

# 复杂网络中的社团发现算法综述\*

李辉, 陈福才, 张建朋<sup>†</sup>, 吴铮, 李邵梅, 黄瑞阳  
(信息工程大学, 郑州 450002)

**摘要:** 社团发现方法能够用来挖掘网络中隐藏的聚簇结构信息,对复杂网络结构与功能的分析具有重要意义。近些年来,随着网络数据的爆炸式增长,网络演化的多样性,涌现出了大量能够处理不同场景的社团发现方法和框架。为了深入了解社团发现领域的研究现状和发展趋势,对复杂网络中的社团发现算法进行综述。首先,对这些算法进行了分类;其次,详细介绍了每一类算法,并进行了分析和对比;之后,介绍了一些常用的评价指标,并阐述了社团发现的应用场景;最后,对该领域未来研究方向进行了展望。

**关键词:** 社团发现; 动态网络; 重叠社团; 深度学习

中图分类号: TP391 文献标志码: A 文章编号: 1001-3695(2021)06-002-1611-08

doi:10.19734/j.issn.1001-3695.2020.06.0211

## Survey of community detection algorithms in complex network

Li Hui, Chen Fucui, Zhang Jianpeng<sup>†</sup>, Wu Zheng, Li Shaomei, Huang Ruiyang  
(Information Engineering University, Zhengzhou 450002, China)

**Abstract:** The community detection method can mine hidden cluster structure in the network, which is of great significance for the analysis of the structure and function of complex network. In recent years, with the explosive growth of network data and the diversity of network evolution, a large number of methods and frameworks for community detection that can handle different scenarios have emerged. In order to deeply understand the research status and development trend, this paper reviewed the community detection methods. This paper firstly classified these algorithms, introduced each kind of algorithm in detail and made analysis and comparison, and then introduced some commonly used evaluation indicators and expounded the application scenarios. Finally, this paper looked forward to the future research directions of this field.

**Key words:** community detection; dynamic network; overlapping community; deep learning

### 0 引言

复杂网络普遍存在于物理自然世界和人类社会之中,物理自然系统、生命系统以及社会管理系统均是复杂性科学的重要研究对象。现实生活中存在的大量复杂系统都可以通过形形色色的网络加以描述,无论是互联网、电话交互网、人际关系网还是生态系统中的食物链网都可以抽象为复杂网络。若将复杂系统中的个体抽象为节点,将个体之间的关系抽象为边,则可用复杂网络描述复杂系统的行为特性。虽然复杂网络的具体结构形态各异,涌现出许多复杂多变的网络现象,但是在统计意义上呈现出了惊人的相似属性,例如在真实网络中大多数都呈现出明显的社团结构;核心节点和普通节点共存且普通节点占据绝大多数,节点度分布呈现出明显的幂律特性;在网络中,任意两个节点间的最短路径通常较小,即著名的小世界效应。然而,随着社交媒体等新兴服务推动,网络空间的数据种类和规模出现前所未有的增长,特别是以脸书(Facebook)、微信(WeChat)为代表的各类社交网络蓬勃发展,每秒钟数以亿计的信息流入社交媒体平台。这些复杂网络中的网络结构承载着人们在社会生活中形成的各种关系模式,蕴涵着丰富的有价值的结构信息(如图1所示)。如何对这些当代复杂网络进行有效的社团划分是当今数据挖掘领域的研究热点。

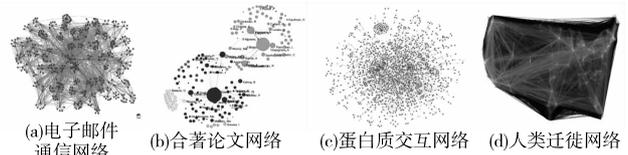


图1 真实世界中的复杂网络

Fig.1 Real complex networks

社团划分旨在利用网络拓扑中所蕴涵的结构信息从网络中发现模块化的社团结构,通常也被称为社团检测和图聚类。社团通常被认为内部连接相对紧密,外部连接相对稀疏的子图,广泛地存在于现实世界的各种网络中。它对复杂网络的拓扑分析、功能分析和行为预测具有重要的理论意义和实践价值。社团发现在推荐系统、网络监控、生物医学、公共安全等诸多领域有着非常多的应用。

自Newman等人<sup>[1]</sup>首次提出基于边介数的GN算法以来,社团发现的研究引起了国内外众多学者的广泛关注,来自各个领域的学者围绕此课题开展了深入研究,并取得了一系列重要研究成果。本文根据复杂网络动态变化与否,将社团发现算法分为基于静态网络的社团发现和基于动态网络的社团发现两大类。针对静态网络的社团发现研究相对较多,研究者们提出了基于不同理论框架(如模块度,标签传播等)的社团划分算法。然而,现实世界中很多网络随着时间的演化在动态发生变

收稿日期: 2020-06-22; 修回日期: 2020-08-18 基金项目: 国家自然科学基金群体项目; 国家重点研发计划项目; 郑州市协同创新重大

专项

**作者简介:** 李辉(1996-),男,湖北丹江口人,硕士研究生,主要研究方向为网络大数据分析;陈福才(1974-),男,江西高安人,研究员,硕导,硕士,主要研究方向为网络安全;张建朋(1988-),男(通信作者),河北廊坊人,助理研究员,博士,主要研究方向为大数据分析(zjp@ndsc.com.cn);吴铮(1992-),男,江苏徐州人,博士研究生,主要研究方向为链路预测;李邵梅(1982-),女,湖北钟祥人,副研究员,硕导,博士,主要研究方向为知识图谱;黄瑞阳(1986-),男,福建漳州人,副研究员,硕导,博士,主要研究方向为文本挖掘、图挖掘。

化。相对于静态网络,在动态网络中进行社团发现是一个艰巨的挑战,目前已经有一些算法被提出,但是还有许多问题亟待研解决。因此,本文对现有的社团发现算法进行了归纳,分析了各类算法的优点和不足;总结了社团发现结果的评价指标,并对未来研究方向进行了展望。

### 1 基本定义

**定义 1 静态网络。**定义  $G = \{V, E\}$  为一个基本的静态网络,其中  $V$  代表网络中的节点集,  $E \subseteq V \times V$  代表网络中的边集,  $N = |V|$  表示网络中节点的数量,  $M = |E|$  表示边的数量。每一条边被定义为  $e = (u, v, w)$ , 其中  $u$  和  $v$  是网络中的两个节点且  $u \neq v, w$  表示边的权重。

**定义 2 动态网络。**动态网络  $g = \{G_t | t = 1, 2, \dots, T\}$  包含了一系列时刻的网络快照  $G_t = \{V_t, E_t\}$ , 每一个网络快照  $G_t$  是由节点集  $V_t$  和边集  $E_t$  组成。

**定义 3 多层网络。**一个多层网络可以定义为  $G = \{G_1, \dots, G_l, E_m\}$ , 其中  $G_l = (V_l, E_l)$  表示第  $l$  层的网络,  $V = \{v_1, \dots, v_n\}$  表示多层网络中所有节点集合,  $V_l \subseteq V$  表示第  $l$  层网络中节点的集合,  $E_l \subseteq V_l \times V_l$  表示第  $l$  层内连边集合,  $E_m$  表示层与层之间的连边集合。

**定义 4 社团结构。**网络中内部连接紧密,外部连接稀疏的一组节点定义为社团,这些节点通常具有相同的属性或者相似的功能。如果网络中的每个节点仅属于一个社团,则称为非重叠社团如图 2(a) 所示,如果一个节点同时属于多个社团,则称为重叠社团如图 2(b) 所示。

**定义 5 社团发现。**也被称为社团检测或图聚类,旨在发现网络中的社团结构,即根据某种准则来划分网络中的节点,给每个节点分配到相应的社团。

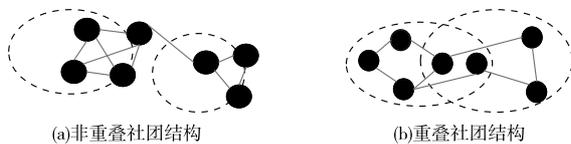


图2 非重叠社团和重叠社团结构  
Fig.2 Disjoint and overlapping community structure

### 2 社团发现算法概述

自从基于边介数的 GN 算法<sup>[1]</sup>问世以来,众多社团发现算法被相继提出来解决各个领域的社团发现问题。本文首先对这些算法进行了分类,如图 3 所示,并对每类算法进行概述和简要分析。

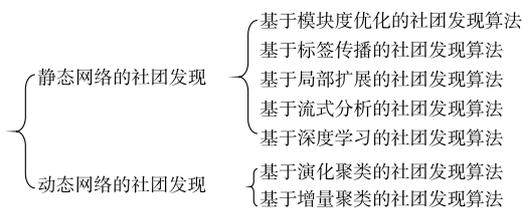


图3 社团发现的算法分类  
Fig.3 Algorithm classification of community detection

#### 2.1 静态网络的社团发现算法

现有的大多数算法都是用于处理静态网络,本文将这些算法分为基于模块度优化、标签传播、局部扩展、流式分析和深度学习五种类型,下面对每类算法进行分析概括。

##### 2.1.1 基于模块度优化的社团发现算法

2004 年,Newman 等人<sup>[1]</sup>首次给出了模块度  $Q$  的定义,用以衡量社团划分结果的好坏。基于模块度优化的社团发现法通过优化目标函数  $Q$  使模块度最大化,从而得到最优的社团划分结果。Newman 等人<sup>[2]</sup>提出快速模块度优化 FN 算法,它

将网络中的每个顶点初始化为一个社团,选择使模块度增值最大的社团进行合并,从而在此过程中形成树图,树的每一层划分对应着网络的某个具体划分,此算法通过模块度的值来判断迭代停止条件,并在树图的所有层次中选择模块度最大的社团划分作为最终划分。FN 算法在性能上较 GN 算法提升较多,但由于大多数网络较为稀疏,导致了邻接矩阵中存在着大量无意义的零值,使得 FN 算法在维护邻接矩阵和计算  $\Delta Q$  的过程中非常耗时。为了解决这个问题,Clauset 等人<sup>[3]</sup>直接构建了一个模块度增量矩阵,采用了平衡二叉树和最大堆的数据结构来存储该矩阵,能够在  $O(\log n)$  的时间复杂度下更新矩阵以及在恒定时间内找到最大的  $\Delta Q$ ,大大地节省了时间和空间消耗。Schuetz 等人<sup>[4]</sup>为模块度优化引入了贪婪算法的多步扩展,算法在每次迭代中合并  $l$  对社团 ( $l > 1$ ),可以避免过早的聚合成超大社团,并通过大量实验给出了用于选择步长  $l$  的经验公式。Blondel 等人<sup>[5]</sup>提出了采用贪婪方法进行局部优化的层次凝聚方法 Louvain 算法,该算法结合层次聚类与模块度优化算法进行迭代更新,从而形成层次化的聚类结构,是目前最为高效且被广泛应用的社团发现算法之一。SLM (smart local moving) 算法<sup>[6]</sup>是一种更加高效的模块度优化方法,该算法通过引入更高级的局部移动启发式算法,进一步扩展了两步 Louvain 算法以及 Rotta<sup>[7]</sup>的多级细化方法,可以在大规模网络中得到高质量的聚类结果。以上方法仅能处理单层网络,然而,随着网络的发展,逐渐出现了节点之间存在着多种关系、每一种关系构成一层网络的多层网络。Tang 等人<sup>[8]</sup>将单层网络上的模块度最大化扩展到多层网络上,为了消除直接扩展带来的对噪声数据敏感的问题,提出了主模块度最大化 (PMM) 来克服噪声数据的影响。算法从网络的每一层提取结构特征,然后对所构造的数据进行跨层次分析以找到低维度的嵌入,从而使所有维度提取的特征高度相关。算法可以得到一个较好的社团结构,对噪声数据有一定的容忍度,但需要预先输入社团的数量。Mucha 等人<sup>[9]</sup>基于单层网络模块度  $Q$  直接定义了多层网络模块度  $Q_{multiscale}$  用以挖掘多层网络中隐藏的社团结构,该函数带有可以控制每层网络社团规模和数量的参数,通过优化该函数可以进行复合社团发现,最后每一层网络都可以得到一个社团划分,该方法考虑了层与层之间的耦合关系,但是参数太多,且计算复杂度高。Pramanik 等人<sup>[10]</sup>提出了一种新的多层模块度指标  $QM$ ,用以衡量多层网络中社团的质量,通过最大化该指标来对多层网络进行社团划分,该方法不需要输入参数且能够获得较好的社团结构。然而,Fortunato 等人<sup>[11]</sup>通过研究发现,以模块度为目标函数的优化算法存在着分辨率限制的问题,即算法无法发现规模小于  $\sqrt{N/2}$  的较小社团。为了缓解分辨率限制的问题,Arenas 等人<sup>[12]</sup>提出了一种允许对社团结构进行多分辨率筛选的方法,可以发现预定义的不同粒度的社团结构。张聪等人<sup>[13]</sup>在网络密度的基础上提出了多分辨率的密度模块函数,能够降低社团结构的误划分率。Ma 等人<sup>[14]</sup>在文献[9]的基础上,提出了多层模块度密度,用以解决多层网络上的分辨率限制问题。但这些方法都只能在一定程度上缓解分辨率问题,当网络中的社团规模较小时,会使结果变得不可靠。基于模块度优化的算法能够较为准确地识别网络中的社团结构,但是计算复杂度偏高,当网络的规模变大时,搜索空间将会变得非常大。

##### 2.1.2 基于标签传播的社团发现算法

标签传播从信息传播的角度出发,认为社团内的信息传播速度要比社团之间的传播速度快,因此,每个节点应该和其周围大部分节点位于同一社团。2007 年,Raghavan 等人<sup>[15]</sup>将标签传播思想<sup>[16]</sup>应用于社团发现任务,提出基于标签传播的社团发现算法,该算法具有线性时间复杂度,能够快速收敛,适用于大规模网络。算法为每个节点初始化唯一标签,以随机的顺

序逐轮更新节点标签。在每次更新过程中,扫描节点所有的邻居节点,把邻居节点中出现次数最多的标签赋予当前节点,直至所有节点的标签保持稳定。然而,算法是以随机的顺序进行更新,使得初始节点的选取会对最终结果产生较大的影响,产生的结果不稳定,社团划分的质量受到较大影响。为了解决这个问题,很多学者做了大量的尝试。Barber 等人<sup>[17]</sup>将 LPA 转换为目标优化问题,给定目标函数,将模块度引入到目标函数中,通过最大化模块度来获得社团划分结果,同时保持了 LPA 算法的计算复杂度。然而,该算法的划分结果中,所有社团内部节点度之和基本相同,因此很容易陷入局部最优,Liu 等人<sup>[18]</sup>为了解决这个问题,提出了多步贪婪凝聚算法,可以一次合并多个社团,能够有效地避免陷入局部最优,并且提高了算法的精度,但是增加了时间复杂度。Zhao 等人<sup>[19]</sup>通过引入信息熵,定义了一种新的度量标签熵,提出了以熵为序的标签传播算法,利用标签熵从小到大的顺序更新标签,在一定程度上提高了 LPA 算法的稳定性。上述文献都只是针对非重叠社团,对于重叠社团,刘世超等人<sup>[20]</sup>将标签传播算法用来发现重叠社团,根据节点的影响力大小对节点进行排序,并综合网络的传播特性和节点的属性特征来更新节点的标签,能够快速地发现网络中的重叠社团结构。Kato 等人<sup>[21]</sup>为了挖掘多层网络上的社团,将多层网络进行聚合,通过图拉普拉斯矩阵的线性组合来优化各层网络的权重,然后在聚合之后的网络上使用单层网络标签算法来对节点进行社团划分,但是该方法无法得到节点在每一层上的网络划分,只能得到在聚合网络上的社团划分。Alimadadi 等人<sup>[22]</sup>将单层网络标签传播算法扩展到多层网络中,该方法定义了多层网络上的节点相似性度量,通过该度量可以筛选出节点  $i$  在所有层中的邻居节点,统计节点  $i$  邻居节点的标签,将出现次数最多标签的赋予当前节点,可以快速地挖掘出多层网络中的社团结构。

可以看到,基于标签传播的算法不需要输入任何参数,而且具有线性时间复杂度,能够快速收敛,可以很好地适用于大规模网络,但是结果具有较大的随机性,算法多次运行会产生不同的结果,稳定性较差。

### 2.1.3 基于局部扩展的社团发现算法

局部方法主要是利用节点的局部信息对网络进行社团发现。与全局方法相比,局部方法可以在没有完整网络结构的情况下发现社团。基于局部的方法通常包括种子节点的选取以及节点局部扩展两个部分。在种子选择阶段,主要基于节点某些指标来寻找社团的核心成员。在社团扩展阶段,根据种子的影响力或具有目标函数的贪婪过程从种子节点扩展社团。种子节点的选取会对社团发现结果产生较大的影响,常见种子节点的选取包括随机选择、根据全局排名、完全子图、混合方法等。随机选取种子方法简单、易于实现、计算复杂度小,但是没有考虑节点的重要性以及网络的拓扑结构,会对结果造成较大的影响。根据全局排名选择种子主要是根据节点在整个网络中的权重值大小来进行选择<sup>[23]</sup>。如 Baumes 等人<sup>[24]</sup>通过不断地移除度最大的节点以找到在给定规模内连接较少的结构作为种子。 $k$  完全子图方法是通过在网络中选择大小为  $k$  的完全子图(也被称为  $k$  派系)作为种子节点,如 Shang 等人<sup>[25]</sup>通过在网络中选取  $k \geq 4$  的完全子图作为种子节点。该方法认为网络中完全子图内的节点相互连接紧密,有着极大的概率是社团的核心节点。用  $k$  完全子图选取种子节点可以使社团发现结果较为稳定,但是计算复杂度高,难以适用于大规模网络。为了克服单一方式的缺陷,也可以通过多种方式相结合来选择种子。种子节点选择好之后,就需要对种子进行局部扩展。局部扩展包括无监督扩展和半监督扩展。当无法得到有效的节点标签信息时,采用无监督扩展,通过最大化/最小化目标函数或者社团的某个特征对种子进行扩展。如在局部自适应算法

LFM<sup>[26]</sup>中,根据连通度(conductance)和模块度(modularity)定义了目标函数,通过最大化目标函数来对种子进行扩展;Luo 等人<sup>[27]</sup>在种子节点的邻居节点中寻找具有最大中心性的节点来扩张社团,这些都是在无法获取先验知识的情况下进行的无监督扩展。基于半监督方法利用部分节点的先验信息来指导种子扩展,先验信息主要包括部分节点的社团标签和成对节点是否同一社团两种类型。Liakos 等人<sup>[28]</sup>通过在每个真实社团中寻找三个节点,以这三个节点的真实社团标签作为先验知识来进行社团扩展。选择合适的种子,定义局部扩展策略,就能以局部扩展的方式搜索网络中的社团。Yin 等人<sup>[29]</sup>通过寻找具有局部最小模体连通度(motif conductance)的节点作为种子节点,然后通过随机游走对种子节点进行扩展。算法引入了网络中的局部高阶信息,将基于边的连通度扩展到基于模体的连通度,提高了社团划分的质量。Tasgin 等人<sup>[30]</sup>通过寻找边界点作为种子节点,然后利用最大化社团指标来进行局部扩展,定义了七个指标,通过实验验证选择出一个最优的基于共同邻居的 G-CN 指标作为扩展指标,在真实数据集和合成数据集上都取得了较好的效果。Guo 等人<sup>[31]</sup>通过度中心性和 Jaccard 系数相结合来挑选初始种子节点,然后对种子节点进行筛选减少种子数量,定义适应度函数  $F_c$  对种子节点进行扩展从而得到高质量的社团,但是算法需要提前设定三个参数,这使得在实际应用中很难获取最优的结果。李慧等人<sup>[32]</sup>通过寻找具有最大聚集系数的节点来作为种子,最大化局部适应度函数对社团进行扩展,最后合并冗余社团,算法能够取得较好的结果,但是时间复杂度较高。Interdonato 等人<sup>[33]</sup>提出了一种在多层网络上进行局部社团发现的方法 ML-LCD。在种子节点周围,通过不断地最大化局部扩展函数(社团内部连接密度和社团外部连接密度之比)来对社团进行扩展,然后将各层的贡献度与连接密度比线性组合来挖掘出多层网络中的社团结构。Li 等人<sup>[34]</sup>提出了一种融合层间和层内影响关系的多层网络局部社团检测模型,算法定义了一个半局部中心性指标,可以有效地识别多层网络中局部信息下节点的重要程度,以此来找到局部核心节点,并且通过基于直接影响和间接影响的度量来对核心节点进行扩展以获取局部社团结构。该算法计算速度快,能够获得较高质量的社团,且算法在稀疏网络中仍然具有较好的表现。

由此可见,基于局部扩展的算法只需要节点的局部信息,并且种子扩展过程可以并行,因此可以适用于大规模网络。由于每个种子扩展的过程是相互独立的,得到的社团是有交叉的,所以可以用来发现网络中的重叠社团。

### 2.1.4 基于流式分析的社团发现算法

随着互联网的发展,网络的规模越来越大,现有的很多网络已经发展到具有数千万个节点和数十亿条边如社交网络 Friendster<sup>[35]</sup>,已有的很多算法在计算时间和内存消耗上都无法很好地处理如此大规模的网络,于是一种新的以流的方式对网络进行处理解决方案被提出。基于流式分析方法的基本思路是以流式的方法对网络中的每条边进行单遍扫描处理,每条边仅被处理一次,不需要将整个网络读入内存,时间复杂度和空间复杂度都较低,适用于处理大规模的网络。Yun 等人<sup>[36]</sup>提出了一种空间复杂度随网络线性增长的算法来解决内存限制的问题,该算法首先构建了网络的邻接矩阵,然后顺序地读取邻接矩阵的列进行处理,通过对邻接矩阵的部分信息进行局部谱聚类,从而对节点进行社团划分。该算法无须一次性处理整个邻接矩阵,邻接矩阵的每一列只处理一次即可丢弃,也不需要额外的空间存储中间结果,但是算法需要提前构建网络的邻接矩阵。Nguyen 等人<sup>[37]</sup>首先在网络中选取部分节点和边构成子图,利用 Louvain 算法在子图上确定初始的社团,然后对剩下的边进行流式处理,每次只处理一条边,将节点分配到

已有社团或者新社团,或者从一个社团移动到另外一个社团,这样可以快速地处理剩下的网络,但初始社团的划分具有较大的随机性。Holloccou 等人<sup>[38]</sup>认为随机均匀地读取网络中的边,社团内边应该比社团间边先到达,因为社团内的边的数量要远大于社团间的边的数量,提出了基于边的流式处理的社团检测算法 SCoDA。该算法以流的方式处理每一条边且每条边仅处理一次,算法时间复杂度和空间复杂度都是线性的,在合成网络和真实网络中都表现非常好。但是该算法要求随机均匀地读取每条边,这就要求在处理之前要随机指定网络中边的顺序,而且需要得到网络中所有节点的度并统计出现次数最多的度数。其仅通过节点度大小来判断节点是否移动到另外一个社团,这样过处理规则过于简单,会造成很多错误的移动。Liakos 等<sup>[28]</sup>提出了基于种子集扩展的流式社团检测算法,首先,在每个真实的社团中选择三个节点作为初始社团,对剩下的边进行局部扩展。算法对边的读取顺序没有限制,流式地对每一条新来的边进行处理,在时间窗口内,社团随意地进行扩展,通过节点对社团的贡献值对社团进行修剪以保证社团的质量。可以看到,基于流式分析的方法不需要将整个网络存储到内存中,对每条边仅处理一次,大大降低了计算时间和内存消耗,可以处理超大规模的网络,但由于每条边仅处理一次,能够利用的信息有限,所以算法的精度还有待提高。

### 2.1.5 基于深度学习的社团发现算法

近些年,随着深度学习的发展,一些研究者利用神经网络来处理社团发现任务。其核心思想是通过深度学习框架来提取网络中节点的潜在特征,从而利用更多的网络信息来提高社团发现的质量。2014年,Tian 等人<sup>[39]</sup>首次将深度学习应用到社团发现任务当中,算法将图的拉普拉斯矩阵作为自编码器的输入进行训练来提取网络节点特征,从而将节点表示为向量,然后用 K-means 算法<sup>[40]</sup>进行聚类得到社团的划分。为了进一步提升模型效率,Yang 等人<sup>[41]</sup>将拉普拉斯矩阵替换为模块度矩阵进行训练,模型还扩展到了半监督学习,利用已知邻接节点是否同属一个社团的信息来提升社团发现的结果。DeepWalk<sup>[42]</sup>算法通过在网络中随机游走产生节点序列,借助 word2vector<sup>[43]</sup>中的 skip-gram 模型将节点表示为向量,再利用 K-means 算法<sup>[40]</sup>进行聚类。以上几种算法都只利用了网络拓扑结构来进行特征表示,而忽略了网络中节点的内容信息。Wang 等人<sup>[44]</sup>提出了一种用于图聚类的边缘化自动编码器,允许节点内容与网络结构进行交互,能够有效地集成结构和内容信息来更好地提取网络的特征,然后将学习到的特征送到光谱聚类算法中进行聚类分析。

然而,上述算法都只是专注于一般性的网络表示学习,所得到的网络向量表示都是普适性的;对于社团检测这个特定的任务,还需要在学习到的向量上采用向量聚类算法如 K-means、EM 算法等才能得到社团划分结果,导致了网络表示学习和社团划分成了相互独立的两个任务。后来研究者将网络表示学习和社团发现这两个问题相结合,提出了统一的框架;Sun 等人<sup>[45]</sup>提出一种 vGraph 的概率生成模型,以协作的方式学习网络表示和社团划分两个任务。算法允许节点表示和社团发现两个任务以互惠互利的方式进行交互,从而得到更优的社团划分结果。Cavallari 等人<sup>[46]</sup>为了更好地对网络特征进行提取,提出了对网络中的社团结构进行低维表达,将社团嵌入定义为节点的多元高斯分布。其提出了一种联合框架,将社团发现、节点表示、社团嵌入三个任务相结合,形成一个增益环,定义了统一的目标函数使三者得到最优化的结果,尽管将社团结构和节点表示一起进行了优化,但仍通过基于节点表示向量的聚类算法来推断社团划分,这意味着其无法处理密集的社团重叠问题。Shechur 等人<sup>[47]</sup>提出了一种能够检测重叠社团的端到端深度学习模型。此方法利用图神经网络和伯努利-泊

松概率模型来生成网络的从属关系矩阵并不断对其进行优化,从而能够较好地检测重叠社团。Jia 等人<sup>[48]</sup>提出了一种新颖的社团发现框架 CommunityGAN,可以解决网络表示学习和重叠社团发现两个子问题。与传统网络表示学习算法学习到的向量(向量值没有特殊含义)不同,此模型学习到的向量代表了顶点对社团的隶属强度,可以通过设定阈值来发现网络中的重叠社团,算法通过特定的生成对抗网络对学习到的向量进行优化。通过模体级(motif-level)生成器和判别器之间的竞争,迭代地进行优化并输出最终的社团结果。Ghorbani 等人<sup>[49]</sup>利用图卷积神经网络(graph convolutional network,GCN)对多层网络中的节点进行向量表示,该模型考虑了图层之间的依赖性,通过定义适当的损失函数来重建层内和层间的连接。Ni 等人<sup>[50]</sup>提出了一种深度多层网络嵌入模型 DMNE(deep multi-network embedding),该模型将多个神经网络(多层网络的每一层对应一个神经网络)与一个协同正则化函数相协调,通过多层非线性映射,逐步将每一层网络转换为一个高度非线性的潜在空间并通过协同正则化将各层的空间相适应,能够更好地捕获多层网络的特征,将节点表示为向量,而对得到的向量进行聚类就可以获取多层网络中的社团划分。Luo 等人<sup>[51]</sup>在多层网络上利用来自不同域的信息提高社团划分的精度,并提出了具有注意机制的深度深度学习模型 DMGC(deep multi-graph clustering),可以同时解决多层网络中的社团划分和社团关联问题。由此可见,基于深度学习的社团发现算法利用神经网络来提取网络的特征,使得社团发现的精度比传统方法提高较多;但由于复杂的推论以及模型的预训练,使得算法的计算复杂度较高,且该类算法大都需要 GPU 等图形服务器的计算单元。

综上,本小节对静态网络的社团发现算法进行了阐述和分析,表 1 对每一类算法进行了简要的概括。

表 1 静态网络社团发现算法总结

Tab. 1 Summary of community detection algorithms on static network

方法	优点	不足
基于模块度优化	算法精度较高,稳定性好	时间复杂度较高,具有分辨率限制
基于标签传播	算法效率高,速度快	存在较大的随机性,稳定性较差
基于局部扩展	无须整个网络信息,可并行,可探索局部社团结构	需要提前选取种子节点,社团的边界难确定
基于流式分析	处理速度快,所需空间小,能够快速处理大规模网络	利用的网络信息有限,精度较低
基于深度学习	能够利用更多的信息,算法的精度高,划分的社团更为准确	需要预训练,计算复杂度高

## 2.2 动态网络的社团发现算法

在 2.1 小节中,主要介绍了用于发现静态网络中的社团方法,但在真实网络中,很多网络是随着时间的推移在动态演化的。网络中不断有新的节点和边的插入和删除。节点间的关系随时间变化导致了网络中的社团也在不停的演变,传统的静态网络的分析方法无法很好地应用于动态网络,这使得在动态网络中进行社团挖掘变得更具挑战性。近些年,也相继出现了一些针对动态网络的社团发现算法来解决上述问题。

动态网络是由各个时刻  $t_1, t_2, \dots, t_m$  所对应的静态网络快照  $G = \{G_1, G_2, \dots, G_m\}$  构成,每个时刻  $t$  有着与之对应的社团划分。随着时间的变化,动态网络中的社团结构也会发生变化,包括社团的诞生、合并、分裂以及消亡。在动态网络中进行社团发现任务的方法主要分为基于演化聚类的社团发现算法和基于增量聚类的社团发现算法两大类。

### 2.2.1 基于演化聚类的社团发现算法

基于演化聚类的社团发现算法认为相邻时刻的社团结构

变化较为缓慢,一般不会发生突变,即当前时刻  $t$  的社团结构受到  $t-1$  时刻的社团结构的影响。因此演化聚类会综合考虑当前快照的社团划分结果与历史时刻的社团划分。即分别对  $t$  和  $t-1$  时刻的快照进行静态聚类,并要求  $t$  时刻的社团结构应尽可能与  $t-1$  时刻保持一致。Chakrabarti 等人<sup>[52]</sup>首先提出了基于演化聚类的社团检测框架,该框架将演化聚类问题转换为双目标优化问题,一是尽可能捕获当前时刻网络快照高质量的社团结构,二是尽可能使得当前社团划分与上一时刻社团划分保持一致。其目标函数为

$$COST = \alpha \times SC(C_t, G_t) + (1 - \alpha) \times TC(C_t, C_{t-1}) \quad (1)$$

其中: $\alpha$  是用户指定的权重系数; $SC(C_t, G_t)$  表示  $t$  时刻团划分的质量; $TC(C_t, C_{t-1})$  表示  $t$  时刻的社团划分结果与  $t-1$  时刻社团划分结果之间的差异程度。基于此框架,Lin 等人<sup>[53]</sup>提出概率生成模型 FaceNet 用于分析社团和社团演化,通过当前网络拓扑结构和历史社团结构给出的先验分布来估计当前的社团结构,但是需要提前设定好社团的个数。Chi 等人<sup>[54]</sup>将谱聚类引入到演化聚类的框架中,考虑了当前时刻的社团结构和历史快照来挖掘动态网络中的社团结构。文献<sup>[55]</sup>提出了一种自适应的演化聚类算法,通过最大化模块度来快速识别动态网络中的社团结构。Niu 等人<sup>[56]</sup>采用遗传算法来优化上述目标函数,引入标签传播算法来初始化社团结构,增加了初始社团结构的多样性,提高了聚类质量,但是增加了搜索空间,使得计算复杂度增加。Messaoudi 等人<sup>[57]</sup>提出了多目标蝙蝠算法来优化动态网络中的社团发现问题,该算法利用均值漂移技术产生初始社团以解决参数收敛问题和减少搜索空间,利用蝙蝠算法来自动寻找最优的社团划分结果。

## 2.2.2 基于增量聚类的社团发现算法

增量聚类主要是将网络在历史时刻的社团划分结果作为当前时刻网络社团划分的初始条件,在历史时刻社团划分的结果之上进行增量更新。此类方法可以自适应地更新社团变化,而不需要重复计算,在  $t$  时刻检测到的社团是基于  $t-1$  时刻检测到的社团结构。也就是说,在增量社团检测方法中仅考虑网络中发生变化的节点和边,保留了原有大部分社团结构,对发生变化的数据进行增量聚类更新。蒋盛益等人<sup>[58]</sup>提出一种基于增量式谱聚类的动态社团自适应发现算法,通过调整线性谱聚类算法从而对下一时刻继续增量聚类。Zakrzewska 等人<sup>[59]</sup>提出一种基于局部扩展的动态网络算法,该算法随着初始图的变化逐步地更新社团,对变化的节点计算其适应度来划分社团,该算法可以获得较高质量的社团,但是时间效率较低。在此基础上,Guo 等人<sup>[60]</sup>利用节点对社团的贡献度来增量的更新社团结构,有效提高了算法的计算效率。Sarıyüce 等人<sup>[61]</sup>提出了一种基于增量的并查集(merge-find)方式用于社团发现,可以挖掘出动态网络中的重叠社团。Wu 等人<sup>[62]</sup>在初始社团的基础上对新的节点和边利用 K-clique 算法进行增量更新,得到了较好的社团划分,但是该算法无法处理节点和边的删除操作。Hu 等人<sup>[63]</sup>提出了一种 LDM-CET 算法,利用近似个性化 PageRank 算法来处理发生变化节点的局部子图。当网络变化较小时,该算法可以有效地跟踪社团的演变,但当社团边界模糊时,LDM-CET 很难获得高质量的社团结构。Javadi 等人<sup>[64]</sup>提出一种利用社团领导者的重要性和持久性指标来增量检测动态网络中的社团结构的算法,它可以有效地发现真实动态网络和人工动态网络中的社团结构。但当网络的规模增加时,其执行效率会急剧下降。除此之外,还可以将动态网络转换为多层时间网络,每个时刻的快照对应多层网络中的一层,即  $G = (G_1, \dots, G_t, E_M)$ 。其中, $G_t$  对应时刻  $t$  的网络快照, $G_t = (V_t, E_t)$  层之间的连边  $E_M = \{E_{t,t+1} | (E_{t,t+1} = \{(v, v); v \in V_t \cap V_{t+1}\})\}$ 。Esraa 等人<sup>[65]</sup>提出了一种基于张量的多层时间网络社团发现方法,该方法将频谱聚类扩展到多层时间网络,将多

层时间网络的邻接矩阵表示为三阶张量,利用张量 Tucker 分解识别社团并跟踪其随时间的活动,该方法能够想充分利用各层的信息,但当网络较为稀疏时,结果就变得不准确。

基于演化聚类的处理方式,综合考虑了当前时刻和历史时刻的社团结构,将历史快照信息和当前拓扑结构有效地进行融合,从而获得相对合理的社团划分。由于演化聚类需要在每一时刻的网络快照上进行聚类操作,计算复杂度较高,但是综合考虑了历史信息,使得社团发现结果更为精准。而基于增量聚类的处理方式,以历史时刻的网络划分结果作为初始条件,只对因节点和边发生变化而受影响的社团进行更新,不需要对所有的节点和边进行社团划分,计算效率较高,能够对网络的变化进行及时更新,但是其是以历史时刻划分结果为基础,这样会造成当前的社团划分质量非常依赖历史时刻的社团质量,一定程度上存在误差累计的效应,从而影响整体的社团划分效果。表2对动态网络社团发现的两种方法进行了总结。

表2 动态网络社团发现算法总结

Tab.2 Summary of community detection algorithms on dynamic network

方法	优点	不足
基于演化聚类	能够综合历史信息和当前社团结构,社团划分较为准确	时间复杂度较高
基于增量聚类	在历史社团结构的基础上进行增量聚类,处理速度快,能够满足及时更新的要求	依赖历史社团质量,更新过程中存在误差累计

## 3 评价指标

在社团发现的研究中,研究者们提出了许多评价指标用于评判社团发现结果的优劣。但是目前,学术界还没有任何一个单一指标能够全面准确地评估社团划分结果,不同的评价指标对于同一个社团发现结果给出的评价也不尽相同。按照是否已知数据的真实标签,本文将其分为有监督的评价指标和无监督的评价指标。

### 3.1 有监督的评价指标

有监督评价指标是已知网络真实社团划分的情况下,将检测到的社团结果与网络的真实社团结构进行对比,两者越相近,表明算法效果越好;反之亦然。目前最为常用的有监督的评价指标为标准互信息<sup>[26]</sup>(normalized mutual information, NMI)和  $F_1$ -measure 指标<sup>[66]</sup>。

a) 标准互信息 NMI。NMI 是一个基于信息熵相似性度量,具体定义如下:

$$NMI(X, Y) = \frac{H(X) + H(Y) - H(X, Y)}{(H(X) + H(Y)) / 2} \quad (2)$$

其中: $H(X)$  是与检测到的社团相关联的随机变量  $X$  的熵; $H(Y)$  是与真实社团相关联的随机变量  $Y$  的熵; $H(X, Y)$  是联合熵。为了评价社团发现的结果与真实结果的相似度,本文将其更加直观地描述为

$$NMI = \frac{-2 \sum_{i=1}^{C_i} \sum_{j=1}^{C_j} N_{ij} \log \left( \frac{N_{ij} N}{N_i N_j} \right)}{\sum_{i=1}^{C_i} N_i \log \left( \frac{N_i}{N} \right) + \sum_{j=1}^{C_j} N_j \log \left( \frac{N_j}{N} \right)} N_i N_j \quad (3)$$

其中: $N_{ij}$  表示社团  $i$  和  $j$  中共同拥有的节点集合; $N_i$  表示社团发现结果中社团  $i$  中的节点集合; $N_j$  表示社团  $j$  中的节点集合; $N$  表示检测到的社团结构与真实社团结构的并集。NMI 的值在  $0 \sim 1$ , 值越大,表明检测结果与真实结果越相近,说明算法精确度越高。此评价指标也可以用于评价重叠社团。

b)  $F_1$ -measure。 $F_1$ -measure 通常用来评价分类结果的好坏。它综合考虑了精确率(precision)和召回率(recall),可以用来衡量检测到的社团与真实社团之间的差异。给定一个检测到的社团  $C_f$  和真实社团  $C_t$ ,精确率和召回率的定义为

$$precision(C_f, C_t) = \frac{|C_f \cap C_t|}{C_f} \quad (4)$$

$$recall(C_f, C_t) = \frac{|C_f \cap C_t|}{C_t} \quad (5)$$

$F_1$ -measure 定义为精确率和召回率的调和平均值。

$$F_1(C_f, C_t) = 2 \frac{precision(C_f, C_t) \times (recall(C_f, C_t))}{precision(C_f, C_t) + recall(C_f, C_t)} \quad (6)$$

对于社团发现算法将网络划分为  $K$  个社团,  $C_f = \{C_1, C_2, \dots, C_k\}$ , 以及真实的社团划分  $C_t = \{C_1, C_2, \dots, C_L\}$ , 本文定义算法划分社团的  $F_1$ -measure 为

$$F_1(C_f, C_t) = \frac{1}{K} \sum_{k=1}^K \max_{1 \leq l \leq L} F_1(C_f, C_l) \quad (7)$$

$F_1$ -measure 越大, 说明社团划分算法的精度越高, 检测到的社团与真实社团越相近。此评价指标既可以评价非重叠社团也可以评价重叠社团。

### 3.2 无监督的评价指标

当网络的真实社团结构无法获取时, 只能利用社团结构应该满足内部连接紧密, 外部连接稀疏这个特性来定义相关的质量函数来量化社团划分的质量, 应用较为广泛的两个无监督评价指标是模块度<sup>[1]</sup>和连通度<sup>[61]</sup>。

a) 模块度  $Q$  (modularity)。为了衡量社团结构, Newman 等人<sup>[1]</sup>提出了模块度用于量化社团划分结果的好坏, 定义为

$$Q = \sum_{i=1}^k (e_{ii} - a_i^2) \quad (8)$$

其中:  $e_{ii}$  代表第  $i$  个社团内的边的数量占网络总的边数的比例;  $a_i^2$  代表社团  $i$  与其他社团之间的连边数量与网络总边数的比值。由式(8)可知  $Q \in [-0.5, 1)$ , 模块度越大, 说明网络的社团结构越明显, 算法的效果越好。在真实网络中, 模块度  $Q$  一般在  $0.3 \sim 0.7$ 。

为了使计算更加清晰和直观, 式(8)可以改写为

$$Q = \frac{1}{2m} \sum_{C_i \in C} \sum_{u,v} \delta_{C_i,u} \delta_{C_i,v} \left[ A_{uv} - \frac{k_u k_v}{2m} \right] \quad (9)$$

其中:  $m$  为网络中的边的数量;  $C_i \in C$  表示第  $i$  个社团;  $u, v$  表示网络中的节点;  $\delta_{C_i,u}$  表示节点  $u$  是否属于社团  $C_i$ , 如果属于, 则为 1, 反之则为 0。  $A_{uv}$  代表邻接矩阵中节点  $u, v$  对应的值, 如果  $u, v$  之间有连边, 则  $A_{uv} = 1$ , 反之则为 0。  $k_u$  表示节点  $u$  的度。

在重叠社团中, 一个节点可能属于多个社团, 为了使得模块度  $Q$  也适用于评价重叠社团, Nicosia 等人<sup>[67]</sup>对模块度  $Q$  做了扩展, 将  $\delta_{C_i,u}$  修改为隶属系数  $\alpha_{C_i,u}$ , 表示节点  $u$  对社团  $C_i$  的隶属程度, 定义为

$$Q_0 = \frac{1}{2m} \sum_{C_i \in C} \sum_{u,v} \alpha_{C_i,u} \alpha_{C_i,v} \left[ A_{uv} - \frac{k_u k_v}{2m} \right] \quad (10)$$

$$\alpha_{C_i,u} = \frac{n_{C_i,u}}{\sum_{C_i \in C} n_{C_i,u}} \quad (11)$$

其中:  $n_{C_i,u}$  是节点  $u$  的邻居节点在社团  $C_i$  内的个数。

b) 连通度 (conductance)。连通度衡量的是社团  $C_i$  与其他社团之间连边和社团  $C_i$  内的连边的比值, 定义为

$$conductance(C_i) = \frac{cut(C_i, \overline{C_i})}{\min\{vol(C_i), vol(\overline{C_i})\}} \quad (12)$$

其中:  $cut(C_i, \overline{C_i})$  表示社团之间的边;  $vol(C_i)$  表示社团  $C_i$  内边的数量。社团的连通度的取值范围是  $(0, 1]$ , 连通度越小, 表示社团内部连接越紧密, 社团之间连接越稀疏, 说明社团划分结果越好。

## 4 社团发现的应用场景

社团发现技术在商业、公共安全、生物医学等领域有着非常广泛的应用, 研究者们将社团发现技术应用于各个领域的网络中来发现社团结构以解决不同领域的实际问题。

1) 商业 在商业领域, 可以在商品推荐以及股票市场等

方面利用社团发现技术为人们服务。侯强等人<sup>[69]</sup>基于中心节点对用户加权网络进行重叠社团划分, 在划分后的社团内部采用加权相似度与 top- $n$  算法相结合进行个性化推荐。可以更加精准地为用户推荐合适的商品, 提高用户体验和商品交易量。Wu 等人<sup>[70]</sup>为了研究上证指数成份股的群落结构, 建立了以股票为顶点, 股票价格对数收益率的相关系数为边的股票相关网络。利用 GN 算法揭示社团结构, 从而找到股票市场的一些产业特征, 有助于全面了解股价波动, 对股票预测和财务监督具有重要的参考价值。

2) 公共安全 在公共安全领域, 社团发现技术可以用在网络安全、舆情管控、反恐等方面。Wang 等人<sup>[71]</sup>提出了一种基于链接聚类的社团检测方法用于社交网络中来抑制计算机蠕虫的传播。丁晟春等人<sup>[72]</sup>将社团检测技术应用于微博热点主题识别研究中, 从海量的网络数据中准确高效地发现当前热点, 可以帮助政府和企业监控和管理舆情。Waskiewicz 等人<sup>[73]</sup>利用社团检测技术很好地分析了社交网络网站上的恐怖团体活动, 可以找到现有恐怖分子团体的领导者和成员以及潜在的受影响者。

3) 生物医学 在生物医学领域, 可以将社团发现技术应用在人类基因组网络、蛋白质网络等生物网络中。利用社团发现算法可以确定生物网络的社团结构, 这有助于分析拓扑结构和预测生物网络的行为, 也可以帮助患者分类以及疾病的识别和诊断。Becker 等人<sup>[74]</sup>将重叠社团检测技术应用于蛋白质相互作用网络中, 在社团的重叠处找到多功能蛋白质。Harenberg 等人<sup>[75]</sup>将社团发现应用于人类功能关联网络中进行疾病关联分析。Liu 等人<sup>[76]</sup>基于主成分分析提出了一种新颖的社团发现方法, 并将其成功地应用于疾病-基因网络中, 找到了很多具有生物学意义的社团结构, 对于疾病的预防和医学诊断具有重要的价值。Esraa 等人<sup>[65]</sup>通过在构建的 fMRI 多层时间网络上进行社团发现, 可以识别和跟踪脑网络中的社团结构。

## 5 社团发现的研究展望

社团发现的研究方法随着网络数据的复杂化而逐渐变化, 从网络中提取到更多、更深层次的有价值信息来指导社团划分一直是社团发现的发展趋势。现阶段, 在社团发现方面已经做了大量的工作, 出了很多成果, 但还有一些问题值得进一步的深入探索和研究。

a) 随着网络规模的发展, 对算法的计算效率提出了更高的要求。现有的网络规模已经发展到了亿级, 在计算资源有限的情况下, 很多静态网络社团发现算法已经无法处理如此大规模的网络。虽然现有的一些算法已经接近线性复杂度, 但是由于在划分过程中为了加快速度并节省存储空间, 使得能够利用的网络信息有限, 导致社团划分结果质量往往不理想。如何在有限的计算资源的条件下提高大规模网络中社团发现的质量是一个亟待解决的问题。

b) 神经网络由于其强大的提取特征的能力, 近两年被广泛地用于处理图数据。特别是随着图神经网络的兴起, 可以很好地用于解决社团发现等问题。目前很多基于神经网络的社团发现算法都是先将网络表示为向量, 然后利用向量聚类算法进行社团划分, 这样得到的向量是一个一般性的网络表示, 而没有很好地针对社团发现这个具体任务来进行网络表示学习。如何根据社团发现这个具体的任务来更好地优化网络的向量表示也值得进一步的探索。

c) 现有的社团发现算法主要针对有向、无向、带权重图等简单网络, 随着社交网络的发展, 出现了节点带有属性内容的属性图如在微博社交网络中, 除了网络拓扑结构外, 每个用户还包含年龄、性别、身高等属性信息以及文本、标签信息。如果在社团划分的过程中能够把这些外部信息也加以利用, 将会使

得社团划分结果更加的精准和具有现实意义。如何融合网络的拓扑信息和节点的属性等外部信息来共同指导社团的划分也是一个很有意义的研究方向。

d)随着智能终端和移动网络的普及,使得社交网络等许多真实网络随着时间的推移在不断地发生变化。这对动态网络社团发现和社团演化的实时性和准确性提出了更高的要求,如何建立高效的模型值得进一步的探究。另外,现有的一些方法是将单层的动态网络转换为多层网络进行社团挖掘,但是针对多层动态网络的社团发现方法还很欠缺。

## 6 结束语

近些年来,随着社交网络、物联网的发展,从复杂网络中挖掘出有效的社团结构对研究网络具有非常重大的意义。社团发现引起了国内外众多学者的广泛关注和研究,并取得了一系列重要研究成果。本文对复杂网络中的社团发现方法进行了分析、归纳和总结。目前,针对传统的静态网络已经有了非常深入的研究,但在网络的大规模、网络的多样性以及网络的动态性等方面还需要进一步的探索和研究。

### 参考文献:

- [1] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2):026113.
- [2] Newman M E J. Fast algorithm for detecting community structure in networks[J]. *Physical Review E*, 2003, 69(6):066133.
- [3] Clauset A, Newman M E, Moore C, et al. Finding community structure in very large networks[J]. *Physical Review E*, 2004, 70(6):066111.
- [4] Schuetz P, Cafilisch A. Multistep greedy algorithm identifies community structure in real-world and computer-generated networks[J]. *Physical Review E*, 2008, 78(2):026112.
- [5] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 2008(10):P10008.
- [6] Waltman L, Van Eck N J. A smart local moving algorithm for large-scale modularity-based community detection[J]. *The European Physical Journal B*, 2013, 86(11):article No. 471.
- [7] Rotta R, Noack A. Multilevel local search algorithms for modularity clustering[J]. *Journal of Experimental Algorithmics*, 2011, 16: article No. 2.3.
- [8] Tang Lei, Wang Xufei, Liu Huan. Uncovering groups via heterogeneous interaction analysis[C]//Proc of Ninth IEEE International Conference on Data Mining. Piscataway, NJ:IEEE Press, 2009:503-512.
- [9] Mucha P J, Richardson T, Macon K, et al. Community structure in time-dependent, multiscale, and multiplex networks[J]. *Science*, 2010, 328(5980):876-878.
- [10] Pramanik S, Tackx R, Navelkar A, et al. Discovering community structure in multilayer networks[C]//Proc of IEEE International Conference on Data Science and Advanced Analytics. Piscataway, NJ: IEEE Press, 2017:611-620.
- [11] Fortunato S, Barthélemy M. Resolution limit in community detection[J]. *Proceedings of the National Academy of Sciences*, 2007, 104(1):36-41.
- [12] Arenas A, Fernández A, Gómez S, et al. Analysis of the structure of complex networks at different resolution levels[J]. *New Journal of Physics*, 2008, 10(5):053039.
- [13] 张聪,沈惠璋,李峰,等.复杂网络中社团结构发现的多分辨率密度模块度[J]. *物理学报*, 2012, 61(14):506-514. (Zhang Cong, Shen Huizhang, Li Feng, et al. Multi-resolution density modularity of community structure discovery in complex network[J]. *Acta Physica Sinica*, 2012, 61(14):506-514.)
- [14] Ma Xiaoke, Dong Di, Wang Quan. Community detection in multi-layer networks using joint nonnegative matrix factorization[J]. *IEEE Trans on Knowledge and Data Engineering*, 2019, 31(2):273-286.
- [15] Raghavan U N, Albert R, Kumara S R, et al. Near linear time algorithm to detect community structures in large-scale networks[J]. *Physical Review E*, 2007, 76(3):036106.
- [16] Zhu Xiaojin, Ghahramani Z. Learning from labels and unlabeled data with label propagation, CMU-CALD-02-107[R]. Pittsburgh, PA: Carnegie Mellon University, 2002.
- [17] Barber M J, Clark J W. Detecting network communities by propagating labels under constraints[J]. *Physical Review E*, 2009, 80(2):026129.
- [18] Liu Xin, Murata T. Advanced modularity-specialized label propagation algorithm for detecting communities in networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2010, 389(7):1493-1500.
- [19] Zhao Yuxin, Li Shenghong, Chen Xiuzhen. Community detection using label propagation in entropic order[C]//Proc of the 12th International Conference on Computer and Information Technology. Piscataway, NJ:IEEE Press, 2012:18-24.
- [20] 刘世超,朱福喜,甘琳.基于标签传播概率的重叠社区发现算法[J]. *计算机学报*, 2016, 39(4):717-729. (Liu Shichao, Zhu Fuxi, Gan Lin. A label-propagation-probability-based algorithm for overlapping community detection[J]. *Chinese Journal of Computers*, 2016, 39(4):717-729.)
- [21] Kato T, Kashima H, Sugiyama M, et al. Robust label propagation on multiple networks[J]. *IEEE Trans on Neural Networks*, 2009, 20(1):35-44.
- [22] Alimadadi F, Khadangi E, Bagheri A, et al. Community detection in facebook activity networks and presenting a new multilayer label propagation algorithm for community detection[J]. *International Journal of Modern Physics B*, 2019, 33(10):1950089.
- [23] 史艳翠,王娜,赵青,等.基于局部扩展的社区发现研究现状[J]. *通信学报*, 2019, 40(1):149-162. (Shi Cuiyan, Wang Yuan, Zhao Qing, et al. Research status of community detection based on local expansion[J]. *Journal on Communications*, 2019, 40(1):149-162.)
- [24] Baumes J, Goldberg M K, Krishnamoorthy M S, et al. Finding communities by clustering a graph into overlapping subgraphs[J]. *IADIS AC*, 2005, 5:97-104.
- [25] Shang Changxing, Feng Shengzhong, Zhao Zhongying, et al. Efficiently detecting overlapping communities using seeding and semi-supervised learning[J]. *International Journal of Machine Learning and Cybernetics*, 2017, 8(2):455-468.
- [26] Lancichinetti A, Fortunato S, Kertesz J, et al. Detecting the overlapping and hierarchical community structure in complex networks[J]. *New Journal of Physics*, 2009, 11(3):19-44.
- [27] Luo Wenjian, Lu Nnanan, Ni Li, et al. Local community detection by the nearest nodes with greater centrality[J]. *Information Sciences*, 2020, 517:377-392.
- [28] Liakos P, Ntoulas A, Delis A. Coeus: community detection via seed-set expansion on graph streams[C]//Proc of International Conference on Big Data (Big Data). Piscataway, NJ:IEEE Press, 2017:676-685.
- [29] Yin Hao, Benson A R, Leskovec J, et al. Local higher-order graph clustering[C]//Proc of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017:555-564.
- [30] Tasgin M, Bingol H O. Community detection using boundary nodes in complex networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2019, 513:315-324.
- [31] Guo Kun, He Ling, Chen Yuzhong, et al. A local community detection algorithm based on internal force between nodes[J]. *Applied Intelligence*, 2019, 50:328-340.
- [32] 李慧,杨青泉,王慧慧.一种基于局部扩展优化的重叠社区发现算法[J]. *计算机工程与科学*, 2018, 40(12):2258-2264. (Li Hui, Yang Qingquan, Wang Huihui. An overlapping community detection algorithm based on local expansion optimization[J]. *Computer Engineering & Science*, 2018, 40(12):2258-2264.)
- [33] Interdonato R, Tagarelli A, Ienco D, et al. Local community detection in multilayer networks[J]. *Data Mining and Knowledge Discovery*, 2017, 31(5):1444-1479.
- [34] Li Xiaoming, Xu Guangquan, Lian Wenjuan, et al. Multi-Layer Network Local Community Detection Based on Influence Relation[J]. *IEEE Access*, 2019, 7:89051-89062.
- [35] Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth[J]. *Knowledge & Information Systems*, 2012, 42(1):181-213.

- [36] Yun Seyoung, Proutiere A. Streaming, memory limited algorithms for community detection[C]//Advances in Neural Information Processing Systems,2014;3167-3175.
- [37] Nguyen N P, Dinh T N, Shen Y, *et al.* Dynamic social community detection and its applications[J]. *PLoS One*,2014,9(4):e91431.
- [38] Hollocou A, Maudet J, Bonald T, *et al.* A linear streaming algorithm for community detection in very large networks[EB/OL]. (2017-05-08). <https://arxiv.org/abs/1703.02955>.
- [39] Tian Fei, Gao Bin, Cui Qing, *et al.* Deep learning for learning graph representations[M]//Pedrycz W, Chen S M. Deep Learning: Concepts and Architectures. Cham;Springer,2020;169-210.
- [40] MacQueen J. Some methods for classification and analysis of multivariate observations[C]//Proc of the 5th Berkeley Symposium on Mathematical Statistics and Probability,1967;281-297.
- [41] Yang Liang, Cao Xiaochun, He Dongxiao, *et al.* Modularity based community detection with deep learning[C]//Proc of the 25th International Joint Conference on Artificial Intelligence,2016;2252-2258.
- [42] Perozzi B, Alrfou R, Skiena S. Deepwalk;Online learning of social representations[C]//Proc of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2014;701-710.
- [43] Mikolov T, Chen K, Corrado G S, *et al.* Efficient estimation of word representations in vector space[C]//Proc of the 1st International Conference on Learning Representations,2013;1-12.
- [44] Wang Chun, Pan Shirui, Long Guodong, *et al.* MGAE: marginalized graph autoencoder for graph clustering[C]//Proc of ACM on Conference on Information and Knowledge Management,2017;889-898.
- [45] Sun Fanyun, Qu Meng, Hoffmann J, *et al.* VGraph: a generative model for joint community detection and node representation learning[C]//Advances in Neural Information Processing Systems,2019;512-522.
- [46] Cavallari S, Zheng V W, Cai Hongyun, *et al.* Learning community embedding with community detection and node embedding on graphs[C]//Proc of ACM on Conference on Information and Knowledge Management,2017;377-386.
- [47] Shchur O, Günnemann S. Overlapping community detection with graph neural networks[EB/OL]. (2019-09-26). <https://arxiv.org/abs/1909.12201>.
- [48] Jia Yuting, Zhang Qinjin, Zhang Weinan, *et al.* CommunityGAN: community detection with generative adversarial nets[C]//Proc of The World Wide Web Conference,2019;784-794.
- [49] Ghorbani M, Baghshah M S, Rabiee H R. Multi-layered graph embedding with graph convolutional networks[EB/OL]. (2018-11-21). <https://arxiv.org/abs/1811.08800v2>.
- [50] Ni Jingchao, Chang Shiyu, Liu Xiao, *et al.* Co-regularized deep multi-network embedding[C]//Proc of World Wide Web Conference,2018;469-478.
- [51] Luo Dongsheng, Ni Jingchao, Wang Suhang, *et al.* Deep multi-graph clustering via attentive cross-graph association[C]//Proc of the 13th International Conference on Web Search and Data Mining,2020;393-401.
- [52] Chakrabarti D, Kumar R, Tomkins A. Evolutionary clustering[C]//Proc of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining,2006;554-560.
- [53] Lin Yuru, Chi Yun, Zhu Shenghuo, *et al.* Analyzing communities and their evolutions in dynamic social networks[J]. *ACM Trans on Knowledge Discovery from Data*,2009,3(2):article No. 8.
- [54] Chi Yun, Song Xiaodan, Zhou Dengyong, *et al.* Evolutionary spectral clustering by incorporating temporal smoothness[C]//Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2007;153-162.
- [55] Dinh T N, Nguyen N P, Thai M T. An adaptive approximation algorithm for community detection in dynamic scale-free networks[C]//Proc of International Conference on Computer Communications,2013;55-59.
- [56] Niu Xinzheng, Si Weiyu, Wu C Q. A label-based evolutionary computing approach to dynamic community detection[J]. *Computer Communications*,2017,108;110-122.
- [57] Messaoudi I, Kamel N. A multi-objective bat algorithm for community detection on dynamic social networks[J]. *Applied Intelligence*,2019,49(6):2119-2136.
- [58] 蒋盛益,杨博泓,王连喜.一种基于增量式谱聚类的动态社区自适应发现算法[J]. *自动化学报*,2015,41(12):2017-2025.( Jiang Shengyi, Yang Bohong, Wang Lianxi. An adaptive dynamic community detection algorithm based on incremental spectral clustering[J]. *Acta Automatica Sinica*,2015,41(12):2017-2025.)
- [59] Zakrzewska A, Bader D A. A dynamic algorithm for local community detection in graphs[C]//Advances in Social Networks Analysis and Mining,2015;559-564.
- [60] Guo Kun, He Ling, Huang Jiangsheng, *et al.* A local dynamic community detection algorithm based on node contribution[C]//Proc of Conference on Computer Supported Cooperative Work,2019;363-376.
- [61] Sarryüce A E, Gedik B, Jacques-Silva G, *et al.* SONIC: streaming overlapping community detection[J]. *Data mining and knowledge discovery*,2016,30(4):819-847.
- [62] Wu Zhenyu, Chen Jiaying, Zhang Yinuo, *et al.* An incremental community detection method in social big data[C]//Proc of International Conference on Big Data,2018;136-141.
- [63] Hu Yanmei, Yang Bo, Lyu Chenyang, *et al.* A local dynamic method for tracking communities and their evolution in dynamic networks[J]. *Knowledge Based Systems*,2016,110:176-190.
- [64] Javadi S H S, Gharani P, Khadivi S. Detecting community structure in dynamic social networks using the concept of leadership[M]//Amini M, Boroojeni K, Iyengar S, *et al.* Sustainable Interdependent Networks. Cham;Springer,2018;97-118.
- [65] Al-Sharoua E, Al-Khassawneh M, Aviyente S. Tensor based temporal and multilayer community detection for studying brain dynamics during resting state fmri[J]. *IEEE Trans on Biomedical Engineering*,2019,66(3):695-709.
- [66] Prat-Pérez A, Dominguez-Sal D, Larriba-Pey J L. High quality, scalable and parallel community detection for large real graphs[C]//Proc of the 23rd International Conference on World Wide Web,2014;225-236.
- [67] Shi Jianbo, Malik J. Normalized cuts and image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,2000,22(8):888-905.
- [68] Nicosia V, Mangioni G, Carchiolo V, *et al.* Extending the definition of modularity to directed graphs with overlapping communities[J]. *Journal of Statistical Mechanics: Theory and Experiment*,2009,2009(3):P03024.
- [69] 侯强,孙瑜.基于中心节点重叠社区发现的协同过滤推荐系统[J]. *沈阳工业大学学报:社会科学版*,2013,6(4):358-361.( Hou Qiang, Sun Yu. Collaborative filtering recommendation system based on central node overlapping community discovery[J]. *Journal of Shenyang University of Technology: Social Science Edition*,2013,6(4):358-361.)
- [70] Wu Sen, Tuo Mengjiao, Xiong Deying. Community structure detection of Shanghai stock market based on complex[M]//Zhang Z, Shen Z, Zhang J, *et al.* Liss. Berlin,Heidelberg;Springer,2015;1661-1666.
- [71] Wang Yibing, Fang Jie, Wu Fuhu. Application of community detection algorithm with link clustering in inhibition of social network worms[J]. *IJ Network Security*,2017,19(3):458-468.
- [72] 丁晟春,王楠,吴靓媛.基于关键词共现和社区发现的微博热点主题识别研究[J]. *现代情报*,2018,38(3):10-18.( Ding Shenchun, Wang Nan, Wu Jingchanyuan. Hot topic detection of weibo based on keyword co-occurrence and community discovery[J]. *Journal of Modern Information*,2018,38(3):10-18.)
- [73] Waskiewicz T. Friend of a friend influence in terrorist social networks[EB/OL]. (2012). <https://www.hsdl.org/?view&did=744909>.
- [74] Becker E, Robisson B, Chapple C E, *et al.* Multifunctional proteins revealed by overlapping clustering in protein interaction network[J]. *Bioinformatics*,2012,28(1):84-90.
- [75] Harenberg S, Seay R G, Ranshous S, *et al.* Memory-efficient query-driven community detection with application to complex disease associations[C]//Proc of SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics,2014;1010-1018.
- [76] Liu Wei, Chen Ling. Community detection in disease-gene network based on principal component analysis[J]. *Tsinghua Science & Technology*,2013,18(5):454-461.