

## 复杂网络社团发现综述

潘雨, 王帅辉, 张磊, 胡谷雨, 邹军华, 王田丰, 潘志松

### 引用本文

潘雨, 王帅辉, 张磊, 胡谷雨, 邹军华, 王田丰, 潘志松. [复杂网络社团发现综述](#)[J]. 计算机科学, 2022, 49(11A): 210800144-11.

PAN Yu, WANG Shuai-hui, ZHANG Lei, HU Gu-yu, ZOU Jun-hua, WANG Tian-feng, PAN Zhi-song. [Survey of Community Detection in Complex Network](#)[J]. Computer Science, 2022, 49(11A): 210800144-11.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于核心节点影响力的社区发现方法](#)

Community Discovery Method Based on Influence of Core Nodes

计算机科学, 2022, 49(11A): 211100002-7. <https://doi.org/10.11896/jsjcx.211100002>

#### [基于k-shell熵的影响力节点的排序与识别](#)

Ranking and Recognition of Influential Nodes Based on k-shell Entropy

计算机科学, 2022, 49(11A): 210800177-5. <https://doi.org/10.11896/jsjcx.210800177>

#### [一种基于局部随机游走的标签传播算法](#)

Local Random Walk Based Label Propagation Algorithm

计算机科学, 2022, 49(10): 103-110. <https://doi.org/10.11896/jsjcx.220400145>

#### [一种基于节点稳定性和邻域相似性的社区发现算法](#)

Community Detection Algorithm Based on Node Stability and Neighbor Similarity

计算机科学, 2022, 49(9): 83-91. <https://doi.org/10.11896/jsjcx.220400146>

#### [基于双目叠加仿生的微换衣行人再识别](#)

Moderate Clothes-Changing Person Re-identification Based on Bionics of Binocular Summation

计算机科学, 2022, 49(8): 165-171. <https://doi.org/10.11896/jsjcx.210600140>

# 复杂网络社团发现综述

潘雨<sup>1,2</sup> 王帅辉<sup>3</sup> 张磊<sup>1</sup> 胡谷雨<sup>1</sup> 邹军华<sup>1</sup> 王田丰<sup>1</sup> 潘志松<sup>1</sup>

1 陆军工程大学指挥控制工程学院 南京 210007

2 中国人民解放军第 31436 部队 沈阳 110000

3 海军航空大学第三飞行训练基地 河北 秦皇岛 066000

(pan\_yu31@163.com)

**摘要** 在复杂网络中,社团结构是广泛存在的重要潜在结构。挖掘复杂网络中的社团结构,对探索网络潜在特性、理解网络组织结构、发现网络隐藏规律和交互模式等具有重要的理论和现实意义,是网络分析任务的关键研究内容。介绍了社团发现的背景和意义,并从静态网络社团发现和动态网络社团发现两个方面对社团发现的方法进行了总结和梳理。其中,静态网络的社团发现包括基于划分的社团发现方法、基于层次聚类的社团发现方法、基于模块度的社团发现方法、基于非负矩阵分解的社团发现方法和基于深度学习的社团发现方法。动态网络社团发现包括增量聚类的社团发现方法和演化聚类的社团发现方法。另外介绍了常用的社团发现评价指标,并在最后讨论了社团发现所面临的一些挑战及未来的发展方向。

**关键词:** 复杂网络; 社团结构; 社团发现; 动态网络

**中图法分类号** TP393

## Survey of Community Detection in Complex Network

PAN Yu<sup>1,2</sup>, WANG Shuai-hui<sup>3</sup>, ZHANG Lei<sup>1</sup>, HU Gu-yu<sup>1</sup>, ZOU Jun-hua<sup>1</sup>, WANG Tian-feng<sup>1</sup> and PAN Zhi-song<sup>1</sup>

1 College of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China

2 The 31436 Unit of the Chinese People's Liberation Army, Shenyang 110000, China

3 The Third Flight Training Base of Naval Aeronautical University of PLA, Qinhuangdao, Hebei 066000, China

**Abstract** Community structure is an important potential feature that exists widely in complex networks. As a key task of network analysis, mining the community structure has important theoretical and practical significance for exploring the potential characteristics, understanding the network organization structure, and discovering the hidden rules and interaction pattern. This paper introduces the background and significance of community detection, and summarizes and combs the methods of community detection from two aspects: static network community detection and dynamic network community detection. Among them, the community detection methods of static network include community detection based on division, community detection based on hierarchical clustering, community detection based on modularity, community detection based on non-negative matrix factorization and community detection based on deep learning. Dynamic network community detection methods include incremental clustering community detection and evolutionary clustering community detection. This paper also introduces the commonly used evaluation metrics of community detection. Finally, some challenges faced by community detection and the future development direction are discussed.

**Keywords** Complex network, Community structure, Community detection, Dynamic network

## 1 引言

在现实世界中,许多个人或组织之间的复杂关系都以复杂网络的形式存在,如社交网络、引文网络、通信网络和蛋白质网络等。其中,网络中的节点代表个体(实体),边代表个体之间的关系。复杂网络科学作为一门跨越社会和计算机科学的现代学科,逐渐成为科学研究的引擎和跨学科的活跃话题。在复杂网络中,普遍存在 3 个重要的统计特性,分别为小世界特性<sup>[1]</sup>、无标度特性<sup>[2]</sup>和强社团结构特性<sup>[3]</sup>。如图 1 所示,小世界特性反映了复杂网络短平均路径和高聚类系数的特征;

无标度特性反映了网络中节点服从幂律分布的特征;强社团结构特性反映了网络呈现出明显的社团结构特征,即社团内节点比社团之间节点交互更为紧密。

社团结构作为复杂网络中重要的潜在结构,广泛存在于现实网络。例如,在社交网络中,用户之间可以相互关注和聊天,还可以对其他用户发布的内容进行转发和评论。相互关注并且频繁互动的用户构成一个小“团体”,“团体”内的用户往往交流和互动频繁,并且关注相同主题,拥有相同兴趣和爱好。这一个个小“团体”就是网络中最广泛且重要的结构特征——社团结构。

基金项目:国家自然科学基金面上项目(62076251)

This work was supported by the National Natural Science Foundation of China(62076251).

通信作者:潘志松(zhisong\_pan@163.com)

网络分析任务为探索复杂网络的这些潜在特性提供了有效的工具。其中,挖掘复杂网络中的社团结构作为网络分析的重要研究内容之一,对揭示网络内部结构、理解复杂网络潜在特性、发现网络隐藏规律等具有重要意义<sup>[4]</sup>。例如,在微博等社交网络中,给拥有共同好友的用户推荐好友,有助于提高用户的忠诚度,对拥有相同兴趣爱好的用户推送感兴趣的内容会提高用户的依赖度<sup>[5]</sup>;在引文网络中,对引文网络进行社团挖掘可以发现相同研究领域的关键性文章,以及历年来研究热点和学科建设的演变过程,有助于预测未来研究方向和前沿学科<sup>[6]</sup>;在舆论网络中,对网络热点话题进行挖掘和分析有助于舆情控制和舆论导向,从而起到净化网络环境的作用<sup>[7]</sup>;在犯罪网络中,对嫌疑人进行社团发现可以侦查潜在的犯罪团伙或恐怖组织,对国家安全具有重要意义<sup>[8]</sup>;在疾病蛋白质网络中,挖掘社团结构有助于找到药物靶,对疾病的治疗起到关键的作用<sup>[9]</sup>;在云计算中,通过分析服务间的流量挖掘虚拟网元之间的社团结构,可以为数据中心的微服务部署和调度提供指导意见,进一步优化运营商的效率并减少运营代价<sup>[10]</sup>;在IP网络中,对IP网络的社团结构进行挖掘和分析有利于理解网络流量,从而为网络优化和安全管理提供有用的信息和决策支持<sup>[11]</sup>;在通话网络中,对用户类别进行识别有助于分析用户通信模式和特征,从而提供定向的推荐服务和安全布控<sup>[12]</sup>。

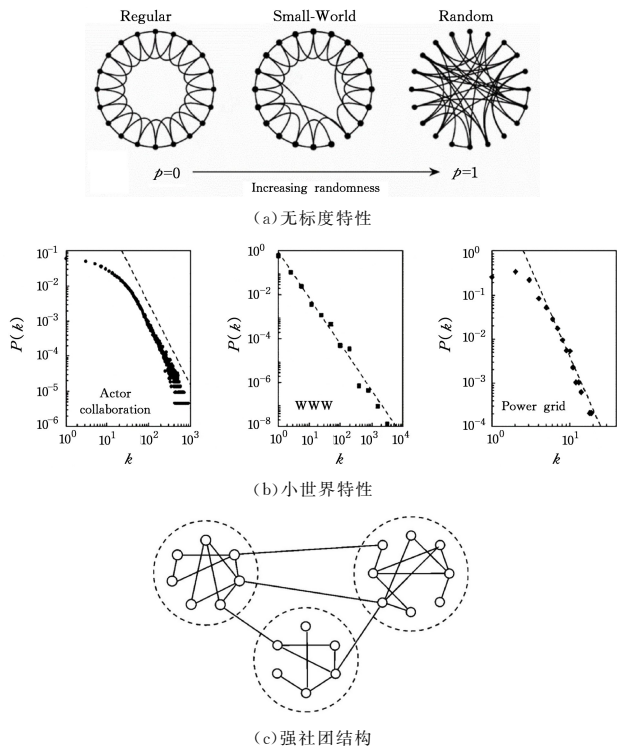


图1 复杂网络的统计特性

Fig. 1 Statistical properties of complex networks

## 2 社团发现

社团结构是普遍存在于复杂网络的重要介观结构。对于社团结构,至今没有一个的定义,但已有很多研究尝试将社团的定义形式化。Fortunato<sup>[13]</sup>将社团从3个层次进行定义:局部定义、全局定义和基于节点相似性的定义。之后 Fortunato 又进一步将社团定义为:可能共享相同属性或在网络中扮演类似角色的节点集合,并将其称为“群”。Porter<sup>[14]</sup>通过回顾

社会学和人类学领域中关于社团研究的起源,将社团定义为紧密相连的节点组,组内节点间的连接比不同组节点间的连接更紧密。Yang 等<sup>[15]</sup>也提出了被广泛认可的社团定义:社团是一组网络的节点,社团内节点之间连接紧密,不同社团节点之间连接稀疏。虽然至今对社团仍然没有一个确定的概念和定义,但是研究员对社团结构普遍拥有一个统一的共识,如图2所示。网络中的15个节点被划分为3个簇,分别用红色、黄色和紫色表示,在每个簇中节点之间边的密度远高于不同簇之间边的密度。

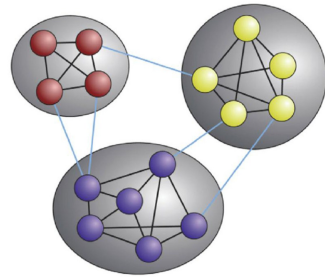


图2 社团结构示意图(电子版为彩图)

Fig. 2 Schematic diagram of community structure

社团发现(Community Discovery)又称为社团检测(Community Detection),是指通过分析网络中节点之间的相互作用和潜在信息,从介观角度挖掘网络隐藏的社团结构的过程<sup>[16]</sup>,具体可将其定义如下。

社团发现:给定一个网络  $G=(V, E)$ , 社团发现的目标是设计一个映射函数  $F$  将网络中的节点划分为  $K$  个社团,使社团内节点之间的连接紧密,而社团之间节点连接稀疏。

可见,社团发现是根据网络中节点之间的相互作用对社团结构进行识别,从而达到挖掘网络中连接紧密、具有相同性质节点集合的目的<sup>[17]</sup>。一个理想的社团结构应该是社团内部的节点之间连接紧密并且共享相同的属性或扮演类似的角色,不同社团的节点之间连接稀疏。

社团发现因为在交叉学科中具有重要的意义,近年来受到研究者的广泛研究。本文根据复杂网络的演化,从静态网络社团发现方法和动态网络社团发现方法两个方向对现有的社团发现方法进行梳理和总结,如图3所示。

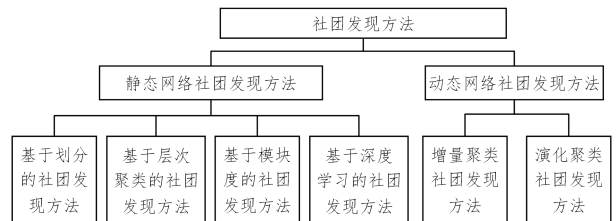


图3 社团发现方法的分类

Fig. 3 Classification of community detection methods

其中,针对静态网络的社团发现方法,研究者们陆续提出了基于划分的社团发现方法、基于层次聚类的社团发现方法和基于非负矩阵分解的社团发现方法等。现实中的复杂网络是随时间发生演化的。网络中节点或节点之间边的动态变化通常会带来网络中社团结构的动态变化,从而导致网络的整体功能、物理特性等宏观特征随时间发生演化。因此,近年来针对动态网络的社团发现研究也吸引了越来越多的关注和研究。

### 3 社团发现的方法

社团发现与计算机科学中的图分割(Graph Partition)和社会学中的层次聚类(Hierarchical Clustering)存在密切的联系。进入 21 世纪之后,随着复杂网络研究的深入,社团发现问题得到了更多关注。随着社团发现在这两个领域研究的深入以及模块度概念的提出,社团发现方法呈现井喷式的发展,从此拉开了社团发现研究蓬勃发展的序幕。

#### 3.1 静态网络的社团发现方法

##### 3.1.1 基于划分的社团发现方法

基于划分聚类的社团发现算法是根据节点之间的距离最大化或最小化一个损失函数,将网络节点分割成  $K$  个簇。 $K$ -means 聚类是当前最流行的划分聚类方法<sup>[18]</sup>,在应用于社团发现问题时,它最小化社团内节点到社团中心节点距离的平方损失函数:

$$E = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (1)$$

其中,  $C_i$  为第  $i$  个社团的节点的集合,  $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$  为第  $i$  个社团的中心节点。 $K$ -means 聚类算法的流程如下:

- (1) 随机或者根据经验选择  $K$  个质心;
- (2) 分别计算所有节点与质心的距离,然后根据计算出来的距离将每个节点分配到最近的质心,形成  $K$  个簇;
- (3) 再对分配后的簇重新计算质心的位置;
- (4) 重复以上的步骤,直到计算的质心不再发生变化。

常用的损失函数还有如下几种。

最小  $K$ -聚类 Minimum  $K$ -clustering: 最小  $K$ -聚类的代价函数为社团的直径,即社团的两点之间的最大距离。

和  $K$ -聚类  $K$ -clustering sum: 与最小  $K$ -聚类相似,和  $K$ -聚类在代价函数中用社团节点对之间的平均距离。

中心  $K$ -聚类  $K$ -center: 该方法为每个社团定义一个质心,计算每个节点到质心的最大距离(直径)作为代价函数。选择簇和质心,使直径最大值最小。

中值- $K$  聚类  $K$ -median: 与  $K$ -center 类似,只不过计算的是社团内每个点到中心点的均值。

##### 3.1.2 基于层次聚类的社团发现方法

社团层次普遍存在于网络,即每个社团可以是不同层次的小集群的集合<sup>[19]</sup>。根据网络图的层次化特点,层次聚类技术可用于识别网络的多级社团结构。层次聚类算法是基于节点相似性进行迭代聚类,不需要预定义社团的大小和数量。基于层次聚类的社团发现方法主要分为两类:自底向上的凝聚式层次聚类方法和自顶向下的分裂式层次聚类方法。

自底向上的凝聚式层次聚类方法是将网络中的每个节点都视为一个社团,然后通过相似度高的社团进行迭代组合获得最终的社团结构<sup>[8,20-30]</sup>。Newman 提出的 Fast Newman (FN)算法<sup>[20]</sup>首先计算任意两个社团之间合并后的模块度增量,然后将模块度增量最大的两个社团进行合并,直至获得网络的社团结构。标记传播算法(Label Propagation Algorithm, LPA)<sup>[21]</sup>在初始时为每个节点分配一个唯一标识,不断将多数邻居的标识更新为自己的标识,直至节点标识分布稳定,具有统一标识的节点构成一个社团,如图 4 所示。Zhang 等<sup>[22]</sup>提出了一种基于真实连接概念的凝聚式方法来发现网络中的重叠社团,该方法通过对原始网络进行预处理

得到“真实连接”图,然后通过不停地重复迭代,将相似度高的社团进行合并,从而获得最终的网络社团结构。Bahulkar 等<sup>[8]</sup>提出了一种凝聚式层次聚类方法用于发现犯罪网络中的社团结构,该方法首先捕捉网络中隐藏的边并在划分社团前将其添加到网络中,然后采用自底而上的搜索方法,通过优化社团的局部模块度来发现社团结构。Riedy 等<sup>[23]</sup>首先将网络中的每个节点看作一个社团,然后根据两个相邻社团合并后优化指标的增量对其进行评价,最后通过迭代将社团进行合并直到模块度目标被最大化。Blondel 等<sup>[24]</sup>提出的凝聚式层次聚类算法 Louvain 首先采用局部搜索来选择小社团,然后对社团进行不断的聚和,直到模块度停止增加。Shang 等<sup>[25]</sup>提出了一种基于增量模块度的动态网络社团发现算法,算法首先采用静态社团发现方法初始化得到初始社团,然后执行增量更新策略来发现动态的社团结构。

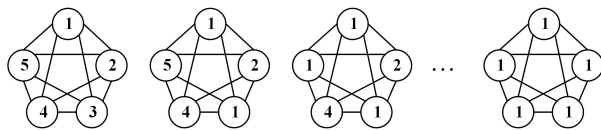


图 4 LPA 算法示意图

Fig. 4 Schematic diagram of LPA algorithm

在基于层次聚类的社团发现方法中,另一种为自顶向下的分裂式方法<sup>[31-33]</sup>。该类方法将整个网络看作一个社团,然后计算网络中节点之间的相似度,通过不断迭代删除网络中相似度最“弱”的连接,得到最终的社团结构。分裂式方法的典型代表 GN 算法<sup>[3]</sup>通过迭代删除边数最大的边,获得社团划分的结果。Ni 等<sup>[31]</sup>将网络的社团发现过程视为几何分解,并采用离散里奇流(Ricci Flow)的基本原理,通过迭代地在里奇流过程中移动边来确定网络的社团结构。Gorke 等<sup>[32]</sup>提出了一个分裂式分层聚类算法来发现时序网络的社团结构,该方法首先将网络中的所有节点看作一个社团,然后将合并后模块度增长最大的两个社团进行合并,重复此过程,直到获得最终的社团结构。

基于层次聚类社团发现算法的优点是能揭示网络的层次结构并直观地显示出来,可直观地对网络结构进行分析和研究。但是每个步骤完成都不能撤销,无法对次优或者错误的划分结果进行调整。

##### 3.1.3 基于模块度的社团发现方法

随着模块度的提出,基于模块度优化的社团发现算法也逐渐成为最流行的社团发现方法之一<sup>[20,33-45]</sup>。其思想是将社团发现问题转化为优化问题,从而得到社团结构。其中,模块度  $Q$  是目前最常使用的优化目标。模块度函数  $Q$  由 Newman 和 Girvan 提出,通过对比随机网络中连接密度和实际情况下社团中连接密度之间的差异来评价社团划分的质量<sup>[20]</sup>。模块度的值越大,代表社团划分的结果越好。模块度  $Q$  的形式化定义如下:

$$Q = \frac{1}{2m} \sum_{ij} (a_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j) \quad (2)$$

其中,  $A$  为网络的邻接矩阵,  $m$  为网络中的总边数,  $k_i$  为节点  $v_i$  的度,即  $k_i = \sum a_{ij}$ ,  $k_i k_j / 2m$  代表在随机图中节点  $v_i$  和  $v_j$  之间有边的概率,  $C_i (C_j)$  分别为节点  $v_i (v_j)$  所属的社团。  $\delta(\cdot)$  为克罗内克函数,当节点  $v_i$  和  $v_j$  属于同一个社团时,  $\delta(C_i, C_j)$  为 1,否则为 0。求解模块度最大化问题被证明是 NP 问题,

已经有很多工作通过贪心算法<sup>[20]</sup>、模拟退火<sup>[34]</sup>等启发式算法进行求解。Zhe 等<sup>[35]</sup>提出了一种针对大规模网络的社团发现算法 AGGMMR,其采用基于贪婪的最大化模块度算法对网络拓扑和属性信息图进行划分,从而获得网络的社团结构。Combo 算法是一种可基于多种优化目标的社团发现算法,其不仅适用于优化模块度函数,还可以应用于优化代码长度等社团发现目标。

更进一步,对于符号网络、重叠网络和多层网络,也有对应的模块度函数相继被提出。为了评价符号网络中的社团发现质量,Gómez 等<sup>[37]</sup>提出了可用于符号网络的模块度指标:

$$Q_s = \frac{1}{2\omega^+ + 2\omega^-} \sum_{i,j \in V} \left( \omega_{ij} + \frac{\omega_i^- \omega_j^-}{2\omega^-} - \frac{\omega_i^+ \omega_j^+}{2\omega^+} \right) \delta(C_i, C_j) \quad (3)$$

其中,  $\omega^+$  和  $\omega^-$  分别表示网络中正边和负边的数量,  $\omega_i^+$  和  $\omega_i^-$  分别表示与节点  $v_i$  相连的正边和负边的数量。基于符号网络模块度, Anchuri 等<sup>[38]</sup>提出了一种迭代优化增强的谱方法来最大化符号网络模块度,从而获得网络的社团结构。Cai 等<sup>[39]</sup>设计了一种新的离散粒子群优化算法来实现符号网络中社团结构的准确挖掘。对于重叠社团发现, Shen 等<sup>[40]</sup>对传统的模块度进行了扩展,定义了重叠模块度函数 EQ 用于评价重叠社团发现的质量。

$$EQ = \frac{1}{2m} \sum_{ij} \frac{1}{O_i O_j} \left( a_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \quad (4)$$

其中,  $O_i$  和  $O_j$  是节点  $v_i$  和  $v_j$  所属社团的总数量。随后, Waltman 等<sup>[41]</sup>针对多层网络的社团发现问题提出了多层模块化指数  $Q_M$ , 并在其基础上最大化  $Q_M$ , 进而得到准确的社团结构。

基于模块度的算法得到了广泛的应用,但是由于分辨率限制的问题,基于模块度算法无法准确的识别更小的社团结构<sup>[42]</sup>。现在已经有工作尝试通过引入分辨率参数解决基于模块度算法对分辨率的限制。Zhu 等<sup>[43]</sup>提出了一种基于 k-plex 的模块度优化算法 MOKP, 该算法使用 k-plex 从整个

网络生成社团种子,然后通过模块度优化算法将剩余节点分配给合适的社团种子,可以准确识别规模较小的社团结构。

### 3.1.4 基于非负矩阵分解的社团发现方法

非负矩阵分解 (Nonnegative Matrix Factorization, NMF)<sup>[46]</sup> 因具有明确的物理意义和良好的可解释性,被广泛应用于图像处理、语音识别和文本分析等领域。近年来,基于非负矩阵分解的社团发现算法也受到了研究者的广泛关注<sup>[47-52]</sup>。

NMF 可以看作是一种数据的表示方法,即将原始数据矩阵中的数据用几个基向量的线性相加组合来表示。给定一个矩阵  $V \in R^{m \times n}$ , NMF 旨在将矩阵  $V$  分解为两个非负矩阵  $W \in R^{m \times k}$  和  $H \in R^{n \times k}$  的乘积,使两个非负矩阵的乘积尽可能地逼近原来的数据矩阵  $V$ , 即  $V \approx WH^T$ 。形式化地, NMF 模型可表示为如下的优化问题:

$$\min D(V, WH^T) \quad \text{s. t. } W \geq 0, H \geq 0 \quad (5)$$

其中,  $\min D(V, WH^T)$  表示矩阵  $V$  与  $WH^T$  之间的误差损失函数。常用的误差损失函数有以下两种。

(1) 基于平方损失函数:

$$L_{LSE} = D(V, WH^T) = \|V - WH^T\|_F^2 \quad \text{s. t. } W \geq 0, H \geq 0 \quad (6)$$

(2) 基于广义 Kullback-Leibler 散度函数:

$$L_{KL} = D(V, WH^T) = KL(V \| WH^T) \quad \text{s. t. } W \geq 0, H \geq 0 \quad (7)$$

在基于 NMF 的社团发现方法中,对邻接矩阵  $A$  进行非负矩阵分解,得到的非负矩阵  $W$  为基矩阵,  $H$  为系数矩阵,矩阵  $H$  的每一列可以看作节点属于每个社团的隶属度<sup>[53]</sup>。图 5 为基于 NMF 的社团发现方法的流程图。图左侧表示由 34 名成员构成的 Zachary 空手道俱乐部社交网络,在对其进行社团发现时,首先将网络表示为邻接矩阵的形式,然后对邻接矩阵进行非负矩阵分解,得到每个节点隶属于社团的隶属度,最终将节点分配到隶属度最高的社团,获得最终的社团结构。

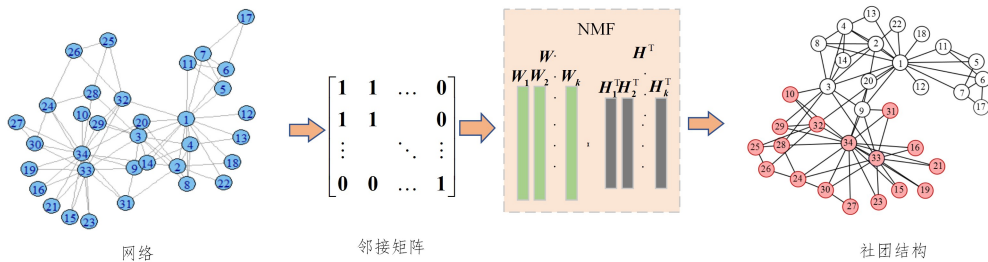


图 5 基于 NMF 的社团发现算法示意图

Fig. 5 Schematic diagram of communitydetection algorithm based on NMF

针对社团发现的不同场景和需求, NMF 的一些变体也陆续应用于社团发现问题, 如对称非负矩阵分解 (Symmetric Nonnegative Matrix Factorization, SNMF)<sup>[54]</sup> 和非负矩阵三因子分解 (Nonnegative Matrix Tri-factorization, NMTF)<sup>[55]</sup>。SNMF 被用于划分无向网络中的社团结构, 将对称的邻接矩阵  $A$  分解为两个对称的低维矩阵乘积的形式  $HH^T$ , 即  $A = HH^T$ 。由于传统的二因子分解只能捕捉两种类型的关系, Ding 等<sup>[56]</sup>将二因子的非负矩阵分解进行扩展, 提出了可以捕捉更多类型关系的三因子非负矩阵分解 NMTF, 即  $A \approx WXH^T$ 。其中, 矩阵  $X$  可以看作是社团之间的关系矩阵, 用于描述社团之间的关系强度。更进一步, 对于无向网络

的三因子分解, 可以得到  $A \approx HXH^T$ 。

近年来, 因为 NMF 算法具有可解释性和模型简单等优点, 越来越多基于 NMF 的社团发现算法被提出。Cao 等<sup>[57]</sup>提出了一种基于 NMF 的社团发现方法 CLNCCD, 算法同时考虑了网络拓扑和属性信息, 并且认为在网络中拥有相同属性的节点有较大概率属于同一个社团。首先, 基于 NMF 模型对邻接矩阵  $A$  进行分解, 得到节点的社团指示矩阵  $H$ :

$$O'(X) = \|A - HH^T\|_F^2 \quad (8)$$

矩阵  $H$  的每一行为节点隶属于每个社团的概率。然后基于内容和节点所属社团的一致性, 来实现网络连接和节点内容的结合。算法通过引入图正则化项来惩罚拥有相同属性

却被划分到不同社团的节点:

$$O(\mathbf{X}) = \|\mathbf{A} - \mathbf{H}\mathbf{H}^T\|_F^2 + \beta \text{tr}(\mathbf{H}^T \mathbf{L}\mathbf{H}) \quad (9)$$

其中,拉普拉斯矩阵  $\mathbf{L} = \mathbf{D} - \mathbf{S}$ ,  $\mathbf{S}$  为通过余弦相似度计算得到的节点属性相似度矩阵,  $\mathbf{D}$  为矩阵  $\mathbf{S}$  的对角矩阵。更进一步,考虑到网络中节点度的不均衡性,为了减少节点度的非均匀分布对图正则化的负面影响,算法引入参数矩阵  $\mathbf{W}$  对社团指示矩阵  $\mathbf{H}$  进行归一化,得到最终的目标函数:

$$O(\mathbf{X}, \mathbf{W}) = \|\mathbf{A} - (\mathbf{W}\mathbf{H}) \cdot (\mathbf{W}\mathbf{H})^T\|_F^2 + \beta \text{tr}(\mathbf{H}^T \mathbf{L}\mathbf{H}) \quad (10)$$

s. t.  $\mathbf{H} \cdot \mathbf{1}_C^T = \mathbf{1}_N^T$

其中,向量  $\mathbf{1}_C^T$  是所有元素都为 1 的  $C$  维列向量。类似地,Pei 等<sup>[58]</sup>提出了一个基于 NMTF 的社团发现框架,通过在框架中引入图正则化项来捕捉社交网络中用户之间相似性、消息之间相似性以及用户之间的交互,完美地结合连接关系和内容。Pei 等提出的方法虽然同时利用了网络拓扑和内容信息来发现社团结构,但没有考虑如何利用挖掘的语义信息来解释社团的含义。为了解决这个问题,Wang 等<sup>[59]</sup>提出了社团发现方法 SCI,算法同时集成了表示网络拓扑的社团成员矩阵以及表示语义信息的社团属性矩阵。SCI 不仅能有效地进行社团发现,还能利用语义信息对社团进行标注,使社团发现结果具有较强的解释性。利用有效的监督信息,Lu 等<sup>[60]</sup>提出了一个基于 SNMF 的半监督社团发现算法,将 NMF 框架与成对约束相结合,分别提出两种社团发现方法来处理线性可分数据和非线性可分数据。Shi 等<sup>[61]</sup>提出了一种成对约束的 SNMF 社团发现方法 PCSNMF,算法利用基于真实社团信息的成对约束来提高社团发现的性能。随后,针对多维网络的社团发现问题,Zhang 等<sup>[62]</sup>提出了一种基于 NMTF 的多维网络社团发现方法 JoNMTF,算法首先结合网络连接关系和属性内容构成统一的图表示,然后利用 NMTF 模型对统一的图表示进行分解,得到准确的社团结构。对于重叠社团结构的发掘,GNMTF<sup>[63]</sup>是一种基于图正则化的 NMTF 模型,用于发现网络中的重叠社团结构,算法不仅可以准确地获得每个节点隶属于社团的情况,还可以捕捉不同社团之间的相互作用。针对符号网络的社团发现,Li 等<sup>[64]</sup>提出了一种基于 NMTF 的符号网络社团发现算法,算法通过引入正则化项和稀疏项获得准确的符号网络社团结构,同时捕获社团之间的“积极”和“消极”关系。针对多层网络的社团发现问题,Ma 等<sup>[65]</sup>提出了一种基于联合非负矩阵分解的多层社团发现算法 MjNMF,该算法通过用公共基矩阵同时分解所有层的邻接矩阵来提取每层的节点特征,并分解所有相关层的节点相似度矩阵来获得网络的社团结构。

NMF 算法在社团发现中得到了广泛的应用,主要源于以下原因:1)NMF 模型可以发现隐含的模块结构;2)因为分解矩阵的非负特性,分解的形式和分解结构都具有明确的物理含义,并且具有非常好的可解释性;3)模型简单,是可加性模型,可扩展性强,适用于不同场景下的社团发现问题;4)NMF 分解的结果具有天然的稀疏性。

### 3.1.5 基于深度学习的社团发现方法

近年来,随着深度学习在视觉、自然语言处理等领域的发展,基于深度学习的社团发现方法也逐渐成为研究的热点。现有的基于深度的社团发现算法主要分为:基于自编码器(Autoencoder)的社团发现方法<sup>[67-75]</sup>、基于生成对抗网络

(Generative Adversarial Networks, GANs)的社团发现方法<sup>[76-81]</sup>和基于图卷积网络(Graph Convolutional Networks, GCNs)的社团发现方法<sup>[82-88]</sup>。基于自编码器的方法利用无监督自编码器,将网络编码成潜在空间中的低维表示,并在低维表示的基础上划分社团结构。生成对抗网络的社团发现方法采用对抗学习的思想,通过生成器和判别器之间的相互博弈学习来发现社团结构。

在基于自编码器的方法中,通过最小化重构误差,在隐层的输出学习到最优的编码表示,可以将其看作输入数据的低维表示。在基于自编码器的方法中有两种获得社团结构的方式:一种是在低维表示向量上运行聚类策略,获得社团结构;另一种是将社团信息整合到模型中用于直接发现社团结构。根据使用自编码器类型的不同,又可以将模型分为栈式、稀疏式、去噪式和变分式自编码器。Tian 等<sup>[66]</sup>基于自编码器和谱聚类之间的相似性,提出了一种基于稀疏自编码器的图聚类方法 SAE。算法将归一化后的图相似度矩阵输入到稀疏自编码器,并通过在目标函数上引入  $L_1$  正则化项得到稀疏的非线性节点低维表示;然后对低维的节点表示运行  $K$ -means 策略,得到最终的聚类结果。随后,Yang 等<sup>[67]</sup>证明了随机生成模型和最大化模块度模型与自编码器之间的等价性,两个模型都是通过寻找网络的低秩嵌入来重构网络拓扑,这与自编码器的目标不谋而合;然后提出了一个半监督的深度神经网络社团发现方法 DNR,该算法将模块度矩阵作为自编码器的输入生成节点表示,然后利用  $K$ -means 对低维空间的节点向量表示进行社团划分。Jin 等<sup>[68]</sup>受其启发,将保存了结构信息的模块度矩阵和保存了属性信息的马尔可夫矩阵进行串联作为自编码器的输入,得到融合两种信息的网络低维表示,然后在节点表示向量上运行聚类算法得到社团结构。Cao 等<sup>[69]</sup>提出了一种基于栈式自编码器的社团发现算法 DANEP,该算法通过结合网络拓扑和节点属性进行社团发现;进一步针对网络中拓扑结构和内容不匹配的情况,提出了一种基于自编码器的自适应社团发现方法 AAGR<sup>[70]</sup>,通过在自编码器中加入图正则化项来合并内容信息,并引入自适应参数实现网络拓扑和节点属性权重的自动调整。SCD<sup>[71]</sup>是一种基于网络节点嵌入聚类的社团发现方法,将社团发现和网络表示学习相结合,通过优化轮廓度量将社团发现问题转化为网络嵌入聚类任务。

生成对抗网络(GANs)的核心思想是零和博弈,通常由生成模块  $G$  和判别模块  $D$  组成。生成模块的作用是捕获数据分布,生成尽可能接近真实数据的样本,而判别模块则是判断一个样本是真实数据而不是由生成器生成的合成数据的概率。将 GANs 应用于社团发现的灵感来源于 GANs 通常是无监督的,并且生成的新数据具有与真实数据相同的分布。CommunityGAN<sup>[76]</sup>是一种基于生成对抗网络的社团发现和图表示的联合框架。算法首先采用隶属关系图模型(Community-Affiliation Graph Model, AGM)的思想,为每个节点社团对分配一个非负因子用于表示节点隶属于社团的概率,每个节点隶属于所有社团的概率构成了该节点的表示向量,然后通过一个专门设计的 GAN 来发现社团结构。此外,Zhang 等<sup>[77]</sup>提出了基于生成对抗性学习的启发式社团发现算法 SEAL,算法使用判别器来预测一个社团是否真实,并

使用生成器通过隐式拟合真实社团的特征来欺骗判别器,从而学习社团的结构。Chen 等<sup>[78]</sup>提出基于对抗学习的重叠社团发现算法 ACNE,算法不仅能够划分重叠的社团结构,还得到一个网络的低维表示。

图卷积网络(GCNs)<sup>[82]</sup>是图神经网络方法中学习图数据表示最具代表性的分支,其在节点监督和半监督分类上的成功吸引了大量研究者的关注。最近,一些基于 GCNs 的算法被提出用于对高维复杂网络数据进行建模和推断,从而获得网络中的社团结构<sup>[83-85]</sup>。He 等<sup>[83]</sup>设计了一种基于 GCN 的社团发现算法,利用以社团为中心的双编码器分别重构网络拓扑和节点属性,从而实现无监督的社团发现。Jin 等<sup>[84]</sup>指出 GCNs 得到的嵌入结果并不是面对社团信息的,而社团发现任务在本质上是无监督任务,并对此提出了一个基于卷积神经网络的无监督社团发现模型 JGECD,模型由网络嵌入模块、社团发现模块和网络重构模块组成。Sattar 等<sup>[85]</sup>针对基于 GCN 的半监督学习在大型图中的社团发现仍然存在可扩展性和准确性问题,提出一种基于 GCN 的半监督节点分类的可扩展社团发现方法。

### 3.2 动态网络的社团发现方法

在现实生活中,随着网络中节点和边的增加与减少,网络的拓扑结构和社团结构都会不断发生演化。例如,在学术网络中,拥有相同研究领域的学者往往构成一个社团,研究热点的变化和研究者兴趣的改变,使得社团具有复杂的动态性。网络的不断演化可能导致社团结构的巨大变化和被重新发现的需求。文献<sup>[89]</sup>中定义了 8 种社团演化行为,分别为社团的新生、消亡、新增、减少、合并、分裂、持续和重生,如图 6 所示。

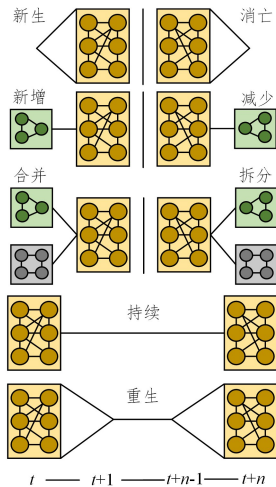


图 6 社团动态演化示意图

Fig. 6 Schematic diagram of community dynamic evolution

(1) 新生: 新社团的形成, 由任意数量的节点组成新社团的第一次出现。

(2) 消亡: 社团的消失, 属于消失社团的所有节点都变为无连接。

(3) 新增: 有新节点新增到社团中, 社团规模变大。

(4) 减少: 有一些节点离开了社团, 社团规模变小。

(5) 合并: 两个或多个社团合并为一个社团。

(6) 拆分: 由于节点或者节点之间的边消失, 社团被拆分为两个或多个社团。

(7) 持续: 社团保持不变。

(8) 重生: 一个社团消失一段时间, 然后又重生。这个过

程可以被看作一个消亡又重生的过程。重生的社团拥有和之前社团相同的节点与边。

相比于静态网络的社团发现, 动态社团发现算法的设计更具有挑战性, 主要有以下几点原因。首先, 动态网络社团发现不仅要考虑社团结构的划分是否准确, 还要充分考虑动态网络的演化过程。其次, 是解的不稳定性问题, 即无法确定划分结果的变化是由社团的自然演化引起的, 还是由于网络中噪声或算法不稳定性造成的。对此, 研究者们提出了大量的解决方案, 其最终目标是使社团的演化更加平滑。为了解决这个问题, Chakrabarti<sup>[90]</sup>提出了演化聚类框架, 该框架是目前应用最广泛的动态社团发现方法之一。其假设在短时间内聚类的突然变化可能是由噪声引起的, 并且不期望聚类发生突然变化。Palla 等<sup>[91]</sup>通过研究科学家合作网络的动态演变过程, 发现在短时间内社团不会发生剧烈的变化, 并且提出社团发生大规模演化后, 社团的生命周期将会更长。这证明了演化聚类模型的假设与现实网络的特质相吻合。

动态网络拓扑结构的不可预测和快速变化的特性为社团发现提出了严峻的挑战, 传统的静态社团发现算法已经不能满足在动态变化的网络中准确挖掘社团结构的需求。Hopcroft 等<sup>[92]</sup>最早提出了动态社团发现问题, 将动态的网络分为若干个静态图, 采用余弦相似度的方法得到每个静态图的层次聚类结果, 找到其中变动较小的部分, 称之为自然社团, 从而以自然社团为参照来分析动态网络的演化规律。近些年相继出现了一些针对动态网络的社团发现算法来解决动态网络的社团发现问题, 这些方法主要可以分为增量聚类的社团发现方法和演化聚类的社团发现方法。

#### 3.2.1 增量聚类社团发现方法

增量聚类算法<sup>[93-100]</sup>尝试将静态网络的社团发现方法扩展到动态网络中, 其主要思想是利用静态社团发现方法在第一个网络快照中发现社团结构, 然后根据连续两个时刻快照之间节点和边的动态变化来划分后续快照的社团结构, 当前时刻的社团结构是从前一时刻的社团结构中推导而来。代表算法有 IA-MCS<sup>[93]</sup>, GraphScope<sup>[94]</sup>等。增量聚类方法可以直接使用静态的社团发现算法, 比较容易实现, 能够保证短期内社团发现结果的稳定性, 并且计算复杂度低, 计算速度快, 非常适合用于具有大序列的演化网络中。然而, 基于增量聚类的方法在不考虑历史社团结构的情况下提取每个快照的社团, 而历史社团结构通常对噪声敏感。增量聚类算法忽略了和前一时刻的联系, 通常会导致不稳定的社团结构和随之而来的、毫无根据的社团进化。从长期和全局角度来看, 这种方法不能保证社团发现结果的一致性。这种方法只适用于局部变化、无噪声的动态网络。

#### 3.2.2 演化聚类社团发现方法

演化聚类算法<sup>[101-113]</sup>是目前最流行的一种动态网络社团发现方法, 最早被 Chakrabarti 等<sup>[90]</sup>提出用于流数据的聚类, 通过将其应用于  $K$ -means 和凝聚层次聚类算法以处理不断变化的数据。演化聚类框架假设短时间内的聚类结果不会发生剧烈变化, 网络的突变多是由于噪声引起的, 并引入时间损失函数对下一时刻突变的社团划分进行惩罚。Chakrabarti 等人提出了用来考查两个相互冲突的准则的概念: 快照质量 (Snapshot Cost, SC) 和时间开销 (Temporal Cost, TC)。快照质量用来衡量当前聚类结果  $C$  在当前网络结构  $G$  下的聚类

质量,而时间开销用来衡量当前时刻聚类结果  $C_t$  与前一时刻聚类结果  $C_{t-1}$  的相似性。所以,同时满足  $SC$  最大和  $TC$  最小的聚类结果被认为是当前时刻最优的聚类结果。以此为基础,提出了时间平滑(Temporal Smoothness)框架,这个框架假设社团在短时间内的突变是不可取的,每个社团在连续的时间内应该是平滑的,定义如下:

$$Cost = \alpha \cdot SC + (1 - \alpha) \cdot TC \quad (11)$$

其中, $SC$  是评测某个网络快照的聚类结果的质量; $TC$  校准当前聚类结果对于历史数据或者历史聚类结果的符合度; $\alpha$  为权重系数,用来平衡  $SC$  和  $TC$  的权重比例。Chi 等<sup>[101]</sup> 首次将演化聚类算法用于动态网络的社团发现问题,通过快照代价(Snapshot Cost, CS)来评价当前时刻社团发现的准确度,通过时间代价(Temporal Cost, CT)来度量连续两个时刻社团发现结果的相似性。

以演化聚类框架为基础, Lin 等<sup>[102]</sup> 提出了时间平滑(Temporal Smoothness)框架用于动态网络的社团发现,并提出了研究社团动态演化的统一框架 FacetNet,该算法是目前最经典且广泛使用的动态社团发现算法。FacetNet 算法采用随机块模型生成社团,并通过一个健壮的统一过程来分析社团及其演化,该过程考虑了社团演化过程和演化的时间平滑性。随后, Folino 等<sup>[103]</sup> 提出了一种基于多目标优化的演化聚类算法 DYNMOGA,算法将快照代价和时间代价看作一个多目标函数,在最大化快照代价的同时最小化时间代价。ECGNMF 算法<sup>[104]</sup> 通过 NMF 框架拟合每个时间片的社团发现情况,从而获得每个时刻的社团结构。对于时间代价,算法不仅考虑了介观社团历史信息,还考虑了节点的微观变化信息,通过引入正则化项对两个连续时间的微观节点变化进行约束,平滑连续时间内微观节点的变化情况。Ma 等<sup>[106]</sup> 提出了一种基于共同正则化非负矩阵分解的动态网络社团发现算法 Cr-ENMF,算法利用前一时刻的网络和社团结构来表征时间代价,并将其通过正则化项加入到目标函数中,成功挖掘动态网络中的社团结构。Ma 等<sup>[107]</sup> 提出两种用于动态社团发现的非负矩阵分解框架,并证明了所提框架的最优问题和模块化密度的优化的等价性。在等价性的基础上,提出了一种半监督 sE-NMF 算法。与传统的半监督算法不同, sE-NMF 将先验信息集成到算法的目标函数中。该算法的主要优点是在不增加时间复杂度的情况下避免了局部最优解。Jiao 等<sup>[108]</sup> 从全概率的角度出发,提出了一种基于非负矩阵分解的聚类方法,引入了重要节点矩阵、社团成员关系矩阵、每个快照的相似性矩阵和两个连续快照之间的社团概率转换矩阵。该算法可处理有权无权、有方向无方向、重叠非重叠网络。基于演化聚类框架的社团发现方法的主要挑战是如何平衡快照质量  $CS$  和时间代价  $CT$ ,以及如何量化时间代价  $CT$  并对社团演化模式进行建模。

## 4 社团发现的评价指标

为了评估不同社团发现算法所划分社团结果的质量,研究人员提出了多种评价指标,常用的主要有模块度  $Q$ 、标准互信息(Normalized Mutual Information, NMI)、准确率(Accuracy)、纯度(Purity)等,其中标准互信息、准确率和纯度是基于已知真实社团划分结果的评价指标,模块度  $Q$  是基于未知真实社团结构的评价指标。

### (1) 模块度

在真实社团结构未知的情况下,通常采用模块度  $Q$  来评价划分社团结构的优异。模块度  $Q$  在之前已经对其进行了介绍,其是指在社团内部边的比例与随机网络中边的比例的差值。模块度  $Q$  的值介于 0 到 1 之间,其值越大代表社团的划分结果越好。

### (2) 标准化互信息

NMI 评价指标来源于信息理论,其利用信息熵来衡量算法划分的社团结构与已知的真实社团划分之间的差异<sup>[114]</sup>。NMI 的定义如下:

$$NMI(T, C) = \frac{-2 \sum_{i=1}^{F_T} \sum_{j=1}^{F_C} F_{ij} \log(F_{ij} \cdot n / F_{i \cdot} \cdot F_{\cdot j})}{\sum_{i=1}^{F_T} F_{i \cdot} \log(F_{i \cdot} / n) + \sum_{j=1}^{F_C} F_{\cdot j} \log(F_{\cdot j} / n)} \quad (12)$$

其中,  $T = \{T_1, T_2, \dots, T_k\}$  是真实社团划分,  $C = \{C_1, C_2, \dots, C_k\}$  为算法给出的社团划分。 $F$  表示混合矩阵,其元素  $F_{ij}$  表示既属于社团  $C_i$  又属于社团  $T_j$  的节点的数量,  $F_{i \cdot}$  和  $F_{\cdot j}$  分别是矩阵  $F$  的第  $i$  行元素的总和、第  $j$  列元素的总和,表示社团  $C_i$ 、社团  $T_j$  的节点数量。如果社团划分  $T$  与  $C$  完全相同,则  $NMI(T, C) = 1$ , 否则  $NMI(T, C) = 0$ 。

### (3) 准确率

准确率 Accuracy 与 NMI 类似,也是基于真实社团划分来衡量算法的社团划分结果,通过计算划分正确的节点数占网络总节点数的比例来评价算法性能。网络中任意两个相似的节点应被认作同一社团的成员,但在实际的社团划分中,会发生以下 4 种情况:1) TP(True Positive),相似的成员被划分到同一个社团中;2) TN(True Negative),不相似的成员被划分到不同的社团中;3) FN(False Negative),相似成员被划分到了不同的社团中;4) FP(False Positive),不相似的成员被划分到了相同的社团中。其中我们认为情况 1) 和 2) 是正确的划分,情况 3) 和 4) 是错误的划分。因此,准确率可以表示为:

$$Accuracy = \frac{TP}{TP + FP} \quad (13)$$

准确率 Accuracy 衡量算法划分的社团与真实社团的匹配程度,其值介于 0 到 1 之间,值越大表示划分正确的节点数越多,当所有节点都划分正确时,  $Accuracy = 1$ 。

### (4) 纯度

Purity 也是基于真实社团划分来衡量算法社团发现性能的指标<sup>[115]</sup>。给定真实的社团划分纯度的定义如下:

$$Purity = \frac{1}{n} \sum_{i=1}^k \max_j |C_i \cap T_j| \quad (14)$$

其中,  $C = \{C_1, C_2, \dots, C_k\}$  为算法给出的社团划分,  $T = \{T_1, T_2, \dots, T_k\}$  为真实的社团划分。Purity 值越大表示算法的社团划分结果与真实结果越相似。

## 5 社团发现面临的挑战

尽管复杂网络社团结构研究取得了令人鼓舞的成果,但是随着研究的不断深入,数据的不断增多,类型的不断丰富,社团发现呈现出新的要求和挑战。

(1) 网络表示学习相结合:近年来,基于深度学习的网络表示学习方法作为网络分析任务的上游表征,受到了越来越多研究者的关注。在对节点进行低维表示后,可以通过机器学习方法对节点进行聚类,从而得到社团结构。因此,网络

表示学习为深度模型在社团发现中应用提供了可能性。深度学习在处理大规模网络数据时,可以保持高效的性能和可行的计算速度,同时也拥有较强的可移植性和特征学习能力,能够捕捉网络底层的非线性结构。因此,如何将网络表示学习与社团发现相结合,实现深度神经网络模型在社团发现问题中的应用,是当前社团发现研究的一个新挑战。

(2)稀疏网络的社团发现:复杂网络存在无标度特征,只有少数的节点拥有大量的连接。对于现实世界中的许多信息网络,由于隐私或法律的限制,网络结构往往存在稀疏性问题。如何在稀疏网络中准确挖掘社团结构,是当前亟需解决的问题。

(3)动态网络社团发现:大多数社团发现算法都基于网络是静态的假设,但在真实网络中,网络的拓扑结构是不断发生变化的,网络的社团结构也随着拓扑的变化而发生演化。相比于静态网络,动态网络的社团结构挖掘任务更具有挑战性,不仅要在每个时刻获得最贴合真实网络的社团结构,还要考虑社团的演化过程。如何在动态网络中准确地发现社团结构并捕捉社团的动态演化模式,仍是社团发现中待解决的难点问题。

(4)超大规模网络的结构挖掘:随着互联网的飞速发展,网络数据规模呈现指数级增长,现有的社团发现算法难以对亿量级的超大规模网络数据进行有效的挖掘。针对超大规模网络,设计一个可扩展性的社团发现算法并且在准确性和效率方面远远超过现行基准方法,是亟待解决的问题。

(5)复杂结构网络的社团发现:真实世界的网络多是异构的、有符号的或者是不完整的。异构网络是指包含不同类型的节点和边的网络。符号网络是指节点之间有积极和消极关系的网络,网络的边有“正”“负”属性。不完整的网络是指拓扑、节点或边信息缺失的网络。对于异构网络,不同类型实体的分布和属性是不同的,如何挖掘异构网络中的社团结构是非常有挑战性的问题。对于符号网络,不仅要将联系紧密的节点划分到一个社团,同时要考虑节点之间边的正负属性,因此符号网络的建模和社团挖掘也是一个有挑战性的课题。对于不完整网络,如何运用不完整的信息,通过网络的多尺度信息进行互补或补全网络后对社团结构进行划分,也是一个棘手的问题。

**结束语** 社团发现作为基础并且重要的网络分析任务,在过去的十多年已经得到了深入的研究。挖掘复杂网络中的社团结构对理解网络潜在结构和功能等具有重要的意义。本文首先阐述了社团发现的研究背景和意义,分别从静态网络和动态网络两个方面总结和梳理了社团发现的主要方法和评价社团划分的主要性能指标,并分析了其面临的困难和挑战。

## 参 考 文 献

- [1] WATTS D J, STROGATZ S H. Collective Dynamics of ‘small-world’ Networks[J]. *Nature*, 1998, 393(6684): 440-442.
- [2] BARABASI A L, ALBERT R. Emergence of Scaling in Random Networks[J]. *Science*, 1999, 286(5439): 509-512.
- [3] GIRVAN M, NEWMAN M E J. Community Structure in Social and Biological Networks[J]. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826.
- [4] FORTUNATO S, HRIC D. Community Detection in Networks: A User Guide[J]. *Physics Reports*, 2016, 659: 1-44.
- [5] LI G, CAI Z, YIN G, et al. Differentially Private Recommendation System Based on Community Detection in Social Network Applications[J]. *Security and Communication Networks*, 2018, 2018: 1-18.
- [6] JIN T, WU Q, OU X, et al. Community Detection and Co-author Recommendation in Co-author Networks[J]. *International Journal of Machine Learning and Cybernetics*, 2020(2): 1-13.
- [7] LI C, BAI J, TANG H, et al. Opinion Community Detection and Opinion Leader Detection Based on Text Information and Network Topology in Cloud Environment [J]. *Information Sciences*, 2019, 504(C): 61-83.
- [8] BAHULKAR A, SZYMANSKIB K, BAYCIK N O, et al. Community Detection with Edge Augmentation in Criminal Networks[C]// *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 2018: 1168-1175.
- [9] BAZAGA A, VELLIDO A. Network Community Cluster-Based Analysis for the Identification of Potential Leukemia Drug Targets [J]. *WSOM 2019: Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*. 2019: 314-323.
- [10] ZU J, HU G, YAN J, et al. A Community Detection Based Approach for Service Function Chain Online Placement in Data Center Network[J]. *Computer Communications*, 2021, 169(1): 168-178.
- [11] ZHANG S, ZHANG Y, ZHOU M, et al. Community Detection Based on Similarities of Communication Behavior in IP Networks[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2020(7307): 1-11.
- [12] SHI D X, LI H, YANG R S, et al. User Daily Frequent Behavior Pattern Mining[J]. *Journal of National University of Defense Technology*, 2017, 39(1): 74-80.
- [13] FORTUNATO S. Community Detection in Graphs[J]. *Physics Reports*, 2009, 486: 3-5.
- [14] PORTER M A, ONNELA J P, MUCHA P J. Communities in Networks[J]. *Notices of the American Mathematical Society*, 2009, 56(9): 4294-4303.
- [15] YANG B, LIU D, LIU, J, et al. Discovering Communities from Social Networks[M]// *Methodologies and Applications*. Springer, 2010: 331-346.
- [16] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2): 026113.
- [17] KELLEY S, GPLDBERG M, MAGDON I M, et al. Defining and Discovering Communities in Social Networks[M]// *Handbook of Optimization in Complex Networks*. Springer, 2012: 139-168.
- [18] MACQUEEN J B. Some methods for classification and analysis of multivariate observations[C]// *The Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1967.
- [19] YANG B, LIU D, LIU, J, et al. Discovering Communities from Social Networks[M]// *Methodologies and Applications*. Springer, 2010: 331-346.
- [20] NEWMAN M E. Fast Algorithm for Detecting Community Structure in Networks[J]. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2004, 69(6 Pt 2): 066133.
- [21] RAGHAVAN U N, REKA A, KUMARA S. Near Linear Time

- Algorithm to Detect Community Structures in Large-Scale Networks[J]. *Physical Review E*, 2007, 76(3 Pt 2):036106.
- [22] ZHANG Y, ZHANG Y, CHEN Q, et al. True-link Clustering Through Signaling Process and Subcommunity Merge in Overlapping Community Detection[J]. *Neural Computing & Applications*, 2018, 30(50):3613-3621.
- [23] RIEDY J, BADER D A, MEYERHENKE H. Scalable Multi-threaded Community Detection in Social Networks[C]// *IEEE International Parallel & Distributed Processing Symposium Workshops & Phd Forum*. 2012:1619-1628.
- [24] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast Unfolding of Communities in Large Networks[J]. *Journal of Statistical Mechanics Theory & Experiment*, 2008(10):10008.
- [25] SHANG J, LIU L, XIE F, et al. A Real-Time Detecting Algorithm for Tracking Community Structure of Dynamic Networks [J]. *arXiv*:1407.2683, 2014.
- [26] CLAUSET A, NEWMAN M E, MOORE C. Finding Community Structure in Very Large Networks[J]. *Physical Review E*, 2004, 70(6 Pt 2):066111.
- [27] LEUNG X, HUI P, LIO P, et al. Towards Real-time Community Detection in Large Networks[J]. *Physical Review E*, 2008, 79(6 Pt 2):066107.
- [28] WU F, HUBERMAN B A. Finding Communities in Linear Time: A Physics Approach[J]. *European Physical Journal B-Condensed Matter*, 2004, 38(2):331-338.
- [29] BAGROW J P, BOLLT E M. Local Method for Detecting Communities[J]. *Physical Review E*, 2005, 72(4):046108.
- [30] LI N, CHEN G. Analysis of a Location-Based Social Network [C]// *International Conference on Computational Science & Engineering*. 2009:263-270.
- [31] NI C, LIN Y, LUO F, et al. Community Detection on Networks with Ricci Flow[R]. *Scientific Reports*, 2019.
- [32] GORKE R, MAILLARD P, STAUST C, et al. Modularity-Driven Clustering of Dynamic Graphs[C]// *International Symposium on Experimental Algorithms*. 2010:436-448.
- [33] FLAKE G W, LAWRENCE S, GILES C L, et al. Self-organization and Identification of Web Communities [J]. *Computer*, 2002, 35(3):66-70.
- [34] GUIMERA R, SALESPARDO M, AMARAL L A, et al. Modularity from Fluctuations in Random Graphs and Complex Networks[J]. *Physical Review E*, 2004, 70(2):025101-025101.
- [35] ZHE C, SUN A, XIAO X. Community Detection on Large Complex Attribute Network[C]// *The 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*. 2019:2041-2049.
- [36] SOBOLEVSKY S, CAMPARI R, BELYI A, et al. General Optimization Technique for High-quality Community Detection in Complex Networks [J]. *Physical Review E*, 2014 ( E 90 ): 012811.
- [37] GOMEZ S, JENSEN P, ARENAS A. Analysis of Community Structure in Networks of Correlated Data[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2009, 80(1):016114.
- [38] ANCHURI P, MAGDONISMAIL M. Communities and Balance in Signed Networks: A Spectral Approach[C]// *Advances in Social Networks Analysis and Mining*. 2012:235-242.
- [39] CAI Q, GONG M, BO S, et al. Discrete Particle Swarm Optimization for Identifying Community Structures in Signed Social Networks[J]. *Neural Networks*, 2014, 58(10):4-13.
- [40] SHEN H, CHENG X, CAI K, et al. Detect Overlapping and Hierarchical Community Structure in Networks[J]. *Physica A Statistical Mechanics & Its Applications*, 2008, 388(8):1706-1712.
- [41] WALTMAN L, VAN N J. A Smart Local Moving Algorithm for Large-scale Modularity-Based Community Detection[J]. *European Physical Journal B*, 2013, 86(11):471.
- [42] FORTUNATO S, BARTHELEMY M. Resolution Limit in Community Detection[J]. *Proceedings of the National Academy of Sciences*, 2007, 104(1):36-41.
- [43] ZHU J, CHEN B, ZENG Y. Community Detection Based on Modularity and K-plexes[J]. *Information Sciences*, 2020, 513:127-142.
- [44] DINH T N, THAI M T. Community Detection in Scale-Free Networks: Approximation Algorithms for Maximizing Modularity[J]. *IEEE Journal on Selected Areas in Communications*, 2013, 31(6):997-1006.
- [45] PRAMANIK S, TACKX R, NAVELKAR A, et al. Discovering Community Structure in Multilayer Networks [C]// *IEEE International Conference on Data Science and Advanced Analytics*. 2017:611-620.
- [46] LEE D D, SEUNG H S. Algorithms for Non-negative Matrix Factorization[J]. *Neural Information Processing Systems*, 2000, 13(6):556-562.
- [47] PENG S, SER W, CHEN B, et al. Robust Semi-supervised Nonnegative Matrix Factorization for Image Clustering[J]. *Pattern Recognition*, 2021, 111:107683.
- [48] LUO H, HAN J. Nonnegative Matrix Factorization Based Transfer Subspace Learning for Cross-Corpus Speech Emotion Recognition[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, 28:2047-2060.
- [49] CAI H, LIU B, XIAO Y, et al. Semi-Supervised Multi-view Clustering Based on Orthonormality-Constrained Nonnegative Matrix Factorization[J]. *Information Sciences*, 2020, 536:171-184.
- [50] LI D, LIN Q, MA X. Identification of Dynamic Community in Temporal Network via Joint Learning graph Representation and Nonnegative Matrix Factorization[J]. *Neurocomputing*, 2021, 435:77-90.
- [51] MA X, DONG D, WANG Q, et al. Community Detection in Multi-layer Networks Using Joint Nonnegative Matrix Factorization[J]. *IEEE Trans on Knowledge and Data Engineering*, 2019, 31(2):273-286.
- [52] LU H, SHEN Z, SANG X, et al. Community Detection Method Using Improved Density Peak Clustering and Nonnegative Matrix Factorization[J]. *Neurocomputing*, 2020, 415:247-257.
- [53] YANG L, CAO X, JIN D, et al. A Unified Semi-Supervised Community Detection Framework Using Latent Space Graph Regularization [J]. *IEEE Transactions on Cybernetics*, 2015, 45(11):2585-2598.
- [54] WANG F, LI T, WANG X, et al. Community Discovery Using Nonnegative Matrix Factorization[J]. *Data Mining and Knowledge Discovery*, 2011, 22(3):493-521.
- [55] ZHANG Y, YEUNG D. Overlapping Community Detection via Bounded Nonnegative Matrix Tri-factorization[C]// *The 18th*

- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD). 2012;606-614.
- [56] DING C, TAO L, WEI P, et al. Orthogonal Nonnegative Matrix Tri-factorizations for Clustering[C]//12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD). 2006;126-135.
- [57] CAO J, WANG H, JIN D, et al. Combination of Links and Node Contents for Community Discovery Using a Graph Regularization Approach[J]. Future Generation Computer Systems, 2019, 91(FEB);361-370.
- [58] PEI Y, CHAKRABORTY N, SYCARA K. Nonnegative Matrix tri-factorization with Graph Regularization for Community Detection in Social Networks[C]//The 24th International Conference on Artificial Intelligence. AAAI, 2015;2083-2089.
- [59] WANG X, JIN D, CAO X, et al. Semantic Community Identification in Large Attribute Networks[C]//The 30th AAAI Conference on Artificial Intelligence. AAAI, 2016;265-271.
- [60] LU H, SANG X, ZHAO Q, et al. Community Detection Algorithm Based on Nonnegative Matrix Factorization and Pairwise Constraints[J]. Physica A: Statistical Mechanics and its Applications, 2019, 545;123491.
- [61] SHI X, LU H, HE Y, et al. Community Detection in Social Network with Pairwisely Constrained Symmetric Non-Negative Matrix Factorization[C]//IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining(ASONAM). 2015;541-546.
- [62] ZHANG L, YANG L, GONG Y, et al. Community Discovery on Multi-View Social Networks via Joint Regularized Nonnegative Matrix Triple Factorization[J]. IEICE Transactions on Information & Systems, 2017, E100. D(6);1262-1270.
- [63] JIN H, YU W, LI S. Graph Regularized Nonnegative Matrix Tri-factorization for Overlapping Community Detection[J]. Physica A: Statistical Mechanics and its Applications, 2019, 515;376-387.
- [64] LI Z, CHEN J, FU Y, et al. Community Detection Based on Regularized Semi-Nonnegative Matrix Tri-Factorization in Signed Networks[J]. Mobile Networks and Applications, 2017, 23(2);1-9.
- [65] MA C, LIN Q, Y LIN, et al. Identification of Multi-layer Networks Community by Fusing Nonnegative Matrix Factorization and Topological Structural Information[J]. Knowledge-Based Systems, 2021;213:106666.
- [66] TIAN F, GAO B, CUI Q, et al. Learning Deep Representations for Graph Clustering[C]//The 28th AAAI Conference on Artificial Intelligence. AAAI, 2014;1293-1299.
- [67] YANG L, CAO X, HE D, et al. Modularity Based Community Detection with Deep Learning[C]//International Joint Conference on Artificial Intelligence(IJCAI). 2016;2252-2258.
- [68] JIN D, GE M, LI Z, et al. Using Deep Learning for Community Discovery in Social Networks[C]//2017 IEEE 29th International Conference on Tools with Artificial Intelligence(ICTAI). 2017;160-167.
- [69] CAO J, 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.
- [70] CAO J, JIN D, DANG J. Autoencoder Based Community Detection with Adaptive Integration of Network Topology and Node Contents[C]//International Conference on Knowledge Science, Engineering and Management(KSEM). 2018;184-196.
- [71] SKRLJ B, KRALJ J, LAVRA N. Embedding-Based Silhouette Community Detection[J]. Machine Learning, 2020, 109(1);161-219.
- [72] HU L, GUO G. An Augmented Lagrangian Alternating Direction Method for Overlapping Community Detection Based on Symmetric Nonnegative Matrix Factorization[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(6755);403-415.
- [73] VANDANA B, RANI R. A Distributed Overlapping Community Detection Model for Large Graphs Using Autoencoder[J]. Future Generation Computer Systems, 2019(94);16-26.
- [74] SUN H, HE F, HUANG J, et al. Network Embedding for Community Detection in Attributed Networks[J]. ACM Transactions on Knowledge Discovery from Data, 2020, 14(3);1-25.
- [75] XU R, CHE Y, WANG X, et al. Stacked Autoencoder-Based Community Detection Method via an Ensemble Clustering Framework [J]. Information Sciences, 2020, 526;151-165.
- [76] JIA Y, ZHANG Q, ZHANG W, et al. CommunityGAN: Community Detection with Generative Adversarial Nets[C]//The World Wide Web Conference(WWW). 2019;784-794.
- [77] ZHANG Y, XIONG Y, YE Y, et al. SEAL: Learning Heuristics for Community Detection with Generative Adversarial Networks [C]//The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD). 2020;1103-1113.
- [78] CHEN J, GONG Z, DAI Q, et al. Adversarial Learning for Overlapping Community Detection and Network Embedding [C]//The 24th European Conference on Artificial Intelligence. 2020;1071-1078.
- [79] YANG L, WANG Y, GU J, et al. JANE: Jointly Adversarial Network Embedding[C]//The 29th International Joint Conference on Artificial Intelligence(IJCAI). 2020;1381-1387.
- [80] HONG H, LI X, WANG M. GANE: A Generative Adversarial Network Embedding [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7);2325-2335.
- [81] YAO Z, YUN X, YUN Y, et al. SEAL: Learning Heuristics for Community Detection with Generative Adversarial Networks [C]//The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD). 2020;1103-1113.
- [82] KIPFT N, WELLING M. Semi-Supervised Classification with Graph Convolutional Networks[C]//International Conference on Learning Representations. 2017.
- [83] HE D, SONG Y, JIN D, et al. Community-Centric Graph Convolutional Network for Unsupervised Community Detection [C]//Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence(IJCAI). 2020;3515-3521.
- [84] JIN D, LI B, JIAO P, et al. Community Detection via Joint Graph Convolutional Network Embedding in Attribute Network [C]//The 28th International Conference on Artificial Neural Networks. 2020;594-606.
- [85] SATTAR N S, ARIFUZZAMAN S. Community Detection using Semi-supervised Learning with Graph Convolutional Network on GPUs[C]//The IEEE International Conference on Big Data. 2020;5237-5246.
- [86] WANG Y, DUAN Z, LIAO B, et al. Heterogeneous Attributed Network Embedding with Graph Convolutional Networks

- [C]//The AAAI Conference on Artificial Intelligence. AAAI, 2019;10061-10062.
- [87] JIN D, LIU Z, LI W, et al. Graph Convolutional Networks Meet Markov Random Fields: Semi-supervised Community Detection in Attribute Networks[C]//The AAAI Conference on Artificial Intelligence. AAAI, 2019;152-159.
- [88] SHCHUR O, GUNNEMANN S. Overlapping Community Detection with Graph Neural Networks[J]. arXiv:1909.12201, 2019.
- [89] ROSSETTI G, CAZABET R. Community Discovery in Dynamic Networks: A Survey [J]. ACM Computing Surveys, 2017, 51(2):1-37.
- [90] CHAKRABARTI D, KUMAR R, TOMKINS A. Evolutionary Clustering[C]//The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2006:554-560.
- [91] PALLA G, BARABASI A L, VICSEK T. Quantifying Social Group Evolution[J]. Nature, 2007;446(7136):664-667.
- [92] HOPCROFT J, KHAN O, KULIS B, et al. Tracking evolving communities in large linked networks[J]. National Academy of Sciences, 2004, 101(1):5249-5253.
- [93] ZHUANG D. Modularity-Based Dynamic Community Detection [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 33(5):1934-1945.
- [94] YANG B, LIU D Y. Force-Based Incremental Algorithm for Mining Community Structure in Dynamic Network[J]. Journal of Computer Science and Technology, 2006, 21(3):393-440.
- [95] GUO K, HE L, HUANG J, et al. A Local Dynamic Community Detection Algorithm Based on Node Contribution [C]//Conference on Computer Supported Cooperative Work. 2019;363-376.
- [96] WU Z, CHEN J, ZHANG Y. An Incremental Community Detection Method in Social Big Data[C]//2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies(BDCAT). 2018;136-141.
- [97] HU Y, YANG B, LV C. A Local Dynamic Method for Tracking Communities and Their Evolution in Dynamic Networks[J]. Knowledge-Based Systems, 2016, 110(5):176-190.
- [98] AL-SHAROA, E, ALKHASSAWENEH M. Tensor Based Temporal and Multilayer Community Detection for Studying Brain Dynamics During Resting State fMRI[J]. IEEE Transactions on Biomedical Engineering, 2019, 66(3):695-709.
- [99] SARIYUCE A E, GEDIK B, JACQUES-SILVA G, et al. SON-IC: Streaming Overlapping Community Detection[J]. Data Mining and Knowledge Discovery, 2016, 30(4):819-847.
- [100] LI X, WU B, QIAN G, et al. Dynamic Community Detection Algorithm Based on Incremental Identification[C]//IEEE International Conference on Data Mining Workshop (ICDMW). 2015:900-907.
- [101] CHI Y, SONG X, ZHOU D, et al. On Evolutionary Spectral Clustering[J]. ACM Transactions on Knowledge Discovery from Data, 2009, 3(4):1-30.
- [102] LIN Y, CHI Y, ZHU S, et al. Analyzing Communities and Their Evolutions in Dynamic Social Networks[C]//International Conference on World Wide Web(WWW). 2008;685-694.
- [103] FOLINO F, PIZZUTI C. An Evolutionary Multiobjective Approach for Community Discovery in Dynamic Networks [J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 26(8):1838-1852.
- [104] YU W, WANG W, JIAO P, et al. Evolutionary Clustering via Graph Regularized Nonnegative Matrix Factorization for Exploring Temporal Networks[J]. Knowledge Based Systems, 2019, 167(MAR. 1):1-10.
- [105] YIN Y, ZHAO Y, LI H, et al. Multi-Objective Evolutionary Clustering for Large-scale Dynamic Community Detection[J]. Information Sciences, 2020, 549(3C5):269-287.
- [106] MA X, ZHANG B, MA C, et al. Co-regularized Nonnegative Matrix Factorization for Evolving Community Detection in Dynamic Networks [J]. Information Sciences, 2020, 528 (6224):265-279.
- [107] MA X, DONG D. Evolutionary Nonnegative Matrix Factorization Algorithms for Community Detection in Dynamic Networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(5):1045-1058.
- [108] JIAO P, YU W, WANG W, et al. Exploring Temporal Community Structure and Constant Evolutionary Pattern Hiding in Dynamic Networks[J]. Neurocomputing, 2018, 314(7):224-233.
- [109] YU W, WANG W, JIAO P, et al. Modeling the Local and Global Evolution Pattern of Community Structures for Dynamic Networks Analysis[J]. IEEE Access, 2019, 7(99):71350-71360.
- [110] WANG S, LI G, HU G, et al. Community Detection in Dynamic Networks Using Constraint Non-negative Matrix Factorization [J]. Intelligent Data Analysis, 2020, 24(1):119-139.
- [111] KIM M S, HAN J. A Particle-and-density Based Evolutionary Clustering Method for Dynamic Networks [J]. Proceedings of the Vldb Endowment, 2009, 2(1):622-633.
- [112] JIAO P, LYU H, LI X, et al. Temporal Community Detection Based on Symmetric Nonnegative Matrix Factorization[J]. International Journal of Modern Physics B, 2017, 31(13):1750102.
- [113] ZENG X, WANG W, CHEN C, et al. A Consensus Community-Based Particle Swarm Optimization for Dynamic Community Detection[J]. IEEE Transactions on Cybernetics, 2020, 50(6):2502-2513.
- [114] DANON L, DIAZGUILERA A, DUCH J, et al. Comparing community structure identification [J]. Journal of Statistical Mechanics: Theory and Experiment, 2005, 2005(9).
- [115] LIN W, KONG X, YU P S, et al. Community detection in incomplete information networks[C]//International World Wide Web Conferences. 2012;341-350.



**PAN Yu**, born in 1990, Ph.D candidate. Her main research interests include data processing and mining in social networks and machine learning.



**PAN Zhi-song**, born in 1973. Ph.D, professor, Ph.D supervisor. His main research interests includes computer vision and machine learning.