函数, 基于聚合函数, 每个节点可以表示为周围节点和自身的信息叠加。因此, 该模型通过定义通用的聚合函数提出图卷积网络的通用框架。消息传播网络分为两个步骤, 首先将聚合函数作用在每个节点及其邻近节点上, 得到节点的局部结构表达; 然后, 将更新函数作用在自身和局部结构表达上, 得到当前节点的新表达,

$$m_x^{t+1} = \sum_{y \in N(x)} M_t(h_x^t, h_y^t, e_{x,y}), \quad h_x^{t+1} = U_t(h_x^t, m_x^{t+1}) \quad (16)$$

其中,  $h_x^t$  表示第  $t$  步节点  $x$  的隐层表示,  $e_{x,y}$  表示节点  $x, y$  的连边特征,  $M_t$  表示第  $t$  步的聚合函数,  $m_x^{t+1}$  表示节点  $x$  通过聚合函数后得到的局部结构

表达,  $U_t$  表示第  $t$  步的更新函数。通过将神经网络的每一层设计成上述的聚合函数和更新函数, 每个节点可以不断以自身和邻近节点为源信息更新自身, 进而得到依赖于节点局部结构的新表达。

在以上空间框架下, 一些方法不再依赖于拉普拉斯矩阵, 而是设计神经网络来学习聚合函数。这些方法学到的聚合函数可以自适应于任务和具体的图结构, 有更大的灵活性。下文我们将给出这类方法的具体分析。

### (2) 基于聚合函数的图卷积神经网络

图神经网络(GNNs)<sup>[27]</sup>是最早提出在图上搭建神经网络的模型。在图神经网络中, 聚合函数被定义成循环递归函数的形式, 每个节点以周围节点和连边作为来源信息更新自身的表达,

$$h_x = f_w(l_x, l_{c0[x]}, l_{ne[x]}, h_{ne[x]}) \quad (17)$$

其中,  $l_x, l_{c0[x]}, l_{ne[x]}, h_{ne[x]}$  分别表示节点  $x$  的标签, 与节点  $x$  相连的边的标签,  $x$  的邻居节点的标签, 以及  $x$  的邻居节点上一个时间步的表达,  $f_w$  是聚合函数, 其在文章中被定义成递归函数, 根据  $f_w$  迭代更新节点  $x$  的表达直到收敛。此外, 图神经网络定义全局输出函数并作用在收敛后每个节点的表达上得到最终输出结果, 我们用  $g_w$  表示全局输出函数, 最终结果为

$$o_x = g_w(h_x, l_x) \quad (18)$$

近期, 注意力机制引起广泛关注, 图注意力网络(GAT)<sup>[28]</sup>正是通过注意力机制定义聚合函数, 但是和以往关心边上信息的模型不同, 在图注意力网络中, 邻接矩阵仅被用来定义相关节点, 而关联权重的计算则是依赖节点的特征表达。图注意力网络每一层的结构如图 4 所示, 图 4(a)以节点  $i, j$  的特征表达作为输入, 计算  $i, j$  之间的注意力权重并归一化, 图 4(b)利用注意力权重将周围节点的表达以加权的形式聚合到自身, 对于多种注意力机制下的计算结果, 图注意力网络提供了拼接和均值两种计算方式。注意力权重和表达更新的计算公式分别如下

$$\alpha_{i,j} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}[\mathbf{W}h_i \parallel \mathbf{W}h_j]))}{\sum_{k \in N(i)} \exp(\text{LeakyReLU}(\mathbf{a}[\mathbf{W}h_i \parallel \mathbf{W}h_k]))},$$

$$h_i = \sigma\left(\sum_{j \in N(i)} \alpha_{i,j} \mathbf{W}h_j\right) \quad (19)$$

其中, 参数  $W$  用于完成每个节点的特征维度变换, 参数  $\mathbf{a}$  用于计算节点间的注意力权重,  $\parallel$  表示向量拼接,  $\alpha_{i,j}$  表示在  $\mathbf{a}$  下计算得到的  $i, j$  节点间的权重,  $\sigma$  表示非线性激活函数。

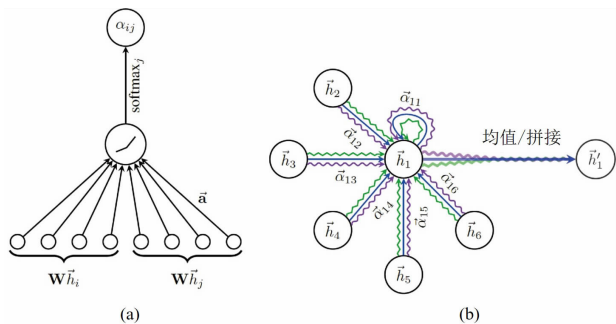


图 4 图注意力网络结构<sup>[28]</sup> ((a) 利用注意力机制计算两个节点之间的权重; (b) 根据计算的权重更新目标节点)

从图注意力网络开始,节点之间的权重计算开始从依赖于网络的结构信息转移到依赖于节点的特征表达,但是以上模型在处理时需要加载整个网络的节点特征,这阻碍了模型在大规模网络上的应用.基于此,Hamilton 等人<sup>[29]</sup>提出图采样聚合网络(GraphSAGE),不同于以往模型考虑所有邻近节点,图采样聚合网络对邻近节点做随机采样,使得每个节点的邻近节点都小于给定的采样个数.图采样聚合网络的结构如图 5 所示.

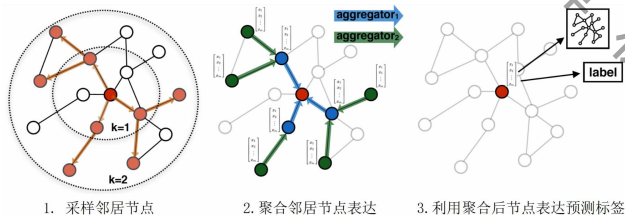


图 5 图采样聚合网络结构<sup>[29]</sup> ((a) 以每个节点为中心,对其邻近节点采样; (b) 利用采样节点更新目标节点的表达; (c) 利用更新后的节点表达预测节点的 Label)

以红色节点为目标节点,图采样聚合网络首先对其一阶邻居和二阶邻居做随机采样,并仅把采样到的节点作为相关节点,然后,模型将聚合函数作用在相关节点的特征表达上,并用聚合结果更新红色节点的特征表达来完成对应任务.此外,图采样聚合网络给出多种聚合函数的形式,分别是基于最大值的聚合,基于均值的聚合和长短时记忆力网络(LSTM).最大值聚合和均值聚合分别指取相关节点的最大值和均值作为聚合结果;长短时记忆力网络指将相关节点输入 LSTM,并把输出作为聚合结果.图采样聚合网络也提出用分批量(Mini-batch)处理数据的方法训练模型,在每个批量输入数据下只需要加载对应节点的局部结构,避免了整张网络的加载,这使得在大规模数据集上搭建图卷积神经网络成为可能.

此外,DCNN<sup>[30]</sup>利用随机行走后得到的  $K$  跳转概率定义节点间的权重,第  $m$  层的结构如下:

$$H^{m+1} = h(P^K H^m W) \quad (20)$$

其中,  $P^K$  表示两个节点在随机行走下  $K$  跳可达概率,  $W$  是需要学习的参数.此类模型刻画了节点之间的高阶信息,但是由于  $P^K$  的计算复杂度为  $O(n^2 K)$ ,难以扩展到大图上.

不同于以往认为节点只属于唯一标签的方法,基于置信度的图卷积网络(ConfGCN)<sup>[31]</sup>认为节点是以一定的置信度为一个标签,因此 ConfGCN 给每个节点学习置信度函数,并将其作用在节点相关性上,修正聚合函数.图 6 给出了 ConfGCN 在应用到节点二分类问题上的结构,在置信度作用后的聚合函数上,异配连接更容易被识别出来,进而降低相关系数.超图卷积网络(HGNN)<sup>[32]</sup>则认为节点间的相关性不应该是节点两两之间产生,而是一组节点相互影响进而构建组内节点相关性.基于此想法, HGNN 将边拓展到连接多个节点的超边,在超边上定义聚合函数进行节点特征传播.

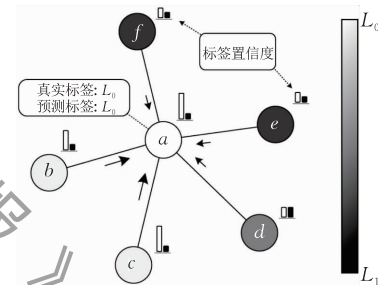


图 6 基于置信度的图卷积网络<sup>[31]</sup>

以上基于聚合函数的空间方法着眼于空间方法中的根本问题,即聚合函数的构造.随着图卷积神经网络的发展,研究人员开始考虑更复杂的场景,并涌现出一类建模信息更为丰富的空间方法,包括如何在具备连边信息的网络上搭建图卷积神经网络,如何建模高阶信息等,我们在第 4 节对这类空间方法做详细介绍.

### (3) 深度理解图卷积网络

上文已经介绍了图卷积神经网络的通用框架和建模方法,在这一节中,我们将分析图卷积模型的内部机制及不同模型的建模能力.

Li 等人<sup>[45]</sup>在图卷积神经网络的协同训练(Co-Train GCN)一文中对一阶图卷积神经网络(GCN)展开详细分析,并指出 GCN 的本质是拉普拉斯平滑. Li 等人指出单层的 GCN 也显著优于全连接网络(FCN),而 FCN 和 GCN 的区别仅在于式(13)中



的  $\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}}$ . 对图做拉普拉斯平滑, 其形式如下

$$(\mathbf{I} - \gamma \hat{\mathbf{D}}^{-1} \hat{\mathbf{L}}) \mathbf{X} \quad (21)$$

其中,  $\hat{\mathbf{L}} = \hat{\mathbf{D}} - \hat{\mathbf{A}}$ ,  $0 < \gamma < 1$  是超参数, 用以调节源节点和周围节点的权重. 令  $\gamma = 1$ , 则拉普拉斯平滑形式为  $\hat{\mathbf{D}}^{-1} \hat{\mathbf{A}} \mathbf{X}$ , 此时如果用对称标准化的拉普拉斯算子  $\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{L}} \hat{\mathbf{D}}^{-\frac{1}{2}}$  替代  $\hat{\mathbf{D}}^{-1} \hat{\mathbf{L}}$ , 则式(21)等价为  $\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}$ . 由此可见, GCN 的本质是在网络上做拉普拉斯平滑. 拉普拉斯平滑使用每个节点的邻居及自身表达的加权和作为当前节点的新特征, 由于在分类任务中, 同类节点倾向于稠密连接, 平滑使得同类节点的特征更为相似, 进而提升了下游的分类任务. 同时文章指出, 多层的 GCN 在更新目标节点表达时, 会混合其他类节点的信息, 进而导致效果降低.

Xu 等人<sup>[51]</sup>把 Weisfeiler-Lehman 测试泛化到图卷积模型上并分析了不同图卷积网络的能力. 他们在文章中指出图卷积网络都是在更新节点的特征表达. 当一个模型满足以下性质, 仅当两个节点的特征一致且局部结构一致时, 这两个节点才会被映射到相同位置, 该模型被称为能力最强的图卷积网络. 如图 7 所示, 基于这种定义, 均值聚合的能力大于最大值聚合的能力, 小于求和聚合的能力.

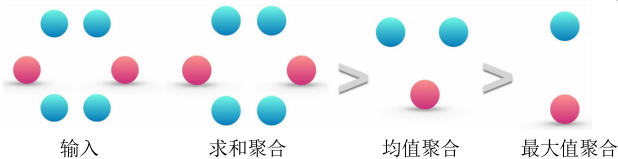


图 7 聚合函数能力比较<sup>[51]</sup> (均值聚合的能力大于最大值聚合的能力, 小于求和聚合的能力)

在这种定义下, 基于均值聚合的函数能力并不是最强大的, 但是 GraphSAGE, GCN 等基于均值或加权均值聚合的模型也取得了很好的效果. 文章指出, 均值聚合关心节点的特征分布, 当节点特征各异时, 均值聚合的能力等同于求和聚合. 因此, 在节点分类等特征丰富的任务上, GraphSAGE, GCN 等也得到了非常有效的结果.

### 3.2 图池化操作

前文介绍了在图上进行的卷积算子, 而在传统的卷积神经网络中, 卷积会和池化相结合, 池化算子一方面能够减少学习的参数, 另外一方面能反应输入数据的层次结构. 而在图卷积神经网络中, 在解决节点级别的任务如节点分类、链接预测时, 池化算子并非必要. 因此, 在图卷积神经网络领域, 池化算子受到的关注较少. 近期, 也有一些研究人员为了解决图级别的问题引入池化算子. 在图结构中使用池化

操作, 主要目的是刻画出网络的等级结构.

图上的池化操作通常对应的是图分类任务, 对于图  $G = (\mathbf{A}, \mathbf{X})$ , 其中  $\mathbf{A}$  为邻接矩阵,  $\mathbf{X}$  为节点的特征矩阵, 给定一些标注的图数据  $\mathbf{D} = \{(G_1, y_1), (G_2, y_2), \dots\}$  和图对应的标签集合  $Y$ , 通过一个映射函数  $f: G \rightarrow Y$ , 能够将图结构映射到对应的标签.

切比雪夫网络 (ChebyNet)<sup>[18]</sup> 利用完全二叉树实现池化算子, 其提出基于 Graclus 贪婪准则为每个节点计算最为匹配的节点, 并将此对节点池化为一个节点. 同时, ChebyNet 通过增加虚假节点保证整个池化过程是一个完全二叉树, 图 8 展示了 ChebyNet 将八节点的图池化为三个节点的过程.

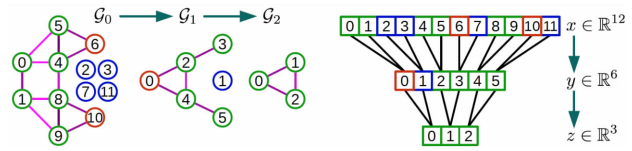


图 8 切比雪夫网络利用完全二叉树实现池化算子<sup>[18]</sup>

由输入图  $G_0$  经过 Graclus 贪婪准则得到  $G_1$  包含 5 个节点, 为保证池化过程是一棵完全二叉树, 则  $G_2$  包含 3 个节点, 并为  $G_1$  增加一个蓝色虚假节点, 为  $G_0$  增加 4 个虚假节点.

不同于 ChebyNet, Ying 等人<sup>[52]</sup>抛弃了预定义的贪婪准则, 并提出微分池化 (DIFFPOOL) 模型, 其提出将图神经网络应用于节点嵌入和池化操作, 模型结构如图 9 所示. 模型在第  $l$  层学习到一个类别分配矩阵  $\mathbf{S}^{(l)} \in \mathbb{R}^{n_l \times n_{l+1}}$ ,  $\mathbf{S}^{(l)}$  的每一行对应于  $l$  层次  $n_l$  类, 每一列对应于  $l$  层的  $n_{l+1}$  节点. 对于给定  $l$  层, 输入的节点表达矩阵为  $\mathbf{Z}^{(l)}$ , 产生新的粗粒度的  $l$  层邻接矩阵  $\mathbf{A}^{(l+1)}$  和新的表达矩阵  $\mathbf{X}^{(l+1)}$ , 即  $\mathbf{A}^{(l+1)}, \mathbf{X}^{(l+1)} = \text{DIFFPOOL}(\mathbf{A}^{(l)}, \mathbf{Z}^{(l)})$ , 具体的计算过程如式(22)所示:

$$\begin{aligned} \mathbf{X}^{(l+1)} &= \mathbf{S}^{(l)T} \mathbf{Z}^{(l)}, \mathbf{A}^{(l+1)} = \mathbf{S}^{(l)T} \mathbf{A}^{(l)} \mathbf{S}^{(l)}, \\ \mathbf{Z}^{(l)} &= \text{GNN}_{l, \text{embed}}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)}), \\ \mathbf{S}^{(l)} &= \text{softmax}(\text{GNN}_{l, \text{pool}}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)})) \end{aligned} \quad (22)$$

其中,  $\mathbf{X}^{(l+1)} \in \mathbb{R}^{n_{l+1} \times d}$ ,  $\mathbf{A}^{(l+1)} \in \mathbb{R}^{n_{l+1} \times n_{l+1}}$ .

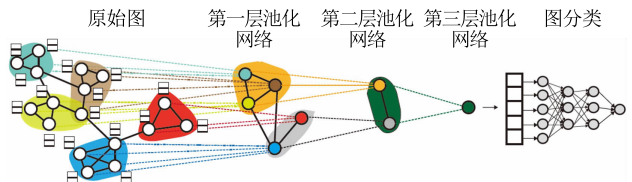


图 9 微分池化模型结构<sup>[52]</sup> (对于每个等级层, 使用图神经网络模型获取节点的表达, 然后使用学习的表达对节点进行聚类, 并使用另一个图神经网络模型作用于这个更加粗粒度的图上. 整个过程重复  $L$  层, 使用最后的输出进行分类)

该模型被用来做软聚类和网络节点表示学习,其需要存储分配矩阵,因此空间复杂度为  $O(kV^2)$ ,  $k$  为池化比例. Cangea 等人<sup>[53]</sup>提出丢掉  $N-kN$  个节点代替聚合这些节点表达,该方法能够减少内存的开销,从而可以建模更大规模的网络结构.

为了在池化过程中充分利用节点特征和局部结构,谱池化(EigenPooling)<sup>[54]</sup>利用谱聚类将整个大图划分成几个不存在重叠的子图,而每个子图即作为池化后的一个新节点,新节点间的连边则基于原子图连边产生. EigenPooling 可以控制每次划分后的子图个数,进而控制每一层的池化比例. 图 10 展示了将 EigenPooling 池化算子和一阶图卷积神经网络结合,完成图分类任务的框架.

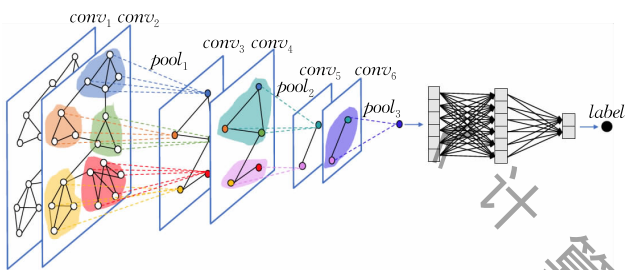


图 10 池化算子和一阶图卷积神经网络结合,完成图分类任务<sup>[54]</sup>(每种颜色表示一个子图,在池化后成为一个新的节点)

无独有偶,基于注意力机制的池化算子(SAG-Pool)<sup>[55]</sup>同样着眼于在池化过程中同时考虑节点属性信息和结构信息,SAGPool 基于注意力机制通过结构和属性信息为每个节点学到一个标量,以此标量表征对应节点在整个图上的重要性,并对此标量进行排序,根据排序结果保留最重要的一部分节点及其连边进而完成池化操作.

池化算子是为了学到图的等级结构,进而完成图级别的任务. 起初的池化算子基于图的拓扑结构,启发式的定义一些节点的舍弃或者融合方式,近期,池化算子不仅依赖于拓扑结构,同样依赖于节点的属性信息,同时池化过程也开始通过注意力机制参数学习等由模型指导完成.

## 4 图卷积神经网络的新进展

上一节介绍了经典的图卷积神经网络算法,他们在很多任务上取得了显著的提升. 然而将这些方法应用到实际的网络化数据上时,仍就面临着一些挑战,例如:真实的网络化数据除了节点和连边还有很多额外的信息,如何在卷积操作中建模这些额外

的信息;真实的网络化数据规模巨大,如何降低图卷积算子的空间和时间消耗使得其能应用在大规模网络化数据上等. 本节将介绍图神经网络的最新进展,并且依据它们解决的问题不同,分成了建模网络额外信息的图卷积网络和适应大规模网络化数据的图神经网络训练方法两部分内容.

### 4.1 建模网络额外信息的图卷积网络

图卷积算子(按特定权重聚合邻居节点的特征)是图卷积神经网络空间方法的核心算子. 部分空间方法虽然能够利用节点的特征来区分不同邻居的重要程度,但依然具有局限性. 它们忽略了网络上除了连边和节点特征之外的其他重要信息,例如:边上的属性,高阶网络结构信息等. 本节将介绍显式建模网络上各种丰富信息的图卷积神经网络.

#### 4.1.1 建模边上信息的图卷积网络

边是网络重要的组成部分,刻画了节点之间的关联关系. 在实际中,不同网络的边蕴含的信息的类型也大不相同:低维离散的类型信息、低维连续的权重信息、高维连续的属性信息等. 根据建模边上信息方式的不同,本文将现有的方法分成了三类:

##### (1) 子图拆解法

此类方法将包含复杂额外信息的网络化数据拆解成多个不同的子图,拆解后每个子图只包含单一的连边类型,对拆解后的每个子图利用传统的图卷积神经网络建模,最后将不同子图上得到的结果按特定方式进行聚合.

关系图神经网络(R-GCNs)<sup>[33]</sup>根据连边的方向、边上的标签类型将原来的网络拆分成不同的子网络,在每个子图上独立地进行邻居特征的聚合. 每一层在聚合操作结束后,将节点在不同子图上得到的结果相加,作为下一层网络的输入. 关系图神经网络的卷积操作可以形式化成如式(23):

$$h_i^{l+1} = \sigma \left( \sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right) \quad (23)$$

其中,  $c_{i,r}$  是归一化因子,可以根据不同任务进行选择,常见的方式是用邻居个数作为归一化因子即  $c_{i,r} = |N_i^r|$ .  $R$  表示不同的边类型,  $W_r^{(l)}$  表示第  $l$  层对于边类型为  $r$  的子图需要学习的参数. 当边类型是高维向量时,连边的类型随维度增长呈指数增长. 此时直接应用关系图神经网络,需要分解成大量的子图,同时每个子图非常稀疏,在这种情况下节点很难学习到有效地表达. 为了减少参数关系,图神经网络使用两种不同的方式,偏置分解(basis-decomposition)和

区块对角化分解 (block-diagonal decomposition). 关系图注意力网络 (R-GAT)<sup>[34]</sup> 在关系图神经网络的基础上引入了注意力机制, 作者提出了两种引入注意力机制的邻居聚合函数替换关系图神经网络中只利用网络结构的聚合函数, 进一步提升了实验结果. 关系图神经网络和关系图注意力网络更适用于具有离散类型的边特征的网络.

符号网络 (signed network) 是一种特殊的边上包含类型信息的网络. 除了包括“朋友关系 (positive link)”, “敌对关系 (negative link)”两种类型的连边, 符号网络还具有结构平衡 (structural balance) 等特殊的社会学性质. 其中“敌人的敌人是朋友” (距离中心节点两跳“敌对关系”的节点和具有“朋友”关系) 这一性质使得仅根据边类型划分聚合函数不可行. 符号图卷积网络 (SGCN)<sup>[35]</sup> 针对符号网络的特殊性质进行设计. 在符号图卷积网络中, 每个节点包含“朋友表达 (friend representation)”和“敌人表达 (enemy representation)”两部分, 分别利用“朋友”聚合器 (friend aggregator) 和“敌人”聚合器 (enemy aggregator) 学习得到. 卷积层结构如图 11 所示. 由于第一层网络只能刻画到一阶邻居, 因此不存在结构平衡问题. 第一层的“朋友表达”通过聚合直接“朋友”的表达即可得到, “敌人表达”同理. 从第二层网络开始, 由于结构平衡性的存在, “朋友”集合和“敌人”集合得到了拓展. 当前节点  $i$  的“朋友表达”将由自身“朋友”的“朋友”, 自身“敌人”的“敌人”, 以及自身的“朋友”三部分通过朋友聚合器得到. 同理“敌人表达”由自身的“敌人”、自身“敌人”的“朋友”以及自身“朋友”的“敌人”通过“敌人聚合器”得到.

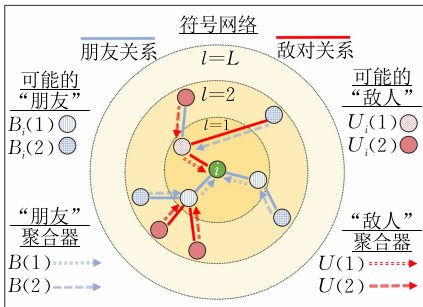


图 11 符号图卷积网络卷积操作示意图<sup>[35]</sup>

## (2) 对偶图构建法

第二类方法为对偶图构建法. 此类方法通过构建对偶网络的方式, 将原网络中的边转换成对偶网络中的节点, 将原图中边上的特征转移到了对偶图的节点上. 对偶图的连边不再具有特征, 因此可以在

对偶图上直接应用传统的图卷积神经网络, 同时原图中边上特征变成了对偶图中节点上的特征, 可以通过在对偶图上定义的图卷积刻画.

原始对偶图卷积神经网络 (DPGCNN)<sup>[36]</sup> 提出了一种构建对偶网络的方法. 以图 12 为例, 原始对偶图卷积神经网络将原网络中的边映射成对偶网络中的节点 (例如右侧图中  $(0, 4)$  节点表示左侧图中 0 节点和 4 节点之间的连边). 同时, 如果原网络中的两条连边有共同节点 (如左侧图中边  $\langle 0, 1 \rangle$  和  $\langle 0, 2 \rangle$  有共同节点 0), 那么这两条边对应的对偶网络中的节点  $((0, 1), (0, 2))$  之间存在一条连边, 此外, 原网络边上的特征被保存在对偶网络的节点上.

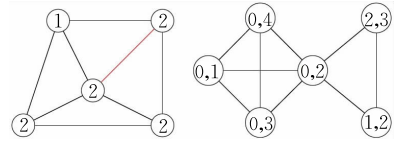


图 12 左侧为原始网络, 右侧为对应的对偶网络<sup>[36]</sup>

原始对偶图卷积神经网络分别在对偶网络和原始网络上定义了对应的卷积操作. 对偶卷积的形式如式 (24) 和 (25):

$$\tilde{f}'_{ij} = \xi_d \left( \sum_{r \in N_i} \tilde{\alpha}_{ij,ir} \tilde{f}_{ir} \tilde{W} + \sum_{t \in N_j} \tilde{\alpha}_{ij,tj} \tilde{f}_{tj} \tilde{W} \right) \quad (24)$$

$$\tilde{\alpha}_{ij,ik} = \frac{e^{\eta(\tilde{a}([\tilde{f}_{ij} \tilde{W}, \tilde{f}_{ik} \tilde{W}]))}}{\sum_{r \in N_i} e^{\eta(\tilde{a}([\tilde{f}_{ij} \tilde{W}, \tilde{f}_{ir} \tilde{W}]))}} + \sum_{t \in N_j} e^{\eta(\tilde{a}([\tilde{f}_{ij} \tilde{W}, \tilde{f}_{tj} \tilde{W}]))}} \quad (25)$$

其中,  $\xi_d$  表示对偶网络中的激活函数 (例如 ReLU),  $f_i$  表示原图中节点  $i$  的特征,  $\tilde{f}'_{ij}$  表示对偶网络中节点  $(i, j)$  的特征 (对应原网络中的边  $\langle i, j \rangle$ ) 当边上缺少特征时可以通过拼接两段节点的特征代替:  $\tilde{f}'_{ij} = [f_i, f_j]$ , 也可以同时使用两者. 其中  $\tilde{W}$  表示需要学习的参数,  $\eta$  表示 Leaky ReLU 激活函数,  $N_i$  表示原网络中节点  $i$  的邻居节点构成的集合.

原始卷积与传统图注意力网络的形式类似, 唯一的区别在于计算注意力系数时输入特征采用的是由对偶卷积计算得到的边上的特征,

$$f'_i = \xi_p \left( \sum_{r \in N_i} \alpha_{ij} f_r W \right), \quad \alpha_{ij} = \frac{e^{\eta(\tilde{a}(\tilde{f}'_{ij}))}}{\sum_{k \in N_i} e^{\eta(\tilde{a}(\tilde{f}'_{ik}))}} \quad (26)$$

直观上理解, 对偶图上的图卷积网络, 为原始图中的每一条边学到了一个表示, 该表示能反映节点特征沿着这条边传播的权重. 而原始图中的图卷积网络依据边的传播权重, 进行特征的传播, 为节点学习表达.



该方法能适用于具有高维边特征的场景,但是需要将原网络转换成对偶网络.在实际中,网络的边数往往远大于网络的节点数,因此对偶网络的规模往往会远大于原网络.在对偶网络中应用卷积操作会有很大的计算代价.故此方法难以应用到大规模的网络中.

### (3) 权重重调法

第三类方法为权重重调法.此类方法认为边上的特征信息只会影响中心节点聚合邻居节点特征的权重,因此这类方法在计算聚合权重时引入了边上的特征信息.

边注意力网络(EGAT)<sup>[37]</sup>定义卷积操作如下所示:

$$x'_i = \sigma \left[ \sum_{j \in N_i} \alpha(x_i, x_j, e_{ijp}) g(x_j) \right], g(x_j) = W x_j \quad (27)$$

其中,  $P$  表示边上特征的维度,  $e_{ijp}$  表示连边  $\langle i, j \rangle$  特征在第  $p$  维上的值,  $g(x_j) = W x_j$  表示节点  $j$  经过线性变换后的特征,  $W$  是需要学习的参数,  $x_j$  是输入特征.边注意力网络改变了传统图注意力网络计算节点之间注意力权重的方式,将边上的特征也作为一个重要的输入参与权重的计算.具体形式如式(28):

$$\alpha(x_i, x_j, e_{ijp}) = \frac{1}{C_{ip}} f(x_i, x_j) e_{ijp}, C_{ip} = \sum_{j \in N_i} f(x_i, x_j) e_{ijp} \quad (28)$$

其中,  $C_{ip}$  是归一化因子.而  $f(x_i, x_j)$  可以采用计算注意力权重时常用的一些形式,例如单层全连接网络等.该方法可以理解成在传统图注意力网络基础上,利用边上的特征对注意力的权重再额外加权,同时认为边上的特征的每一个维度之间是独立的.

边约束卷积网络(ECC)<sup>[38]</sup>,将图上连边的类型用低维向量表示,并将这些向量通过单层线性神经网络转化为该连边对应两个节点间的聚合权重.边约束卷积网络每一层的结构如图 13 所示,其中  $L(i, j)$  表示  $i, j$  节点连边的低维向量表示,  $F^l(\cdot)$  表示第  $l$  层的参数,用于将输入的边向量转化为节点间的权重,  $X^l(i)$  表示  $i$  节点在  $l$  层的特征表达.该模型通过学习参数将边关系转化为权重,并用该权重做节点的聚合,使得为每个节点学到局部结构相关的表达.但该方法仅考虑了边上特征各个维度之间的关系,忽略了两端节点特征之间的相似性.

#### 4.1.2 建模网络高阶信息的图卷积网络

除了节点特征和边特征,网络的高阶结构特征也是筛选邻居,以及区分邻居节点重要性的有效依

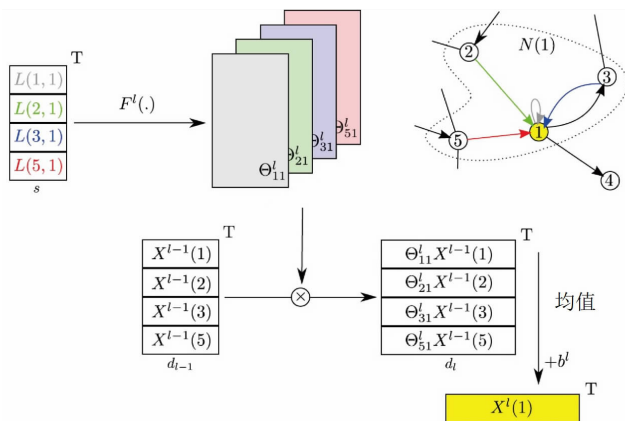


图 13 边约束卷积神经网络结构<sup>[38]</sup>(图上连边的类型用低维向量表示,并将这些向量通过参数转化为该连边对应两个节点间的权重,基于以上权重做节点表达更新)

据.在图 14 中,方形色块的颜色表示对应节点的类别,以中心节点为目标节点,目标节点的类别为机器学习(绿色类别).当仅考虑一阶邻居时,与目标节点连接最多的点是可视化(蓝色类别)的节点,因此 GCN 会将目标节点误分成可视化.但是当引入高阶结构(三角形模体)时,可以发现目标节点只剩下三个标签为机器学习的邻居了,因此目标节点在高阶结构下可以被正确分类.

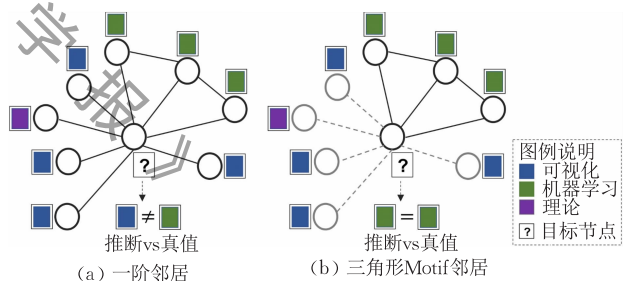


图 14 对节点分类时,仅用一阶邻居和考虑高阶结构对比<sup>[42]</sup>(a)将所有一阶邻居视为同等重要的节点,可能会导致推测错目标节点真实的标签;(b)利用三角模体来筛选邻居节点,可以找到与中心点关系更密切的节点,使得中心点更可能被推测出真实的标签)

为了在引入高阶关系的同时更好地区分每个“邻居”节点的重要性,高阶图卷积神经网络(HA-GCN)<sup>[39]</sup>用  $k$  阶邻接矩阵  $\hat{A}^k$  替换 GCN<sup>[19]</sup> 中的  $\hat{A}$ .同时为了区分每个节点的重要性,HA-GCN 引入了元素级的自适应权重矩阵.高阶图卷积神经网络的卷积层形式化如式(29):

$$\hat{L}_{HA}^{(k)} = (\tilde{W}_k \odot \hat{A}^k) X + B_k \quad (29)$$

其中,  $\hat{A}^k$  表示  $k$  阶邻接矩阵,  $B_k$  表示第  $k$  阶的偏置(biases)矩阵,  $\tilde{W}^k$  表示自适应权重矩阵,根据节点的特征、邻接关系以及原始权重矩阵计算得到,

$$\tilde{\mathbf{W}}_k = g \odot \mathbf{W}_k, \quad g = f_{adp} = (\hat{\mathbf{A}}^k, \mathbf{X}) \quad (30)$$

自适应权重矩阵由式(30)计算得到, 其中  $g$  是非线性算子(单层神经网络), 输入特征由节点的特征和高阶邻接矩阵的拼接得到。

在异质信息网络中, 节点和边具有多种类型, 单纯使用  $k$  阶邻接矩阵丢失了重要的语义信息。例如在图 15 所示的 DBLP 网络中, 节点具有三种不同的类型: 作者(A), 论文(P), 会议(V)。  $v_1$  和  $p_6$  都是中心节点  $a$  的二阶邻居, 无法用  $k$  阶邻居矩阵区分, 但是显然两者对  $a$  提供的语义信息不相同。因此异质信息网络上的图卷积神经网络需要能刻画具有更加复杂的高阶语义信息。

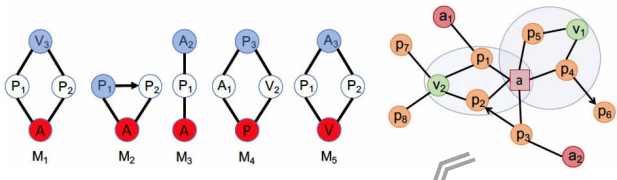


图 15 DBLP 数据集中的模体类型<sup>[40]</sup> (a) 列举了 DBLP 数据集中出现的模体类型, 节点包含三种类型作者(A), 论文(P), 会议(V); (b) 展示了以  $a$  为中心节点的  $M_1$  模体实例)

模体卷积网络(Motif-CNN)<sup>[40]</sup>通过显式定义模体(motif), 来区分邻居节点对中心节点不同的语义作用。模体卷积网络利用模体来定义网络上的卷积算子, 根据节点的语义角色信息来共享参数。其卷积形式如式(31)所示:

$$h^M(v_i) = \sigma \left( \omega_0 x_i + \frac{1}{D_{ii}^M} \sum_{j=1}^N \sum_{k=1}^{K_M} \omega_k (\mathbf{A}^M)_{kij}^M x_j \right) \quad (31)$$

其中,  $\mathbf{D}^M$  是对角矩阵, 对角元素表示每个节点参与的不同的模体的个数。  $\mathbf{A}^M$  是邻接模体张量,  $\mathbf{A}_{kij}^M$  表示节点  $j$  以  $k$  角色出现在以节点  $i$  为中心的模体实例  $M$  的次数。  $\omega_k$  是共享参数, 相同语义角色的节点之间共享参数。不同的任务需要的语音信息不一样, 即在不同任务下不同模体的重要程度不同。模体卷积网络采用注意力机制给不同的模体学习重要程度的权重, 形式如式(32)和(33),

$$e_{k,i} = a(h^k(v_i), z_k) = \frac{\mathbf{z}_k^T h^k(v_i)}{\sqrt{|\mathbf{z}_k|}} \quad (32)$$

$$\alpha_{k,i} = \text{softmax}_k(e_{k,i}) = \frac{\exp(e_{k,i})}{\sum_{j=1}^U \exp(e_{j,i})} \quad (33)$$

模体卷积网络采用点积形式计算注意力权重, 其中  $\mathbf{z}_k$  表示模体  $M_k$  的共享注意力向量, 不同的模体独立计算注意力权重, 不共享参数。  $\alpha_{k,i}$  表示节点  $i$

在模体  $k$  下的注意力权重。

$$h(v_i) = \sum_{k=1}^U \alpha_{k,i} h^k(v_i) \quad (34)$$

最终目标节点的表达由不同模体下表达的加权和得到。

和模体卷积网络类似, 异质注意力网络(HAN)<sup>[41]</sup>利用元路径(meta-path)来定义卷积操作, 对于同一种元路径共享参数。其模型结构如图 16 所示。

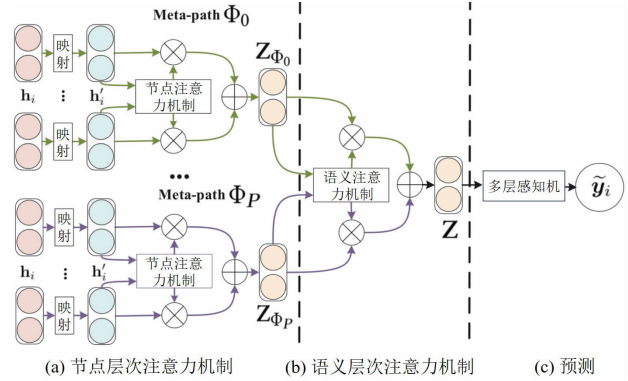


图 16 HAN 模型结构<sup>[41]</sup>

聚合邻居节点的特征是图卷积网络的关键步骤, 此问题的关键是如何根据任务选择合适的邻居节点, 同时区分各个邻居节点的重要性。前文提到的几个方法利用网络的高阶结构筛选候选邻居节点, 用注意力机制计算邻居节点的重要程度。而模体卷积决策网络(MCN)<sup>[42]</sup>将这个关键点形式化成了决策问题: 如何选择合适的邻居, 以及对选择的邻居赋予多大的聚合权重? 模体卷积决策网络认为选择邻居节点集合就是为每个节点选择最合适的模体关系, 而共同在模体中出现的频率作为聚合权重。与之前的方法类似, 模体卷积决策网络也通过定义  $k$  步模体矩阵的方式构建候选的邻居节点集合。模体卷积决策网络采用  $T$  种不同的模体结构, 每一种模体计算  $k$  个不同步长的模体矩阵。  $k$  阶模体定义和  $k$  阶邻接矩阵类似, 为  $k$  个相同模体矩阵的乘积。

$$\Psi(\mathbf{A}_i^k) = \Psi(\underbrace{\mathbf{A}_i \cdots \mathbf{A}_i}_k) \quad (35)$$

模体卷积决策网络的决策过程分为两步: 第 1 步选择与目标节点最相关的模体, 第 2 步选择最相关的步长。模体卷积决策网络采用注意力机制计算模体被选中的概率以及各个步长的相关程度。模体卷积决策网络的目标函数包含两部分: 分类的损失函数以及决策的损失函数。分类的损失函数使用交叉熵度量节点分类的准确性, 而决策的损失函数用来衡量决策的准确性, 对于每个节点  $v$ , 如果分类准

确则回报(reward)为正,否则回报为负.由于每一层决策的结果都会影响最终的结果,此方法通过为每一层的所有节点都计算回报来约束每一层都做出正确决策.

除了作为选择邻居的依据,模体矩阵由于其对称性,可以用来解决图卷积神经网络谱方法不能用于有向网络的缺点.模体网络(Motif-Net)<sup>[43]</sup>提出利用模体矩阵替换基于谱方法卷积中的拉普拉斯矩阵,就可以将图卷积神经网络谱方法应用在有向网络上.

根据算法适用的网络类型,将本节提到的方法总结如表 3.

表 3 本节方法总结

模型	节点类型	边属性	邻居范围
RGCN <sup>[33]</sup>	多种类型	高维离散	直接邻居
SGCN <sup>[35]</sup>	单一类型	离散标量	高阶邻居(结构平衡)
DPGCN <sup>[36]</sup>	单一类型	高维连续	直接邻居
EGAT <sup>[37]</sup>	单一类型	连续标量	直接邻居
ECC <sup>[38]</sup>	多种类型	高维连续	高阶邻居(模体)
HA-GCN <sup>[39]</sup>	多种类型	单一类型	高阶邻居(模体)
Motif-GCN <sup>[40]</sup>	多种类型	单一类型	高阶邻居(模体)
HAN <sup>[41]</sup>	多种类型	单一类型	高阶邻居(模体)
MCN <sup>[42]</sup>	多种类型	单一类型	高阶邻居(模体)
Motit-Net <sup>[43]</sup>	多种类型	单一类型	高阶邻居(模体)

4.2 图卷积网络的训练技术

图卷积网络已经在很多场景下取得了显著的结果,但是依然面临着一些问题:难以直接应用在大规模网络上,堆叠多层图卷积网络会造成效果的下降等.本节介绍了将图卷积网络应用在大规模网络上时采用的训练技术.

4.2.1 深层图卷积神经网络

残差网络解决了传统神经网络在网络层数很深时拟合能力反而下降的问题.在堆叠多层图卷积层后,由于节点之间的特征变得过于平滑缺少区分性,也会导致网络的效果变差.简单的应用残差连接并不能解决这个问题,因为在 GCN 中每一层节点只将特征传递给自己的直接邻居,不同的节点传播特征的速度并不相同:中心节点(hub)可能经过一两层图卷积就将特征传递给了整个网络中的大部分节点,而对于网络中的边缘节点,需要很多次的传播才能影响网络中的部分节点.跳跃知识网络(Jumping Knowledge Network)<sup>[44]</sup>使用跳跃连接(jump connections)和注意力机制为每个节点选择合适的传播范围.跳跃知识网络的模型结构如图 17 所示.

跳跃连接将每一层图卷积的结果都连接到网络

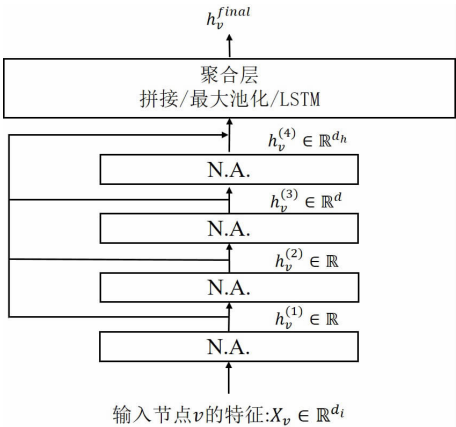


图 17 4 层 Jumping Knowledge Network 的示意图<sup>[44]</sup> (N.A. 表示从邻居节点聚合特征的操作)

的最后一层,并且在最后一层引入自适应的聚合机制,为每个节点选择合适的邻居范围.跳跃知识网络可以使用多种聚合方式,包括拼接、最大池化以及递归神经网络.

4.2.2 大规模网络图卷积技术

传统的机器学习方法,例如:多层感知机(MLP)等,认为样本之间是独立的,因此可以使用分批量(Mini-Batch)处理数据的方式来应对大规模的训练的数据.而在图卷积神经网络中,由于卷积操作依赖于邻居节点,因此直接使用分批量的训练方式需要引入大量的相关节点.即对于某个中心节点,更新其表达需要用到的邻居节点个数随着网络层数增加呈现指数增长.同时在大规模网络中,某些大度节点即使只考虑二阶邻居计算量也过于庞大.这两者导致直接使用分批量的训练方法不能解决 GCN 难以应用在大规模网络上的问题.

GraphSAGE<sup>[29]</sup>使用分批量的训练方法,结合采样邻居节点 Neighbor Sampling(NS)的方法,将每次计算所需要的节点数目控制在一定范围之内.例如,假设模型采用两层卷积,同时限定第一层采样邻居个数为  $D_1$ ,第二层采样邻居个数为  $D_2$ ,那么对于每一个节点的感受野范围为  $D_1 \times D_2$ .若批量大小(Batch size)为  $k$ ,那么对于一个批量内的数据可以将需要计算的节点上限控制为  $k \times D_1 \times D_2$ .

GraphSAGE 使用随机采样邻居节点的方式降低了每次卷积需要计算的节点数目,但是这种估计方式是有偏差的,同时无法保证收敛性.基于方差控制的算法 Control Variate Based Algorithm(CV)<sup>[47]</sup>在 NS 的基础上利用那些没有被采样到的节点的历史表达(historical activation)来控制方差.该方法认为节点  $v$  的表达  $h_v$  和它的历史表达  $\bar{h}_v$  在参数变化不大的情况下,应该很接近.因此对于没有采样到的



节点,用它的历史表达来近似。

CV, NS 以及原始 GCN 的区别如图 18 所示,红色的节点表示最新的表达,蓝色的节点表示历史的表达。如图所示,NS 和 CV 方法相比于传统的 GCN 明显减少了每次卷积操作需要用到的邻居个数。NS 仅仅使用采样的邻居节点来估计整个邻居的表达,方差大。而 CV 方法每次需要采样的邻居节点数目和 NS 方法相同,同时利用上一轮迭代的邻居节点的表达来估计那些没有被采样到的节点的表达,可以减少对中心节点表达估计的方差。

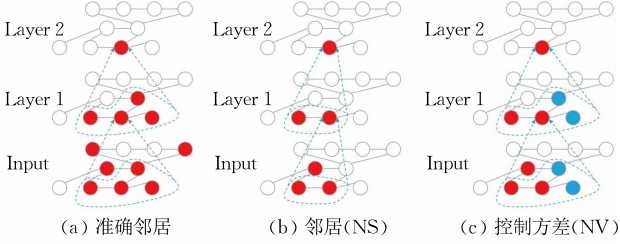


图 18 邻居采样方法对比示意图<sup>[47]</sup> (CV, NS 以及原始 GCN 在更新节点表达时需要依赖节点的对比,红色的节点代表更新后的表示,蓝色的节点是更新前的表达 ((a)原始的 GCN 更新所有依赖的节点,(b) NS 采样的方法只依赖于部分邻居节点,(c) CV 方法需要依赖的节点))

CV 和 NS 是节点级的采样方法,以 GCN<sup>[19]</sup>为例,其图卷积操作可以写成期望形式:

$$h^{(l+1)}(v_i) = \sigma_w^{(l)}(N(v_i) \mathbb{E}_{p(u_j|v_i)}[h^{(l)}(u_j)]) \quad (36)$$

$\sigma_w^{(l)}$  表示线性特征变换以及非线性激活函数,  $p(u_j|v_i)$  是给定节点  $v_i$  的情况下,采样到节点  $u_j$  的概率。此类方法的关键在于如何很好地估计概率  $p(u_j|v_i)$ 。虽然此类方法可以将每个节点需要考虑的节点数限定在常数级别,但是对于每个节点需要考虑的节点数随着网络层数增加依旧是呈指数增长的。同时各个节点之间的采样过程是独立的,因此在邻居节点之间没法复用上一层出现的节点。

层级采样方法不再以节点为单位进行采样,而是对于每一层卷积网络采样需要用到的节点,即每一层从原图中采样出一个子图,只对采样到的节点进行卷积操作,同时所有的节点共享上一层的父亲节点。层级采样方法可以理解成一种重要性采样方法(importance sampling),用全局的概率  $q(u_j|v_1, \dots, v_n)$  估计采样概率  $p(u_j|v_i)$ 。FastGCN<sup>[48]</sup> 和 Adapt<sup>[49]</sup> 均采用层级采样方法来加速 GCN。两者的区别在于 FastGCN 认为层与层之间的采样是独立的,而 Adapt 认为低层网络在采样时应该受到上层采样结果的影响(越低层越靠近输入)。在这种模式下,需要采样的节点个数和网络层数变成了线性关系,同时

能重复共享父亲节点减少了重复计算。

节点级的采样和层级采样的区别如图 19 所示。在节点级的采样中,每个父节点的邻域不会被其他父节点看到,因此邻居和其他父节点之间的连接未被使用。相反,对于层级采样,所有邻居由上一层的节点共享,因此所有层间连接都被利用了<sup>[49]</sup>。

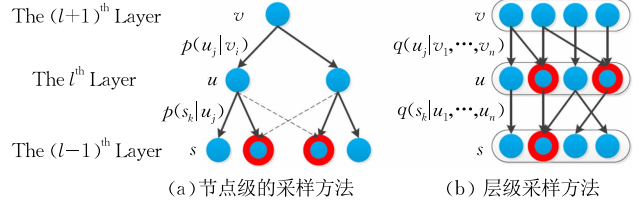


图 19 不同采样方法示意图<sup>[49]</sup>

#### 4.2.3 半监督节点分类问题的训练技术

GCN 在半监督节点分类问题中取得了有效地结果,但是由于所采用的卷积算子隐式地认为所有一阶邻居同等重要,导致 GCN 对于处于网络数据中心的节点的分类效果不佳<sup>[56]</sup>。DGCN<sup>[56]</sup> 指出当去除掉这类节点中的某些连边后能提升 GCN 的效果,说明只利用一阶邻居的邻近性具有一定的局限性,需要引入额外的信息来区分一阶邻居。DGCN 通过引入 PPMI 矩阵的方式减少了一阶邻居对节点的影响,同时引入了相比一阶邻居更加丰富的关系。利用 PPMI 进行卷积能够将 GCN 中分错的一些点纠正,但同时也引入了新的错误,因此作者采用集成学习的方式结合两种不同卷积的优势,提升分类效果。其模型结构如图 20 所示:两个卷积算子共享降维参数  $W$ ,  $L_0$  表示交叉熵,希望图卷积网络  $A$  能够将节点正确分类,  $L_{reg}$  是无监督的约束,希望对于同一个节点,经过两个卷积神经网络得到的表示相同。

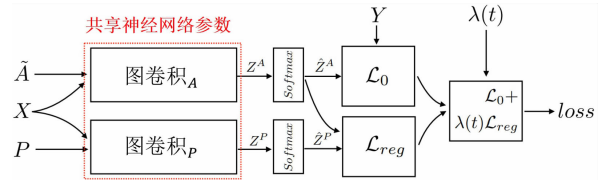


图 20 DGCN 模型结构示意图<sup>[56]</sup>

## 5 应用

图卷积神经网络自提出以来,受到了研究人员的大量关注,主要集中于以下几个领域:网络分析、推荐系统、生物化学、交通预测、计算机视觉、自然语言处理等。图卷积神经网络的应用领域既包括计算

机科学、人工智能、信号处理等传统机器学习领域，也包括物理、生物、化学、社会科学等跨学科的研究。不同的领域包含各种不同的图数据，节点和连边关系也各有不同，如何结合领域知识对给定的图数据利用图卷积神经网络进行建模是图卷积神经网络应用的关键问题. 我们在表 4 总结了常见的应用分类.

表 4 图卷积神经网络应用领域总结表

应用领域	节点	连边	应用问题	任务	相关论文
网络分析	用户	社交关系	用户影响力预测	图回归	DeepInf <sup>[57]</sup>
	论文	引用关系	半监督节点分类	节点分类	GCN <sup>[19]</sup> , GAT <sup>[28]</sup> , GWNN <sup>[21]</sup>
交通预测	道路中的传感器	物理距离	道路流量预测	节点回归	DCRNN <sup>[58]</sup> , TGC-LSTM <sup>[59]</sup> , GGRU <sup>[60]</sup> , Yu 等人 <sup>[61]</sup> , GCN-GAN <sup>[62]</sup>
推荐系统	用户, 商品	购买某商品的可能性分数	用户偏好推荐	矩阵补全/链接预测	MGCNN <sup>[25]</sup> , GC-MC <sup>[63]</sup> , PinSage <sup>[46]</sup> , RippleNet <sup>[64]</sup> , GraphRec <sup>[65]</sup>
生物化学	小分子	化学键	化学功能预测	图分类	Duvenaud 等人 <sup>[66]</sup> , Kearnes 等人 <sup>[67]</sup> , MPNNs <sup>[26]</sup> , DIFFPOOL <sup>[52]</sup> , GAM <sup>[68]</sup>
	小分子	化学键	分子发现	图生成	MolGAN <sup>[69]</sup> , GCPN <sup>[70]</sup>
	蛋白质	相互作用	蛋白质预测	图分类	Fout 等人 <sup>[71]</sup> , Graph-CNNs <sup>[72]</sup>
	蛋白质/药物	相互作用	副作用预测	链接预测	Decagon <sup>[73]</sup>
计算机视觉	实体概念	语义关系	图片分类	少样本学习/零样本学习	Garcia 等人 <sup>[74]</sup> , Marino 等人 <sup>[75]</sup> , Lee 等人 <sup>[76]</sup> , Wang 等人 <sup>[77]</sup> , ADGPM <sup>[78]</sup>
	3D 点云图	距离信息	语义分割/形状分类	语义分割	3DGNN <sup>[79]</sup> , Wang 等人 <sup>[80]</sup> , SPG <sup>[81]</sup> , RGCNN <sup>[82]</sup>
	对象	对象之间的关联	场景图生成/视觉推理	图生成	Graph VQA <sup>[83]</sup> , GRM <sup>[84]</sup> , Chen 等人 <sup>[85]</sup>
	概念实体	语义关系	知识推理	链接预测	GNN-for-OKB <sup>[86]</sup>
自然语言处理	词	依赖关系	语义角色标注/抽象含义表达	关系提取	Semantic GCN <sup>[87]</sup> , LSTM+GCN <sup>[88]</sup> , C-GCN <sup>[89]</sup> , Song 等人 <sup>[90]</sup>
	词	依赖关系	事件提取	序列标注	JMEE <sup>[91]</sup> , Nguyen 等人 <sup>[92]</sup>
	词	关系标记	机器翻译	序列生成	Bastings 等人 <sup>[93]</sup> , Beck 等人 <sup>[94]</sup>
	词	共现关系	文本分类	图分类	HR-DGCNN <sup>[95]</sup> , TextGCN <sup>[96]</sup>

5.1 网络分析

在社会网络分析领域,引文网络是最为常见的数据,即节点为论文,连边关系为引用关系,常见的数据集包括 Cora, DBLP, Citeseer 等. 这些常见的网络的数据集的描述如表 5. 一个典型的分类任务是给定每篇文章的内容信息和文章之间的引用关系,将每篇文章分类到对应的领域中. 例如,在节点的非监督分类场景下,已知节点的属性信息包括文章的标题或摘要信息,以及节点之间的引用关系构成的网络信息,给定少量的数据标签,通过机器学习的方式,对网络中的每个节点的所属领域进行划分. 在该任务中,图卷积神经网络将节点文本属性和引用网络结构有效的建模,取得了巨大的成功. 如表 6 所示,相比直接使用内容信息(如 MLP),仅使用结构信息(如 DeepWalk<sup>[14]</sup>)和传统图上半监督节点分类方法(如 Planetoid<sup>[97]</sup>),以 GCN 为代表的图卷积神经网络算法的分类准确率远高于传统的方法,其中, GAT 相比 Planetoid 模型在 Cora、Citeseer、Pubmed 上,分别提升 7.3%, 7.8% 和 1.8%. 该任务也通常被视为衡量一个图卷积神经网络模型是否有效的基准任务. GCN<sup>[19]</sup>, GAT<sup>[28]</sup>, GWNN<sup>[21]</sup>等都使用该任务验证模型的有效性.

表 5 半监督节点分类常用数据集

数据集	Cora	CiteSeer	PubMed	NELL
类别	引用网络	知识图谱		
节点数	2708	3327	19717	65 755
边数	5429	4732	44 338	266 144
类别数	7	6	3	210
特征	1433	3703	500	5414
标签比例	0.052	0.036	0.003	0.003

表 6 半监督节点分类不同模型结果 (单位: %)

方法	数据集属性		
	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
MoNet	81.7±0.5	70.3	79.0
GCN	81.5	—	78.8±0.3
GAT	83.0±0.7	72.5±0.7	79.0±0.3
GWNN	82.8	71.7	79.1

另外一个较为常见的基准任务为图分类任务,对于给定的图结构,每个图有一个标签,通过学习图结构的模式,达到图分类的任务. 图分类的数据集<sup>[98]</sup>

主要包括生物化学结构分类(如 Enzymes、D&D、Proteins 等)、社交网络分类(如 Reddit-Multi-12k)和科学合作网络(如 Collab)等. 相关的数据集描述见表 7. 而不同的模型的分类结果见表 8, 图卷积神经网络相比传统的图核的方法, 也取得了明显的提升. 其中, DIFFPOOL 相比 Graphlet, 在 Enzymes、D&D、Proteins、Reddit-Multi-12k 和 Collab 上, 分别提升 21.5%、5.80%、3.34%、25.35% 和 10.82%. 图卷积神经网络在图节点分类和图分类的机器学习任务上相比传统的图表示学习和图核方法取得了效果上的明显提升.

表 7 图分类任务常用数据集

数据集	数据集属性			
	图的数量	类别数	图平均节点数	图平均边数
Enzymes	600	6	32.63	62.14
D&D	1178	2	284.32	715.66
Proteins	1113	2	39.06	72.83
Reddit-Multi-12k	11929	11	391.41	456.89
Collab	5000	3	74.49	2457.78

表 8 图分类不同模型结果 (单位: %)

方法		数据集				
		Enzymes	D&D	Proteins	Reddit-Multi-12k	Collab
Kernel	Graphlet	41.03	74.85	72.91	21.73	64.66
	Shortest-path	42.32	78.86	76.43	36.93	59.10
	1-WL	53.43	74.02	73.76	39.03	78.61
	WL-OA	60.13	79.04	75.26	44.38	80.74
GNN	PatchySan	—	76.27	75.00	41.32	72.60
	GraphSAGE	54.25	75.42	70.48	42.24	68.25
	ECC	53.50	74.10	72.65	41.73	67.79
	Set2set	60.15	78.12	74.29	43.49	71.75
	SortPool	57.12	79.37	75.54	41.82	73.76
	DIFFPOOL	62.53	80.64	76.25	47.08	75.48
	EigenGCN	64.50	78.60	76.60	—	—

此外,在社区发现问题上,过去的算法主要是将其显式的定义和优化图分区最小割问题. Chen 等人<sup>[99]</sup>提出了一个新的图神经网络模型用于社区发现, 该方法是纯数据驱动而无需基础生成模型, 其在社区发现任务上取得了良好的结果. 其他的网络分析例如信息传播<sup>[57,100]</sup>、社交网络地理位置信息预测<sup>[101]</sup>等都有相关研究者引入图卷积神经网络, 使得网络的结构信息和节点属性信息得到了有效的建模.

5.2 推荐系统

人们通过将其视为矩阵补全或者链接预测的方式, 能够有效地建模商品和用户之间的联系如图 21. Monti 等人<sup>[26]</sup>将多图卷积神经网络和循环神经网络相结合, 其中多图卷积神经网络被用来提取局部静止的特征, 而循环神经网络能够扩散分数值, 重建矩阵. Berg 等人<sup>[63]</sup>将推荐系统建模为图上的链接预

测问题, 提出基于不同消息传播的图自编码框架对推荐系统的二部图进行建模, 在包含社交网络的数据上取得了最好的结果. Leskovec 等人<sup>[46]</sup>将卷积神经网络应用到推荐系统中, 其提出一个数据高效的图卷积神经网络算法 PinSage, 对商品节点产生嵌入表达. 这些表达包含了图结构和节点特征信息, 相比传统的图卷积方式, 其提出了一个高效的随机游走策略建模卷积, 设计了一个新的训练策略, 成功地将图卷积神经网络应用到节点数为 10 亿级的超大规模推荐系统中. Wang 等人<sup>[64]</sup>提出 RippleNet 框架, 引入知识图谱信息, 提高了推荐系统的性能. Fan 等人<sup>[65]</sup>提出 GraphRec 框架, 包括用户建模, 商品建模和打分预测 3 个部分, 使用注意力机制, 有效地建模了用户的交互信息和用户的社交网络信息.

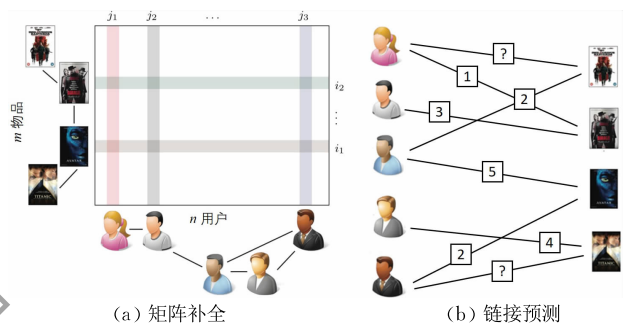


图 21 推荐系统中的矩阵补全和链接预测建模示意图<sup>[25]</sup>

图卷积神经网络被认为能够很好地建模图的结构属性和节点特征信息, 而推荐系统既可以被视为一个矩阵补全问题, 也可以被视为是二部图(用户和商品)的链接预测问题. 相比传统的方法, 图卷积神经网络能够更好地利用在推荐系统中普遍存在的用户属性和商品属性信息, 这也是图卷积神经网络能够在推荐系统任务上引起人们广泛关注的原因.

5.3 交通预测

交通预测问题也是图卷积神经网络得到广泛应用的 任务之一. 其目的是在给定历史交通速度和路线图的情况下, 预测未来的交通的速率. 在交通预测问题中, 如图 22 所示, 节点表示在道路中放置的传感器, 而边则表示节点对的物理距离. 每个节点包含一个时序的特征. 相比于传统的图分析问题, 交通预测问题中包括时间和空间两个方面的建模, 而如何利用图卷积神经网络更好地建模交通中的路网带来了机遇和挑战.

Li 等人<sup>[58]</sup>提出扩散卷积循环神经网络(DCRNN)建模交通预测. 在该模型中, 其将交通流量视为一个有向图上的扩散问题, 提出通过使用扩散卷积的方式来建模图结构化数据. 使用循环神经网络来建模

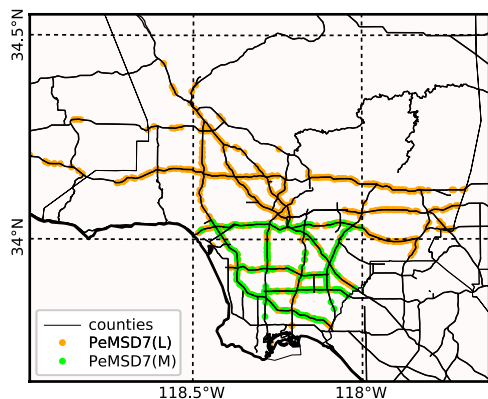


图 22 加利福利亚州第七街区 PeMS 传感器网络示意图<sup>[61]</sup>(其中每个点表示传感器,图中包含两组数据 PeMSD7(M)和 PeMSD7(L))

时间依赖.在两个大规模道路网络交通数据集上取得了 12%~15% 的提升. Cui 等人<sup>[59]</sup> 提出一个交通图卷积长短时记忆网络(TGC-LSTM)学习道路网络和时变的交通模式.其定义图卷积神经网络在物理的路网拓扑结构上.实验结果表明该方法能够捕获有效存在于车辆交通网络中的复杂的时空依赖. Zhang 等人<sup>[60]</sup> 提出图门递归单元(GGRU)解决交通流量预测问题,其将图门递归单元应用于循环神经网络的编码解码模型,应用在洛杉矶高数公路数据集上. Yu 等人<sup>[61]</sup> 提出一个新的深度学习框架,空间时间图卷积神经网络(STGCN),解决在交通领域中的时序预测问题.在该框架中,其首先形式化问题到图上使用卷积结构进行建模,由于更好地利用了拓扑结构,其在短期和中长期交通预测上取得了相比传统机器学习方法显著的提升.

在交通预测相关的场景下,如何解决时空依赖是重要的研究方向,由于图卷积神经网络提供了一种解决图数据问题建模的方案,其通过和循环神经网络等时序模型的结合,给出了一个建模交通预测问题的良好的解决思路.而如何进一步更细粒度考虑时空数据建模依然是未来研究的热点.

#### 5.4 生物化学

除了传统的图数据的建模外,图卷积神经网络在生物化学等领域也受到了研究人员的大量关注.相比传统的图数据的研究,在生物化学领域,人们通常将一个化学结构或一个蛋白质视为一个图,图中的节点是更小的分子,边代表键或者相互作用.其中图 23(a)是布洛芬分子图,节点为碳氢氧原子,连边为化学键.图 23(b)是 FAA4 蛋白质交互网络,节点代表蛋白质,连边为相互作用.研究人员关注于一个图的化学功能,即研究对象不再是图中的节点,而是整个图本身.

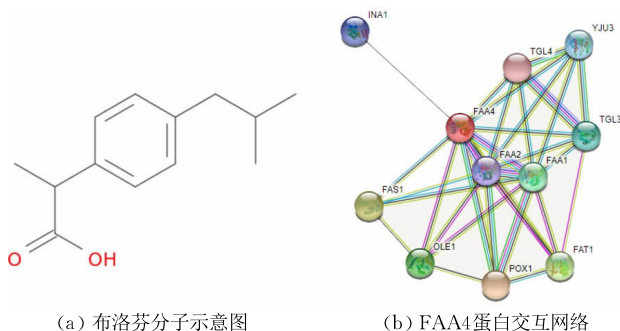


图 23 分子图<sup>[67]</sup>和蛋白质交互网络示意图

Duvenaud 等人<sup>[66]</sup> 直接在图上定义卷积神经网络.该神经网络模型输入为任意大小或形状 of 分子,通过端到端地学习分子指纹.该模型能够更好地帮助实现特定功能的分子设计. Kearnes 等人<sup>[67]</sup> 使用图卷积神经网络对原子、键和距离进行编码,能够更好地利用图结构中的信息.其提供了新的基于配体的虚拟筛选的范式. Gilmer 等人<sup>[26]</sup> 提出消息传播模型 MPNNs 预测给定分子的化学性质. Lee 等人<sup>[68]</sup> 提出图注意力模型(GAM),自适应地选择一些“信息节点”进而收集整个图的信息,用于图分类问题.

Such 等人<sup>[72]</sup> 直接将滤波器定义为图邻接矩阵的函数的多项式,提出 Graph-CNNs 模型,能够处理异构和同质的图数据.在分子分类问题上,表现出了最好的实验结果. Zitnik 等人<sup>[73]</sup> 使用图卷积神经网络建模多种药物副作用.其首先构建蛋白质-蛋白质交互,药物-蛋白质靶标相互作用和多种药物相互作用的多模态图.在图中,每种副作用被视为一个不同类型的边缘.进而将对与药物副作用的建模转化为一个链接预测问题,为进一步研究药理学提供了新的研究思路. Fout 等人<sup>[71]</sup> 提出将图卷积神经网络应用到蛋白质交互预测中.在该任务中,蛋白质是氨基酸残基链,折叠成三维结构,赋予它们生化功能.蛋白质通过与其他蛋白质相互作用的复杂网络发挥其功能. Cao 等人<sup>[69]</sup> 提出 MolGAN,该模型通过生成对抗网络结合图卷积神经网络,设计出包含特定化学性质的分子结构. You 等人<sup>[70]</sup> 提出图卷积策略网络(GCPN),一个基于通用图卷积和强化学习来生成目标图的模型.该模型通过消息传播的方式令隐藏状态为节点表达,然后产生策略  $\pi$ . 通过采样的方式选择策略  $\alpha$ ,然后环境给出化学检测的状态和奖励  $r_t$ . 实验表明该方法相比基线方法,在化学性质优化上有 61% 的提升.

生物化学领域主要集中于对于分子拓扑结构的建模,这些问题中,许多的化学结构和性质体现在图本身的结构特性上.使用图卷积神经网络对这些分



子结构的刻画能够显著地帮助到包括新药发现、药物分类等任务。

5.5 计算机视觉

在计算机视觉中，图卷积神经网络的应用主要集中于少样本学习(Few-Shot Learning)、零样本学习(Zero-Shot Learning)、点云建模(Point Clouds)、

场景图(Sence graph)等，如图 24。

其中，少样本学习旨在使用较少的样本训练能够识别出一个全新的样本。其通常包含两个阶段：元训练和元测试，如图 25(a)所示。在该任务中，数据集包括：训练集，支持集和测试集。支持集和测试集共享相同的标签空间，但训练集有单独的标签空间，且与支持/测试集不相交。如果支持集包含每个类包含  $K$  个标签样本， $C$  个类别，则该问题被称为  $C$ -way  $K$ -shot 问题<sup>[103]</sup>。常见的数据集包括：Omniglot 和 miniImageNet。其中 Omniglot 包含 Omniglot 数据集包含来自 50 个不同字母的 1623 个不同手写字符。miniImageNet 则包含 100 个不同类别，每个类别 600 样本的  $84 \times 84$  的 RGB 图片。少样本学习由于存在较少的训练样本，因此需要进一步刻画出不同的物体或者概念之间的语义关系，常见的方法包括引入知识图谱，构建图片之间的全链接图等方式。

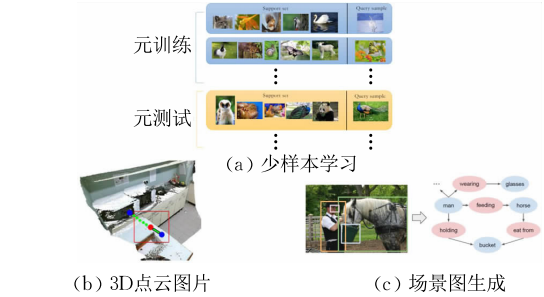


图 24 计算机视觉中常见的图卷积神经网络应用场景：3D 点云图片和场景图生成<sup>[79, 102]</sup>

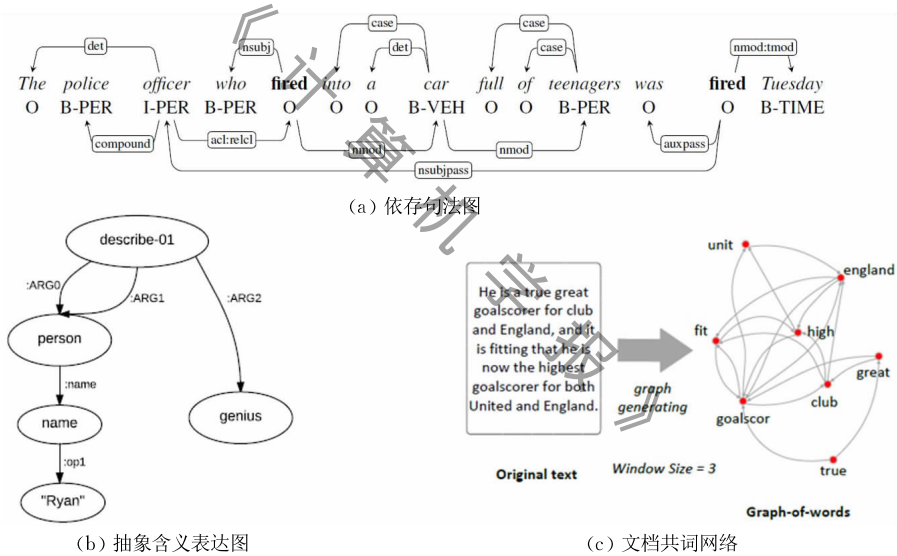


图 25 自然语言处理中常见的图结构场景：依存句法图，抽象含义表达和文档共词矩阵<sup>[90, 92, 95]</sup>

Garcia 等人<sup>[74]</sup>定义一个全连接的图，其中节点是图片，连边是图片和图片之间的相似度，他们使用图神经网络对节点进行编码，使用神经消息传播模型能够更好地利用图片之间的关联结构信息，其在少样本、半监督和主动学习等任务上取得了较好的实验结果，其结果见表 9。Marino 等人<sup>[75]</sup>将知识图谱引入到图片分类任务中，其使用图卷积神经网络更好地利用在知识图谱中的先验知识。在 COCO 数据集的多标签分类任务上取得了提升。Lee 等人<sup>[76]</sup>同样将知识图谱引入到零样本学习任务中，在多标签分类任务中取得了提升。Kampffmeyer 等人<sup>[78]</sup>在使用知识图谱时，定义先祖和后继传播两种方式，能够更好地利用图谱中的语义信息。

点云图是指 3D 扫描器产生的，某个坐标系下的点的集合，其包含了 3D 的坐标信息、颜色等相比 2D 图片更多的几何信息。Qi 等人<sup>[79]</sup>使用图神经网络实现了 RGBD 图片的语义分割任务。Wang 等人<sup>[80]</sup>在点云上使用图卷积神经网络，提出通过边卷积的方式收集边的特征，既包含了局部领域的信息，也通过堆叠或循环的方式学习到全局的几何属性。该模型在形状分类(Shape Classification)和局部分割(Part Segmentation)任务上取得了不错的结果。Landrieu 等人<sup>[81]</sup>使用消息传播机制在点云图上进行建模。

场景图是另一类在计算机视觉领域较为常见的图结构数据，其节点是物体，边的特征代表其空间关

表 9 不同模型在少样本数据集 Omniglot 和 Mini-Imagenet 的结果 (单位: %)

模型	Omniglot				miniImageNet	
	5-Way		20-Way		5-Way	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Pixels	41. 70	63. 20	26. 70	42. 60	—	—
Siamese Net	97. 30	98. 40	88. 20	97. 00	—	—
Matching Networks	98. 10	98. 90	93. 80	98. 50	43. 60	55. 30
N. Statistician	98. 10	99. 50	93. 20	98. 10	—	—
Prototypical Networks	97. 40	99. 30	95. 40	98. 80	46. 61±0. 78	65. 77±0. 70
ConvNet with Memory	98. 40	99. 60	95. 00	98. 60	—	—
Agnostic Meta-learner	98. 7±0. 4	99. 9±0. 3	95. 8±0. 3	98. 9±0. 2	48. 70±1. 84	63. 1±0. 92
Meta Networks	98. 90	—	97. 00	—	49. 21±0. 96	—
TCML	98. 96±0. 20	99. 75±0. 11	97. 64±0. 30	99. 36±0. 18	55. 71±0. 99	68. 88±0. 92
Garcia 等人 GNN	99. 20	99. 70	97. 40	99. 00	50. 33±0. 36	66. 41±0. 63

系. 相比传统描述的句子线性结构,图结构包含了更多有价值的语义信息. 如何使用图卷积神经网络建模场景图受到了大量的关注. Teney 等人<sup>[83]</sup>通过场景图和句子依存句法图的建模,在视觉问答上得到了有效的应用. Chen 等人<sup>[85]</sup>则提出 3 个图模块:知识图谱、图片区域空间关联图和区域类别分布图,在可视化回答上进行了有效地建模.

总的来说,在计算机视觉领域,人们在完成了包括物体识别,图片分类,语义分割后,计算机视觉更关注物体在少量样本,复杂语义情况下的建模和学习. 在这些场景下,图是重要的数据结构,而图卷积神经网络是建模该图数据有效的方法.

5.6 自然语言处理

图卷积神经网络在自然语言处理领域有大量的应用. 在该领域中,较为常见的图数据为知识图谱、句法依赖图和抽象含义表达图、词共现图以及其他方式构建的图. 抽象含义表达 Abstract Meaning Representation(AMR)是一种将一个句子的含义编码为有根有向图. Bastings 等人<sup>[93]</sup>将图卷积神经网络作用于依存句法树上,应用在英语和德语、英语和捷克语的机器翻译任务. Beck 等人<sup>[94]</sup>使用门限图神经网络(GGNN)在抽象含义图上,作用于基于语法的机器翻译任务.

依存句法图或树,节点是单词,连边是语义关系. Liu 等人<sup>[91]</sup>和 Nguyen 等人<sup>[92]</sup>使用图卷积神经网络应用于事件提取,这里使用的图是依存句法树. Song 等人<sup>[104]</sup>将图卷积神经网络作用于阅读理解、抽象含义图到文本的生成任务和关系提取等任务上. 语义角色标注 Semantic Role Labeling(SRL)的任务是给定一个句子,识别出句子中的谓语和对应的对象. Marcheggiani 等人<sup>[87]</sup>提出使用图卷积神经网络作用于句法依赖图,并且和长短时记忆网络叠

加使用,应用于语义角色标注上.

除了上述的图谱外,词共现网络也被应用于文本分类任务上. 其中节点是非停用词,连边是在给定窗口下的词共现关系. Defferrard 等人<sup>[18]</sup>提出了一个在图谱理论上定义的卷积神经网络,它提供了必要的数学背景和有效的数值方案来设计图上的快速局部卷积滤波器. Henaff 等人<sup>[20]</sup>使用图卷积神经网络在 Reuters 数据集上的文本分类任务. Yao 等人<sup>[96]</sup>通过构建共词网络和文档关系网络,将图卷积神经网络应用到文本分类任务上,在不使用外部知识和单词表达的情况下,取得了最好的结果. Peng 等人<sup>[95]</sup>从原始文本基于词共现网络和一个给定的窗口大小,构建了一个图. 然后使用图卷积操作进而实现对于文本的分类任务.

大量的研究表明,在使用图卷积神经网络模型后,各项自然语言处理任务的结果都出现了一定的提升. 图结构的使用,使得对象之间的复杂的语义关系得到了有效地挖掘. 相比传统的对于自然语言处理的序列化建模,使用图卷积神经网络能够挖掘出非线性的复杂语义关系.

5.7 其他

在除了上述的应用领域外,包括程序推断、优化求解<sup>[105-106]</sup>等任务上,图卷积神经网络都开始被人们使用. 由于其可以建模在现实生活中常见的图数据,并且通过卷积、注意力或消息传播等机制,能够将网络的拓扑结构和节点属性等信息以神经网络进行捕获和建模,因此图卷积神经网络有广泛的应用前景.

6 未来研究方向展望

虽然图卷积网络在近些年取得了成功,但仍然



有些没有克服的问题和值得深入研究的方向。

## 6.1 深层的网络结构

传统的深度学习模型,在堆叠了大量网络层后,由于其强大的表示能力,在很多问题上取得了显著地效果。但是在图卷积神经网络模型中,在堆叠了较少的层数后,网络就达到最好效果,同时再增加图卷积层反而会使得结果变差。因为图卷积包含了聚合邻居节点特征的操作,当网络堆叠多层后使得节点之间的特征过于平滑,缺乏区分性。GCN 的实验结果显示<sup>[19]</sup>,当网络层数超过两层后,随着层数增加,GCN 在半监督节点分类问题上的效果反而会下降。同时随着网络的不断叠加,最终所有的节点会学到相同的表达。图神经网络是否需要深层的结构,或者说能否设计出一种深层的网络结构能避免过于平滑的问题,是一个迫切需要解决的研究问题。

## 6.2 大规模数据

在实际场景中,网络的规模往往非常大,比如新浪微博, Twitter 等社交关系网络,往往包含了数亿级的节点和边。而目前绝大部分的图卷积神经网络模型都不适用于这种大规模的网络。比如基于谱方法的图卷积神经网络需要计算图拉普拉斯矩阵的特征向量矩阵,而这个操作的计算复杂度和空间复杂度都很高,难以用于大规模网络。而空间方法,在更新节点表达时依赖于大量的邻居节点,也使得计算代价过大,不适用于大规模网络。虽然近些年已经有一些基于采样的方法来处理大规模网络数据的问题<sup>[29,46,48]</sup>,但是这一问题仍然没有得到有效解决。

## 6.3 多尺度的图上任务

图挖掘任务,根据主体对象的不同可以分成节点级的问题<sup>[19]</sup>,图以及子图级的问题<sup>[66]</sup>以及信号级的问题<sup>[58]</sup>。节点级任务的关键点在于为每个节点学习有效地表达,而为图学习表达是图级别任务的关键。信号级任务的关键点在网络结构不变的情况下,为不同的图信号学习有效地表达。目前绝大部分的图卷积神经网络是针对节点级任务设计的,对于图级别和信号级别的任务关注较少。

## 6.4 动态变化的图数据

在实际场景中,网络往往具有动态性。这种动态性包括不断随时间变换的节点与边上的特征,不断变换的网络的结构(有新的边,节点加入网络也有节点和边从网络中消失)。考虑网络的动态性也是图挖掘的算法的趋势<sup>[107-109]</sup>。而目前的图卷积神经网络都是针对静态的网络设计的,因此设计能建模网络动态变化的图卷积神经网络也是未来的一个重要方向。

## 6.5 图数据的复杂性

在实际场景中,网络往往具有复杂的结构特性<sup>[110]</sup>。例如节点的类型,边上的复杂特征<sup>[111]</sup>,网络的社区结构<sup>[112]</sup>等等。虽然目前有不少工作提出了一些解决方案,但它们都是针对某一种特性设计的网络。能否设计一种网络能同时建模网络的各种复杂特性,也是一个值得探讨的问题。

## 6.6 图神经网络上的对抗攻击

神经网络在各项任务中大放异彩,但仍然具有不稳定的问题<sup>[113]</sup>。例如对图片增加一定噪声,在人眼看来图片的类型没有发生变化,但是神经网络已经将其判断成其他的类型。通过设计一种有针对性的样本从而让机器学习模型做出误判,这便被称为对抗性攻击。在图神经网络领域,利用节点自身的特征和网络结构构造对抗样本<sup>[114]</sup>,以及设计能防御对抗攻击的图神经网络,都是未来发展的一个重要方向。

## 6.7 图神经网络的可解释性

深度学习模型的可解释性与可视化一直是深度学习领域备受关注的方向,图结构给图神经网络的可解释性与可视化带来了新的挑战。如何可视化图神经网络学到的结构模式对于理解图神经网络的工作原理有重要意义<sup>[115]</sup>。图神经网络已经在很多场景下取得了显著的效果, Xu 等人<sup>[51]</sup>尝试给出了一种解释,但是如何从理论上说明图神经网络为什么能取得显著的提升仍然是一个没有解决的问题。

# 7 总 结

图卷积神经网络是处理图数据的高效模型,在过去几年受到研究者的广泛关注,本文充分总结了近些年的图卷积神经网络模型。

图卷积神经网络主要面临的三个挑战:图数据是非欧空间数据,图数据具有多样的特性,以及图数据的规模很大。

在第3节中重点介绍了该领域的主流方法,现有的方法主要分为谱方法和空间方法两类,谱方法利用图上傅里叶变换和卷积定理从谱域定义图卷积,空间方法在节点域定义加权函数在聚合中心节点及其邻居节点的特征。我们同时还总结了最近新涌现出的图卷积神经网络模型,并且分析了其适用场景优缺点。对于图卷积神经网络适用的应用问题进行了整理分类,讨论了未来可能的发展方向。总的来说,目前图卷积神经网络取得了一定的效果,但是

仍然有一些问题需要解决,相信图卷积神经网络会在图数据的场景中被充分地使用。

## 参 考 文 献

- [1] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324
- [2] Zhou Fei-Yan, Jin Lin-Peng, Dong Jun. Review of convolutional neural network. *Chinese Journal of Computers*, 2017, 40(6): 1229-1251(in Chinese)  
(周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. *计算机学报*, 2017, 40(6): 1229-1251)
- [3] Zhang Shun, Gong Yi-Hong, Wang Jin-Jun. The development of deep convolution neural network and its applications on computer vision. *Chinese Journal of Computers*, 2019, 42(3): 453-482(in Chinese)  
(张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用. *计算机学报*, 2019, 42(3): 453-482)
- [4] Hu Baotian, Lu Zhengdong, Li Hang, Chen Qingcai. Convolutional neural network architectures for matching natural language sentences//*Proceedings of the Advances in Neural Information Processing Systems*. Montreal, Canada, 2014: 2042-2050
- [5] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep residual learning for image recognition//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 770-778
- [6] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 2012, 29(6): 82-97
- [7] Shuman D I, Narang S K, Frossard P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 2013, 30(3): 83-98
- [8] Zhou Jie, Cui Ganqu, Zhang Zhengyan, Yang Cheng, et al. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*, 2018
- [9] Wu Zonghan, Pan Shirui, Chen Fengwen, et al. A comprehensive survey on graph neural networks. *arXiv preprint arXiv:1901.00596*, 2019
- [10] Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the Web. *Stanford Digital Libraries Working Paper*, 1998, 9(1): 1-14
- [11] Kleinberg J M. Hubs, authorities, and communities. *ACM Computing Surveys*, 1999, 31(4es): 5
- [12] Ng A Y, Jordan M I, Weiss Y. On spectral clustering: Analysis and an algorithm//*Proceedings of the Advances in Neural Information Processing Systems*. Vancouver, Canada, 2002: 849-856
- [13] Qi Jin-Shan, Liang Xun, Li Zhi-Yu, et al. Representation learning of large-scale complex information network: Concepts, methods and challenges. *Chinese Journal of Computers*, 2018, 41(10): 2394-2419(in Chinese)  
(齐金山, 梁循, 李志宇等. 大规模复杂信息网络表示学习: 概念、方法与挑战. *计算机学报*, 2018, 41(10): 2394-2419)
- [14] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations//*Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA, 2014: 701-710
- [15] Tang Jian, Qu Meng, Wang Mingzhe, et al. Line: Large-scale information network embedding//*Proceedings of the 24th International Conference on World Wide Web*. Florence, Italy, 2015: 1067-1077
- [16] Grover A, Leskovec J. node2vec: Scalable feature learning for networks//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016: 855-864
- [17] Bruna J, Zaremba W, Szlam A, LeCun Y. Spectral networks and locally connected networks on graphs//*Proceedings of the International Conference on Learning Representations (ICLR2014)*. Banff, Canada, 2014: URL <http://arxiv.org/abs/1312.6203>
- [18] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering//*Proceedings of the Advances in Neural Information Processing Systems*. Barcelona, Spain, 2016: 3844-3852
- [19] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016
- [20] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015
- [21] Xu Bingbing, Shen Huawei, Cao Qi, et al. Graph wavelet neural network//*Proceedings of the International Conference on Learning Representations*. New Orleans, USA, 2019
- [22] Xu Bingbing, Shen Huawei, Cao Qi, et al. Graph convolutional networks using heat kernel for semi-supervised learning //*Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, China, 2019: 1928-1934
- [23] Klicpera J, Bojchevski A, Günnemann S. Predict then propagate: Graph neural networks meet personalized PageRank //*Proceedings of the International Conference on Learning Representations*. New Orleans, USA, 2019
- [24] Wu F, Zhang T, de Souza Jr A H, et al. Simplifying graph convolutional networks. *arXiv preprint arXiv:1902.07153*, 2019
- [25] Monti F, Boscaini D, Masci J, et al. Geometric deep learning on graphs and manifolds using mixture model CNNs//*Proceedings of the Conference on Computer Vision and Pattern Recognition*. Honolulu, USA, 2017: 5425-5434

- [26] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 1263-1272
- [27] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model. *IEEE Transactions on Neural Networks*, 2009, 20(1): 61-80
- [28] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018
- [29] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 1024-1034
- [30] Atwood J, Towsley D. Diffusion-convolutional neural networks //Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 1993-2001
- [31] Vashishth S, Yadav P, Bhandari M, Talukdar P. Confidence-based graph convolutional networks for semi-supervised learning //Proceedings of the International Conference on Artificial Intelligence and Statistics. Naha, Japan, 2019: 1792-1801
- [32] Feng Yifan, You Haoxuan, Zhang Zizhao, et al. Hypergraph neural networks//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 3558-3565
- [33] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks//Proceedings of the European Semantic Web Conference. Heraklion, Greece, 2018: 593-607
- [34] Busbridge D, Sherburn D, Cavallo P, Hammerla N Y. Relational graph attention networks. *arXiv preprint arXiv:1904.05811*, 2019
- [35] Derr T, Ma Y, Tang J. Signed graph convolutional networks //Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM). Singapore, 2018: 929-934
- [36] Monti F, Shchur O, Bojchevski A, et al. Dual-primal graph convolutional networks. *arXiv preprint arXiv:1806.00770*, 2018
- [37] Gong Liyu, Cheng Qiang. Adaptive edge features guided graph attention networks. *arXiv preprint arXiv:1809.02709*, 2018
- [38] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 3693-3702
- [39] Zhou Zhenpeng, Li Xiaocheng. Graph convolution: A high-order and adaptive approach. *arXiv preprint arXiv:1706.09916*, 2017
- [40] Sankar A, Zhang X, Chang K C C. Motif-based convolutional neural network on graphs. *arXiv preprint arXiv:1711.05697*, 2017
- [41] Wang Xiao, Ji Houye, Shi Chuan, et al. Heterogeneous graph attention network//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 2022-2032
- [42] Lee J B, Rossi R A, Kong X, et al. Higher-order graph convolutional networks. *arXiv preprint arXiv:1809.07697*, 2018
- [43] Monti F, Otness K, Bronstein M M. MotifNet: A motif-based graph convolutional network for directed graphs//Proceedings of the IEEE Data Science Workshop. Lausanne, Switzerland, 2018: 225-228
- [44] Xu K, Li C, Tian Y, et al. Representation learning on graphs with jumping knowledge networks//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2018: 5449-5458
- [45] Li Qimai, Han Zhichao, Wu Xiao-Ming. Deeper insights into graph convolutional networks for semi-supervised learning//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 3558-3565
- [46] Ying R, He Ruining, Chen Kaifeng, et al. Graph convolutional neural networks for web-scale recommender systems//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, United Kingdom, 2018: 974-983
- [47] Chen Jianfei, Zhu Jun. Stochastic training of graph convolutional networks//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2017: 941-949
- [48] Chen Jie, Ma Tengfei, Xiao Cao. FastGCN: Fast learning with graph convolutional networks via importance sampling//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018
- [49] Huang Wenbing, Zhang Tong, Rong Yu, Huang Junzhou. Adaptive sampling towards fast graph representation learning//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2018: 4563-4572
- [50] Hammond D K, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 2011, 30(2): 129-150
- [51] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks?//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2019
- [52] Ying Z, You J, Morris C, et al. Hierarchical graph representation learning with differentiable pooling//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2018: 4805-4815
- [53] Cangea C, Veličković P, Jovanović N, et al. Towards sparse hierarchical graph classifiers. *arXiv preprint arXiv:1811.01287*, 2018
- [54] Ma Y, Wang S, Aggarwal C C, Tang J. Graph convolutional networks with eigenpooling//Proceedings of the 25th ACM

- SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, USA, 2019: 723-731
- [55] Lee J, Lee I, Kang J. Self-attention graph pooling//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA, 2019: 3734-3743
- [56] Zhuang Chenyi, Ma Qiang. Dual graph convolutional networks for graph-based semi-supervised classification//Proceedings of the 2018 World Wide Web Conference on World Wide Web. Lyon, France, 2018: 499-508
- [57] Qiu Jiezhong, Tang Jian, Ma Hao, et al. DeepInf: Social influence prediction with deep learning//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 2110-2119
- [58] Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting//Proceedings of the International Conference on Learning Representations. Vancouver, BC, Canada, 2018
- [59] Cui Z, Henrickson K, Ke R, Wang Y. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. arXiv: 1802.07007, 2018
- [60] Zhang Jiani, Shi Xingjian, Xie Junyuan, et al. GaAN: Gated attention networks for learning on large and spatiotemporal graphs//Proceedings of the International Conference on Uncertainty in Artificial Intelligence. California, USA, 2018: 339-349
- [61] Yu Bing, Yin Haoteng, Zhu Zhanxing. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting//Proceedings of the International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 3634-3640
- [62] Lei Kai, Qin Meng, Bai Bo, et al. GCN-GAN: A non-linear temporal link prediction model for weighted-dynamic networks //Proceedings of the IEEE International Conference on Computer Communications. Paris, France, 2019: 388-396
- [63] van den Berg R, Kipf T N, Welling M. Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263, 2017
- [64] Wang Hongwei, Zhang Fuzheng, Wang Jialin, et al. Exploring high-order user preference on the knowledge graph for recommender systems. ACM Transactions on Information Systems, 2019, 37(3): 32
- [65] Wang Xiao, Ji Houye, Shi Chuan, et al. Graph neural networks for social recommendation//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 417-426
- [66] Duvenaud D K, Maclaurin D, Iparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Quebec, Canada, 2015: 2224-2232
- [67] Kearnes S, McCloskey K, Berndl M, et al. Molecular graph convolutions: Moving beyond fingerprints. Journal of Computer-Aided Molecular Design, 2016, 30(8): 595-608
- [68] Lee J B, Rossi R, Kong X. Graph classification using structural attention//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 1666-1674
- [69] De Cao N, Kipf T. MolGAN: An implicit generative model for small molecular graphs. arXiv preprint arXiv: 1805.11973, 2018
- [70] You Jiaxuan, Liu Bowen, Ying Zhitao, et al. Graph convolutional policy network for goal-directed molecular graph generation//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2018: 6412-6422
- [71] Fout A, Byrd J, Shariat B, Ben-Hur A. Protein interface prediction using graph convolutional networks//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 6530-6539
- [72] Such F P, Sah S, Dominguez M A, et al. Robust spatial filtering with graph convolutional neural networks. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(6): 884-896
- [73] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 2018, 34(13): i457-i466
- [74] Garcia V, Bruna J. Few-shot learning with graph neural networks//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018
- [75] Marino K, Salakhutdinov R, Gupta A. The more you know: Using knowledge graphs for image classification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017: 20-28
- [76] Lee Chung-Wei, Fang Wei, Yeh Chih-Kuan, Wang Yu-Chiang Frank. Multi-label zero-shot learning with structured knowledge graphs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 1576-1585
- [77] Wang X, Ye Y, Gupta A. Zero-shot recognition via semantic embeddings and knowledge graphs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 6857-6866
- [78] Kampffmeyer M, Chen Y, Liang X, et al. Rethinking knowledge graph propagation for zero-shot learning//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 11487-11496
- [79] Qi Xiaojuan, Liao Renjie, Jia Jiaya, et al. 3D graph neural networks for rgb-d semantic segmentation//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 5209-5218
- [80] Wang Yue, Sun Yongbin, Liu Ziwei, et al. Dynamic graph cnn for learning on point clouds. arXiv preprint arXiv: 1801.07829, 2018
- [81] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs//Proceedings of the IEEE

- Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4558-4567
- [82] Te G, Hu W, Guo Z, Zheng A. RGCNN: Regularized graph CNN for point cloud segmentation//Proceedings of the ACM Multimedia Conference on Multimedia Conference. Seoul, Korea, 2018: 746-754
- [83] Teney D, Liu L, van den Hengel A. Graph-structured representations for visual question answering//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 3233-3241
- [84] Wang Z, Chen T, Ren J, et al. Deep reasoning with knowledge graph for social relationship understanding//Proceedings of the International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 1021-1028
- [85] Chen X, Li L J, Fei-Fei L, Gupta A. Iterative visual reasoning beyond convolutions//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7239-7248
- [86] Hamaguchi T, Oiwa H, Shimbo M, Matsumoto Y. Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach//Proceedings of the International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 1802-1808
- [87] Marcheggiani D, Bastings J, Titov I. Exploiting semantics in neural machine translation with graph convolutional networks //Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. New Orleans, USA, 2018: 486-492
- [88] Marcheggiani D, Titov I. Encoding sentences with graph convolutional networks for semantic role labeling//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017: 1506-1515
- [89] Zhang Y, Qi P, Manning C D. Graph convolution over pruned dependency trees improves relation extraction//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, System Demonstrations. Brussels, Belgium, 2018: 2205-2215
- [90] Song L, Zhang Y, Wang Z, Gildea D. A graph-to-sequence model for amr-to-text generation//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 1616-1626
- [91] Liu Xiao, Luo Zhunchen, Huang Heyan. Jointly multiple events extraction via attention-based graph information aggregation//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, System Demonstrations. Brussels, Belgium, 2018: 1247-1256
- [92] Nguyen T H, Grishman R. Graph convolutional networks with argument-aware pooling for event detection//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 5900-5907
- [93] Bastings J, Titov I, Aziz W, et al. Graph convolutional encoders for syntax-aware neural machine translation//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017: 1957-1967
- [94] Beck D, Haffari G, Cohn T. Graph-to-sequence learning using gated graph neural networks//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 273-283
- [95] Peng H, Li J, He Y, et al. Large-scale hierarchical text classification with recursively regularized deep graph-CNN//Proceedings of the International Conference on World Wide Web. Lyon, France, 2018: 1063-1072
- [96] Yao Liang, Mao Chengsheng, Luo Yuan. Graph convolutional networks for text classification//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 7370-7377
- [97] Yang Z, Cohen W, Salakhudinov R. Revisiting semi-supervised learning with graph embeddings//Proceedings of the 33rd International Conference on Machine Learning. New York, USA, 2016: 40-48
- [98] Shervashidze N, Schweitzer P, van Leeuwen E J, et al. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 2011, (2011): 2539-2561
- [99] Chen Zhengdao, Li Lisha, Bruna J. Supervised community detection with line graph neural networks//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2018
- [100] Cao Qi, Shen Huawei, Gao Jinhua, et al. Popularity prediction on social platforms with coupled graph neural networks//Proceedings of the 13th International Conference on Web Search and Data Mining. Houston, Texas, UAS, 2020: 70-78
- [101] Rahimi A, Cohn T, Baldwin T. Semi-supervised user geolocation via graph convolutional networks//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 2009-2019
- [102] Xu Danfei, Zhu Yuke, Choy C B, Li Fei-Fei. Scene graph generation by iterative message passing//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 3097-3106
- [103] Wang Yaqing, Yao Quanming. Few-shot learning; A survey. *arXiv preprint arXiv:1904.05046*, 2019
- [104] Song Linfeng, Wang Zhiguo, Yu Mo, et al. Exploring graph-structured passage representation for multi-hop reading comprehension with graph neural networks. *arXiv preprint arXiv:1809.02040*, 2018
- [105] Khalil E, Dai H, Zhang Y, et al. Learning combinatorial optimization algorithms over graphs//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 6348-6358

- [106] Li Z, Chen Q, Koltun V. Combinatorial optimization with graph convolutional networks and guided tree search//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2018: 537-546
- [107] Du Lun, Wang Yun, Song Guojie, et al. Dynamic network embedding: An extended approach for skip-gram based network embedding//Proceedings of the International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 2086-2092
- [108] Zhou Lekui, Yang Yang, Ren Xiang, et al. Dynamic network embedding by modeling triadic closure process//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 571-578
- [109] Nguyen N P, Dinh T N, Xuan Y, Thai M T. Adaptive algorithms for detecting community structure in dynamic social networks//Proceedings of the IEEE International Conference on Computer Communications. Shanghai, China, 2011: 2282-2290
- [110] Benson A R, Gleich D F, Leskovec J. Higher-order organization of complex networks. *Science*, 2016, 353(6295): 163-166
- [111] Huang Junjie, Shen Huawei, Hou Liang, Cheng Xueqi. Signed graph attention networks//Proceedings of the International Conference on Artificial Neural Networks. Munich, Germany, 2019: 566-577
- [112] Fortunato S. Community detection in graphs. *Physics Reports*, 2010, 486(3-5): 75-174
- [113] Akhtar N, Mian A. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, 2018, 6: 14410-14430
- [114] Zügner D, Akbarnejad A, Günnemann S. Adversarial attacks on neural networks for graph data//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 2847-2856
- [115] Ying R, Bourgeois D, You J, et al. GNN explainer: A tool for post-hoc explanation of graph neural networks. *arXiv preprint arXiv:1903.03894*, 2019



**XU Bing-Bing**, Ph. D. candidate. Her research interests include graph neural network and graph-based semi-supervised learning.

**CEN Ke-Ting**, Ph. D. candidate. His research interests include graph neural network and network embedding.

**HUANG Jun-Jie**, M. S. candidate. His research interests include social media computing and graph neural network.

**SHEN Hua-Wei**, Ph. D., professor. His research interests include social network analysis and social media computing.

**CHENG Xue-Qi**, Ph. D., professor. His research interests include big data analysis and mining, network science, network and information security, web search and data mining.

## Background

Many real-life scenarios, such as transportation networks, social networks and citation networks, are located in the form of graph data. In recent years, the powerful modeling capabilities of convolutional neural network and the ubiquity of graph data have inspired researchers to transfer convolutional neural network to graphs. This paper surveys and summarizes the research and developments of graph convolution networks. The classical methods in this area are divided into two groups: spectral methods and spatial methods. Spectral methods implement graph convolution via convolution theorem, while spatial methods define aggregator function in vertex space directly to implement graph convolution. The latest developments in this field are focus on how to construct

graph convolution in huge or complicated graph, e. g., graph with million nodes, signed graph and heterogeneous graph. Graph convolution networks have shown powerful modeling capabilities in many applications, e. g., traffic network, molecular network.

This paper is funded by the National Natural Science Foundation of China under Grant Nos. 61425016 and 91746301. Shen Huawei is also funded by K. C. Wong Education Foundation and Beijing Academy of Artificial Intelligence (BAAI). These projects focus on graph convolution, information diffusion and online social media. Our group has been working on this area for several years and published a number of papers.