

复杂网络社区发现研究进展



赵卫绩^{1,2} 张凤斌¹ 刘井莲^{2,3}

1 哈尔滨理工大学计算机科学与技术学院 哈尔滨 150080

2 绥化学院信息工程学院 黑龙江 绥化 152061

3 东北大学计算机科学与工程学院 沈阳 110819

(sdzhaoweiji@163.com)

摘要 近年来,随着现代网络通信和社会媒体等技术的飞速发展,复杂网络成为多学科交叉研究的热点之一,社区发现是复杂网络中的一个重要问题,对其进行研究具有重要的理论意义和应用价值。该问题吸引了多个学科领域的众多学者的关注,并且已有许多社区发现算法被提出。已有的社区发现综述多是侧重某一方面或特定领域展开,基于此,文中在之前工作的基础上,对国内外社区发现工作进行了深入调研,较全面地阐述了复杂网络社区发现的研究现状。首先,针对不同网络结构,给出社区发现的问题定义和主要的评价指标。然后,介绍了不同网络结构中的经典社区发现算法,包括同质网络中的全局社区发现、局部社区发现算法,异质网络中的二分网络、三分网络和多分网络中的社区发现,结合节点内容和连接结构的社区发现算法,以及动态网络中的社区发现和社区演化工作。最后,简要介绍了社区发现的典型应用,包括影响最大化、链路预测和情感分析领域的应用。此外,探讨了当前社区发现研究面临的主要挑战,试图为社区发现研究领域勾画一个较为清晰和全面的轮廓,为初学者提供指引。

关键词: 社区发现;社区结构;同质网络;异质网络;动态网络

中图法分类号 TP391

Review on Community Detection in Complex Networks

ZHAO Wei-ji^{1,2}, ZHANG Feng-bin¹ and LIU Jing-lian^{2,3}

1 School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China

2 School of Information Engineering, Suihua University, Suihua, Heilongjiang 152061, China

3 School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

Abstract In recent years, with the rapid development of modern network communication and social media technologies, complex network has become one of the frontier hotspots of interdisciplinary research. As an important problem in the research of complex network, community detection has important theoretical significance and application value, and has attracted increasing attention. Many community detection algorithms and reviews have been proposed. However, most of the existing reviews on community detection focus on a particular direction or field. On the basis of previous work, this paper did deep research in the community detection algorithms, and gave a review on the research progress of community detection. Firstly, this paper gave the definition of community detection and evaluation measurements for different network structure. Then, this paper introduced the classic community detection algorithms on different network structure, including the global community detection and local community detection algorithms on homogeneous networks, community detection on heterogeneous network, and community detection on link structure combined with node content, as well as the dynamic network community detection and community evolution. Finally, this paper briefly introduced the typical applications of community discovery in the real world, including impact maximization, link prediction and emotion analysis application. In addition, this paper discussed the challenges in the current community discovery field. This paper try to draw a clearer and more comprehensive outline for the community detection research field, and provide a good guide for beginners in the community detection.

Keywords Community detection, Community structure, Homogeneous network, Heterogeneous network, Dynamic network

到稿日期:2019-01-25 返修日期:2019-03-11 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61172168,61772122);黑龙江省省属高校基本科研业务费科研项目(KYYWF10236180104, KYYWF10236180207)

This work was supported by the National Natural Science Foundation of China (61172168,61772122) and Fundamental Research Funds for the Universities of Heilongjiang Provincial Department of Education of China (KYYWF10236180104,KYYWF10236180207).

通信作者:张凤斌(zhangfb@hrbust.edu.cn)

1 引言

近年来,随着电子通信技术、社交媒体等新技术的快速发展,产生了大量社交媒体网络平台,给人们提供了多样化的虚拟交互环境和丰富的信息资源,形成了复杂、多样的网络大数据。这些网络数据呈现出大规模、富含内容标签、信息异构、动态变化等特性。复杂网络还普遍存在一些统计特征,如反映网络节点服从幂律分布特点的“无标度特性”^[1](见图1(a))、反映网络短路径长度和高聚类系数特点的“小世界效应”^[2](见图1(b)),以及反映网络中普遍存在的同一社区中节点连接紧密、不同社区间节点连接稀疏的“社区结构”等^[3](见图1(c))。近年来,社区结构作为复杂网络大数据的一个重要特征,吸引了国内外众多学者的关注。亚利桑那大学、斯坦福大学、康奈尔大学、卡耐基梅隆大学、清华大学、吉林大学、天津大学、中科院等国内外许多大学和研究机构的众多学者在此领域都开展了深入研究,并取得了很好的研究成果,其成果发表在 *Science*、*Nature*、*PNAS*、*WWW*、*KDD*、*AAAI*、*Physical Review E*、*Physical A*、《计算机学报》、《软件学报》、《自动化学报》、《计算机研究与发展》、*CCML*、*ICDM* 等国内外的重要期刊和会议上。

所谓社区,是指具有“相似连接结构的节点集合”,社区内的节点连接紧密,与社区外的节点连接稀疏^[3]。在社会学中,社区概念类似于社群概念。社区发现类似于传统的图划分、聚类,Cheng 等对社区发现进行了问题界定^[4],澄清了它们之间的区别与联系^[4]。社区发现和图划分处理的都是网络,图划分是按指定条件对网络进行切分,把网络划分成给定个数、大小相同的节点集合,而社区发现是将网络划分成内部连接紧密、与外部连接稀疏的节点集合。社区发现在于寻找网络固有的自然划分,没有指定个数的限制。此外,不同于图划分,社区之间允许相互重叠或嵌套。在这点上,社区发现与聚类是相似的,聚类的各个类之间也允许相互重叠或嵌套,但聚类与社区发现处理的对象一般不同,聚类处理的是高维向量的属性数据,而社区发现处理的是网络的关系数据。

社区发现是复杂网络分析中的一项重要研究工作,早期的研究更多集中在同质网络上,目前在同质网络上的研究较为成熟,产生了许多好的算法^[3-29]。相比之下,异质网络上的研究相对较少。然而,真实世界中存在着很多多关系多节点类型的异质网络,例如社交媒体网络。在这些网络中,同类型节点间和不同类型节点间形成了复杂多样的关系网络。如何获取这些网络中的社区结构,是社区发现研究面临的一个新挑战。目前,面向二分网络的社区发现算法^[30-36]、面向三分和多分网络的社区发现算法^[37-48]陆续被提出,这些算法大多是基于网络的拓扑结构,忽略了网络中节点所附带的属性信息。在实际应用中,无论是网络的拓扑结构,还是节点属性所附带的内容信息,对社区发现都有着重要的影响。而如何将网络的连接结构与节点的属性信息有机地结合起来,寻找更准确的社区结构^[49-61],也是社区发现研究面临的一个新挑战。此外,现实中的很多网络是动态变化的,传统方法无法探

测网络的演化过程,因此出现了一些面向动态网络的社区发现算法^[62-71]。也有学者对各种类型的社区发现算法进行了综合分析,发表了一些较好的综述性文献^[4,15,17],为初学者提供了很好的指引。然而,这些社区发现综述文献主要针对同质网络,有的是针对某一个特定方向和特定问题进行介绍,有的是结合作者所在课题组所做的工作进行介绍。基于此,本文对不同网络结构的社区发现研究进展进行了全面介绍,对同质网络、异质网络以及动态网络中的社区发现研究进行了较为全面的阐述和分析,试图为社区发现研究领域描绘一个较为清晰和全面的轮廓,为后续初学者提供有益的参考。

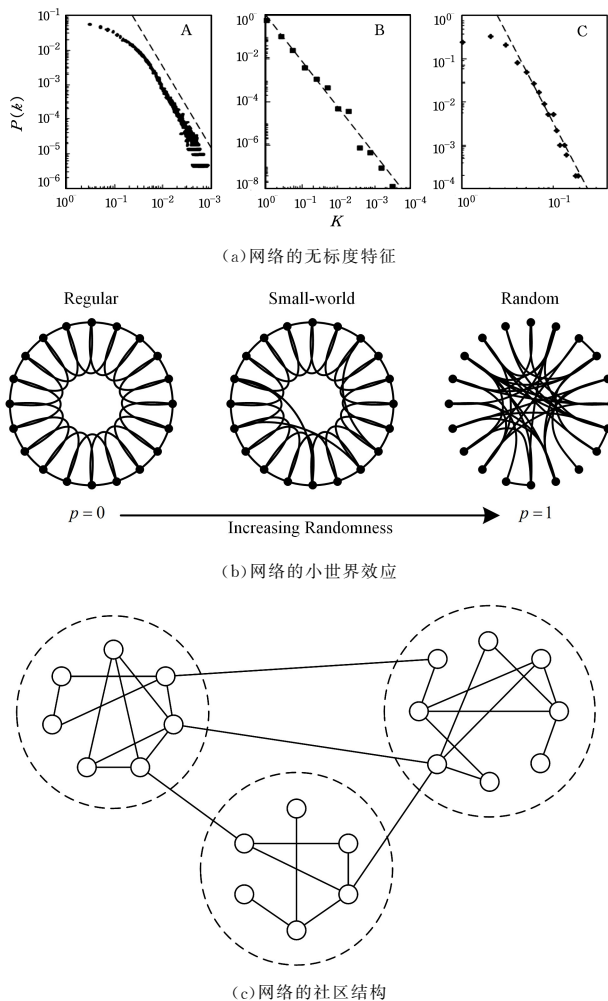


图1 复杂网络的一些基本特性

Fig.1 Basic characters of complex network

本文第2节给出面向不同网络结构的社区发现的问题定义;第3节介绍了评价社区发现算法优劣的度量指标;第4节结合不同网络结构,对社区发现算法进行了分类和分析,具体算法包括同质网络中的全局非重叠/重叠社区发现算法,局部社区发现算法,面向异质网络中的二分网络、三分网络、多分多维网络的社区发现算法和结合节点内容和连接结构的社区发现算法;第5节介绍动态网络中的社区发现算法;第6节介绍了社区发现的典型应用,包括影响最大化、链路预测和情感分析等领域的应用;第7节探讨了社区发现研究面临的挑战,并对下一步的研究工作进行了展望。

2 问题定义

社区结构是复杂网络中的一个重要研究问题,该问题吸引了多个学科领域的众多学者的关注。对于社区结构,在同质网络上开展了全局社区发现、重叠社区发现的研究,也有一些基于局部网络结构信息的局部社区发现研究。此外,还有在异质网络、动态网络上的各种算法的研究。在 DBLP 上,以“community detection”为主题搜索到的文献从 2004 年的每年几篇递增到 2017 年的每年近 300 篇,呈逐年增长的趋势。

2.1 基本定义

定义 1(信息网络^[72]) 给定一个有向图 $G=(V, E; \phi, \varphi; A, R)$ 。 V 表示节点集, E 表示边集, ϕ 表示节点类型映射函数, φ 表示边类型映射函数, A 表示节点类型集, R 表示边类型集。 $\phi(v) \in A$ 表示任意节点 $v \in V$ 都属于一个特定的节点类型, $\varphi(e) \in R$ 表示任意关系 $e \in E$ 都属于一种特定的关系类型。当节点类型数量 $|A| > 1$ 或边的类型数量 $|R| > 1$ 时,这样的信息网络称为异质网络。反之则为同质网络。

定义 2(社区结构) 社区结构刻画的是信息网络中节点间连接边的关系的局部聚集特性,网络中的社区通常由功能相近或性质相似的网络节点组成^[4]。

2.2 社区发现问题的定义

社区发现是基于网络的拓扑结构信息识别出具有相似特征或相似行为的节点组。在过去的十几年里,社区发现受到了国内外学者的广泛关注,并取得了很多开创性的研究成果。早期社区发现研究的是非重叠社区结构(见图 2(a)),一个节点只属于一个社区。而在实际生活中,常常存在重叠社区结构(见图 2(b)),如同一个人可能既属于篮球俱乐部,也属于乒乓球俱乐部。在大规模复杂的社会网络中,不仅存在重叠社区结构,同时还常常存在基于某些中心性节点的社区结构(见图 2(c)),例如明星粉丝团、传销团、恐怖分子团等。这些中心性节点也称为领袖节点,相对网络中其他参与者节点,中心节点具有一定的领导功能。

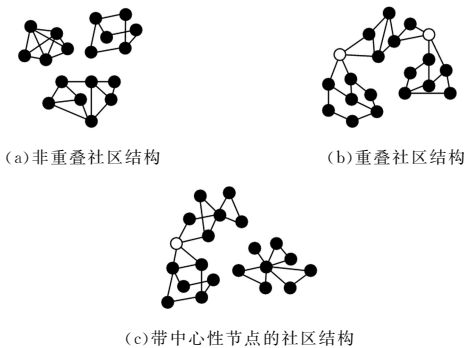
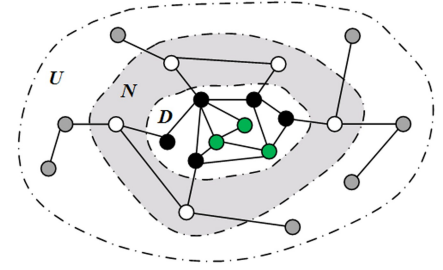


图 2 典型的网络社区结构

Fig. 2 Typical community structures in networks

还存在一些真实网络,由于规模太大或动态变化,难以获取其整体网络结构。为了解决该问题,Clusset 提出了局部社区发现^[21]。局部社区发现,是在网络整体结构未知的假设下,从给定节点出发寻找其所在社区的方法。在局部社区发现过程中,可以将网络划分为 3 部分,即局部社区 D 、 D 的外壳节点集合 N 以及未知节点集合 U ,具体如图 3 所示。



注:改自文献[28]

图 3 局部社区发现

Fig. 3 Division of a network into local community

自然界中许多网络都呈现出自然的二分结构,形成二分网络。二分网络由两种类型的节点组成,不同类型的节点间形成了许多边,同类型的节点间大多没有连接。例如常见的学术信息网络包含作者和论文两种类型的节点,作者与论文之间形成了写作关系。在二分网络中,有两种形式的社区结构定义:1)基于同质网络的社区定义,即将不同类型的节点划分到同一个社区,社区内不同类型节点间的连接比较紧密,社区间的连接比较稀疏,异质社区间形成一对一关系,如图 4(a)所示;2)基于社区内节点具有某种相似性质来考虑,将二分网络中同类型节点划分到一个社区,社区内的节点间没有连接,异质社区间形成多对多关系,如图 4(b)所示。

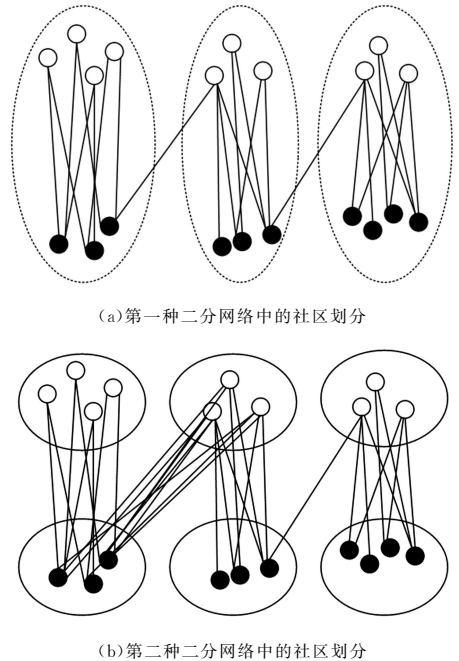
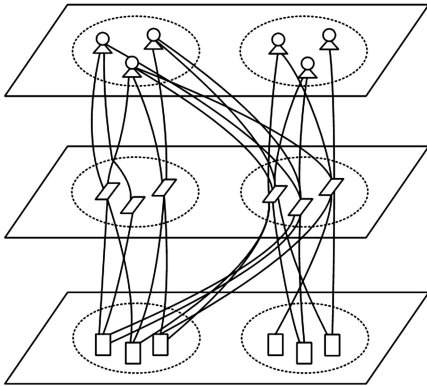


图 4 二分网络社区划分

Fig. 4 Community division of bipartite network

随着在线社交媒体平台的不断涌现,出现了复杂的多类型节点和多关系组成的异质网络。这些复杂的异质网络包括用户个人信息、文本、视频、图片等资源实体信息,还包括用户与资源实体之间形成的复杂多样的关系。例如,在复杂的学术信息网络中,除了作者类型的节点,还有论文、论文出处(期刊或会议)、作者单位等类型的节点。此外,作者与论文之间形成了写作关系,论文与期刊(会议)之间形成了发表关系,作

者之间形成了合作关系,论文之间形成了引用与被引用关系等。在多元多维的异质网络中,社区结构一般定义为同种类型节点组成的社区,社区间形成多对多关系。例如当前研究较多的社交媒体标注系统,将类型相同且具有相似特征或行为的用户节点、照片节点、标签节点分别组成社区,如图5所示。



注:改自文献[49]

图5 三分网络的社区划分

Fig. 5 Community division of tripartite network

3 社区发现的评价指标

本节主要介绍适用于不同类型网络的评价社区发现算法的度量指标。

3.1 模块度

在复杂网络中,模块度是衡量社区划分质量优劣的一个重要评价指标,已经成为最常用的社区发现质量评价指标之一。本节介绍了适用于不同网络结构的社区模块度量指标。

3.1.1 同质网络的社区模块度

在同质网络中,节点的类型和关系都只有一种。早期的研究主要关注同质网络中的社区发现,这方面的研究也相对成熟。

2004年,Newman等提出了模块度 Q ^[8],其成为第一个度量社区划分结果优劣的评价指标。该指标主要用于衡量无向无权的同质网络中的社区划分结果的优劣,社区内节点间的边数越多,模块度 Q 值越大;反之, Q 值越小。模块度 Q 的计算方法如下:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \delta(C_i, C_j) \quad (1)$$

或

$$Q = \sum_{c=1}^k \left[\frac{l_c}{L} - \left(\frac{d_c}{2L} \right)^2 \right] \quad (2)$$

式(1)中, m 为网络中的边数; A_{ij} 为邻接矩阵 A 中的元素,若节点 i, j 相连,则 $A_{ij}=1$,否则 $A_{ij}=0$; C_i 为节点 i 所属的社区;节点 j 所属的社区 C_j 与节点 i 所属的社区 C_i 相同时, $\delta(C_i, C_j)$ 取值为1,否则为0; d_i 是节点 i 的度; $\frac{d_i d_j}{2m}$ 表示在随机网络中节点 i 和节点 j 之间存在连边的可能性。式(2)中, L 为网络中连接的总边数, k 为社区个数, l_c 为社区 c 内部的连接边数, d_c 为社区 c 内所有节点的度之和。从公式中不难发现,节点 i 和节点 j 之间没有边连接,如果误划入到一个社区,则会导致 Q 值减小;而如果节点 i 和 j 在一个社区

中且有边连接,则会导致 Q 值越大。因此,当同一社区内的节点连接越紧密,不同社区间的节点连接越稀疏时, Q 取值越大。基于模块度优化的方法能够将连接最紧密的节点划分到同一个社区内。

式(1)、式(2)主要适用于无权网络,不适用于加权网络。为此,Newman又给出了加权网络的模块度 Q_w ^[11],计算方法如下:

$$Q_w = \frac{1}{2W} \sum_{i,j} \left(W_{ij} - \frac{w_i w_j}{2W} \right) \delta(C_i, C_j) \quad (3)$$

或

$$Q_w = \sum_{c=1}^k \left[\frac{W_c}{W} - \left(\frac{e_c}{2W} \right)^2 \right] \quad (4)$$

其中, W 表示整个网络的边权重之和, W_{ij} 表示连接节点 i 和 j 的边的权重, $w_i = \sum_j w_{ij}$ 表示连接节点 i 的边的权重之和, k 是网络中划分的社区数目, W_c 表示社区 c 中所有边的权重之和, e_c 表示社区 c 内所有节点的权重之和。

针对以上模块度仅适用于无向网络的问题,Nicosia等提出了适用于有向网络的社区模块度^[12],具体公式如下:

$$Q_d = \frac{1}{m} \sum_{i,j} \left[A_{ij} - \frac{d_i^{out} d_j^{in}}{m} \right] \delta(C_i, C_j) \quad (5)$$

上述几个模块度假设社区之间是独立的,没有考虑社区的重叠情况,而现实中社区结构常常是重叠的。因此,Shen等于2009年针对现实复杂网络中一个节点可能属于多个社区的情况,对于无向无权网络,给出了重叠社区模块度^[13],具体公式如下:

$$Q_c = \frac{1}{2m} \sum_{k=1}^c \sum_{i \in C_k, j \in C_k} \frac{1}{O_i O_j} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \quad (6)$$

其中, O_i 是包含节点 i 的社区数目。

3.1.2 异质网络社区的模块度

二分网络是相对简单的异质网络,相比多元多维的异质网络获得了更多的研究。Barber充分考虑到二分网络的特性和两种类型的节点,将不同类型节点划分到同一个社区中,给出了Barber二分网络的模块度^[30],具体公式如下:

$$Q_b = \frac{1}{M} \sum_{i,j} \sum_{k=1}^m \left\{ A(i, j) - \frac{t_i \times t_j}{m} \right\} \delta(C_i, C_j) \quad (7)$$

其中, C_i 表示节点 i 所属的社区。

Liu等提出了一种一对多关系的二分模块度,能够表示两个异质类型社区对应关系的强弱度^[34],计算方法如下:

$$Q_m = \sum_l Q_{ml} = \frac{1}{M} \sum_{i=1}^m (e_m - a_i a_m) \quad (8)$$

其中, l 和 m 分别对应两个异质类型社区。 Q 值越大,表示二分网络具有越强的社区结构。

Roger等给出了评估一种类型节点的社区划分质量的二分网络模块度^[37],描述的是同种类型节点集合中的节点对的共同邻居数占异类型节点总数比例的实际值与期望值的差值大小,用社区内两个同类型节点的共同邻居个数来表示该节点对的连通程度,计算方法如下:

$$Q_b = \sum_{s=1}^k \left\{ \frac{\sum_{i,j \in S} c_{ij}}{\sum_a m_a (m_a - 1)} - \frac{\sum_{i,j \in S} t_i t_j}{\left(\sum_a m_a \right)^2} \right\} \quad (9)$$

其中, U 和 V 是节点集合, a 是 U 类型节点, m_a 是连接节点 a 的边数, c_{ij} 是节点 i 和节点 j 连接 V 类型社区的个数, t_i 和 t_j 分别是节点 i 和节点 j 连接 V 类型节点的个数。

Suzuki 等在 Murata 二分模块度的基础上,提出了一种异质社区间多对多关系的二分模块度^[38],即通过任意两个异质社区之间连接紧密程度的加权来表示它们之间连接的紧密程度,计算方法如下:

$$Q_s = \frac{1}{2} \sum_{l,s} (e_{lm}/a_l)(e_{lm} - a_l a_m) \quad (10)$$

其中, e_{lm}/a_l 表示社区 l 连接紧密程度的加权重值。

3.2 其他评价指标

NMI(归一化互信息)是目前广泛使用的一种社区划分评价指标^[14],用于度量算法所得到的社区划分和真实社区划分之间的相似程度,计算方法如下:

$$NMI(R, F) = \frac{-2 \sum_{i=1}^{C_R} \sum_{j=1}^{C_F} N_{ij} \log\left(\frac{N_{ij} S}{N_{i*} N_{*j}}\right)}{\sum_{i=1}^{C_R} N_{i*} \log(N_{i*}/S) + \sum_{j=1}^{C_F} N_{*j} \log(N_{*j}/S)} \quad (11)$$

其中,矩阵 \mathbf{N} 中的行表示真实的社区,列表示算法所得到的社区;矩阵中第 i 行元素表示为 N_{i*} ,第 j 列元素表示为 N_{*j} ; N_{ij} 表示真实社区与算法所得到的社区相同的节点个数;所有元素之和记为 S ; C_R 表示真实社区个数; C_F 表示算法所得到的社区个数。当算法所得到的社区划分与网络中的真实划分完全一致时, NMI 值为 1; 当算法所得到的社区划分与网络的真实划分完全独立时, NMI 值为 0。

4 静态网络社区发现

4.1 同质网络社区发现

4.1.1 全局社区发现

复杂网络社区发现的研究具有重要的理论意义和实际应用价值。2002年, Girvan 等提出了 GN 算法^[3], 通过不断删除边介数最高的边来获取网络中的社区结构, 掀起了社区发现研究的热潮, 得到了计算机学科、社会学科、物理学科等众多领域的关注。GN 算法由于计算边介数的时间复杂度高, 因此被应用到较大规模的网络。针对该问题, 学者们提出了一些改进方法。2003年, Tyler 等提出了一种改进的基于统计方法的 GN 算法^[5], 其采用蒙特卡洛方法近似估计部分边的边介数, 通过牺牲算法的精度来换取计算速度的提高。2010年, Jin 等采用结构化相似度取代 GN 算法中的边介数^[7], 从网络中移除结构化相似度最小的边, 得到层次化的网络聚类结果。

以上算法得到的都是层次化的聚类结果, 为了从中选出一个最佳的聚类, 2004年, Newman 等提出第一个评价社区划分优劣的度量指标——模块度 Q 函数^[8], 并提出了一种基于模块度 Q 的社区挖掘算法^[9], 即从多个聚类结果中选择模块度 Q 最高的聚类作为算法的结果。模块度 Q 作为衡量社区划分优劣的度量指标得到了学者们的广泛认可, 进而一些基于模块度优化的社区发现方法被提出^[10]。针对模块度 Q 仅适用于无向无权网络的问题, 2004年, Newman 提出了加权网络的模块度 Q_w ^[11]; 2009年, Nicosia 等提出了适用于有向网络的社区评价指标 Q_d ^[12], 并给出了相应的社区发现算法。2016年, Liu 等在网络中节点间的连接关系和节点的影响力基础上, 提出了一种基于核心节点扩展的社区挖掘算法^[14]。

上述社区发现算法都是基于网络中的节点只归属于一个

社区的假设。在真实世界中, 我们知道任何一个人都可能归属于不同的圈子。因此, 网络中存在重叠社区。2005年, Palla 充分利用完全图节点间连接紧密的特性, 通过搜索邻接团探测网络中的重叠社区结构, 提出了著名的团渗算法 (CPM)^[6]。2009年, Shen 等基于合并相似度社区来构造层次聚类树, 提出了一种基于极大团凝聚式的层次聚类算法^[13], 该算法同时可以发现网络中的重叠社区。

针对同质网络中的社区发现, 也有一些学者对其进行了综述。2010年, 社区发现研究的著名学者 Fortunato 对分裂式、模块度优化类算法进行了综述^[15]。2013年, Xie 等对重叠社区发现算法进行了综述^[16], 对经典重叠社区发现算法进行了对比分析, 结果表明 SLPA, OSLOM, Game 和 COPRA 算法在低密度重叠网络中的性能更好, SLPA 和 Game 算法在高密度网络中性能相对更好。2013年, Liu 等分别对同质网络中的非重叠社区发现算法 FN, FEC, LPA, infomap, FUA, MABA, 以及重叠社区发现算法 CPM, LFM 和 UEOC 进行了实验比较, 得出非重叠社区发现算法 infomap 和 FUA 算法、重叠社区发现算法 LFM 和 UEOC 的效果较好^[17]。2016年, Yang 等使用 LFR 标准网络比较测试了 8 种重叠社区发现算法^[18]。2016年, Fortunato 等对同质网络中当前的流行方法进行了较为全面的综述^[19], 涉及到社区定义的演化, 用于测试社区发现算法的基准数据集, 基于谱方法、模块度优化、动态网络的社区发现算法, 以及各类社区发现软件包, 为初学者提供了很好的指导。

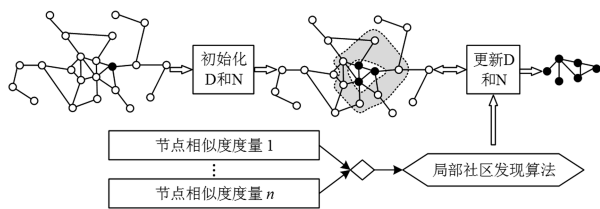
4.1.2 局部社区发现

全局社区发现能够找出网络中的所有社区, 有助于我们理解网络的构成, 但这些算法都是建立在已知网络的整体结构的基础上。然而, 现实中由于一些网络规模太大或动态变化, 很难获取其整体网络结构, 例如万维网、大型社交网络。虽然这些网络的整体结构很难获得, 但是可以获取某个节点周围的局部网络结构信息, 由此基于局部网络结构的社区发现(即局部社区发现)应运而生, 获得了很多学者的关注。

2005年, Bagrow 等提出了一种基于广度搜索的局部社区结构发现算法^[20], 该算法通过寻找起始节点的 l 层外壳节点来获得其所在的局部社区。该算法的缺点是需要一个预定义阈值来停止局部社区的扩张, 此外只有当起始节点处于社区中心时该算法才有效。2006年, Clauset 提出了局部社区模块度 R ^[21], 采用边界节点与社区内节点连接的边数占连接边界节点的总边数的比例来度量一个社区的优劣。其通过优化局部社区模块度 R , 提出了一种局部社区发现算法。该算法从起始节点出发, 通过从邻居节点集中选取能够使得模块度 R 增量最大的节点并将其加入局部社区的方法来计算目标社区。2006年, Luo 等提出了另一种局部社区模块度 M ^[22], 即用社区内节点形成的边数与社区外节点连接的边数的比值来度量一个社区的优劣。由于定义 M 值大于 1 的子图才为社区, 导致在社区内连接相对稀疏时, 返回的子图不满足 M 大于 1, 从而使得算法的查全率低。2013年, Chen 等提出了一种基于局部中心节点的局部社区发现算法^[23], 将起始节点转换为与其关联的局部中心节点, 然后基于局部中心节点进行局部扩展。2014年, Yang 等提出了一种基于边界节点识别的复杂网络局部社区发现算法^[24], 该算法全面比较符合并节

点的连接相似性并进行节点聚类,通过边界节点识别并控制局部社区的规模和范围,来获取给定节点所属的社区。2015年,Li等提出了一种基于谱方法的局部社区发现算法^[25]。

以上局部社区发现算法基于网络的拓扑结构,认为网络中的边具有同等权重。而在真实世界中,实体之间的联系强弱是不同的,例如社交网络中,每个用户与其众多好友的亲密度是不同的。但在将这些网络建模为图的过程中,由于节点对之间的相似度不易度量等原因,忽略了这些信息。因此,研究者利用网络中节点对之间的相似度,将社区定义为相似度高度的节点集合,产生了一类基于相似度的局部社区发现算法。为了更好地度量节点对之间的相似度,2013年,Tabarzad等^[26]和Ma等^[27]提出采用多层邻居节点计算节点间相似度;2016年,Zhao等提出基于带权邻居节点的节点相似度^[28]。2017年,Liu等提出一种基于节点对相似度的局部社区发现框架^[29],通过嵌入更好的节点相似度度量可以获得新的局部社区发现算法。该框架如图6所示。



注:来自文献^[29]

图6 局部社区发现框架

Fig.6 Local community detection framework

4.2 异质网络社区发现

相比同质网络的社区发现研究,异质网络的研究目前相对较少。在现实世界中,很多复杂系统都呈现出异质特征。在复杂多样的异质网络中获取社区结构信息,是社区发现研究面临的一个新挑战。

4.2.1 二分网络社区发现

二分网络是最简单的异质网络,现实世界中存在着大量二分网络,例如科学家论文协作网、演员影视作品合作网、投资者股份公司合作网、俱乐部会员活动参与网、听众歌曲网等。这些网络的一个典型特征是其由两类节点构成,不同类型的节点间有边连接,而同类型的节点间大多没有边连接。本文将两类节点命名为用户节点和资源节点。目前二分网络的社区发现有两种方案:1)将二分网络映射成同质网络,然后应用较为成熟的同质网络社区发现算法进行社区划分,但其在映射成同质网络的过程中容易丢失信息,从而导致社区划分出现偏差;2)直接在二分网络上进行社区识别。当前最主要的两种方法分别是边集聚系数法和双派系社团的划分方法。采用边集聚系数法的层次聚类方法,把二分模块度作为算法的停止条件。2007年,Barber提出了一种基于二分模块度的Brim算法^[30],该算法主要适用于对小规模的二分网络进行社区检测。2008年,Zhang等提出二分网络边集聚系数的定义,并基于此定义提出了二分网络社区聚类方法^[31]。该算法的过程类似于同质网络中的GN算法,重复计算所有边的集聚系数并删除集聚系数最小的边,直到网络的模块度达

到局部最大值。这种算法在网络社区结构比较明显时有较好的效果,但是需要根据模块度来确定社团的个数,且该算法仅考虑了连接超过2次的节点。单依靠模块度指标对节点进行贪婪搜索容易导致社区聚类陷入局部最优,难以有效地揭示二分网络的社区结构。2008年,Schwartzm等提出一种二分团社区划分算法^[32],该算法基于Palla的重叠社团概念及派系过滤算法,定义了二分网上的派系及相应社团。该算法允许社区重叠,具有很好的精确度和稳定性,并且参数可调,具有灵活性;不足之处是对于不同网络,该算法的可调参数不容易确定。2007年,Raghavan等提出了基于标签传播的社区检测算法LPA^[33],其缺点是二分结构会引起节点标签振荡、稳定性差等问题。针对该问题,2010年,Liu等对标签传播的方法进行了改进,对二分网络进行了有效的划分^[34]。2015年,Xu等提出一种基于密度的二分网络模块度,同时考虑了二分网络中社区内连接数目和节点的数量,解决了以往二分网络社区模块度无法识别较小规模社区的问题^[35]。2016年,Alzahrani等概述了该领域的研究进展^[36]。

4.2.2 多分多维网络社区发现

相比二分网络,三分网络或多分多维的异质网络社区发现算法的研究相对更少。在现实世界中,多分多维的异质网络是存在的,这类网络包含多种类型的节点和边,网络结构更为复杂。2005年,Cai等考虑到复杂网络中的多种关系以及不同网络关系的重要性,通过线性合并多关系的方法,首次提出在多维异质网络上进行社区发现^[39]。2010年,Neubauer等提出在社会媒体标注网络中的社区发现算法^[40],该算法是在扩展Murata的二分模块度的基础上,给出了一种基于映射的三分模块度,然后基于三分模块度进行社区识别。2010年,Murata探讨了从三分网络中识别社区结构^[41]。2011年,Murata给出了一种新的三分模块度,主要探讨了由用户、标签和资源组成的三分异质社会媒体网络的模块度,该异构网络模块度不需要映射信息,他们通过实验证明了基于优化的三分模块度可以得到较好的社区结构^[42]。2011年,Murata证明了基于三分模块度,能够从大型真实社会媒体网络中识别出社区^[43]。2011年,Xin等介绍了一种多分多维网络的社区发现算法,提出了一个度量函数来评价多分多维网络社区划分的优劣^[44]。2012年,Papadopoulos等针对社会媒体网络的复杂性和动态性,探讨了社会媒体中社区概念和社区识别的问题,并针对几种典型异质网络社区识别算法从理论和实验上分别进行了对比分析^[46]。2011年,Tang等给出能够整合多个数据源的社区发现算法,首先提出了整合两个数据源的社区发现算法,在此基础上扩展得到整合多个数据源的社区发现算法^[47]。2016年,Bangcharoensap等探讨了多模网络上的基于边界标准化的转化分类方法^[48],该算法在传统边系数的计算上做了两个额外改进。

4.3 结合节点内容和连接结构的社区发现

当前社区发现方法主要基于网络的拓扑结构特征,忽略了网络中的节点属性所附带的内容。而传统聚类技术大多关注的是节点属性。因此,无论是网络的拓扑结构,还是网络中的节点属性,都有助于识别网络中的社区。如何将网络的连接结构与节点属性有机地结合起来,是当前社区发现研究的

热点问题。自 2009 年,多篇结合网络连接结构和节点属性的社区发现方面的文献出现在 VLDB, KDD, WWW 和 ICDM 等重要的国际会议上。2009 年, Zhou 等提出结合连接结构和属性相似性的图分类方法^[49], 通过调整结构相似性和属性相似性的贡献度, 在随机游走中学习不同属性的贡献度。2010 年, Zhou 等在文献^[49]的基础上提出了一种在属性图上聚类的增量算法^[50]。2011 年, Cheng 等给出一种在矩阵计算上的优化技术改进的 SA 算法^[51]。2012 年, Xu 等提出了一种面向属性图的贝叶斯概率模型^[52], 他们将图聚类问题转换成概率推理问题。2012 年, Tang 等结合网络中的结构和属性信息来进行社区发现^[53], 获得了较高的准确度。2013 年, Ruan 等介绍了一种度量方法来融合连接相似性和内容相似性^[54], 其首先将原始网络图进行转换, 然后采用经典社区发现算法 metis 算法或马尔可夫算法进行社区划分。2013 年, Yang 等给出一种结合网络的边和节点属性的 CESNA 算法^[55], 实验表明该算法能够提高社区发现的准确性。2009 年, Yang 等通过合并基于连接的社区发现模型和节点属性社区发现模型, 形成一种基于连接和节点内容的社区发现算法, 提高了社区发现的准确度^[56]。2016 年, Zhang 等结合网络边信息和节点属性, 提出了一种社区发现新算法^[57], 并在真实网络和模拟网络上验证了该方法的有效性。2015 年, Liu 等考虑到社区结构是节点交互的结果, 将影响力传播和随机游走应用到具有节点属性的异构网络社区发现中^[58]。2015 年, Bothorel 等将具有节点属性的网络定义为属性图, 将在属性图的社区发现称为属性图聚类, 并对当前的典型属性图上的聚类方法及评价标准进行了介绍、比较和分析^[59]。2016 年, Newman 等对具有节点属性的网络结构进行了剖析, 为社区发现研究者提供了很好的指引方向^[60]。2016 年, Li 等利用连接和内容信息, 提出了一种基于非负矩阵分解的社区发现算法^[61]。

以上研究成果证明了结合网络中的节点属性和网络结构进行社区划分, 能够得到更加准确的划分结果。但在很多情况下, 由于隐私保护等原因, 网络中节点属性难以获取或不可获取, 即使获取到也常常伴随着大量的噪声数据, 相比之下, 网络结构更容易获得, 也更加可靠^[72-73]。因此, 在社区发现任务中, 是否结合节点属性的内容信息, 要结合实际问题进行选择。如果网络中节点属性内容信息容易获取, 且含有很少或者没有噪声数据, 那么结合网络中的节点属性内容信息, 一般会得到更加准确的社区划分结果。

5 动态网络社区发现

当前, 学者主要致力于静态网络的社区发现研究, 然而现实世界中很多网络都是动态变化的。在动态网络中, 节点和边是可变的, 新节点和关系的产生以及已有节点和关系的消失对网络的社区结构都会产生影响^[62]。然而, 传统的社区发现研究是把现实网络抽象成静态网络, 忽略了不同时间点节点和关系的变化情况。而现实中的网络很多都是动态的, 随着网络中节点和关系的变化, 社区结构也会发生变化。因此, 动态网络的社区发现具有重要的研究意义, 是当前复杂网络社区研究的一个难点问题。

动态社区发现的主要工作是识别出动态复杂网络中不同

时间段的社区结构, 动态网络社区发现在影响力分析、事件监测、舆情传播、网络营销等实际应用中具有重要的应用价值^[67]。2008 年, Tang 等从社区评价方法入手, 将单一关系网络的社区模块度评价扩展为多关系网络的综合模块度评价^[63]。Zhang 等针对 Web 社会网络中存在的确定-不确定关系, 提出了一种给定阈值的 α 关系社区概念, 分别设计了静态和动态 α 关系社区的挖掘算法^[64], 该算法的不足之处是缺少社区结构的质量评价和验证。Lin 等对多关系社会网络动态社区发现的性能问题进行了研究^[65], 提出在动态社区计算中仅对算法生成的上一时间快照上的层次化社区树的分枝进行选择更新, 以减少计算量, 从而提升动态社区发现的时间效率。2014 年, Suo 等提出一种基于信息流动分析的动态社区发现方法^[66]。2015 年, Wang 等从“同构社会网络的动态社区研究”和“异构社会网络的动态社区研究”两个方面, 对国内外相关研究的现状进行了阐述和分析, 并对动态社区领域的难点和发展趋势进行了展望^[67]。当前的动态网络社区发现大多是针对网络的连接结构, 结合节点连接结构和内容信息的社区发现算法主要集中在静态网络, 基于此, Wang 等于 2013 年提出了结合动态网络连接结构和节点内容信息的社区发现算法^[68], 该算法给出了一种新颖的转换方法, 将节点内容和连接内容无缝嵌入到一起, 并基于转移概率矩阵设计了一种 NEI 网络, 进而给出在 NEI 网络上的随机游走 (NEI-Walk) 方法。2017 年, Chen 等为了高效地发现大规模动态社会网络的社区结构, 提出了一种基于邻域跟随关系的社区表示模型 Follow-Community, 并基于该模型提出了一种具有接近线性时间复杂度的邻域跟随算法 NFA, 然后进一步扩展得到增量邻域跟随算法 iNFA, 通过更新网络演化过程中相关节点的邻域跟随关系, 发现动态社会网络的社区结构及社区演化^[69]。2017 年, Meng 等针对以往动态网络社区发现算法需要先验知识的问题, 提出了一种基于模块度优化的新颖的动态网络社区发现算法^[70], 该算法通过考虑历史网络结构, 可以自动确定社区个数。他们将该算法应用于时态金融网络, 证明了其有效性。2017 年, Wang 等结合结构扰动和拓扑特征相似性, 提出了一种基于时间平滑框架的动态聚类算法^[71], 并通过实验证明了该算法相比基于 SP, RA 的传统方法具有优越性。2018 年, Zhou 等针对传统的动态网络社区发现需要输入参数控制快照社区的偏好程序和时间开销问题, 为了解决参数选择的局限性, 提出将多目标离散化蚁群算法 (MDBA) 用于动态网络社区发现^[74], 实验结果表明 MDBA 算法具有较高的准确率。

动态网络社区发现研究的是不同时段不断动态变化的社区结构, 尤其是找出动态网络中稳定的社区结构。而社区演化是观察不同时间片段中社区结构的变化情况。两者在研究目标和方法上有所差异, 但共同点都是发现不同时间点或时间窗口的社区信息和社区变化情况。因此, 当前许多研究工作不对这两类研究进行严格区别。目前, 动态网络社区发现和社区演化的研究还处于起步阶段, 还没有成熟的理论架构, 如何建立有效的模型将成为未来的研究趋势。

6 社区发现的主要应用

当前社区发现的研究更多集中在算法研究, 社区发现的

应用研究相对较少。目前社区发现在影响最大化、疫情分析、商品推荐、链路预测和情感分析等领域有着重要作用。例如在线社会网络舆情的传播过程中,通过分析在线话题社区的内容和传播规律,可以发现消极舆论的传播趋势和领导舆论的核心节点,能够即时控制消极舆论的扩散,降低舆论带来的负面影响。

6.1 基于社区发现的影响最大化

影响最大化是社会网络分析和病毒营销领域研究的经典优化问题。影响最大化研究的是一组节点使影响传播最大化,在现实生活中具有重要意义。影响最大化的目标是在一定的传播模式下,找到能够使影响力传播最大化的关键用户的子集。社会网络中,由于社区结构具有内部节点连接紧密、社区间节点连接稀疏的特性,因此社区内部信息比社区之间的信息扩散得更快,能够更快地实现信息传播。因此,一些学者开始尝试将社区发现与影响最大化相结合进行研究,并取得了较好的效果。

2010年,Cao等结合社区性质和动态规划的思想,提出了一种将影响最大化问题当作最佳资源动态分配问题进行处理的方法。2010年,Wang等利用动态规划方法和Mixgreedy算法相结合的思想,提出了一种解决影响最大化问题的CGA算法^[75]。2011年,Wang等提出了一种新的基于社区结构的影响最大化方法AMICS^[77]。该方法先利用已有的社区挖掘算法识别出隐藏在网络中的社区结构,然后迭代选择跨越社区数最多的 k 个节点作为影响的初始传播者来最大化影响的社区覆盖。该算法在小型网络和中等规模网络数据集上具有更好的效果。2013年,Lv等提出了一种基于社区的影响最大化算法(CIP)^[78],他们认为网络中的信息传播如同具有转移效用的博弈,将其中每个社区看作一个成员,综合社区内的节点个数总和及每个节点在其他社区内影响度之和作为每个社区的沙普利值,据此再进行种子节点的选取。该算法也适用于大型网络。2017年,Shang等提出一种基于社区的影响最大化框架CoFIM^[79],该框架的影响传播过程主要分为种子节点在不同社区间扩展和影响力在社区内传播两个阶段,适用于大规模网络。2017年,Bozorgi等针对现实场景中的竞争影响最大化问题,提出了一种新的传播模型^[80],该模型是线性阈值模型的扩展,并赋予节点关于传入影响的决策能力,以寻找给定社会图中有影响力的节点,利用该图的社区结构计算每个节点在其自身社区内的本地传播。

6.2 基于社区发现的链路预测

链路预测作为数据挖掘领域的重要研究方向之一,具有重要的现实意义,例如可以运用于电子商务平台的个人推荐系统来解决信息过载的问题^[81]。链路预测是指如何通过已知的网络节点以及网络结构等信息,来预测网络中尚未产生连边的节点对之间可能出现连边的可能性^[82]。早期的链路预测工作只考虑连接模式或节点属性。由于社区内的节点比社区之间的节点的连接更加紧密,预测节点对处于所在社区内部受邻居节点的影响要比处于其他社区的邻居节点受到的影响大很多^[83]。近几年,一些学者设计了一些基于社区信息的链路预测算法^[83-86],以提高链路预测的精确度和推荐性能。

6.3 社区发现在情感分析领域的应用

近年来,社交媒体作为一种新的在线交互平台,吸引了海

量的用户,这些用户不断传播和表达各种各样的信息,而这些信息常包含着多样的情感信息,这些情感对网络世界和真实世界都有着重要影响。挖掘与分析这些用户产生的情感信息具有重要的理论意义与应用价值,得到了学术界和工业界的广泛关注。传统的情感分析研究工作主要是从文本中分析情感,但忽略了用户形成的情感网络以及形成的情感社区对情感分析的影响。在线社交网络中节点或群体间相互影响、作用、依赖,不断传递着多样的信息,包括各种情感信息。基于这些多样化的社交关系,常形成各种各样的虚拟网络社区。在虚拟社区内用户之间转发的信息情感往往比不同社区之间用户转发的信息情感更加相似。基于此,吴斌等提出了一种虚拟的情感网络模型^[87]。在该模型中,将以用户群体以及他们之间的“关注”关系为基础的用户情感网络用于发现情感社区;将微博数据以及它们之间的“转发”关系所得到的微博转发情感网络用于验证社区发现结果是否合理。

7 社区发现面临的挑战

本文基于不同类型的网络结构,介绍了社区发现的研究进展。然而,随着大规模在线复杂社会网络的出现,传统的社区发现方法面临着一些新的挑战。

(1)异质性。现实网络往往拥有丰富多样的异质信息,例如在线社会媒体网络,除了包含用户和他们之间的相互关系,还包含用户发布的微博、转发的消息、图片、视频等实体信息;学术网络除了包含研究者和研究者之间的相关关系,还包含论文、期刊等额外的文本、标签信息。当前的社区发现方法大多不适用于这些异质网络。

(2)动态性。现实中复杂网络不仅异质,而且还是动态变化的,常用来模拟复杂系统的演化。在动态网络中,节点和边都是可变的,节点和关系的生成或消失都深深地影响着网络中的社区结构。

(3)计算效率。随着大规模社会网络的出现,人们对社区发现算法的效率有着更高的要求。传统的社区发现方法分为两类,一类是直接基于复杂网络进行计算,另一类是基于矩阵分解或矩阵特征向量的方法,在实际应用场景中,这些方法的存储效率和计算效率都较低,难以适用于大规模的复杂网络。

(4)节点的表示学习。对于复杂网络中节点的表示,传统的方法是采用人工的特征提取方法。近几年,基于深度学习的网络表示学习方法引起了复杂网络分析研究者的关注。该方法是将复杂网络中的节点利用低维的实值向量来表示,可用传统的聚类方法来解决社区发现问题。同时,基于深度学习的网络表示学习方法使用梯度下降优化算法,往往只需要使用有限的存储空间,支持并行或者在线训练。因此,结合网络表示学习进行社区发现任务是当前研究的一个新挑战^[88-90],也是面对大规模网络挑战的一个突破口。未来,将会得到更多学者广泛而深入的研究。

结束语 近年来,随着社交媒体网络的普及和应用,社会网络呈现出大规模、动态变化和富含内容信息等特性,国内外很多学者在社区发现研究上做了大量工作。目前,在社区发现方向进行较多研究的有Fortunato、Newman、刘欢、Murata、于旭、金弟、何东晓、沈华伟等所在的课题组。有的学者还提供了相关研究网站与大家交流学习,如Newman课题组的

研究成果¹⁾、刘欢课题组的研究成果²⁾。此外,还有社区发现常使用的数据集,例如 SNAP 数据集³⁾、Pajek 网络数据集⁴⁾、CCNR 数据集⁵⁾、IMC 2007 数据集⁶⁾、KDD Cup 2003 数据集⁷⁾。

目前,社区发现的研究还处于初期阶段,社区发现算法的评价标准没有统一,对于同一个网络,用不同社区发现算法会得到不同的社区结构,不同评价标准也会得到不同的最优社区结构。本文对国内外社区发现研究进行了较为全面的阐述,但由于能力有限,总结得还不够全面、深入。异质信息网络上的动态社区发现还有待深入研究,同时社区发现应用研究也有待进一步的深入探讨。

参 考 文 献

[1] BARABASI A L, ALBERT R. Emergence of Scaling in Random Networks[J]. *Science*, 1999, 286(5439): 509-512.

[2] WATTS D J, STROGATZ S H. Collective Dynamics of Small-world Networks[J]. *Nature*, 1998, 393(6684): 440-442.

[3] GIRVAN M, NEWMAN M E J. Community Structure in Social and Biological Networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821-7826.

[4] CHENG X Q, SHEN H W. Community Structure of Complex Networks[J]. *Complex Systems & Complexity Science*, 2011, 8(1): 57-70.

[5] TYLER J R, WILKINSON D M, HUBERMAN B A. E-Mail as Spectroscopy: Automated Discovery of Community Structure within Organizations[J]. *The Information Society*, 2005, 21(2): 143-153.

[6] PALLA G, DERENYI I, FARKAS T, et al. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society[J]. *Nature*, 2005, 435(7043): 814-818.

[7] JIN D, LIU J, JIA Z X, et al. K-nearest-neighbor Network based Data Clustering Algorithm[J]. *Pattern Recognition & Artificial Intelligence*, 2010, 23(4): 546-551.

[8] NEWMAN M E J, GIRVAN M. Finding and Evaluating Community Structure in networks[J]. *Physical Review E*, 2004, 69(2): 026113-1-026113-16

[9] NEWMAN M E J. Fast Algorithm for Detecting Community Structure in Networks[J]. *Physical Review E*, 2004, 69(6): 066133-1-066133-5

[10] SHANG R H, BAI J J, LI C, et al. Community Detection based on Modularity and an Improved Genetic Algorithm[J]. *Physica A Statistical Mechanics & Its Applications*, 2013, 392(5): 1215-1231.

[11] NEWMAN M E J. Analysis of Weighted Networks[J]. *Physical Review E*, 2004, 70(5): 056131-1-056131-9.

[12] NICOSIA V, MANGIONI G, CARCIOLO V, et al. Extending the Definition of Modularity to Directed Graphs with Overlapping Communities[J]. *Journal of Statistical Mechanics Theory and Experiment*, 2009, 3: 3166-3168.

[13] SHEN H, CHENG X, CAI K, et al. Detect Overlapping and

Hierarchical Community Structure in Networks[J]. *Physica A*, 2009, 388(8): 1706-1712.

[14] LIU J L, WANG D L, ZHAO W J, et al. A Community Mining Algorithm based on Core Nodes Expansion[J]. *Journal of Shandong university(Natural Science)*, 2016, 51(1): 106-114.

[15] FORTUNATO S. Community Detection in Graphs[J]. *Physics Reports*, 2010, 486(3): 75-174.

[16] XIE J, KELLEY S, SZYMANSKI B K. Overlapping Community Detection in Networks: The state-of-the-art and Comparative Study[J]. *Acm Computing Surveys*, 2013, 45(4): 1-35.

[17] LIU D Y, JIN D, HE D X, et al. Community Mining in Complex Networks[J]. *Journal of Computer Research & Development*, 2013, 50(10): 2140-2154.

[18] YANG Z, ALGESHEIMER R, TESSONE C J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks[J]. *Scientific Reports*, 2016, 6: 30750.

[19] FORTUNATO S, HRIC D. Community Detection in Networks: A User Guide[J]. *Physics Reports*, 2016, 659: 1-44.

[20] BAGROW J P, BOLLT E M. Local Method for Detecting Communities[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2005, 72(2): 046108.

[21] CLAUSET A. Finding Local Community Structure in Networks[J]. *Physical Review E*, 2005, 72(2): 026132.

[22] LUO F, WANG J Z, PROMISLOW E. Exploring Local Community Structures in Large Networks[C]// *IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE, 2006: 233-239.

[23] CHEN Q, WU T T, FANG M. Detecting Local Community Structures in Complex Networks Based on Local Degree Central Nodes[J]. *Physica A Statistical Mechanics & Its Applications*, 2013, 392(3): 529-537.

[24] YANG L, JI X S, LIU C X, et al. Detecting Local Community Structure Based on the Identification of Boundary Nodes in Complex Networks[J]. *Journal of Electronics & Information Technology*, 2014, 36(12): 2809-2815.

[25] LI Y X, HE K, BINDEL D, et al. Uncovering the Small Community Structure in Large Networks: A Local Spectral Approach[J]. *Computer Science*, 2015, 36(6): 658-668.

[26] TABARZAD M A, HAMZEH A. A Heuristic Local Community Detection Method (HLCD) [J]. *Applied Intelligence*, 2017, 46(1): 62-78.

[27] MA L H, HUANG H, HE Q, et al. GMAC: A Seed-Insensitive Approach to Local Community Detection[C]// *The 15th International Conference of Data Warehousing and Knowledge Discovery*. Prague Czech Republic: Springer Press, 2013: 297-308.

[28] ZHAO W J, ZHANG F B, LIU J L. Local Community Detection via Edge Weighting[C]// *Asia Information Retrieval Symposium*. Springer International Publishing, 2016: 68-80.

[29] LIU J L, WANG D L, ZHAO W J, et al. A Unified Framework of Lightweight Local Community Detection for Different Node Similarity Measurement[C]// *Chinese National Conference on*

¹⁾ <http://www-personal.umich.edu/~mejn/>

²⁾ <http://www.public.asu.edu/~huanliu/>

³⁾ <http://snap.stanford.edu/>

⁴⁾ <http://vlado.fmf.uni-lj.si/pub/networks/data/>

⁵⁾ <http://www.nd.edu/networks/resources.htm>

⁶⁾ <http://socialnetworks.mpi-sws.org/data-imc2007.html>

⁷⁾ <http://www.cs.cornell.edu/projects/kddcup/datasets.html>

- Social Media Processing. Springer International Publishing, 2017;283-295.
- [30] BARBER M J. Modularity and Community Detection in Bipartite Networks[J]. *Physical Review E*,2007,76(6):066102.
- [31] ZHANG P,WANG J L,LI X J, et al. Clustering Coefficient and Community Structure of Bipartite Networks[J]. *Physica A*, 2008,387(27):6869-6875.
- [32] SCHWARTZ M L S,HANSEN L K. Bi-clique Communities[J]. *Physical Review E*,2008,78(1):108-118.
- [33] RAGHAVAN U N,ALBERT R,KUMARA S. Near Linear Time Algorithm to Detect Community Structures in Large-scale Networks[J]. *Physical Review E*,2007,76(3):036106.
- [34] LIU X,MURATA T. Evaluating Community Structure in Bipartite Networks[C]// *IEEE Second International Conference on Social Computing*. IEEE,2010;576-581.
- [35] XU Y,CHEN L,LI B, et al. Density-based Modularity for Evaluating Community Structure in Bipartite Networks[J]. *Information Sciences*,2015,317:278-294.
- [36] ALZAHIRANI T,HORADAM K J. Community Detection in Bipartite Networks: Algorithms and Case studies[M]// *Complex Systems and Networks*. Springer Berlin Heidelberg, 2016: 25-50.
- [37] ROGER G,MARTA S P,LUIS A, et al. Module Identification in Bipartite and Directed Networks[J]. *Physical Review E*, 2007,76(3):036102.
- [38] SUZUKI K,WAKITA K. Extracting Multi-facet Community Structure from Bipartite Networks[C]// *Proceedings of the Computational Science and Engineering*. Vancouver, Canada, IEEE Computer Society,2009;312-319.
- [39] CAI D,SHAO Z, He X F, et al. Mining Hidden Community in Heterogeneous Social Networks[C]// *Proceedings of the 3rd international workshop on Link discovery*. ACM,2005;58-65.
- [40] NEUBAUER N, OBERMAYER K. Community Detection in Tagging-Induced Hypergraphs[C]// *Workshop on Information in Networks*. NY, USA: New York University,2010:24-25.
- [41] MURATA T. Detecting Communities from Tripartite Networks[C]// *International Conference on World Wide Web*. Raleigh, North Carolina, USA; DBLP,2010;1159-1160.
- [42] MURATA T. A New Tripartite Modularity for Detecting Communities[J]. *Information & Media Technologies*,2011,28:154-161.
- [43] MURATA T. Detecting Communities from Social Tagging Networks Based on Tripartite Modularity[C]// *Proc. Workshop on Link Analysis in Heterogeneous Information Networks*. 2011;1-4
- [44] XIN L, TSUYOSHI M. Detecting Communities in K-Partite K-Uniform (Hyper) Networks[J]. *Journal of Computer Science and Technology*,2011,26(5):778-791.
- [45] JI M, HAN J, DANILEVSKY M. Ranking-based Classification of Heterogeneous Information Networks[C]// *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM,2011;1298-1306.
- [46] PAPADOPOULOS S,KOMPATSIARIS Y,VAKALI A, et al. Community Detection in Social Media[J]. *Data Mining and Knowledge Discovery*,2012,24(3):515-554.
- [47] TANG J L,WANG X,LIU H. Integrating Social Media Data for Community Detection[C]// *International Workshops MSM and MUSE*. 2011;1-20.
- [48] BANGCHAROENSAP P,MURATA T,KOBAYASHI H, et al. Transductive Classification on Heterogeneous Information Networks with Edge Betweenness-based Normalization[C]// *ACM International Conference on Web Search and Data Mining*. ACM,2016;437-446.
- [49] ZHOU Y,CHENG H,YU J X. Graph Clustering based on Structural/Attribute Similarities[J]. *Proceedings of the Vldb Endowment*,2009,2(1):718-729.
- [50] ZHOU Y,CHENG H,YU J X. Clustering Large Attributed Graphs: An Efficient Incremental Approach[C]// *2010 IEEE International Conference on Data Mining*. IEEE,2010;689-698.
- [51] CHENG H,ZHOU Y,YU J X. Clustering Large Attributed Graphs: A Balance between Structural and Attribute Similarities[J]. *Journal of Information Processing*,2011,5(2):12.
- [52] XU Z,KE Y,WANG Y, et al. A Model-based Approach to Attributed Graph Clustering[C]// *ACM SIGMOD International Conference on Management of Data*. ACM,2012;505-516.
- [53] TANG A,VIENNET E. Community Detection based on Structural and Attribute Similarities[C]// *International Conference on Digital Society (ICDS)*. 2012;7-12.
- [54] RUAN Y,FUHRY D,PARTHASARATHY S. Efficient Community Detection in Large Networks using Content and Links[C]// *International Conference on World Wide Web*. ACM, 2013;1089-1098.
- [55] YANG J,MCAULEY J,LESKOVEC J. Community Detection in Networks with Node Attributes[C]// *IEEE 13th International Conference on Data Mining*. IEEE,2013;1151-1156.
- [56] YANG T B,JIN R,CHI Y, et al. Combining Link and Content for Community Detection: a Discriminative Approach[C]// *KDD*. 2009;927-936.
- [57] ZHANG Y,LEVINA E,ZHU J. Community Detection in Networks with Node Features[J]. *Electronic Journal of Statistics*, 2016,10(2):3153-3178.
- [58] LIU L,XU L,ZHEN W, et al. Community Detection Based on Structure and Content: A Content Propagation Perspective[C]// *IEEE International Conference on Data Mining*. IEEE, 2015;271-280.
- [59] BOTHOREL C,CRUZ J D,MAGNANI M, et al. Clustering Attributed Graphs: Models, Measures and Methods[J]. *Network Science*,2015,3(3):408-444.
- [60] NEWMAN M E,CLAUSET A. Structure and Inference in Annotated Networks[J]. *Nature Communications*, 2016,7(2): 11863.
- [61] LI Z,PAN Z,ZHANG Y, et al. Efficient Community Detection in Heterogeneous Social Networks[J]. *Mathematical problems in engineering*,2016;1-15.
- [62] ROSSETTI G,CAZABET R. Community Discovery in Dynamic Networks: a Survey[J]. *ACM Computer Survey*,2017,51(2).
- [63] TANG L,LIU H,ZHANG J, et al. Community Evolution in Dynamic Multi-mode Networks[C]// *Proc. 14th ACM Sigkdd International Conference on Knowledge Discovery & Data Mining*. Las Vegas, NV, USA,2008;677-685.
- [64] ZHANG C Y, GUO J F. The α Relationship Communities of Set Pair Social Networks and Its Dynamic Mining Algorithms[J]. *Chinese Journal of Computers*,2013,36(8):1682-1692.
- [65] LIN W Q,DENG L,DING Z Y, et al. A new Dynamic Hierarchical Parallel Computing Community[J]. *Chinese Journal of*

- Computers, 2013, 35(8):1712-1725.
- [66] SUO B, LI Z H, CHEN Q, et al. Dynamic Community Detection based on Information Flow Analysis[J]. Journal of Software, 2014, 25(3):547-559.
- [67] WANG L, CHENG X Q. Dynamic Community in Online Social Networks[J]. Chinese Journal of Computers, 2015, 38(2):219-237.
- [68] WANG C D, LAI J H, YU P S. Dynamic Community Detection in Weighted Graph Streams[C]//Proceedings of the 2013 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, 2013:151-161.
- [69] CHEN Y Z, SHI S, ZHU W P, et al. An Incremental Community Discovery Algorithm Based on Neighborhood Following Relationship[J]. Chinese Journal of Computers, 2017, 40(3):570-583.
- [70] MENG X, TONG Y, LIU X, et al. A Novel Dynamic Community Detection Algorithm based on Modularity Optimization[C]//IEEE International Conference on Software Engineering and Service Science. IEEE, 2017:72-75.
- [71] WANG P, GAO L, MA X. Dynamic Community Detection based on Network Structural Perturbation and Topological Similarity [J]. Journal of Statistical Mechanics Theory & Experiment, 2017, 1(1):013401.
- [72] SUN Y Z, HAN J W, YAN X, et al. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks[J]. Proceedings of the VLDB Endowment, 2011, 4(11):992-1003.
- [73] HUANG L B, LI D Y, MA Y T, et al. A Meta-Path -Based Link Prediction Model of Heterogeneous Information Networks[J]. Chinese Journal of Computers, 2014(4):848-858.
- [74] ZHOU X, ZHAO X, LIU Y. A Multi Objective Discrete Bat Algorithm for Community Detection in Dynamic Networks[J]. Applied Intelligence, 2018, 48(9):3081-3093.
- [75] CAO T, WU X, WANG S, et al. OASNET: an Optimal Allocation Approach to Influence Maximization in Modular Social Networks[C]//Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 2010:1088-1094.
- [76] WANG Y, CONG G, SONG G, et al. Community-based Greedy Algorithm for Mining Top-k Influential Nodes in Mobile Social Networks[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010:1039-1048.
- [77] JI J C, HUANG L, WANG Z, et al. A New Approach to Maximizing the Spread of Influence Based on Community Structure [J]. Journal of JiLin University(Science Edition), 2011, 49(1):93-97.
- [78] LV J, GUO J, REN H. A Novel Community-Based Approach for Influence Maximization in Social Network[J]. International Review on Computers & Software, 2013, 8(1):43-49.
- [79] SHANG J, ZHOU S, LI X, et al. CoFIM: A community-based framework for influence maximization on large-scale networks [J]. Knowledge-Based Systems, 2017, 117:88-100.
- [80] BOZORGI A, SAMET S, KWISTHOUT J, et al. Community-based influence maximization in social networks under a competitive linear threshold model [J]. Knowledge-Based Systems, 2017, 134:149-158.
- [81] XIANG P. Link Prediction of Community Network[D]. Xi'an: Xidian University, 2015:1.
- [82] GUIMERÁ R, SALES-PARDO M. Missing and Spurious Interactions and the Reconstruction of Complex Networks[J]. Proceedings of the National Academy of Sciences, 2009, 106(52):22073-22078.
- [83] VALVERDE R J, ANDRADE L A. Link Prediction in Complex Networks-based on Cluster Information[C]//Proc of the 21st Brazilian Conference on Advances in Artificial Intelligence. 2012:92-101.
- [84] VALVERDEREBAZA J, LOPES A D A. Structural Link Prediction Using Community Information on Twitter[C]//Fourth International Conference on Computational Aspects of Social Networks. IEEE, 2013.
- [85] BISWAS A, BISWAS B. Community-based Link Prediction[J]. Multimedia Tools and Applications, 2017, 76(18):18619-18639.
- [86] SOUNDARAJAN S, HOPCROFT J. Using Community Information to Improve the Precision of Link Prediction Methods [C]//Proceedings of the 21st International Conference on World Wide Web. ACM, 2012:607-608.
- [87] 吴斌, 张云雷. 虚拟社区发现与演化[M]. 北京: 科学出版社, 2018:173-188.
- [88] YANG L, CAO X C, HE D X, et al. Modularity based Community Detection with Deep Learning[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. AAAI Press, 2016:2252-2258.
- [89] DING W, LIN C, ISHWAR P. Node Embedding via Word Embedding for Network Community Discovery[J]. IEEE Transactions on Signal and Information Processing over Networks, 2017, 3(3):539-552.
- [90] CAVALLARI S, ZHENG V W, Cai H, et al. Learning Community Embedding with Community Detection and Node Embedding on Graphs[C]//Proceedings of the 2017 ACM Conference on Information and Knowledge Management. ACM, 2017:377-386.



ZHAO Wei-ji, born in 1980, doctoral student, associate professor, is member of China Computer Federation (CCF). His main research interests include community detection and data mining.



ZHANG Feng-bin, born in 1965, Ph.D, professor, Ph.D supervisor, is member of China Computer Federation (CCF). His main research interests include network security and intrusion detection technology.