

# 图神经网络社区发现研究综述

宁懿昕 谢辉 姜火文

江西科技师范大学数学与计算机科学学院 南昌 330038

(1509257998@qq.com)

**摘要** 社区结构是复杂网络中普遍存在的拓扑特性之一,发现社区结构是复杂网络分析的基本任务。社区发现旨在将网络划分为多个子结构,对于理解网络、揭示网络的潜在功能有着重要作用。图神经网络是一种处理图结构数据的模型,具有从图中对数据进行特征提取和表示的优势,已经成为人工智能和大数据领域的重要研究方向。网络数据就是典型的图结构数据,使用图神经网络模型解决社区发现问题,是社区发现研究的一个新方向。首先对 GNN 模型进行深入探讨,分析 GNN 社区发现过程,并从重叠社区和非重叠社区这两个方面详细讨论现有 GNN 社区发现取得的进展以及未来可研究的方向。

**关键词:** 图神经网络;社区发现;深度学习;重叠社区发现;非重叠社区发现

**中图法分类号** TP331

## Survey of Graph Neural Network in Community Detection

NING Yi-xin, XIE Hui and JIANG Huo-wen

School of Mathematics and Computer Science, Jiangxi Science and Technology Normal University, Nanchang 330038, China

**Abstract** Community structure is one of the universal topological properties in complex networks, and discovering community structure is the basic task of complex network analysis. The purpose of community detection is to divide the network into several substructures, which plays an important role in understanding the network and revealing its potential functions. Graph Neural Network (GNN) is a model for processing graph structure data, which has the advantage of feature extraction and representation from graph, and has become an important research field of artificial intelligence and big data. Network data is a typical graph structure data. Using graph neural network model to solve the problem of community detection is a new direction of community detection research. In this paper, we first discuss the GNN model, analyze the process of GNN community detection, and discuss the progress of existing GNN community detection and the direction of future research in detail from two aspects of overlapping community and non-overlapping community.

**Keywords** Graph neural network, Community detection, Deep learning, Overlapping community detection, Non-overlapping community detection

图神经网络<sup>[1]</sup> (Graph Neural Network, GNN) 是受卷积神经网络和图嵌入的思想启发而来的,目的是扩展现有神经网络,以处理图结构数据。GNN 可以对图结构数据进行特征提取和表示,在学习图数据方面表现出了强大的功能。与传统的深度学习方法相比,图神经网络可以通过构建图模型来反映对象之间的复杂关系。目前,国内外对 GNN 的研究也取得了一定的进展,GNN 模型也被应用到各个领域,如文本分类<sup>[2]</sup>、自然语言处理<sup>[3]</sup>、疾病预测<sup>[4]</sup>和特征关系提取<sup>[5]</sup>等。

复杂网络就是将系统中的对象和关系抽象成节点和连边得到的图,其数据结构为典型的图数据结构。社区结构是许多实际网络中具有的普遍拓扑特性之一。所谓社区结构,即整个网络由若干个社团组成,社区之间连接相对稀疏,社区内部节点的连接相对紧密。社区检测是利用图拓扑结构中隐藏

的信息从网络中解析出其模块化的社团结构。社区发现问题的研究有助于研究整个网络的模块、功能及其演化,对理解复杂系统的组织原则、拓扑结构与动力学特性具有十分重要的意义。传统的社区发现算法大多数基于统计推断和机器学习发展而来,而依赖于机器学习的社区发现算法大多依赖于手工特征提取,具有一定的局限性。深度学习可以端对端地学习到节点的特征信息,使得计算模型可以学习到具有多层次抽象的数据表征。图神经网络作为神经网络用于处理图结构数据的扩展,为人们进一步分析复杂网络社区发现问题提供了更好的表示方式。基于 GNN 的社区发现问题的相关研究受到人们的广泛关注并取得了一定的成果。本文将主要分析基于 GNN 的重叠社区和非重叠社区问题的研究进展,提出 GNN 社区发现当前存在的问题并讨论未来

基金项目:国家自然科学基金项目(71561013,61762044);江西省社会科学研究规划项目(20TQ04);江西省高校人文社会科学研究项目(JC17221,JD18083,JC18109);江西省教育厅科技计划重点项目(GJJ170661)

This work was supported by the National Natural Science Foundation of China (71561013,61762044), Social Science Planning Projects in Jiangxi Province (20TQ04), Fund of Humanities and Social Sciences in Universities of Jiangxi Province (JC17221,JD18083,JC18109) and Key Project of Science & Technology Plan by Education Department of Jiangxi Province(GJJ170661).

通信作者:谢辉(huixie@aliyun.com)

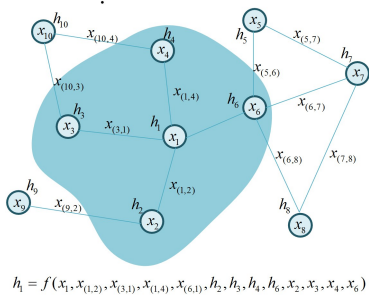
GNN 社区发现的主要研究方向。

## 1 图神经网络

GNN 是借鉴神经网络领域的研究成果,由 Gori 等<sup>[6]</sup>最先提出的一种用于处理图结构数据的模型。2009 年, Scarselli 等<sup>[7]</sup>对此模型进行了详细阐述。为简单起见,我们将讨论 Scarselli 等人提出的模型。

图神经网络是利用一定的方法对节点进行描述,目标是为每个节点学习一个对其邻居信息编码的嵌入状态  $h_v \in \mathbb{R}^s$ , 这个状态  $h_v$  用来产生所需要的输出  $o_v$ 。给定节点和边的输入特征,接下来我们将讨论模型如何获得节点状态  $h_v$  和输出  $o_v$ 。

以节点 1 为例(如图 1 所示),给定节点 1 的状态向量  $h_1$ 。这个状态包含节点的信息以及其邻域的信息。



注:节点 1 的状态  $x_1$  取决于其邻域包含的信息

图 1 图和节点的邻域

Fig. 1 Graph and neighborhood of nodes

定义一个局部转换函数  $f$ , 这个函数在所有节点之间参数共享,描述节点对其邻域的依赖性。为了产生节点的输出,设置一个参数函数  $g$ , 称为局部输出函数。节点状态和输出定义如下:

$$h_v = f(x_v, x_{co[v]}, h_{ne[v]}, x_{ne[v]}) \quad (1)$$

$$o_v = g(h_v, x_v) \quad (2)$$

其中,  $x$  表示输入特征,  $h$  表示隐藏状态,  $co[v]$  表示与节点  $v$  相连的边的集合,  $ne[v]$  表示节点  $v$  的邻居节点的集合。因此,  $x_v, x_{co[v]}, h_{ne[v]}, x_{ne[v]}$  分别表示节点  $v$  的输入特征、与节点  $v$  相连的边的输入特征、节点  $v$  邻居节点的状态和输入特征。

令  $H, O, X$  和  $X_N$  分别为堆叠所有状态、输出、特征和节点特征所构造的矩阵,因此可以得到一个更简洁的形式:

$$H = F(H, X) \quad (3)$$

$$O = G(H, X_N) \quad (4)$$

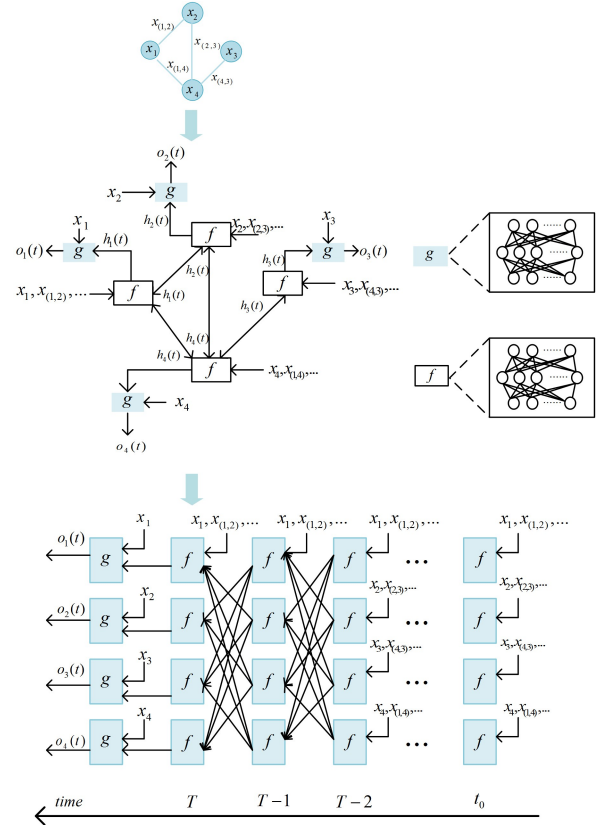
其中,  $F$  和  $G$  分别表示全局转换函数和全局输出函数。假设函数  $F$  是唯一定义的,则  $H$  的值是等式(3)的固定点。根据 Banach 不动点定理, GNN 使用以下迭代来计算节点的状态:

$$H^{t+1} = F(H^t, X) \quad (5)$$

GNN 信息传播及其迭代计算的具体过程如图 2 所示。

近年来,图神经网络引起了众多邻域学者的注意,图神经网络模型层出不穷。根据卷积实现方法的不同,大致可以分为两类:基于频谱方法的图神经网络和基于空间方法的图神经网络。频谱方法的基本思想是根据图谱理论和卷积定理,通过傅里叶变换将图数据由空域转换到频域做处理,处理完再由傅里叶逆变换将频域数据转换到空域,从而完成对图数据的数据处理。主要的模型有: Spectral Network<sup>[8]</sup>, ChebNet<sup>[9]</sup>, GCN<sup>[10]</sup> 和 AtGCN<sup>[11]</sup> 等。其中, ChebNet 模型采用

Chebyshev 多项式代替谱域的卷积核; GCN 模型为 ChebNet 的进一步简化,它仅考虑一阶切比雪夫多项式,且每个卷积核仅有一个参数。空间方法的基本思想是不依赖图谱理论和卷积定理,直接在空间上定义卷积操作。空间方法的主要挑战是如何构建邻域,即怎么找固定数量的邻居节点以及找到的邻居节点如何排序和定义不同邻域大小的卷积核。主要的模型有: Niepert 等<sup>[12]</sup> 提出的 PATCHY-SAN 模型, Hamilton 等<sup>[13]</sup> 提出的 GraphSAGE 模型, Atwood 等<sup>[14]</sup> 提出的 DCNN 模型, Zhuang 和 Ma 等<sup>[15]</sup> 提出的 DGCN 模型等。其中, GraphSAGE 模型通过采样得到邻域节点,使用聚合函数来聚合邻居节点的信息以获得目标节点的嵌入。



注:给定原始图  $G$ , 对每个节点进行状态嵌入和更新,最后计算得到输出

图 2 GNN 信息传播流程图及其迭代计算过程

Fig. 2 GNN information propagation process and its iterative calculation process

## 2 社区发现

社区结构是许多复杂网络尤其是社会关系网络所具有的共同性质之一。学术界对社区结构的定义并不是很明朗。一般来说,社区结构可以描述为:将网络中的节点按照一定的规则进行分组,使得同一组内的节点连接相对稠密,不同组之间的节点连接相对稀疏。社区发现就是发现网络中的社区结构。研究社区发现的方法有很多,我们大致将其分为 3 类:传统方法、深度学习和基于图神经网络的社区发现方法。传统算法大多数基于统计推断和机器学习发展而来。Luo 等<sup>[16]</sup> 和 Liu 等<sup>[17]</sup> 将社区发现问题分为非重叠网络的社区发现和重叠网络的社区发现,分别介绍了重叠网络和非重叠网络的社区发现算法。Cheng 等<sup>[18]</sup> 以介绍社区发现的研究历程为脉络,从 Newman 等<sup>[19]</sup> 首次将社区结构作为网络普遍具有的拓扑性质提出并给出的基于边介数的分裂式层次聚类算

法开始,介绍了基于层次聚类的社区发现方法、基于模块度优化的社区发现方法、多尺度社区发现、异质网络社区发现、谱方法以及重叠社区发现算法等。Zhao等<sup>[20]</sup>从网络可以分成同质网络、异质网络和动态网络等的角度出发,介绍不同网络中的社区发现算法。然而,随着社区发现问题研究的深入,传统方法在处理当下的复杂数据和社交场景时,将会面临很多问题。处理图及其属性、大规模网络和动态环境中形成的高维数据时需要更好的技术。

与机器学习的方法相比,深度学习能够处理高维数据情形下的社区发现问题。随着深度学习的发展,基于深度学习的社区发现研究也有很多。Liu等<sup>[21]</sup>从技术的角度总结了现有的方法。从深度学习技术来看,Xin等<sup>[22]</sup>提出了一种新型CNN,并提出了一种用于拓扑结构不完整的有监督算法,以检测拓扑不完整的网络中的社区,这些网络从实际网络中观察时往往会丢失一些边信息;为了利用节点属性信息,Cao等<sup>[23-24]</sup>提出了一种堆叠式自动编码器,该编码器结合网络拓扑和节点属性进行社区检测,以增强深度神经网络隐藏层的幻化能力;为进一步解决网络拓扑和节点属性之间的匹配问题,通过引入自适应参数作为匹配性的平衡控制,开发了一种图形正则化自动编码器方法。深度学习技术的出现,大大提升了社区发现算法的性能。然而,社区检测任务需要处理包含节点元素之间丰富的非欧几里得图数据,并不能通过传统的深度学习模型处理得很好。图神经网络作为用于处理图数据的神经网络,可以弥补这一缺陷。

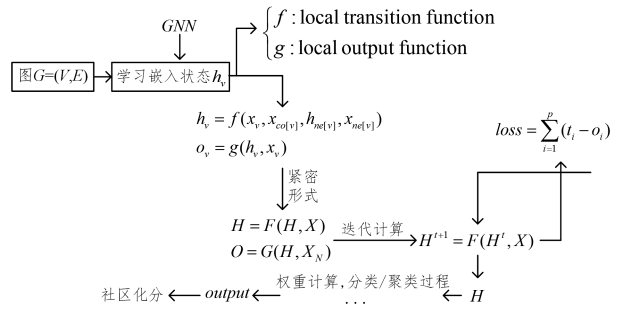
### 3 GNN 社区发现的主要过程

一般来说,GNN 社区发现的主要过程分为以下几个步骤,如图 3 所示。

1)根据实际网络中研究对象和对象之间的关系构造 GNN 模型。主要考虑如何将实际网络中的研究对象映射为 GNN 中图的节点,不同研究对象之间的关系映射为边,如何使用函数去拟合 GNN 中不同的节点。在 GNN 模型中,节点都可以根据相关信息做嵌入化处理,得到其对应节点的向量化描述。如对于两个不直接相连的节点考虑其相似度,可以根据节点间是否有最短路径相连且最短路径的长度等相关信息做嵌入化处理。

2)决定 GNN 模型的信息传播与更新方法。GNN 模型的变体有很多,不同的变体之间,信息的传递和更新方法也不同。文献<sup>[25]</sup>指出,不同的 GNN 变体中选用不同的聚集过程,如:GCN 模型通常选择 mean 聚集函数,GraphSAGE 模型通常采用 max 函数,GIN 模型通常使用 sum 函数等。不同的聚集函数适合于不同的图结构,mean 函数可以反映图中节点的分布情况,max 函数可以反映主要特征,sum 可以考虑到比较全面的图形结构特征等。信息更新过程,主要是将信息聚集之后的结果,与中心节点进行特定的运算,并作为下一层节点的初始状态。这方面,不同的模型差异也很大,如 GCN 模型进行信息更新时,不会考虑中心节点的信息;GraphSAGE 则会将聚集结果和中心节点的向量进行连接操作;GIN 模型会将两者直接相加等。选择合适的信息传播算法进行信息更新,往往可以取得更好的社区结果。

3)提取出更新后的节点特征,使用相关算法实现社区发现,如分类、聚类算法等。



注:对于图  $G=(V,E)$ ,利用 GNN 迭代计算学习得到嵌入状态  $H$ ,最后利用分类或聚类算法得到社区划分结果

图 3 GNN 社区发现过程

Fig. 3 Community detection with GNN

## 4 GNN 社区发现相关研究

现有的图神经网络 GNN 社区检测的相关研究中,我们可以将社区检测问题分为非重叠社区的社区检测问题和重叠社区的社区检测问题。

### 4.1 非重叠社区

非重叠社区发现是指识别出的社区之间互不相交,即每个节点仅属于一个社团。基于 GNN 的社区发现主要是通过图神经网络对网络上的复杂关系进行建模,从而捕获网络中节点和连边的关系。图神经网络在社区检测任务上已经被证明是强大的。Bruma等<sup>[26]</sup>提出了一种基于图神经网络的方法,来解决数据驱动下的社区检测问题;Peter等<sup>[27]</sup>将自表达原则和无监督图神经网络的框架结合,提出了一种无监督社区检测方法;Zhang等<sup>[28]</sup>通过扩展 Gumbel Softmax<sup>[29]</sup>方法提出了图神经网络社区检测新方法。自编码器是一种深度学习模型,它在社区发现任务中表现出了强大的性能<sup>[30]</sup>,文献<sup>[31]</sup>提出了一种优化变分自编码器的方法来解决社区检测问题。使用图神经网络编码方法来解决复杂属性网络中的社区检测问题也已被证明可行<sup>[32]</sup>。

Chen等<sup>[33]</sup>引入非回溯运算符来定义边的邻接性,得到一种 GNN 框架 LGNN(Line Graph Neural Network)以解决有监督环境中的社区检测问题,其具体步骤如下。

1)考虑能够编码  $2^J$ -跳邻域的图邻接矩阵  $A_j = \min(1, A^{2^j})$  定义 GNN 层。对于给定的  $J$ ,有这样的一个矩阵族  $\mathcal{F} = \{I, D, A, A_2, \dots, A_J\}$ ,其中,  $I$  是单位矩阵,  $D$  为度矩阵。定义一个多尺度 GNN 层:  $x^{(k)} \in \mathbb{R}^{|V| \times b_k} \rightarrow x^{(k+1)} \in \mathbb{R}^{|V| \times b_{k+1}}$ 。首先计算  $\underline{z}$  和  $\bar{z}$  的计算公式如下:

$$\underline{z}^{(k+1)} = \rho\left(\sum_{O_i \in \mathcal{F}} O_i x^{(k)} \theta_i\right) \quad (6)$$

$$\bar{z}^{(k+1)} = O_i x^{(k)} \theta_i \quad (7)$$

$$x^{(k+1)} = [\underline{z}^{(k+1)}, \bar{z}^{(k+1)}] \in \mathbb{R}^{|V| \times b_{k+1}}$$

其中,  $\theta_i$  为训练参数,  $\rho(\cdot)$  为非线性激活函数,  $O_i \in \mathcal{F}$ ,  $x^{(k+1)}$  是  $\underline{z}^{(k+1)}$  和  $\bar{z}^{(k+1)}$  的连接。

2)对给定的无向图  $G=(V,E)$ ,引入非回溯线性运算符定义线图  $L(G)=(V_L, E_L)$  来编码图  $G$  的定向边缘邻接结构,其中  $V_L = \{(i \rightarrow j) : (i, j) \in E\}$ ,  $E_L$  由非回溯矩阵  $B \in \mathbb{R}^{2|E| \times 2|E|}$  给定。该非回溯运算符被定义为:

$$B_{(i \rightarrow j), (i' \rightarrow j')} = \begin{cases} 1 & \text{if } j = i' \text{ and } j' \neq i \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

3)由 2)扩展 1)中 GNN 层,类比定义第二个 GNN 层。

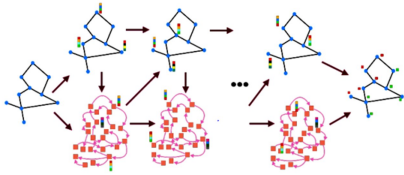
简述如下:类似于 $A_j$ 定义 $B_j = \min(1, B^{2j})$ ,在每个层中使用 $P, \tilde{P} \in \{0, 1\}^{|V| \times 2|E|}$ 来连接节点和边的状态信息,其中: $P_{i,(i \rightarrow j)} = 1, P_{j,(i \rightarrow j)} = 1, \tilde{P}_{i,(i \rightarrow j)} = 1, \tilde{P}_{j,(i \rightarrow j)} = 1$ ,其他情况取0。类似GNN层的定义,每层的更新规则可写为:

$$z^{(k+1)} = \rho \left[ \sum_{O_i \in \mathcal{S}} O_i x^{(k)} \theta_i + \sum_{O_j \in \mathcal{S}'} O_j y^{(k)} \theta_j' \right] \quad (9)$$

$$w^{(k+1)} = \rho \left[ \sum_{O_i \in \mathcal{S}} O_i' y^{(k)} \theta_i' + \sum_{O_j \in \mathcal{S}'} (O_j'')^T x^{(k+1)} \theta_j'' \right] \quad (10)$$

LGNN由上述一系列GNN层组成,其架构如图4所示。

4)令 $y_i$ 为节点 $i$ 的真实社区, $[C] = \{1, 2, \dots, C\}$ 。应用softmax分类器对得到的节点嵌入进行条件概率计算 $o_{i,c} = p(y_i = c | z_i, G)$ ,选择损失函数最小的社区标签对图进行划分,得到最优划分结果。



注:给定一个图 $G$ ,我们用非回溯子图构造它的有线图 $L(G)$ 。在每一层中, $G$ 和 $L(G)$ 中的所有节点的状态都根据式(9)、式(10)进行更新。 $G$ 中节点的最终状态用于预测节点上的标签,并使用具有标签排列不变损失的标准反向传播进行端到端训练

图4 LGNN的体系结构概述

Fig. 4 Architecture overview of LGNN

该模型首先考虑非重叠社区情况下的社区划分,能够考虑节点以及连边所包含的特征信息,从而得到所需的社区划分结果。对于重叠社区划分来说,扩大社区标签的集合定义相应的置换组,也可得到社区划分结果。但是由于网络种类的多种多样,该模型的具体应用还存在局限性,如不能解决异质-时间网络上的社区检测问题。对于异质-时间网络上的社区检测问题,Zheng等<sup>[34]</sup>提出一种异质-时间图卷积网络(HTGCN),该网络模型包含一个能够捕获异构图的特征表示的异构GCN组件以及一个残余压缩聚合组件以捕获“变化”的社区特征。

## 4.2 重叠社区

在许多真实网络中,社区通常是重叠而不是不相交的,文献<sup>[35-36]</sup>中已经全面研究了社区的重叠结构。重叠社区指在网络中的节点可能属于多个社区。重叠社区发现更符合真实世界的组织规律,复杂网络中的重叠社区模式是对传统社区概念的一个重要拓展。近年来,重叠社区检测已经成为了社区检测领域的热点。在现有的少数GNN社区发现研究中,大部分都侧重于检测非重叠社区,即使在实际网络中社区往往是重叠的。针对这一不足,Shchur等<sup>[37]</sup>将图卷积神经网络与允许社区重叠的图生成模型<sup>[38-40]</sup>——Bernoulli-Poisson(BP)模型相结合,提出了一种用于重叠发现的模型NCOD(Neural Overlapping Community Detection)进行重叠社区检测,如图5所示。该模型主要是基于BP模型学习社区隶属关系矩阵从而将节点分配给社区的思想,不同于传统的直接优化隶属矩阵的方法,采用GNN模型,以端对端的方式学习社区隶属关系,具体设置如下。

1)对于给定的无向图 $G, A$ 为其邻接矩阵, $X \in \mathbb{R}^{N \times D}$ 为其属性矩阵,其中 $D$ 为节点的属性个数。使用2层的图卷积神经网络<sup>[10]</sup>(Graph Convolutional Neural Network, GCN)作为

NCOD模型的基础,学习得到社区隶属关系矩阵 $F$ 。GCN定义如下:

$$F := GCN_{\theta}(A, X) = ReLU(\hat{A} ReLU(\hat{A} X W^{(1)})) W^{(2)} \quad (11)$$

其中, $\hat{A} = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}, \tilde{A} = A + I_N$ 。

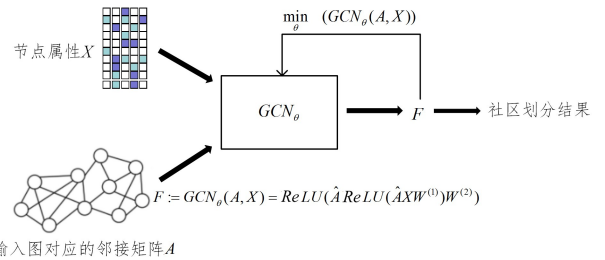
2)类似于BP模型的最大负对数似然估计,考虑真实图情况,类似定义平衡的负对数似然函数:

$$\mathcal{L}(F) = -\mathbb{E}_{(u,v) \sim P_E} [\log(1 - \exp(-F_u F_v^T))] + \mathbb{E}_{(u,v) \sim P_N} [F_u F_v^T] \quad (12)$$

3)最小化平衡的负对数似然函数寻找参数 $\theta^*$ ,使得 $F$ 生成的图和真实图尽量相似,得到最优的社区隶属矩阵。其中:

$$\theta^* = \arg \min_{\theta} \mathcal{L}(GCN_{\theta}(A, X)) \quad (13)$$

4)由得到的最接近真实图的社区隶属关系矩阵 $F$ ,得到社区划分。



输入图对应的邻接矩阵 $A$

注:对于给定无向图 $G$ ,利用GCN学习得到社区隶属关系矩阵,根据社区隶属关系得到社区划分结果

图5 GNN重叠社区发现过程

Fig. 5 Overlapping community detection with GNN

相较于非重叠社区检测模型,该模型允许节点可能未分配给其他社区,也可能属于多个社区,这是由于社区隶属关系的强度决定的。但是,对于网络中节点之间连边的强度,该模型没有考虑,不同节点之间的关系指向也就没有加以区分。

GNN模型不仅可以端对端学习节点的特征表示,也可以用于检测社区划分的结果。许多现有的重叠社区检测算法有一个共同的问题:社区的个数无法确定,必须手动预定义社区个数。文献<sup>[41]</sup>针对这一问题,引入一种允许社区个数 $K$ 随着图数据的增加而变化的非参数贝叶斯模型,建模得到社区分配矩阵 $Z$ ,利用图卷积神经网络的图分类器来帮助总结并估计社区分配矩阵 $Z$ 的社区划分质量。

## 5 现有GNN社区发现的研究难点

### 5.1 社区个数未知

社区发现近年来受到了广泛关注,涌现出了上百种的算法<sup>[42]</sup>。在许多现有的社区发现算法中,在进行社区发现前都需要手动定义社区个数 $k$ 。长久以来,由于社区个数未知而带来的挑战始终没有得到很好的解决。常用的一种解决办法是利用社区质量评价标准确定最好的社区划分,从而解决在社区发现前需要用户指定社区个数的难题。如何确定网络中的社区个数,或者说如何建立起一个合适的社区度量标准,是目前需要解决的一个难题。文献<sup>[41]</sup>中引入了一种非参数贝叶斯模型,该模型允许社区个数 $k$ 随着数据的增加而增大,并且使用GCN对社区划分质量进行评价,这为使用GNN解决社区个数未知,建立社区检查度量标准问题提供了可能。

## 5.2 大规模网络

随着大规模网络的出现,人们对社区发现算法的效率有着更高的要求。传统的社区发现算法在实际应用场景中,存储效率和计算效率都较低,难以适用于大规模网络。适用于大规模复杂网络的社区发现算法是未来的重要研究方向之一。对于大规模网络的社区检测,如何设计或选择合适的 GNN 模型,是未来 GNN 社区发现问题的一个重要挑战。与此同时,降低算法的计算复杂度也是未来的一个重要研究方向。

## 5.3 层次化重叠社区结构的社区发现问题

许多真实复杂网络中的社区往往包含层次结构,具有重叠现象。网络层次反映了网络的分层结构,它将位于独立层上的多个群组连接了起来,从而形成了一个更加复杂的网络。例如,在第  $i$  层中的中心节点在第  $j$  层中可能变为边缘节点。在社区结构中,重叠性和层次性两者联系非常紧密,有必要将两者融合在一起考虑重叠社区结构发现问题。用于层次化重叠社区发现的 GNN 模型必须实现多层表征提取这一挑战,这包括不同层中的稀疏程度。

## 5.4 异质网络

现实网络中往往包含丰富的异质信息,也就是说,网络中含有不止一种类型的节点和连边,而各种各样的节点集合和它们之间的联系形成了异质网络。当前大部分的社区发现算法都不适用于异质网络,因此,我们应该通过不同于同质网络的方法来研究异质网络的社区发现。使用 GNN 模型进行异质网络社区发现一个重要的挑战是如何获得节点和连边的信息表征。

## 5.5 动态网络

真实世界中的复杂网络不仅异质,而且还是动态变化的,常用来模拟复杂系统的动态演变。在动态网络中,节点和连边都是可变的,这些变化往往导致网络拓扑的变化以及在固定拓扑上的属性变化,这些变化会引起社区的演化。如何在动态变化的复杂网络上进行社区发现,需要我们进一步探索研究。

## 5.6 GNN 模型与统计建模相结合

最近的研究已经提出了将统计建模与图神经网络模型相结合的几种方法,如 MRFasGCN<sup>[43]</sup> 将 GCN 和 MRF 集成在一起,以解决网络中的半监督社区检测问题。但将图模型与统计建模相结合仍是一个研究挑战。例如:现有的方法通常是利用统计模型提供的先验知识(如社区)来完善 GCN 的嵌入以改善所得社区,但这些方法没能充分考虑模型的时间复杂性和可解释性,从而在实践中对社区检测提出了巨大挑战。同时,将统计模型集成到图神经网络中仍是一个需要解决的问题。集成统计推断和图神经网络模型来创新新的算法,以产生更多适用于各种网络问题的可解释的网络表示模型是非常有必要的。

**结束语** 作为复杂网络中普遍具有的拓扑性质<sup>[44]</sup>,社区结构的研究得到了众多领域的关注,取得了许多研究成果,尤其是社区发现算法。已有的研究成果为网络的拓扑结构分析提供了强有力的工具,对理解真实世界的复杂系统有着重要意义。但是,传统的社区发现算法仍存在尚需解决的问题。如典型的非欧几里得数据结构的图数据中的节点特征提取。

图神经网络社区发现方法可以通过直接处理图数据捕获网络中节点、边及对应的拓扑结构,在图上对相关信息进行不

断更新计算,相较于传统社区发现算法的特征提取,具有强大的优势,可以有效避免在特征提取过程中信息利用不完全带来的诸多问题。本文通过分析 GNN 社区发现及其过程,对现有的 GNN 社区发现研究从重叠社区发现和非重叠社区发现两个方面进行详细探讨,指出 GNN 社区发现未来的研究方向,对基于图神经网络的社区发现的进一步研究具有一定的借鉴意义。

## 参 考 文 献

- [1] LIU Z Y,ZHOU J. Introduction to Graph Neural Networks [M]. Morgan & Claypool, 2020:1-3.
- [2] HENAFF M,BRUNA J,LECUN Y. DeepConvolutional Networks on Graph-structured Data[J]. arXiv:1506.05163, 2015.
- [3] BASTINGS J,TITOV I,AZIZ W, et al. GraphConvolutional Encoders for Syntax-aware Neural Machine Translation[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017: 1957-1967.
- [4] RHEE S,SEO S,KIM S. Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification[J]. arXiv:1711.05859, 2017.
- [5] ZHANG Y H,QI P,MANNING C D. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2205-2215.
- [6] GORI M,MONFARDINI G,SCARSELLI F. A New Model for Learning in Graph Domains[C]// IEEE International Joint Conference on Neural Networks, 2005:729-734.
- [7] SCARSELLI F,GORI M,TSOI A C, et al. The Graph Neural Network Model[J]. IEEE Transactions on Neural Networks, 2009,20(1):61-80.
- [8] BRUNA J,ZAREMBA W,SZLAM A, et al. Spectral Networks and Locally Connected Networks on Graphs[C]// International Conference on Learning Representations (ICLR). 2014.
- [9] HAMMOND D K,VANDERGHEYNST P,GRIBONAL R. Wavelets on Graphs via Spectral Graph Theory[J]. Applied and Computational Harmonic Analysis, 2011,30(2):129-150.
- [10] KIPF T N,WELLING M. Semi-supervised Classification with Graph Convolutional Networks[C]// International Conference on Learning Representations (ICLR). 2017.
- [11] LI R,WANG S,ZHU F, et al. Adaptive Graph Convolutional Neural Networks[C]// Proc. of AAAI. 2018:3546-3553.
- [12] NIEPERT M,AHMED M,KUTZKOV K. Learning Convolutional Neural Networks for Graphs[C]// Proc. of ICML. 2016: 2014-2023.
- [13] HAMILTON W,YINGR,LESKOVEC J. Inductive Representation Learning on Large Graphs[C]// Advances in Neural Information Processing Systems. Long Beach, US, 2017:1024-1034.
- [14] ATWOOD J,TOWSLEY D. Diffusion-convolutional Neural Networks[C]// Proc. of NIPS. 2016:1993-2001.
- [15] ZHUANG C,MA Q. Dual Graph Convolutional Networks for Graph-based Semi-supervised Classification [C] // Proc. of WWW. 2018:499-508.
- [16] LUO Z G,JIANG X Z,DING F, et al. New Development of Community Discovery Algorithm in complexnetworks[J]. Jour-

- nal of Defense Science and Technology University, 2011, 3(1): 47-52.
- [17] LIU D Y, JIN D, HE D X, et al. Survey of Community Mining in Complex Networks[J]. Computer Research and Development, 2013, 50(10): 2140-2154.
- [18] CHENG X Q, SHEN H W. Community Structure of Complex Network [J]. Complex System and Complexity Science, 2011, 8(1): 57-70.
- [19] GRIVAN M, NEWMAN M E J. Community Structure in Social and Biological Networks[J]. Proc Natl Acad Sci USA, 2002, 99(12): 7821-7826.
- [20] ZHAO W J, ZHANG F B, LIU J L. Research Progress of Community Discovery in Complex Networks[J]. Computer Science, 2020, 47(2): 10-20.
- [21] LIU F Z, XUE S, WU J, et al. Deep Learning for Community Detection: Progress, Challenges and Opportunities[C]//International Joint Conference on Artificial Intelligence. 2020: 4981-4987.
- [22] XIN X, WANG C K, YING X, et al. Deep Community Detection in Topologically Incomplete Networks[J]. Physica A: Statistical Mechanics and its Application, 2017, 469: 342-352.
- [23] CAO J X, JIN D, YANG L, et al. Incorporating Network Structure with Node Contents for Community Detection on Large Networks using Deep Learning[J]. Neurocomputing, 2018, 297: 71-81.
- [24] CAO J, JIN D, DANG J W. Autoencoder Based Community Detection with Adaptive Integration of Network Topology and Node Contents[C]//KSEM. 2018: 184-196.
- [25] XU K, HU W H, LESJOVEC J, et al. How Powerful are Graph Neural Networks[C]//ICLR 2019. 2019: 1-17.
- [26] BRUNA J, XIANG L. Community Detection with Graph Neural Networks[J]. arXiv: 1705. 08415, 2017.
- [27] PETER V, BANDYOPADHYAY S. Self-Expressive Graph Neural Network for Unsupervised Community Detection[J]. arXiv: 2011. 14078, 2020.
- [28] ZHANG H, ACHARYA D B. Community Detection Clustering via Gumbel Softmax[J]. SN Computer Science, 2020, 1(5): 1-11.
- [29] ACHARYA D B, ZHANG H. Feature Selection and Extraction for Graph Neural Networks[C]//Proceedings of the 2020 ACM Southeast Conference. 2020: 252-255.
- [30] YANG L, CAO X, HE D, et al. Modularity Based Community Detection with Deep Learning[C]//International Joint Conference on Artificial Intelligence. AAAI Press, 2016: 2252-2258.
- [31] CHOONG J J, LIU X, MURATA T. Optimizing Variational Graph Autoencoder for Community Detection[C]//2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 5353-5358.
- [32] SUN J, ZHENG W, ZHANG Q, et al. Graph Neural Network Encoding for Community Detection in Attribute Networks [C]//IEEE Transactions on Cybernetics. 2020: 1-14.
- [33] CHEN Z D, LI L S, BRUNA J. Supervised Community Detection with Line Graph Neural Networks[C]//ICLR. 2019.
- [34] ZHENG Y P, CHEN S Y, ZHANG X N, et al. Heterogeneous-Temporal Graph Convolutional Networks: Make the Community Detection Much Better[J]arXiv: 1909. 10248, 2019.
- [35] LU N N, LUO W J, NI L, et al. Extending CDFR for Overlapping Community Detection[C]//Proceedings of the 1st International conference on Data Intelligence and Security. 2018: 200-206.
- [36] LUO W, YAN Z, BU C, et al. Community Detection by Fuzzy Relations[J]. IEEE Transactions on Emerging Topics in Computing, 2017, 1(9): 125-134.
- [37] SHCHUR O, GUNNEMANN S. Overlapping Community Detection with Graph Neural Networks [C] // KDD Workshop DLG'19. 2019.
- [38] TODESCHINI A, MISCOURIDOU X, CARON F. Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities[J]. arXiv: 1602. 02114.
- [39] YANG J, LESKOVEC J. Overlapping Community Detection at Scale: A Nonnegative Matrix Factorization Approach[C]//Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. 2013.
- [40] ZHOU M. Infinite Edge Partition Models for Overlapping Community Detection and Link Prediction[C]//2014 International Conference on Artificial Intelligence and Statistics. 2014: 1135-1143.
- [41] YU Q C, YU Z W, WANG Z, et al. Estimating Posterior Inference Quality of the Relational Infinite Latent Feature Model for Overlapping Community Detection [J]. Frontiers of Computer Science, 2020, 14(6): 1-15.
- [42] KOVACS I A, PALOTAI R, SZALAY M S, et al. Community Landscapes: An Integrative Approach to Determine Overlapping Network Module Hierarchy, Identify Key Nodes and Predict Network Dynamics[J]. Plos One, 2010, 5(9): e12528.
- [43] JIN D, LIU Z, LI W, et al. Graph Convolutional Networks Meet Markov Random Fields: Semi-Supervised Community Detection in Attribute Networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 152-159.
- [44] LIU D Y, JIN D, HE D X, et al. Survey of Community Mining in Complex Networks[J]. Computer Research and Development, 2013, 50(10): 2140-2154.



**NING Yi-xin**, born in 1997. Her main research interests include community detection and graph neural networks.



**XIE Hui**, born in 1978, Ph.D, associate professor. His main research interests include data mining and intelligent recommendation.