

文章编号:1672-3813(2017)03-0008-22;DOI:10.13306/j.1672-3813.2017.03.002

复杂网络模糊重叠社区检测研究进展

肖 婧¹,张永建¹,许小可^{1,2}

(1. 大连民族大学信息与通信工程学院,辽宁 大连 116600;2. 贵州省公共大数据重点实验室贵州大学,贵阳 550025)



摘要:模糊重叠社区检测通过扩展隶属度取值空间,实现了重叠节点与社区之间复杂且模糊隶属关系的精确化测量,不仅能够有效提升重叠社区结构检测的精确性,而且能够深度挖掘出节点和社区的重叠特性。文中首先分析了模糊重叠社区检测与传统离散重叠社区检测的关系;然后对二者的国内外相关研究现状进行阐述和分析,其中在模糊重叠社区检测方法研究中根据模糊隶属度获取方式的不同将当前相关研究分为扩展标签传播、非负矩阵分解、基于边界节点的两阶段检测、模糊聚类、模糊模块度优化五大类进行综述,重点分析了基于进化算法的模糊模块度优化方法;最后对模糊重叠社区检测研究未来的发展趋势进行了分析和展望。

关键词:社区检测;离散重叠;模糊重叠;模糊模块度优化;进化算法

中图分类号:TP399

文献标识码:A

Research Progress of Fuzzy Overlapping Community Detection in Complex Networks

XIAO Jing¹, ZHANG Yongjian¹, XU Xiaoke^{1,2}

(1. College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China;

2. Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China)

Abstract: Through expanding value space, fuzzy overlapping detection redefines the fuzzy membership degree, which can not only improve the detection accuracy of the complicated community structures, but also explore the overlapping features of nodes and communities. In this paper, we firstly give the explanation of the difference between crisp and fuzzy overlapping detection, and then summarize their related researches. To clearly state the fuzzy overlapping detection, we introduce the available work by dividing them into five classes on the acquisition method of fuzzy membership degree, including expanded label propagation, nonnegative matrix factorization, edge nodes based two-phase detection, fuzzy clustering and fuzzy modularity optimization. The advances and challenges of the fuzzy modularity optimization based on evolutionary algorithms are discussed in detail. At last some future research topics are given.

Key words: community detection; crisp overlap; fuzzy overlap; fuzzy modularity optimization; evolutionary algorithm

0 引言

复杂网络社区检测是指从复杂网络中挖掘出有实际意义的模块或层次结构的过程,能够提取出网络拓扑结

收稿日期:2016-12-12;修回日期:2017-03-01

基金项目:国家自然科学基金(61374170,61603073,61773091)、辽宁省自然科学基金(201602200)、辽宁省博士科研启动基金(201601294)、中央高校基本科研业务费专项基金(DC201502060201,DCPY2016002)、黑龙江省博士后科学基金(LBH-Z12073)

作者简介:肖婧(1985-),女,湖北十堰人,博士,讲师,主要研究方向为高维多目标优化和网络社团检测。

通讯作者:许小可(1979-),男,辽宁庄河人,博士,教授,主要研究方向为网络科学和社交网络大数据。

构特征,有助于理解和分析网络的拓扑特性、功能特性及动力学特性,挖掘出网络中蕴含的属性关联性 etc 深层次信息并对网络行为进行预测^[1-5],因此具有重要的理论研究价值。近年来,复杂网络社区检测在互联网基础设施优化、在线销售产品推荐、社交网络及生物网络分析、反恐组织识别、政治选举预测等诸多领域中的应用需求日益凸显^[6-7],因此对其进行研究具有重要的现实意义。

针对重叠社区结构的检测是复杂网络社区检测中的一种特殊形式。由于现实世界网络中节点具有天然的多社区归属特性^[8],因此“重叠”是大多数真实网络的重要特征。研究真实网络社区结构的重叠特性,对于探索网络中重要节点在拓扑及属性功能上的多面性特性,掌握重叠节点与社区之间的拓扑及属性关联性,挖掘社区之间的功能相似性及差异性等均具有重要意义,因此重叠社区检测成为近年来复杂网络社区检测领域的研究热点^[1-8]。

重叠社区结构中存在着一些同时隶属于多个社区的重叠节点,是重叠社区结构的重要特征,也是重叠社区检测的重要内容。一方面,重叠节点在拓扑及属性等特性上具有多样性,该特性是重叠社区结构形成的重要因素;另一方面,重叠节点是社区之间连接的桥梁,对社区之间的连接关系、信息交互及动态演化起到关键性作用^[9-10]。鉴于此,重叠社区检测的任务主要包括两方面:一是如何更加完整、精确地挖掘网络中的重叠节点及其社区分布,提升对于网络中重叠社区结构的检测能力;二是如何更加精确、细致地体现重叠节点与所属各社区之间的拓扑与属性关联性,并由此揭示重叠社区结构中隐含的深层次拓扑信息。

现实世界中复杂网络的重叠社区结构通常具有较强的复杂性和模糊性。复杂性表现在重叠节点可同时隶属多个社区,且不同重叠节点实际隶属的社区数目可能存在较大差异。模糊性主要体现在重叠节点与社区之间的隶属关系以及社区结构的重叠程度上,一方面节点对于不同社区的隶属程度存在较大差异,重叠节点与社区之间的隶属关系具有模糊性;另一方面真实网络中并没有严格的社区划分界限^[10],重叠社区的边界及不同社区之间的重叠程度均存在一定模糊性。实际网络环境下,根据检测分析需求和用户解释的不同,对重叠节点的判定准则和计算依据也存在较大差异,使重叠节点、重叠社区结构及节点和社区的重叠程度通常存在较大差异^[4,8],因此具有较大的模糊性。鉴于此,对于真实世界中复杂网络的重叠社区检测,一方面需要尽可能提升对于重叠社区结构的识别和提取能力,即重叠社区检测的精确性;另一方面,需要侧重于体现出网络中更细致化、多样化的社区结构信息,如量化并区分重叠节点与各社区之间的隶属程度,从而为不同环境及需求下真实网络重叠社区结构的挖掘与分析提供较为精确可靠的测度信息。通过在检测结果中体现出重叠社区结构的复杂性及模糊性,为网络功能及动力学特性分析、链路预测、社区动态演化预测等更深层次的复杂网络挖掘提供参考依据。

根据重叠节点与社区之间隶属程度的量化程度差异,可将复杂网络的重叠社区检测分为两类:即离散重叠(Crisp Overlap)社区检测和模糊重叠(Fuzzy Overlap)社区检测。目前国内外绝大多数重叠社区检测研究仅限于离散重叠社区检测^[1-5],相关检测方法的数量及种类较多,其中典型代表性方法包括:派系过滤、链接聚类、局部扩展、模块度优化、多目标优化及标签传播等。该类检测方法的前提假设为重叠节点对所属社区具有完全且一致的隶属关系(Full Membership),在此假设下虽然能够检测出重叠社区结构,但无法体现出重叠节点对不同社区不完全且非一致的隶属关系。因此,该种假设实质上是简化了对于真实重叠社区结构中复杂且模糊拓扑关系的刻画,限制了对于重叠社区结构内在隐含深层次拓扑信息的探索能力。

近年来,针对重叠社区检测的研究延伸出了新的分支,即模糊重叠社区检测(Fuzzy Overlapping Community Detection)。2011年Gregory针对社交网络的社区检测首次提出了“模糊重叠划分(Fuzzy Overlapping Partition)”的概念^[6]。模糊重叠社区检测与传统离散重叠社区检测的区别在于:允许重叠节点对所属社区具有不完全且不一致的隶属关系,利用 $[0,1]$ 连续区间内分布的模糊隶属度(Fuzzy Membership Degree)量化重叠节点对不同社区的相对隶属程度,同一节点对所有社区的隶属度总和为“1”。此外,模糊重叠社区检测与模糊社区检测的区别在于,不仅前者在检测过程中利用隶属度测度,而且在检测结果中体现所有节点对所有社区完整且相对的隶属度分布信息,从而体现重叠社区结构。

模糊重叠社区检测实质上是通过扩展离散重叠社区检测的隶属度取值空间,更加细致完整地测量了节点与社区之间的隶属关系及隶属程度,因此能够加强对于真实重叠社区结构中复杂且模糊拓扑结构的探索能力。模糊重叠社区检测在针对科学家合作网络、社会网络、生物网络等基于个体关联性构成的复杂网络重叠社区检测中具有重要的理论和应用研究价值。美国密西根理工大学、法国里昂大学、英国布里斯托大学、澳大利亚墨尔本大学、西北工业大学等国内外研究机构都在此领域开展了探索性研究,近期研究成果频频出现在《IEEE TRANSACTIONS ON FUZZY SYSTEMS》、FUZZ-IEEE等国际重要期刊和会议上。

本文对复杂网络中模糊重叠社区检测的研究进行综述,组织结构为:首先对重叠社区检测问题的基本概念进

行阐述,分析了模糊重叠社区检测与传统离散重叠社区检测的关联性,重点分析了模糊重叠社区检测的功能性优势;然后对二者的国内外研究现状进行阐述和分析,其中,在对模糊重叠社区检测研究的介绍中,根据重叠节点模糊隶属度分布获取方式的不同,将当前相关研究分为扩展标签传播、非负矩阵分解、基于边界节点的两阶段检测、模糊聚类及模糊模块度优化五大类检测方法进行综述,重点分析了基于进化算法(Evolutionary Algorithms, EAs)模糊模块度优化方法的功能性优势及面临的技术挑战;最后对模糊重叠社区检测的研究难点及未来发展趋势进行了分析和展望。

1 相关基本概念

1.1 模糊重叠社区检测定义

复杂网络社区检测的社区划分可行搜索空间如图 1 所示。按照检测复杂度及功能性递增的顺序分为非重叠社区划分(或称为硬划分)、离散重叠社区划分和模糊重叠社区划分 3 类,其中硬划分最简单且易实现,可看作离散重叠社区划分和模糊重叠社区划分的特例;模糊重叠社区划分的复杂度最高且检测难度最大,是对传统离散重叠社区划分的功能性补充和扩展。

上述 3 类社区检测及对应社区划分的定义说明如下:

1)非重叠社区划分,又称为硬划分(Hard Partition)。定义网络中任意节点仅能隶属于单个社区,且与所属社区之间具有完全的隶属关系。

2)离散重叠划分(Crisp Overlapping Partition),又称为脆性重叠划分或非模糊重叠划分。定义网络中每个节点可同时隶属于多个社区,且对不同社区的隶属程度只能为完全隶属或完全不隶属,即隶属度在二进制空间 $\{0,1\}$ 内取值。任意重叠节点对所有所属社区具有完全相同的隶属度,且隶属度取值为“1”。因此,离散重叠社区检测过程中仅需考虑节点与社区之间是否存在隶属关系,而不必衡量节点与社区之间的隶属程度。

3)模糊重叠划分(Fuzzy Overlapping Partition)。定义网络中任意节点不仅可同时隶属于多个社区,而且允许节点对所属不同社区具有不同的隶属程度。隶属度在十进制连续空间 $[0,1]$ 内取值,当隶属度取值为“0”时代表完全不隶属,当隶属度取值为“1”时代表完全隶属,当隶属度取值介于“0”和“1”之间时代表部分隶属。隶属度数值越大说明节点对社区的隶属程度越强,约束条件为任意节点对网络中所有社区的隶属度总和恒为“1”。模糊重叠通过扩展离散重叠的隶属度取值空间,大幅细化了节点对于社区的隶属程度,因此能够更精细地描述节点与社区之间复杂且模糊的隶属关系。

1.2 模糊重叠社区检测数学模型

下面对上述 3 类社区检测对应社区划分的数学模型进行说明。不失一般性,以无向加权网络为例,其对应子图可表示为 $G=(V,E,W)$,其中 V 代表网络中 n 个节点组成的集合, E 代表网络中 m 条边组成的集合, W 为规模为 $n \times n$ 边权重矩阵(或邻接矩阵),元素 w_{ij} 代表节点 i 和 j 之间的连接权重, $i,j=1,\dots,n$ 。网络社区检测的过程可视为在子图 G 上找到一个规模为 $n \times c$ 的社区划分 $U=\{u_{ik}\},i=[n],k=[c]$,如式(1)所示, U 中每个元素 u_{ik} 代表网络中任意节点 i 对于任意社区 k 的隶属度。

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1k} & \cdots & u_{1c} \\ u_{21} & u_{22} & \cdots & u_{2k} & \cdots & u_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{i1} & u_{i2} & \cdots & u_{ik} & \cdots & u_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nk} & \cdots & u_{nc} \end{bmatrix} \quad (1)$$

3 类社区检测对应社区划分的数学公式如式(2)~(5)所示,其中式(2)中 M_{pcn} 代表所有类型社区检测对应可行社区划分的集合,物理含义为任意节点对任意社区隶属度的取值空间为 $[0,1]$,任意节点对所有社区的隶属度总和不超过社区总数 c ,而任意社区内所有节点的隶属度总和小于网络节点总数 n 。式(3)中 M_{hcn} 代表社区硬划分, M_{hcn} 是 M_{pcn} 的子集,不仅限制任意节点仅能隶属于一个社区,而且对任意社区的隶属度在 $\{0,1\}$ 空间内取值。

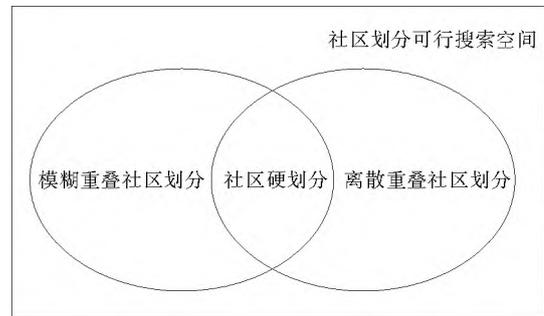


图 1 社区检测分类及可行搜索空间

Fig. 1 Classification of community detection and feasible region of community partition

式(4)中 M_{ccn} 代表离散重叠社区划分集合, M_{ccn} 是 M_{pcn} 的子集, 限制任意节点对任意社区的隶属度在二值空间 $\{0, 1\}$ 内取值。式(5)中 M_{fcn} 代表模糊重叠社区划分集合, M_{fcn} 是 M_{pcn} 的子集, 限制任意节点对所有社区的隶属度总和恒为“1”。

$$M_{pcn} = \left\{ U \in \mathbf{R}^{n \times c}; 0 \leq u_{ik} \leq 1, \forall i, k; \sum_{k=1}^c u_{ik} \leq c, \forall i; \sum_{i=1}^n u_{ik} < n, \forall k \right\} \quad (2)$$

$$M_{hcn} = \{U \in M_{pcn}; \sum_{k=1}^c u_{ik} = 1, \forall i; u_{ik} \in \{0, 1\}, \forall i, k\} \quad (3)$$

$$M_{ccn} = \{U \in M_{pcn}; u_{ik} \in \{0, 1\}, \forall i, k\} \quad (4)$$

$$M_{fcn} = \{U \in M_{pcn}; \sum_{k=1}^c u_{ik} = 1, \forall i\} \quad (5)$$

通过上述数学分析可知 3 种类型社区划分之间存在如下对应关系: 社区硬划分 M_{hcn} 是离散重叠社区划分 M_{ccn} 和模糊重叠社区划分 M_{fcn} 的子集, M_{hcn} 可看作是 M_{ccn} 和 M_{fcn} 的特例, M_{hcn} 等同于每个节点仅隶属于单个社区的 M_{ccn} 或每个节点仅隶属于单个社区且隶属度取值恒为“1”的 M_{fcn} 。此外, M_{fcn} 可简化为相应的 M_{ccn} 和 M_{hcn} 。若将 M_{fcn} 中所有节点的隶属度阈值设置为“0”并进行二值化处理, 可简化为相应的 M_{ccn} ; 若将 M_{fcn} 中每个节点的最大隶属度 $\max_{k \in [1, 2, \dots, c]} (u_{ki})$ 作为阈值并进行二值化处理, 可简化为相应的 M_{hcn} 。

1.3 模糊重叠社区检测功能性说明

为更直观地体现模糊重叠社区检测的定义、特征及功能性优势, 以 Psorakis 等提出的已知真实重叠社区结构的无向加权网络^[11]为例进行说明。图 2a 为示例网络的拓扑连接关系, 其对应真实重叠社区结构如图 2b 所示。

对该网络进行模糊重叠社区检测, 能够得到模糊重叠社区划分 M_{fcn} , 包含所有节点的社区分布及相应隶属度。基于此实现模糊重叠社区检测相比较于非重叠社区检测、传统离散重叠社区检测的特殊功能性优势主要包括以下 4 项。

1.3.1 获得重叠节点及完整社区隶属度分布

通过模糊重叠社区检测, 能够获得所有节点对各社区完整、相对且非一致的模糊隶属度分布 U , 如表 1 所示, 由此精确体现重叠节点的社区分布及对不同社区的隶属程度差异。

根据表 1 所示的节点模糊隶属度分布 U , 模糊重叠社区检测不仅可实现离散重叠社区检测的基本检测功能, 即检测出重叠节点 $\{4, 5, 6, 9, 10, 13, 14\}$, 而且能够更进一步定量体现出所有重叠节点对所有社区 $\{k_1, k_2, k_3, k_4\}$ 的隶属程度差异。

1.3.2 精细化测量重叠节点的重叠程度差异

基于节点的模糊隶属度分布 U , 不仅可获知所有节点的重叠性和社区隶属程度差异, 而且能够更精确地测量节点的重叠程度。节点的重叠程度主要体现在隶属社区数目的多少以及对重叠社区隶属程度的差异性两方面, 节点隶属社区数目越多且同时对所属社区的隶属度越接近, 说明节点的重叠程度越大。模糊重叠社区检测不仅能够与离散重叠社区检测一样, 从重叠节点隶属社区数目的多少衡量节点重叠程度大小, 而且能够从更具体的隶属度数值差异中更精细地测量节点重叠程度。

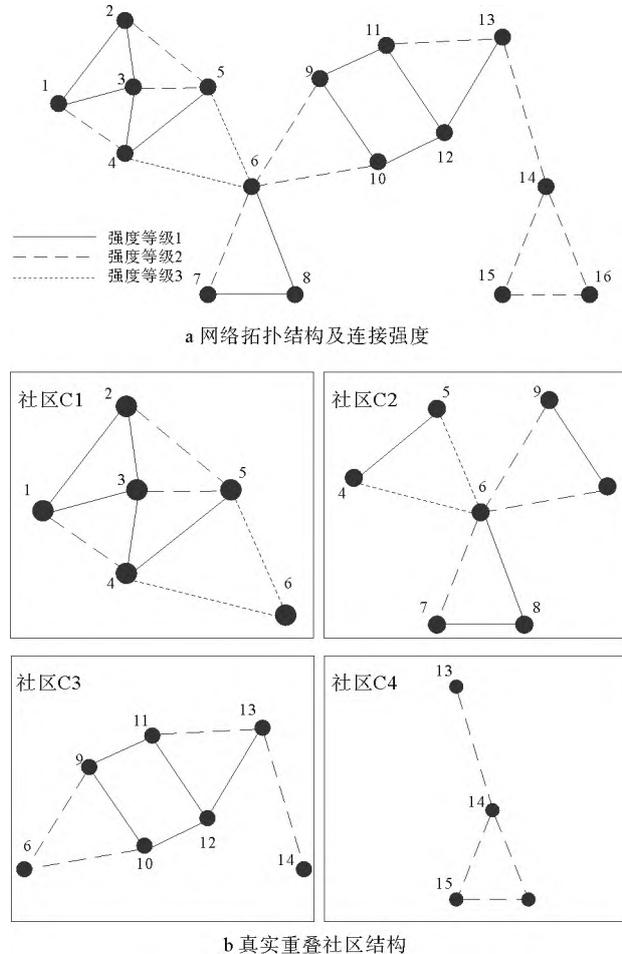


图 2 模糊重叠社区检测网络示例

Fig. 2 Network of fuzzy overlapping community detection

根据表 1 中节点的隶属度数值分布及隶属社区数目,可设计如式(6)所示的测量指标 O_i 全面测量重叠节点的重叠程度。

$$O_i = \frac{\max_{k, k \neq k^*} u_{i,k}}{\max_k u_{i,k}} \times \frac{\|\{k | u_{i,k} \neq 0\}\|}{c} \quad (6)$$

式中 O_i 代表节点 i 的重叠程度测量值, k^* 代表节点 i 最大隶属度对应社区, $\max_k u_{i,k}$ 和 $\max_{k, k \neq k^*} u_{i,k}$ 分别对应节点 i 的最大和次大隶属度, $\|\{k | u_{i,k} \neq 0\}\|$ 和 c 分别对应节点 i 隶属社区数目及真实社区数目。 $O_i \in [0, 1]$, 当 O_i 取值为 0 时, 说明节点 i 仅隶属于一个社区, 因此为非重叠节点, 重叠程度最低; 当 O_i 取值大于 0 时, 说明节点 i 至少同时隶属于 2 个社区, 因此为重叠节点, O_i 数值越大说明节点隶属社区数目越多或对于多个隶属社区的隶属程度越接近, 因此重叠程度越高; 当 O_i 取值为 1 时, 说明节点 i 同时隶属于所有社区, 隶属社区数目达到最高, 且对于隶属程度最大的两个社区的隶属度均等, 因此重叠程度最高。根据式(6)计算结果可得到示例网络中所有节点的重叠程度测量图如图 3 所示, 节点的大小代表其重叠程度的大小。

1.3.3 定量测量社区重叠程度及重叠节点重要性

与离散重叠社区检测不同, 模糊重叠社区检测除了能够更加精细地测量重叠节点的重叠程度, 还能够根据节点隶属度分布 U 定量描述重叠社区之间的重叠程度以及重叠节点对于社区重叠程度的贡献度。

1) 基于 U 可计算社区与社区之间的相关性 $Z = U^T U$, 而 Z 中非对角线元素即代表不同社区之间的重叠程度的测量值, 计算公式如式(7)所示。

$$Z_{k_1 k_2} = \sum_{i=1}^n u_{i k_1} u_{i k_2} \quad (7)$$

根据式(7)测量结果可知, 示例网络的所有重叠社区 $\{k_1, k_2\}, \{k_1, k_3\}, \{k_2, k_3\}, \{k_3, k_4\}$ 中, 社区 k_1 与 k_2 之间重叠程度最高, 而社区 k_3 与 k_4 之间重叠程度最低。

2) 基于 U 可计算重叠节点对于社区重叠程度的贡献度 B , 计算公式如(8)所示。 B 值取值较大的节点在社区之间的连接及通信方面起到较为重要的作用, 对于社区之间的连接稳定性具有重要影响, 因此其重要程度较高。通过重叠节点的贡献度计算, 可精确判定重叠节点的重要性大小, 有助于分析及预测网络社区结构的动态演化特性。

$$B = \frac{u_{i k_1} u_{i k_2}}{\sum_{i=1}^n u_{i k_1} u_{i k_2}} \times 100\% \quad (8)$$

根据示例网络计算结果可知, 对于重叠程度最高的两社区 k_1 与 k_2 , 重叠节点 v_4, v_5 和 v_6 对于社区重叠程度的贡献度分别为 40.06%、36.85% 和 23.06%, 即节点 v_4 的贡献程度最高, 因此其重要性最大, 对于社区结构的演化起到关键性决定作用。

1.3.4 反映重叠节点的连边强度差异

离散重叠社区检测中仅能够根据节点的社区隶属数目反映出重叠节点的拓扑信息, 而模糊重叠社区检测不仅能够更加全面的展现重叠节点的拓扑信息, 还能够体现出重叠节点连边的拓扑信息。基于节点隶属度分布 U 可测量重叠节点连边的平均强度关系。以示例网络中重叠程度最大的节点 v_6 为例, 根据 v_6 对社区 k_1, k_2 和 k_3 的隶属度数值差异及 v_6 在各社区中的连边数量差异, 可计算获知 v_6 涉及连边之间的平均强度存在如下关系

表 1 节点模糊隶属度分布 U

Tab. 1 Fuzzy Membership Degree distribution U of nodes

节点 i	$u_{i,k1}$	$u_{i,k2}$	$u_{i,k3}$	$u_{i,k4}$
1	1.000 0	0	0	0
2	1.000 0	0	0	0
3	1.000 0	0	0	0
4	0.667 0	0.333 0	0	0
5	0.713 8	0.286 2	0	0
6	0.253 4	0.505 1	0.241 5	0
7	0	1.000 0	0	0
8	0	1.000 0	0	0
9	0	0.027 0	0.973 0	0
10	0	0.022 0	0.978 0	0
11	0	0	1.000 0	0
12	0	0	1.000 0	0
13	0	0	0.999 3	0.000 7
14	0	0	0.010 5	0.989 5
15	0	0	0	1.000 0
16	0	0	0	1.000 0

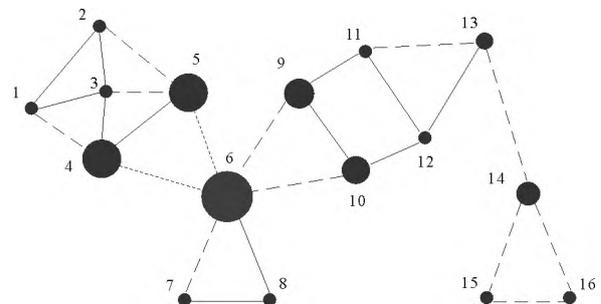


图 3 网络节点重叠程度测量

Fig. 3 Overlapping degree of nodes

$\overline{w_{6,7}} + \overline{w_{6,8}} > \overline{w_{6,9}} + \overline{w_{6,10}} > \overline{w_{6,4}} + \overline{w_{6,5}}$,在一定程度上反映出重叠节点连边的强度差异。

2 离散重叠社区检测方法及其性能分析

目前离散重叠社区检测仍然是复杂网络重叠社区检测领域的研究重点和热点^[1-5]。现有绝大多数重叠社区检测方法所得检测结果均为离散重叠划分,检测目标及评价准则为精确、高效、稳定地提取出网络中的重叠社区结构,并尽可能提升对于规模较大、社区模糊程度较高、重叠程度较高的复杂网络的检测能力。离散重叠社区检测的典型代表性方法主要包括以下几类,包括派系过滤、链接聚类、局部扩展、模块度优化、多目标优化和标签传播等,以下分别予以说明。

2.1 派系过滤

派系过滤(Clique Percolation),如 CPM^[13]、CPMw^[14]、CFinder^[15]、SCP^[16]等。该类方法的核心概念为“全连通子图”,核心思想为网络中密度较高的连边构成 k -团,即包含 k 个节点的全耦合子图,团之间通过共享节点而紧密连接,所有相邻 k -团组成的最大全连通子图构成 k -团社区。网络的重叠社区结构是由多个 k -团社区构成的重叠集合,通过团过滤算法识别出同时隶属于多个不相邻 k -团的重叠节点,从而检测出重叠社区结构。团规模参数 k 取 $[3, 6]$ 之间的较小数值容易得到较好的检测结果。

该类方法中基于 k -团的社区结构定义较为严格,虽然有利于发现网络中有结合力的局部社区结构以及高度重叠的社区结构,但仅能检测到特定的基于 k -团的社区结构,使检测精度受到影响。此外,该类方法需要得到网络中所有的 k -团,近似指数级的时间复杂度使该类方法不适用于大规模网络中的重叠社区检测。

2.2 链接划分

链接划分(Link Partitioning),如 LINK^[17]、LinkComm^[18]、DBLINK^[19]、CDAEO^[20]等。该类方法的核心思想是将链接而不是节点作为聚类对象,利用链接的相似度进行社区划分。通常利用 Jaccard 指数等相似性函数计算链接相似度,再根据相似度进行层次聚类以获得链接社区划分,隶属于不同社区的链接的交点即为重叠节点。

该类方法利用了链路属性的唯一性特性,能够在一定程度上克服基于节点的检测方法难以识别重叠度较高的社区结构的问题,但不能保证比基于节点的检测方法获得更好的社区划分^[21]。此外,该类方法通常将所有链接都划分到特定的链接社区中,容易形成过度重叠现象^[19]。最终的检测结果需要将链接社区划分转化为节点社区划分,增加了该类方法的时间复杂度。

2.3 局部扩展

局部扩展(Local Expansion),如 LFM^[22]、GCE^[23]、DEMON^[24]、OSLOM^[25]、UEOC^[26]及 SLEM^[27]等。该类方法主要基于核心团生长进化的观点,核心思想是从不同种子节点或离散团开始扩张,通过优化局部效益函数探索种子节点或团所在的局部最优社区结构,通过融合局部最优社区结构获得网络重叠社区划分。通常情况下采用离散团作为初始种子能够获得更好的检测效果,有利于发现具有高度重叠特性的社区结构。

该类方法能够同时获得重叠社区划分和多层次社区划分,局部搜索方法使其需要相对较少的局部信息,并具有较快的搜索速度,因此适用于大规模网络重叠社区检测。然而,种子节点的选取、局部效益函数的构建、效益函数优化质量等对于检测结果有较大影响,若种子选取不佳或局部效益函数优化性能不足,容易使检测结果陷入局部最优。

2.4 标签传播

标签传播,如 BMLPA^[39]、MLPAO^[40]等。该类方法的核心思想为节点通过与邻域节点之间交互社区归属标签信息,更新节点自身的社区归属标签,使网络中所有节点对应的标签分布达到动态平衡,具有相同标签的节点构成社区,而具有多个社区标签的节点为重叠节点,由此得到重叠社区结构。

该类方法由于采用了局部搜索的思想,因此通常在计算效率上具有一定优势,适用于大规模复杂网络社区检测。然而,标签更新方式存在较大差异,一方面节点标签选择依据通常为邻居节点标签中出现次数最多社区标签,没有考虑节点标签的历史信息;另一方面标签传播过程中错误传播的风险较大。上述原因使得该类算法检测结果具有较大的不确定性,容易陷入局部最优。

2.5 模块度优化

模块度优化(Modularity Optimization),如 Fast Unfolding Algorithm^[28]、EAGLE^[29]、OMIM^[30]、cob-ai-Net^[31]等。该类方法的核心思想为将重叠社区检测问题转化为重叠模块度优化问题,在可行解空间中按照预定

义的启发式规则搜索使重叠模块度函数达到全局最优的重叠社区划分。现有典型的重叠模块度函数包括 $EQ^{[29]}$ 和 $Q_{ov}^{[10]}$ 等。

该类检测方法的关键在于重叠模块度函数和启发式优化算法,一方面,需构建基于节点或链接拓扑相似性的高性能重叠模块度函数,提升对于重叠社区划分的评价能力以及对于最优重叠社区划分的搜索引导能力;另一方面,需设计高性能的优化算法,保证在大规模社区划分搜索空间中收敛到最优的重叠社区划分。该类方法由于将评价指标即重叠模块度函数直接作为目标函数进行优化,克服了检测目标与评价标准之间的差异,能够在一定程度上提升检测精度和效率。近年来遗传算法^[33]、蚁群算法^[34-35]、粒子群优化算法^[36]等基于群智能的超启发式优化算法相继被用于重叠模块度优化,相比较于传统启发式搜索算法提升了高维复杂搜索空间中的全局收敛性能,有效提升了最优重叠社区划分的检测精度。

2.6 多目标优化

多目标优化(Multi-Objective Optimization),如 MOGA-Net^[37]、iMEA_CDPs^[38]等。该类方法的核心思想为通过增加社区划分质量评价目标函数,提升对社区划分的综合评价能力以及检测过程中对于最优社区划分搜索的多方向引导能力。多目标优化能够在单次运算中获得多样化、多层次、多分辨率的 Pareto 最优重叠社区划分集合,能够使检测所得重叠社区划分在模块性、精确性和重叠性等方面的性能达到最佳平衡,同时为用户提供多样化的检测结果,满足用户多样化检测需求。

该类方法的关键在于构建合理、高效的目标函数集合,前期实践经验表明,具有负相关性的目标函数集合能够获得质量更优的社区划分。然而,目前现有算法的目标函数多凭经验设定,对于目标函数的种类、精确性、稳定性及函数之间的相关性缺乏较为系统的分析和理论性指导。由于重叠社区检测除了需要满足社区划分在精确性和模块性上的性能需求,还要使社区划分的重叠性尽可能达到最优,因此多目标优化中目标函数的选择或构建需要从以下三方面考虑。首先,需要考虑目标函数的精确性、对于社区结构模糊程度的适应性以及函数自身的稳定性,保证在多样且模糊的网络环境下获得尽可能精确的检测结果;其次,需要尽可能选择负相关性较大的目标函数集合,获得尽可能多样化的 Pareto 最优划分集合;最后,需要高质量的重叠特性评价函数,保证社区划分的重叠质量。

目前常用的社区划分拓扑质量评价函数主要包括 $f_1 = \text{Modularity}$ 、 $f_2 = \text{Conductance}$ 、 $f_3 = \text{Community Score}$ 、 $f_4 = \text{Average ODF}$ 、 $f_5 = \text{Cut Ratio}$ 、 $f_6 = \text{Expansion}$ 和 $f_7 = \text{Internal Density}$ 等。对其进行实验测试的统计结果表明, f_1, f_2, f_3 和 f_4 具有相对较好的精确度,能够有效逼近真实社区结构;随着网络社区结构模糊程度的增强,各函数的检测性能均存在不同程度的下降,而 f_1 和 f_3 表现出较为优越的稳定性,对社区结构模糊性的适应能力较强; f_2, f_5 和 f_6 则具有相对良好的稳定性,能够获得较为稳定的社区划分结果;函数之间的相关性如图 4 所示,根据函数之间的皮尔逊

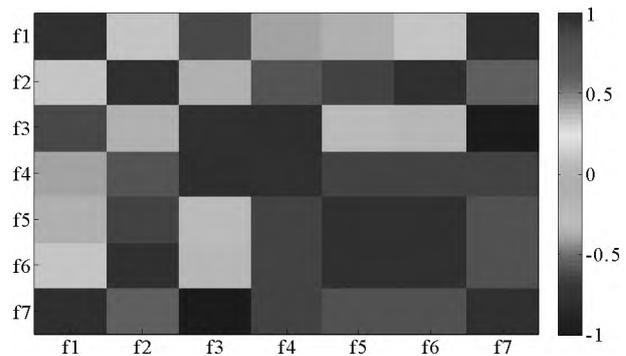


图 4 社区划分拓扑质量评价函数的相关性
Fig. 4 Correlations between topological objective functions

相关系数(Pearson Correlation Coefficient)可看出, f_1 和 f_7, f_3 和 f_4 以及 f_3 和 f_7 之间存在较强的负相关性,因此适合于构建多目标优化函数集合。社区重叠特性包括重叠质量(Overlap Quality)和重叠覆盖(Overlap Coverage)两方面,对应评价指标函数主要有准确率 Precision、召回率 Recall 和综合性指标 F-score 等,其作为目标函数的优化性能以及与社区质量目标函数之间的相关性仍有待于进一步研究。

上述对于离散重叠社区检测方法的分类主要基于识别重叠社区结构时采用的关键技术,不同类别的检测方法之间没有绝对的界限,各类检测方法的分类总结如表 2 所示。近年来的离散重叠社区检测研究倾向于将不同类别的检测方法进行融合,通过结合不同类型方法的优势提高重叠社区结构的检测质量。例如,模块度优化和多目标优化通常与标签传播相融合,检测过程中首先利用标签传播速度快的优势获得具有一定社区结构的优质初始种群,再对其进行优化获得最优重叠社区划分^[30]。

综上所述,目前离散重叠社区检测研究已取得了丰硕的研究成果,在重叠社区结构的检测能力及检测效率上均取得了较大进展。然而,现有检测方法由于算法设计及离散重叠概念本身的限制,在真实复杂网络的重叠社区检测中仍存在一些关键性问题有待解决:

1)检测算法的设计和性能对社区检测结果形成制约。现有离散重叠社区检测算法检测到的重叠节点数量通常不准确,大部分算法检测得到的重叠节点数量偏少,仅占实际数量的 30%左右^[4],而部分算法则存在“过度重叠”现象^[19]。此外,为简化检测问题和算法设计,现有离散重叠社区检测中通常将重叠节点能够隶属的最大社区数目统一设置为某一数值,如通常设置为 2,即每个重叠节点仅能够同时隶属于两个社区,限制了重叠节点的重叠程度,人为降低了重叠社区结构的复杂度,导致其与真实网络社区结构存在较大差异。

2)离散重叠概念对社区检测功能造成限制。在离散重叠概念下,不论何种类型的检测方法检测到的重叠节点对所属社区均具有绝对且完全的隶属关系,即隶属程度全部为“1”。从检测结果中我们无法判别出重叠节点与多个不同所属社区之间的隶属程度,也无法判别出同一社区内不同节点的隶属程度差异。因此,离散重叠社区检测对重叠节点隶属程度以及重叠社区结构内在深层次拓扑特性的探索能力上均存在欠缺,使其无法有效反映出真实复杂网络中重叠社区结构的复杂性和模糊性,不适用于真实世界中较为复杂的重叠网络社区检测问题。

表 2 离散重叠典型社区检测方法

Tab. 2 Typical methods of crisp overlapping detection

检测方法	核心思想	性能优势	性能不足	代表性算法
派系过滤	识别由相邻 k -团组成的最大全连通子图进行社区划分	有利于发现网络中有结合力的局部社区结构以及高度重叠的社区结构	特定的基于 k -团的社区结构使检测精度受限且时间复杂度较高	CPM ^[13] , CPMw ^[14] , CFinder ^[15] , SCP ^[16]
链接划分	利用链路属性唯一性特性,根据链接相似性进行社区划分	能够在一定程度上克服难以识别重叠度较高的社区结构的问题	需要将链接社区划分转化为节点社区划分且容易形成过度重叠现象	LINK ^[17] , Link Comm ^[18] , DBLINK ^[19] , CDAEO ^[20]
局部扩展	利用种子节点或离散团局部扩展搜索获得的局部最优社区结构进行社区划分	能够同时获得重叠社区划分和多层次社区划分,需要相对较少的局部信息,且具有较快的检测速度	种子节点和局部效益函数影响较大,容易陷入局部最优	LFM ^[22] , GCE ^[23] , DEMON ^[24] , OSLOM ^[25] , UEOC ^[26]
标签传播	利用邻域节点之间的社区标签信息交互进行社区划分	具有较快的检测速度	标签传播错误风险较大,容易陷入局部最优	BMLPA ^[39] , MLPAO ^[40]
模块度优化	在可行解空间中按照启发式规则搜索使重叠模块度函数达到全局最优,获得全局最优社区划分	克服了检测目标与评价标准之间的差异,能够获得全局最优社区划分	重叠模块度函数和最优化算法影响较大,容易陷入局部最优	EAGLE ^[29] , OMIM ^[30] , cob-aiNet ^[31]
多目标优化	同时最优化两项以上检测目标,获得多样化、多层次、多分辨率的 Pareto 最优重叠社区划分集合	能够使检测结果在模块性、精确性和重叠性等方面达到最佳平衡	多目标优化目标函数集合构建影响较大且最优设置较为困难	MOGA-Net ^[37] , iMEA — CDPs ^[38]

3 模糊重叠社区检测方法及其性能分析

模糊重叠社区检测从原理上克服了离散重叠社区检测固有的功能性欠缺,其本质上是对离散重叠社区检测的强化和功能性扩展。该类检测方法一方面能够通过模糊隶属度分布精确衡量节点与社区之间精确化、细粒化的隶属强度关系,提升真实复杂网络环境下重叠社区结构的检测精度;另一方面,能够在所得社区划分中完整体现所有重叠节点的社区隶属度分布,有助于揭示重叠社区结构中隐含的深层次拓扑信息。以下分别对模糊重叠社区检测的两项功能优势进行说明:

1)模糊重叠社区检测中模糊隶属度的计算有助于提升重叠社区结构检测的精确性。模糊重叠社区检测将节点对于任意社区的隶属程度归一化到 $[0, 1]$ 十进制连续取值空间内,大幅提升了对于隶属关系的量化程度,以及对于节点与社区之间隶属强度的刻画及区分能力。检测过程中根据节点对社区的隶属度取值分布,可以获知当前社区划分中每个节点对各个社区的隶属程度强弱关系。依据该信息及真实网络环境下的“重叠”判定标准及检测需求,可以更加精准、灵活地控制重叠节点的数目、重叠节点能够隶属的最大社区数目及重叠社区分布,在检测过程中避免对所有重叠节点统一设置的粗放式简化处理。因此,模糊重叠社区检测能够有效增强真实网络中重叠社区结构的识别和检测能力,更能够满足重叠社区检测中的复杂性需求。

2)真正的模糊重叠社区检测,其检测结果不仅能够体现出重叠节点及其社区分布,而且能够体现出所有节点对各个社区完整且相对的隶属度数值分布。根据重叠节点的隶属度分布,一方面能够精确反映出同一重叠节点对不同社区的不同隶属程度,以及同一社区内不同节点对本社区的隶属程度;另一方面,根据重叠节点的隶属度

分布,以及网络拓扑结构中重叠节点与不同社区之间的连接情况,能够揭示出重叠社区结构的深层次拓扑信息,包括重叠节点的重叠程度差异、社区重叠程度差异、重叠节点对社区重叠的贡献度及节点重要性等,并可在一定程度上挖掘出重叠社区结构中隐含的拓扑连接强度信息。因此,模糊重叠社区检测能够更深入地体现重叠社区结构中的拓扑关系。

基于上述分析可知,模糊重叠社区检测能够更好地满足复杂网络重叠社区检测的两项核心任务,使检测结果更好地体现出重叠社区结构的复杂性和模糊性,因而更适用于真实复杂网络环境下的重叠社区检测问题,是目前该领域中最前沿、性能最优且最具发展前景的一类重叠社区检测方法。然而,虽然模糊重叠社区检测相比较于传统重叠社区检测方法在功能性上具备明显优势,但由于其检测过程中需要定量评估每个节点对各社区的隶属度分布,因此检测所需信息量较大、检测结果体现内容较多、检测难度也相对较高,对算法的设计和性能均提出了较高要求。目前国内外对于模糊重叠社区检测方法的研究相对较少^[1-3],虽然部分算法利用了隶属度测度确定重叠节点,但现有隶属度计算大多仅以节点连边数量为依据且难以在检测结果中体现出完整相对的隶属度分布,在模糊重叠的实现能力和实现效果上均存在不足,现阶段模糊重叠社区检测算法的研究仍处于起步阶段。

根据网络分析需求及用户解释的不同,模糊重叠社区检测问题可构造为多种不同形式,对应检测算法之间存在较大差异。根据重叠节点模糊隶属度获取方式的不同,本文将当前国内外现有的基于模糊重叠思想的社区检测方法分为五类进行阐述,包括扩展标签传播、非负矩阵分解、基于边界节点的两阶段检测、模糊聚类及模糊模块度优化,以下分别对各类检测方法进行详细说明。

3.1 基于扩展标签传播的模糊重叠社区检测

扩展的标签传播算法是目前较为常见的一类模糊重叠社区检测方法,其在离散重叠社区检测的基本框架上融入了模糊重叠的思想,在为每个节点维护标签集合的同时计算并保存节点对标签代表社区的模糊隶属度。算法的核心思想为根据节点的邻域节点数目及相应标签数目,计算节点对邻域节点所属社区的隶属度,并通过设定隶属度阈值更新标签集合并确定节点的社区归属,由此得到重叠节点及其重叠社区划分。

该类检测方法的典型代表算法包括 COPRA 算法^[41]、SLPA 算法^[42]及 LPPB 算法^[9]等。COPRA 是 Gregory 于 2010 年提出的首个基于标签传播的模糊重叠社区检测算法,节点标签对中不仅含有社区名称,而且包含节点对该社区的归属系数,任意节点对所有隶属社区的归属系数总和为“1”。每个节点通过平均所有邻居节点的隶属度系数更新隶属度分布,通过设置参数 ν 控制节点能够隶属的最大社区数目。Xie 等 2011 年提出 SLPA 算法,为每个节点提供存储信息(标签)的记忆空间,将从记忆空间中获取标签的概率作为节点隶属度,无需社区数目等先验信息。刘世超等 2015 年提出一种基于标签传播概率的 LPPB 算法,将节点获得社区标签的概率作为隶属度,设置隶属度阈值为节点能够隶属最大社区数目的倒数,综合网络的结构传播特性和节点的属性特征共同计算标签传播的概率。

基于扩展标签传播的模糊重叠社区检测最大优势在于具有较高的运算效率,通常具有接近线性的计算复杂度,当网络社区结构较为清晰时能够获得较好的模糊重叠社区划分,可处理加权、有向及二分网络。然而,该类检测方法也存在一些不足:1)标签更新过程中,通常需要根据预定且统一设置的节点最大标签数确定隶属度阈值,通过阈值限定重叠节点隶属社区个数及社区分布。节点最大标签数即节点隶属最大社区个数取值直接影响社区划分的重叠性和模块性,取值过低将限制重叠节点数量及重叠程度,取值过高则导致过度重叠或重复社区。因此,该参数合理取值十分关键,但目前仍缺乏统一有效的设置方法,对其进行最优化设置通常导致较高的计算复杂度;2)由于标签传播过程通常具有一定的随机性且不同节点的标签传播能力不同,因此当网络社区结构较为模糊或社区间重叠程度较高时检测精度将大幅下降。

3.2 基于非负矩阵分解的模糊重叠社区检测

非负矩阵分解(Non-negative Matrix Factorization, NMF)是机器学习领域中一种常用的特征提取和维度减少算法,近年来被用于求解模糊重叠社区检测问题。算法的核心思想认为社区的形成源于共享群体从属关系,一个节点同时可隶属于多个社区,而两节点之间的交互程度越高,隶属于同一社区的概率越大,共享社区的数目也越多。算法将代表节点之间期望交互次数或相似程度的特征矩阵 V 在非负约束($V \approx WH$)下分解为两个相乘的矩阵 W 和 H ,其中矩阵 W 代表了节点对不同社区的隶属度分布,归一化矩阵 W 中的每个元素 ω 衡量了相应节点与社区的隶属程度。该类检测方法的典型代表算法包括 Zhao 等 2010 年提出的对称 NMF 算法 s-NMF^[43], Psorakis 等 2011 年提出的贝叶斯 NMF 算法^[44],李玉翔提出的基于贝叶斯先验的 NMF 算法 BNMF-CD^[45]。

该类算法的最大优势在于检测结果中能够定性和定量地评价节点对所属社区隶属关系的绝对程度,然而存

在的不足之处包括:1)矩阵乘法使其需要耗费较多的计算时间和存储空间,进行高维矩阵分解时复杂度较高,不适用于规模较大的网络;2)传统 NMF 算法通常需要预先给定或通过优化模块度来确定社区数目^[46],不可避免会带来较高的计算开销;3)为了根据隶属度分布确定重叠社区结构,通常需要人为设定隶属度阈值,以确定各节点与社区之间的最终隶属关系,但由于阈值设置的精确性和稳定性限制,所得重叠社区划分的质量造成影响。

3.3 基于边界节点的两阶段模糊重叠社区检测

基于边界节点的两阶段检测方法是目前研究较多的一类模糊重叠社区检测方法。核心概念为“边界节点”或称“连接节点”,即社区硬划分中不仅与自身所在社区存在连接,而且与其他社区存在连接的节点。核心思想为“边界节点”不论是在拓扑结构还是在功能层面均具有多面性,因而是重叠节点的主要来源。

该类方法在检测过程中不限定具体的检测算法,而是采用一种通用的基于边界节点的两阶段检测模式,流程图如图 5 所示,相应的操作流程说明如下:1)第一阶段采用现有较为成熟的社区检测方法,如快速模块度算法 CNM^[47]、LM 算法^[48]等,进行社区结构的硬划分。识别硬划分中社区内与其他社区存在连接的“边界节点”,以及具有单一社区归属的“孤立节点”,两类节点在图 5 中分别用空心心和实心圆圈表示。2)第二阶段通过影响系数函数 M_d 计算^[47]、基于局部信息的邻居投票机制^[48]、基于混合整数非线性规划模型 OverMod 的局部模块度优化^[49]等方法,度量“边界节点”对邻域社区的贡献或计算“边界节点”对邻域社区的隶属度,根据贡献大小或隶属度数值确定“边界节点”最终的社区归属,由此得到社区硬划分对应的重叠社区软划分。

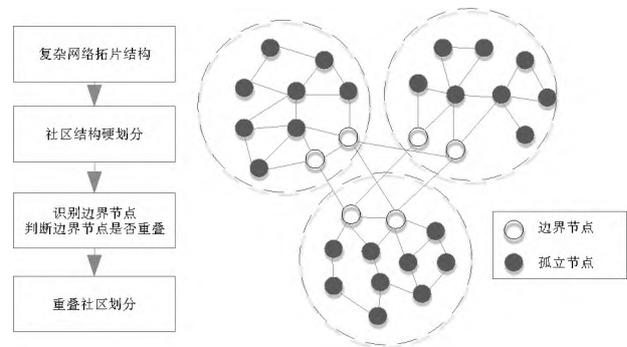


图 5 基于边界节点的两阶段模糊重叠社区检测

Fig. 5 Two-phase of fuzzy overlapping community detection based on edge nodes

该类检测方法的典型代表算法包括 2014 年西安电子科技大学刘勇提出基于快速模块度算法 CNM 和新的影响系数函数 M_d 的两阶段检测算法^[47];2014 年陈俊宇等基于 LM 和新的邻居投票机制两阶段检测算法 LM-NV^[48];2014 年 Laura Bennett 等提出基于新混合整数非线性规划模型 OverMod 的两阶段检测算法^[49]等。

该类算法通常能够快速有效地得到网络的重叠社区结构。一方面,由于只需考虑“边界节点”的重叠性,大幅减少了重叠节点检测的计算复杂度;另一方面,由于有效结合了现有的非重叠社区检测算法,使算法的通用性和可移植性得到提升。然而,该类算法的问题主要包括以下两点:1)重叠节点的检测质量很大程度上依赖于第一阶段的社区硬划分结果,若硬划分准确度较低会严重影响重叠节点的检测质量,容易导致错分点并使重叠节点判断误差增大,使重叠社区划分的精确性和稳定性受到制约。2)该类方法基于一个默认的前提假设,即通过在优质的社区硬划分上识别重叠节点能够获得优质的重叠社区划分。然而,实验结果证明最优的重叠社区划分可能不是由最优的硬划分生成得来的^[44],因此该项假设的合理性和通用性仍有待于进一步研究。

3.4 基于模糊聚类的模糊重叠社区检测

模糊聚类检测算法也是一种常见的模糊重叠社区检测方法。该类方法的基本思想为网络中连接相对紧密的节点构成社区,将社区看作对象(节点或团体)之间紧密连接关系的集合,任意对象对于任意集合的连接紧密程度用介于 0 和 1 之间的模糊隶属度进行度量,根据对象与集合之间的隶属度进行模糊聚类从而获得重叠社区结构。该类算法中较为通用的操作流程如下:首先将网络节点映射为欧式空间中的数据点集,然后通过构建“距离”因子(如余弦距离等),衡量节点与节点、节点与社区、社区与社区之间的相似度,最后依据相似度进行模糊 C 均值聚类 FCM^[50-51]、概率 C 均值聚类 PCM^[5]等聚类操作,得到最终的重叠社区结构。

该类检测方法的典型代表性算法包括 2013 年 Wang 等提出基于 FCM 的模糊聚类算法,通过随机游走和 SRW 局部相似性测度计算并构建节点之间的距离矩阵,将其映射到低维空间并利用 FCM 进行聚类^[51]。该算法虽然能够得到模糊重叠社区划分,但检测所得社区结构的模块度较低。2015 年 Samira 等提出基于改进 PCM 的模糊聚类算法,由于节点对社区的隶属度总和不为“1”,因此无法得到真正的模糊重叠社区划分^[5]。2015 年西安交通大学李刘强等提出基于模糊层次聚类的 CDHC 算法,由于仅在社区内部计算节点对本社区的隶属度,使隶属度具有局部性和绝对性,无法体现节点对于各社区完整且相对的隶属度分布^[52]。2015 年 Suman Kundu 等提出一种新的基于模糊粒度理论(Fuzzy Granular Theory)的知识表示框架用于社交网络的模糊重叠社区检测,即

以网络代表性节点为中心,合并连接紧密的邻域节点形成微粒并将其表示为模糊集,由此将网络表示成为模糊粒子集合。新算法 FRC-FGSN 能够在基于粒度模型的知识表示框架下,检测到以模糊-粗糙形式定义的社区结构^[53]。

该类检测方法中,如何有效衡量节点与社区之间的相似度对检测结果具有重要影响。相似性测度分为局部相似性测度和全局相似性测度两种,其中局部相似性测度计算仅需节点度及最近邻域信息,而全局相似性测度计算则需要网络拓扑全局信息。现有算法通常采用基于“距离”因子的局部拓扑相似性测度^[50-53],计算相对简单但精确度较低,此外由于部分“距离”因子中权重参数设置困难、相似度阈值和隶属度阈值设置困难等原因,“距离”相似性测度测量精度受到限制,使聚类质量即社区划分结果受到影响。

3.5 基于模糊模块度优化的模糊重叠社区检测

模糊模块度优化(Fuzzy Modularity Maximization, FMM)是近年来新兴的一类模糊重叠社区检测方法^[1],扩展至模糊重叠的模块度函数的相继提出为该检测方法实现奠定了重要的理论基础。该类方法的基本思想是将模糊重叠社区检测问题抽象为带约束条件的大规模组合优化问题,在模糊重叠模块度函数引导下,利用优化算法在模糊重叠约束条件对应的可行解空间内进行方向性搜索,收敛至全局最优的重叠社区划分,使模糊模块度达到最优。

该类检测方法的典型代表性算法包括 Nepusz 等提出 VSGD 算法,构建了基于节点相似度的模糊模块度函数 Q_{ov}^{Ne} ,相似度等于两节点隶属度乘积的总和,在此基础上利用模拟退火算法进行优化^[54]。Nicosia 构建了基于链接归属系数的模糊模块度函数 Q_{ov} 并利用遗传算法对 Q_{ov} 进行优化^[10]。Zhang 和 Liu 分别提出两种相似的模糊模块度函数 Q_c 和 Q_l ,并分别利用结合谱映射的模糊 C 均值聚类算法 FCM 和模拟退火算法行优化^[55,56]。Havens 等提出两层搜索框架的 MSFCM,利用模糊模块度函数 Q_c 从所有可能社区数目对应的局部最优划分中选择全局最优社区划分^[3]。Su 等提出 FMM/FBLN 算法,将模糊模块度最大化问题转化为二次规划优化问题,并利用树结构进行全局最优化^[1]。Zhan 等提出 MAGA* 算法,通过重构基于节点隶属度的零模型构建了新的模糊重叠模块度函数,并利用改进的自适应遗传算法进行优化^[57]。

基于模糊模块度优化的模糊重叠社区检测方法区别于其他方法的最大不同在于采用了反向搜索的方式获取节点对社区的最优隶属度分布,即在模糊重叠约束条件对应的大规模可行隶属度分布组合空间内,自动搜索到使模糊模块度函数取得最优值的节点隶属度分布。该过程称为反向搜索是因为社区检测过程中不再需要基于网络的拓扑结构及属性等信息计算节点对社区的隶属度,仅需要依据模糊模块度在决策空间中搜索最佳的节点隶属度分布组合。该种隶属度计算方式具有两项优势:1)避免了隶属度计算过程中的相似度函数构建、相似度计算以及相似度阈值设置带来的复杂性和不稳定性;2)获得的最终隶属度分布是使重叠社区结构评价指标如 Q_{ov} 最优化的结果,能够体现重叠节点的多样化拓扑特性差异,克服了隶属度计算过程中由于计算指标单一(如仅计算节点连边数目)带来的不精确性。

该类检测方法的关键技术包括两项:1)一是选择或构建能够有效评价模糊重叠社区划分质量的扩展模糊模块度函数,提升对模糊重叠社区划分的评价能力,避免评价过程中从模糊重叠社区划分到离散重叠划分或硬划分的转化,此外充分发挥其在目标空间搜索过程中的方向性引导作用。2)二是用于模糊模块度优化的优化算法设计,主要包括启发式优化算法和超启发式优化算法两种,充分发挥优化算法的智能化搜索能力和全局最优化能力,以获得全局最优的模糊重叠社区划分。

3.6 模糊重叠社区检测方法性能分析

汇总上述五类模糊重叠社区检测方法如表 3 所示。

现有检测方法能够通过不同方式获取节点对于不同社区的精确化、细粒化隶属度分布,有效增强对于节点与社区之间归属关系的判定能力,从而获得高质量的重叠社区划分。然而,现有模糊重叠社区检测方法在检测性能及检测功能方面仍存在一定不足:

1)现有算法存在检测过程依赖于社区结构的硬划分、重叠节点挖掘依赖于复杂的相似度和隶属度计算、相似度阈值和隶属度阈值的合理设置困难、缺乏有效的模糊重叠社区划分质量评价指标、检测结果需转化为离散重叠社区划分或硬划分进行评价等问题,使得检测算法普遍存在设计复杂、稳定性不足、通用性和易操作性差等缺陷。

2)现有大部分检测方法仅是借鉴了“模糊重叠”的思想,即虽然在判定节点社区归属及重叠节点时计算了节点对邻域社区的隶属度,但该隶属度计算大多具有局部性和绝对性。社区划分结果中没有体现节点对于各社区满足约束条件的完整且相对隶属度分布,检测结果绝大多数仍为离散重叠社区划分,因此严格意义上并不属于真

正的模糊重叠社区检测,也无法基于隶属度分布进一步揭示重叠社区结构中的深层次拓扑关系,模糊重叠社区检测特有的功能性优势没有得到很好地实现。

表 3 模糊重叠典型社区检测方法
Tab. 3 Typical methods of fuzzy overlapping detection

检测方法	核心思想	性能优势	性能不足	代表性算法
扩展标签传播	邻域节点之间的社区信息交互根据模糊隶属度更新社区归属	具有较快的检测速度	隶属度阈值和节点隶属最大社区个数影响较大且最优设置困难	COPRA ^[41] , SLPA ^[42] , LP-PB ^[9]
非负矩阵分解	将特征矩阵在非负约束下分解得到节点对不同社区的隶属度分布	能够获得节点对所属社区的绝对隶属度	高维矩阵分解复杂度较高,需要预知社区数目,隶属度阈值最优设置困难	s-NMF ^[43] , 贝叶斯NMF ^[44] , BNMF-CD ^[45]
基于边界节点的两阶段检测	通过社区硬划分搜索边界节点,根据其邻域社区的隶属度确定重叠社区结构	具有较快的检测速度和较好的通用性	依赖于第一阶段社区硬划分质量,需严格保证精度	LM-NV ^[48]
模糊聚类	根据对象与集合之间的隶属度进行模糊聚类获得重叠社区结构	能够获得节点对所属社区的绝对隶属度	节点与社区的相似性测量影响较大,快速高精度测量较为困难	CDHC ^[52] , FRC-FGSN ^[53]
模糊模块度优化	全局最优化模糊重叠模块度函数获得重叠社区划分	自动搜索获得全局最优模糊重叠社区划分及完整节点隶属度分布	模糊模块度函数和优化算法性能影响较大,优质函数和高精度算法构建困难	MSFCM ^[3] FMM/FBLN ^[1] MAGA [*] ^[57]

尽管 Gregory 的最新研究成果中提出一种“MakeFuzzy”方法^[6],能够根据重叠节点对所属社区的连边数目计算隶属度,将离散重叠社区划分转化为模糊重叠社区划分。然而,由于一方面隶属度计算仅依据邻居节点数目,计算标准仅采用单一化的度信息,使隶属度估计的精确性受到制约;另一方面仅计算节点对自身所在社区的隶属度或贡献程度,因此该隶属度计算具有局部性和绝对性,对节点与所有社区之间的真实隶属程度评估的完整性受到影响。上述原因使转化所得模糊重叠社区划分并不能很好地体现真实重叠社区结构的复杂性和模糊性。

现有的模糊重叠社区检测方法中,模糊模块度优化方法由于具有检测过程无需获得社区结构硬划分、重叠节点挖掘不依赖于复杂的相似度和隶属度计算、社区划分质量评价函数与优化目标函数无差异、检测结果能够完整体现所有节点对各社区的隶属度分布、操作简单易实现等诸多性能优势,成为目前功能性相对最强且最具发展潜力的一类检测方法。然而,由于模糊模块度函数最优化本质上仍是一个典型的 NP 难问题,一般算法很难在规定时间内求得满意的最优解^[1-3,55-57],而模糊重叠社区检测由于不仅优化节点社区归属而且优化节点隶属度分布,进一步扩张了最优化问题的决策空间及目标空间,加剧了模块度优化时存在的极端退化^[58]现象,使得全局最优解的搜索难度大幅增加。现有模块度优化检测方法中普遍使用的贪心法^[59,60]、模拟退火^[61,62]、谱方法^[63]、极值优化^[64]、数学规划法^[65,66]等最优化算法,均属于基本启发式最优化方法,虽然具有理论比较成熟、设计简单、鲁棒性强、速度快等优势,但通常存在全局寻优能力不足、求解精度低、易陷入局部最优等缺陷,不仅难以得到全局最优的社区划分,而且在求解规模较大、搜索复杂度较高的全局最优化问题时存在搜索效率急剧下降的问题^[10],因而使复杂网络社区检测中模糊模块度函数的最优化性能受到较大影响。

进化算法(Evolutionary Algorithms, EAs)作为智能优化领域中国内外公认的求解 NP 难问题最有效的超启发式最优化方法,不仅对优化问题的种类有较强的鲁棒性,而且具有强大的全局最优化能力及自组织、自学习特性,能够在大规模复杂搜索空间中智能化且高效地收敛到全局最优解,基于种群进化的搜索机制使其具有更强的稳定性,因此更适合于模糊重叠社区检测中复杂的模糊模块度函数最优化问题求解。鉴于此,本文将重点对基于 EAs 的模糊模块度优化模糊重叠社区检测方法进行分析和说明。

4 基于 EAs 的模糊模块度优化模糊重叠社区检测

4.1 基于 EAs 的模糊模块度优化方法框架

基于 EAs 的模糊模块度优化方法框架设计思路为,将模糊重叠社区检测问题抽象为带约束条件的模糊隶属度分布组合优化问题,利用遗传算法(Genetic Algorithm, GA)、免疫克隆算法(Immune Clone Algorithm, ICA)及粒子群优化算法(Particle Swarm Optimization, PSO)等群智能进化算法,在节点隶属度分布对应的连续空间

内进行全局搜索,在模糊模块度引导及模糊重叠约束条件下收敛至全局最优的隶属度分布,由此得到真正的模糊重叠社区划分,并根据隶属度分布确定重叠社区结构。

基于 EAs 的模糊模块度优化社区检测方法的数学模型如下:不失一般性,以一个具有 n 个节点, c 个社区和 m 个质量评价函数的模糊重叠社区检测问题为例,其对应带约束条件的节点隶属度组合优化问题数学模型如式 (9) 所示。

$$\begin{aligned}
&\text{maximize} && y = F(x) = (f_1(x), f_2(x) \cdots f_m(x)) \\
&&& U \in R^{n \times c}, x = \bar{\omega}(U) \\
&\text{subject to} && 0 \leq u_{ik} \leq 1, \forall k, i; \sum_{i=1}^n u_{ik} < n, \forall k \\
&&& \sum_{k=1}^c u_{ik} = 1, \forall i
\end{aligned} \tag{9}$$

式中 $U = \{u_{ik}\} \in R^{n \times c}$ 为节点隶属度分布矩阵, u_{ik} 代表节点 i 对于第 k 个社区的隶属度, $i = [n], k = [c]$ 。 $x = \varphi(U)$ 为隶属度分布 U 的映射,代表 U 对应的模糊重叠社区划分。 $y = F(x)$ 定义了由决策空间到 m 维目标空间的映射函数,包含 m 个模糊重叠社区划分质量评价函数。当 $m = 1$ 时为单目标优化问题,目标函数即为模糊模块度函数 Q ;当 $m > 1$ 时转化为多目标优化问题,目标函数为包含 Q 在内的函数集合,衡量社区划分在多项评价标准上的性能优劣。

基于 EAs 的模糊模块度优化方法中,种群个体在初始化及进化过程中需满足模糊重叠社区检测对应的多项约束条件,具体包括:1)满足可行社区划分对应的一般约束条件以保证个体的合法性。2)满足模糊重叠特性对应的两项特殊约束条件:任一节点 i 对任一社区 k 的隶属度 u_{ki} 在 $[0, 1]$ 范围内取值;任一节点 i 对所有社区的隶属度总和取值恒为“1”。EAs 在上述约束条件限定的节点隶属度分布可行连续空间内进行全局搜索,收敛到使集合 F 中目标函数同时达到最优的隶属度分布 U^* ,对应最优的模糊重叠社区划分为 $x^* = \text{argmax} F(x)$ 。

构建通用的基于 EAs 的模糊模块度优化方法基本流程框架如图 6 所示,对应具体步骤描述如下:

步骤 1 对种群个体进行编码并构建初始种群,解码获得个体对应的初始社区划分。

步骤 2 在不同类型的编码方式下获得种群个体对应的模糊重叠社区划分。若采用非隶属度矩阵编码,按照一定准则计算初始社区划分中每个节点对各社区的隶属度分布;若采用隶属度矩阵编码,则无需隶属度计算。

步骤 3 判断终止条件是否满足?若满足,则输出当前种群中最优模糊重叠社区划分作为检测结果,否则转至步骤 4。

步骤 4 对父代种群执行变异、交叉进化操作生成子代种群,进化过程保留父代个体中优质的社区划分基因。

步骤 5 对子代种群进行 Clean-Up 操作,修正进化操作中社区划分错误的节点,使邻域节点尽可能划分至同一社区。

步骤 6 对修正后的子代种群个体进行适应度值评价。根据构建或选择的 Q 函数的不同,判断是否首先需要设置最优的模糊隶属度阈值。若需要阈值设置,则根据实验经验或一定的优化方法进行设置,根据阈值将模糊重叠社区划分转化为相应的离散重叠社区划分或硬划分,再计算其模块度函数值。若不需要阈值设置,则直接对模糊重叠社区划分进行模糊模块度计算。

步骤 7 判断目标函数个数 m 是否大于 1,即是否为多目标优化。若为单目标优化,则直接根据子代种群个体的 Q 值进行精英选择,保留优质个体进入下一代父代种群,转至步骤 3;若为多目标优化,则首先计算子代种群个体的其他目标函数值,并对合并之后的父代子代种群集合进行 Pareto 非支配排序及个体密度估计,然后根据个体精确性和分布性结果进行环境选择,保留同等规模的 Pareto 最优个体进入下一代父代种群,转至步骤 3。

上述检测流程中需要说明的是,当模糊隶属度函数能够直接对节点隶属度分布对应的模糊重叠社区划分进行质量评价时,无需根据隶属度阈值将模糊重叠社区划分转化为离散重叠社区划分或硬划分,由此可避免隶属度阈值的设置。此外,由于多目标优化中无法根据单个目标函数值确定唯一的最优解,因此需要通过 Pareto 非支配排序和个体密度估计构建 Pareto 最优的非支配解集,其中 Pareto 支配、Pareto 最优等多目标优化概念在复杂网络社区检测中的定义可参见文献[67],文中在此不做详述。

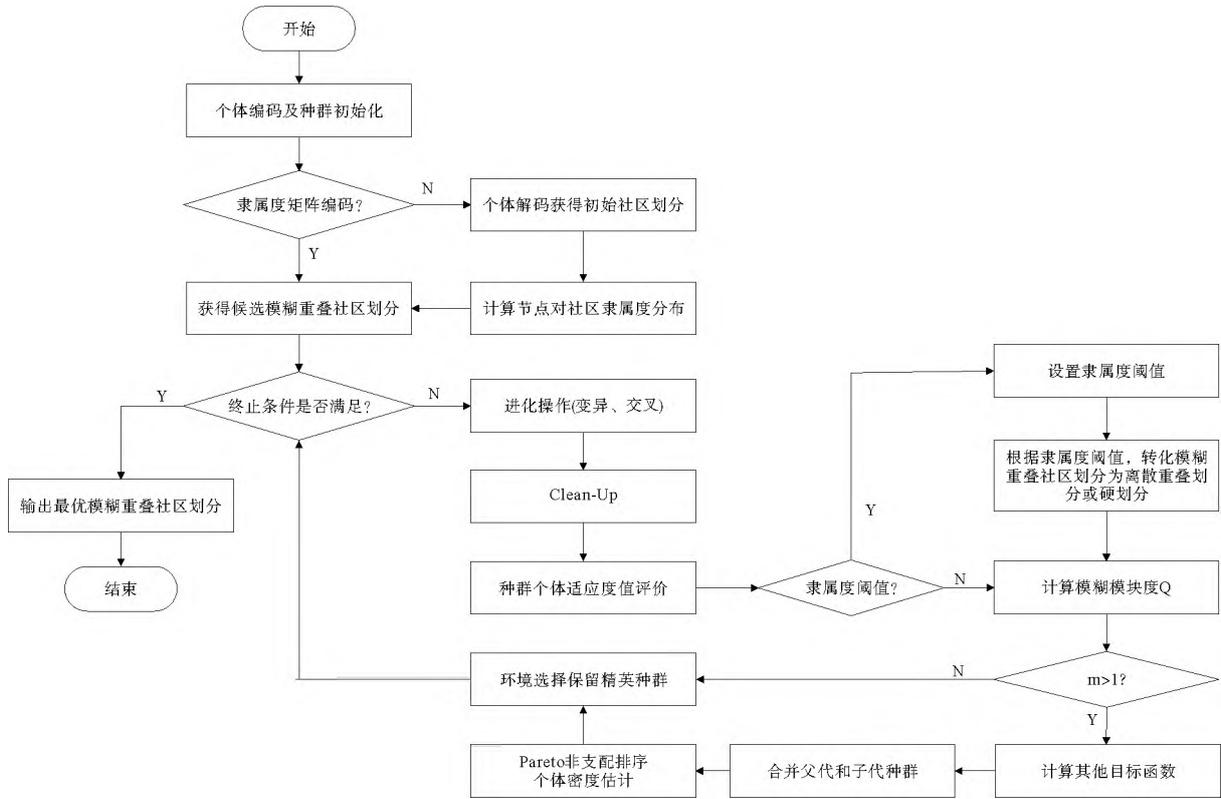


图 6 基于进化算法的模糊模块度优化检测框架

Fig. 6 Flowchart of fuzzy modularity optimization based on evolutionary algorithm

为测试基于进化算法的模糊模块度优化算法性能,本文设计了基于广义重叠模块度函数 GM 和遗传算法 GA 的单目标模糊模块度优化算法 GM_GACD。为验证该类算法的有效性,将其与目前国内外较为先进的多种社区检测算法,在计算机生成网络和真实世界网络上进行性能测试。

1)非重叠社区检测性能测试

将 GM_GACD 在扩展 GN Benchmark 网络上进行社区检测,所得社区划分对应归一化互信息 NMI 和正确节点检测比例统计数据如表 4 所示。扩展 GN Benchmark 网络含有 128 个节点,共被划分为 4 个不同社区,每个社区包含 32 个结点,社区中每个节点的平均度数为 16。随着 Z_{out} 值的增加,网络社区结构模糊程度逐渐增强,精确检测出社区结构的难度大幅提升。从实验结果可以看出,当 $Z_{out} > 7$ 时 GM_GACD 仍然能够获得较好的检测精度,说明算法能够有效实现非重叠网络社区检测且具有较好的稳定性。

表 4 GM_GACD 算法在 GN 网络上检测结果

Tab. 4 Community detection results of GM_GACD on GN network

测试指标	$Z_{out}=0$	$Z_{out}=1$	$Z_{out}=2$	$Z_{out}=3$	$Z_{out}=4$	$Z_{out}=5$	$Z_{out}=6$	$Z_{out}=7$	$Z_{out}=8$	$Z_{out}=9$	$Z_{out}=10$
NMI	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	0.944 5	0.800 2	0.079 1	0.017 1
Fraction_array	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	0.957 0	0.937 5	0.312 5	0.320 3

将 GM_GACD 与多种社区检测算法 FN、GN、VSGD^[41]、MSFCM^[2]、FMM/H1^[1]、MOGA-Net^[22] 和 MA^[48] 等,在斯坦福大学网络数据集中的 4 种真实网络上进行测试,采用标准模块度函数 Q 测量算法所得非重叠社区划分质量,实验统计结果如表 5 所示,表中数据为各算法独立运行 20 次的平均最优值,各项对比实验中的最优结果均用黑体加粗表示。

从表 5 中统计数据可以看出,GM_GACD 相比较于其他算法能够得到相对较优的模块度数值 Q,说明其不仅能够在真实网络上检测出较为精确的社区结构,而且能够有效识别出社区结构中的小团体,实现多尺度社区划分。

表5 GM_GACD算法在真实网络上非重叠社区检测结果
Tab.5 Community detection results of GM_GACD on real-world networks

网络	FN	GN	VSGD	MSFCM	FMM/H1	MOGA-Net	MA	GM_GACD
Karate	0.252 8	0.401 3	0.219 8	0.423 1	0.394 1	0.372 0	0.419 4	0.419 8
Dolphins	0.371 5	0.470 6	0.119 9	0.399 1	0.488 2	0.485 0	0.526 1	0.530 9
PolBooks	0.502 0	0.516 8	0.069 3	0.460 1	0.517 5	0.519 7	0.527 0	0.527 0
Football	0.454 9	0.599 5	0.146 4	0.526 8	0.596 0	0.546 0	0.600 9	0.604 6

2)重叠社区检测性能测试

为测试 GM_GACD 在重叠网络上的社区检测性能,将 GM_GACD 与 LINK、COPRA、CFinder、CPM 和 LFM 几种代表性重叠社区检测算法,在斯坦福大学网络数据集中的 4 种真实网络上进行性能测试,采用具有重叠结构的模块度 Q_{ov} 评价算法对于重叠社区结构的检测性能,实验结果如表 6 所示,表中数据为各算法独立运行 20 次的平均最优值,各项对比实验中的最优结果均用黑体加粗表示。

表6 GM_GACD算法在真实网络上重叠社区检测结果
Tab.6 Overlapping community detection results of GM_GACD on real-world networks

网络	LINK	CFinder	CPM	LFM	COPRA	GM_GACD
Karate	0.146	0.483	0.522	0.648	0.419	0.661
Dolphins	0.351	0.593	0.663	0.636	0.687	0.719
PolBooks	0.254	0.758	0.790	0.807	0.813	0.807
Football	0.568	0.589	0.646	0.629	0.685	0.717

从表 6 中统计数据可以看出,GM_GACD 相比较于其他算法能够得到相对较优的模糊模块度数值 Q_{ov} ,说明其能够有效提取出真实网络中的重叠社区结构。此外,根据 GM_GACD 模糊重叠社区检测结果所得隶属度分布,能够测量节点的重叠程度差异分布,如 Karate 网络和 PolBooks 网络中各节点的重叠程度如图 7-8 所示,节点颜色热度图分布对应节点重叠程度大小。

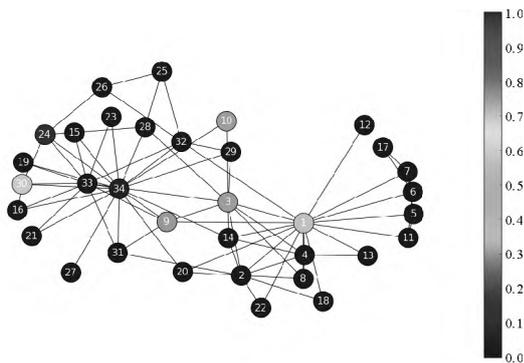


图7 Karate网络节点重叠程度分布图
Fig.7 Overlapping degree distribution of nodes in Karate network

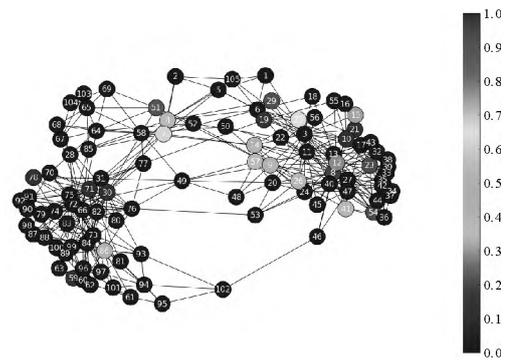


图8 PolBooks网络节点重叠程度分布图
Fig.8 Overlapping degree distribution of nodes in PolBooks network

4.2 基于 EAs 的模糊模块度优化关键技术及挑战

由上述实验分析可以看出,基于 EAs 的模糊模块度优化方法不仅能够获得相对较优的非重叠社区结构,而且能够挖掘出真实社区中存在的小团体;不仅能够实现重叠社区检测,而且能够获得真正的模糊重叠社区划分。基于模糊隶属度分布,能够测得网络节点的重叠程度分布等深层次重叠社区结构信息,加强对于真实重叠社区结构中复杂且模糊拓扑结构的探索能力。尽管模糊重叠社区检测算法在检测功能上具备一定优势,但由于重叠社区检测算法发展尚未成熟,模糊重叠社区检测问题本身的复杂度较高且检测难度较大,因此现阶段针对该类算法的研究仍处于起步阶段,在实际应用中仍然面临若干关键挑战,包括个体隶属度编码问题、社区个数最优设置问

题、模糊模块度函数设计问题、模糊隶属度阈值设置问题、模糊重叠社区划分评价问题、以及 EAs 全局最优化性能问题等。

4.2.1 个体隶属度编码问题

EAs 中种群个体对应候选可行社区划分,编码操作将个体每一维代表的节点与社区分布联系起来,是应用 EAs 实现模糊重叠社区检测的重要环节。现有离散重叠及非重叠社区检测中普遍使用的个体编码方式主要包括两种:字符串编码^[68]和基于基因位的局部邻接编码(Locus-Based Adjacency Representation, LAR)^[69],二者解码后形成的社区划分均为硬划分,从社区划分本身无法体现出社区重叠情况,更无法体现节点对不同社区的隶属程度。为实现模糊重叠社区检测,现有检测算法中主要采取了两种不同的解决方式:

1)对字符串编码及 LAR 编码所得社区硬划分进行转化,使其能够体现出重叠社区及重叠节点对不同社区的隶属程度^[57]。通常采用的处理方法为:针对个体编码所得社区硬划分,首先根据节点的连接数目或邻居节点数目,计算每个节点对自身所在社区及邻域社区的隶属度,然后通过设置隶属度阈值确定每个节点的社区归属数目及社区分布,从而得到个体对应的模糊重叠社区划分。然而,该种编码转化方式存在计算复杂、效率较低且稳定性较差等缺陷,主要原因包括:(1)隶属度计算存在多种方式,不同的计算标准和计算方式下同一节点对不同社区的隶属度分布可能存在较大差异,使得重叠节点的判定存在较大不确定性,同一硬划分可能转化为不同的模糊重叠社区划分,个体编码的质量和稳定性难以保证;(2)隶属度阈值的最优设置没有统一有效的解决方法,且通常设置方法较为复杂。

2)在个体编码时直接对节点的隶属度分布进行数值化编码,即将公式(1)中的隶属度分布矩阵 $U = \{u_{ik}\}$ 作为个体^[10]。根据算法需要也可将个体隶属度编码的矩阵形式转化为向量形式,如式(10)所示。

$$U = [u_{11} \cdots u_{n1}, u_{12} \cdots u_{n2}, \cdots, u_{1k} \cdots u_{nk}, \cdots, u_{1c} \cdots u_{nc}] \quad (10)$$

该种基于隶属度分布的编码方式相比较于字符串编码和 LAR 编码的性能优势包括:1)能够清晰直观地体现所有节点对各社区的隶属度分布,个体本身即为一个真正的模糊重叠社区划分;2)通过种群在隶属度分布组合对应可行空间中的搜索实现模糊重叠社区检测,种群收敛至全局最优解对应个体即为最优模糊重叠社区划分,由此避免了隶属度计算,以及模糊划分与离散划分及硬划分之间的转化问题。然而,该种编码方式也存在较为明显的缺陷:1)个体编码的维度较高,单个个体二维矩阵编码使得种群需要三维矩阵表示,导致进化操作中的计算复杂度增加。2)要求社区数目参数 c 不小于真实社区数目,因此通常需要真实社区数目的相关先验信息,在许多实际网络环境下难以实现。此外,若参数 c 远大于真实社区数目会导致较高的时间和空间计算成本。

由上述分析可知,模糊重叠社区检测中由于需要利用并体现节点对社区的隶属度分布,因此增加了 EAs 中个体编码的复杂度。如何选择和设计个体编码方式,使其在编码功能和编码效率之间达到良好平衡,将是算法研究中需要解决的首个关键问题。

4.2.2 社区个数最优设置问题

社区个数直接关系到检测所得社区划分的合理性和准确性,是检测算法中最重要的参数。社区个数的最优化设置是模糊重叠社区检测中需要考虑和解决的重要问题。

现有模糊重叠社区检测方法中,最优社区数目的确定主要采用以下几种方法:1)根据先验知识预知真实社区个数。该方法最为简单,但实际网络环境下通常无法获得准确的社区个数信息。2)在 LAR 编码方式下,解码过程中能够根据节点社区编号和邻域节点连接情况确定个体对应社区划分,并在此过程中自动获得社区个数,检测结果将根据评价指标确定最优个体及对应社区个数。该方法无需社区个数的先验信息或人为干预,但无法体现隶属度信息。3)社区数目递增的优化方法^[43,51,54]。社区数目从 2 开始以 1 为步长逐步递增,计算每个社区数目下最优社区划分对应的模糊模块度函数值直至模块度函数不再增加,确定最高模块度数值对应社区数目为最优社区数目^[51,54]。该方法能够在一定程度上提升最优社区个数设置的准确性,但模糊模块度函数优化受到包含社区数目在内多种因素的影响,因此容易获得局部最优的社区数目。4)在以 MSFCM 和 FMM/H 为代表的模糊重叠社区检测算法中,通过遍历方法获得最优社区个数。首先根据先验知识预知社区个数的取值范围,然后在所有可能社区个数取值下进行模糊重叠社区检测,获得多个局部最优社区划分,最后根据评价指标确定全局最优社区划分,并确定最优社区个数。该方法既能够获得相对精确的全局最优社区个数,但相应的计算复杂度也较高。因此,如何精确且高效地获得模糊重叠社区划分的最优社区个数也是一项有待解决的关键问题。

4.2.3 模糊模块度函数设计问题

模糊模块度函数的设计与选择是模糊模块度优化方法中的核心关键问题,模糊模块度函数不仅在 EAs 优化

过程中起到最优社区划分搜索的方向性引导作用,而且是社区划分质量的评判标准,因此对于模糊重叠社区检测性能具有决定性影响。

现有的扩展模块度函数大多仅能够用于评估离散重叠社区划分的模块性优劣,能够评价模糊重叠社区划分的模糊模块度函数数量极少。目前已知的模糊模块度函数主要包括:Zhang 提出的基于隶属度阈值 λ 的模块度 $Q_z^{[55]}$;Nepusz 等提出基于节点隶属度相似性的模糊模块度函数 Q_{ov}^{Ne} [54];Nicosia 构建了基于链接归属系数的模糊模块度函数 $Q_{ov}^{[10]}$;Liu 构建了基于节点对社区模糊隶属度的零模型,并提出基于零模型的模糊模块度函数 $Q_l^{[56]}$; Havens 等提出的能够同时评估离散重叠社区划分和模糊重叠社区划分的广义重叠模块度函数 (Generalized Modularity, GM) 或称 Q_g 等[3]。然而,尽管 Q_z 和 Q_l 能够对模糊重叠社区划分进行质量评估,但因为二者需要首先根据预定隶属度阈值将模糊重叠社区划分进行转化,分别获得相应的离散重叠划分和硬划分后才能进行质量评估,因此都不是真正的模糊模块度函数[1],将其作为模糊模块度优化的目标函数,最终得到的最优社区划分不是模糊社区划分。相比之下,广义模糊重叠模块度函数 GM 能够直接对模糊重叠社区划分进行质量评估,无需设定隶属度阈值对模糊重叠社区划分进行转化,适用于模糊重叠、离散重叠及非重叠社区划分的质量评价,因此具有明显的性能优势和通用性。然而,Jianhai Su 等在文献[1]中的实验结果表明,将 Q_g 作为模糊模块度优化的目标函数,可能得到不同类型的社区划分,包括模糊重叠社区划分和硬划分。优化 Q_g 能够在 Jazz、Word 和 Metabolic 网络上获得最优的模糊重叠社区划分,而在 Karate、Dolphin 和 Football 等网络上得到的最优社区划分仍然是硬划分, Q_g 在不同网络上实现模糊重叠社区检测的适用性及其与网络类型及重叠特性之间的关联性仍有待于进一步研究。因此,构建高效通用的模糊模块度函数,提升模糊重叠社区划分评价质量,探索模糊模块度函数与网络类型及重叠特性之间的关联性,是模糊重叠社区检测研究中最关键的重点和难点问题。

4.2.4 模糊隶属度阈值设置问题

模糊重叠社区检测过程中节点对社区的隶属度分布通常仅仅是一系列人工权重系数,没有清晰明确的物理含义[29],通常需要人为指定界限值即隶属度阈值,确定每个节点最终的社区归属。隶属度阈值的设置对社区划分质量具有重要影响,若取值过高则容易丢失重叠节点,导致社区结构重叠程度偏低;若取值较低则产生过多重叠节点,导致社区结构重叠程度偏高,容易产生过度重叠或重复社区,因此最优隶属度阈值的设定成为影响模糊重叠社区划分质量的关键性因素。然而,现有研究中没有通用且高效的最优隶属度阈值设定方法,通常在大量统计实验过程中依据人工经验设定[2]。由于不同网络对应的最优隶属度阈值通常存在差异,该方法容易引入误差且缺乏理论依据,使最优模糊重叠社区划分的质量和稳定性受到影响。因此,如何提高最优隶属度阈值设置的科学性和稳定性,或通过构建性能更优的模糊模块度函数避免隶属度阈值设置,将是模糊重叠社区检测中的重要挑战。

4.2.5 模糊重叠社区划分评价问题

模糊重叠社区划分的评价指标既可用于评估社区划分质量,也可作为 EAs 的优化目标函数,因此对于模糊重叠社区检测结果具有重要影响。由于重叠社区划分的特殊性和复杂性,非重叠社区划分的评价指标并不适用于重叠社区检测,离散重叠社区划分的评价指标也不适用于模糊重叠社区检测。

现有针对重叠社区划分的质量评价标准,从评价功能上看主要包括两类:一类用于评价所得社区划分的模块性,常用于未知真实社区划分的网络,如 Karate、Football、Blogs 等真实网络,评价函数主要为标准模块度函数 Q 、 Q 的重叠扩展形式 $Q_{ov}^{[19]}$ 、 $EQ^{[70]}$ 以及模块密度 $D^{[71]}$ 等;另一类用于评价所得社区划分与真实社区划分的一致性,常用于已知真实社区划分的网络,如 LFR 等人工合成网络,评价函数主要包括归一化互信息 (Normalized Mutual Information, NMI) 及其重叠扩展形式 (Extended Normalized Mutual Information, ENMI), Adjusted Rand Index 及其重叠扩展形式 Omega Index[6] 等。

目前针对模糊重叠社区划分质量评价的研究尚未完善,现阶段存在的问题包括:1)由于目前仅有极少数重叠社区检测方法能够获得真正的模糊重叠社区划分结果,因此能够用于评价模糊重叠社区划分质量的评价标准及评价函数较为缺乏,已知的评价标准仅有 Fuzzy Rand Index[6] 和 Normalized Fuzzy Mutual Information (NFMI)[53],能够测量两个模糊重叠社区划分的相似性。然而,Fuzzy Rand Index 和 NFMI 仅适用于已知真实模糊重叠社区结构的网络,对于未知真实社区结构的真实网络,大部分现有算法仍采用非重叠社区划分评价指标 Q 对转化后的非重叠社区划分进行近似质量评估[39],在一定程度上限制了模糊重叠社区划分评估的精确性。2)绝大部分现有基于 EAs 的模糊重叠社区检测中均将模糊模块度函数作为唯一的优化目标函数进行最优化,对于检测结果应用标准模块度函数进行质量评估,即对于最优社区划分的搜索和评价仅考虑了模块性指标。然而,实验研究

表明,社区划分的模块性和与真实社区结构的一致性之间存在一定的冲突,即模块性最优的社区划分不一定最符合真实社区结构。3)现有检测方法中缺乏完整的质量评价标准,对于模糊重叠社区检测,除了测量模糊重叠社区划分的模块性和与真实社区结构的一致性外,社区划分的重叠性也应得到有效评估,包括检测所得重叠节点的完整性和精确性等,因此评价标准中应增加 Recall 和 Precision 等评价指标。基于上述分析可知,构建或选择高性能的模糊重叠社区划分评价指标,并尽可能保证对于社区划分质量评价的完整性,从而提升对检测算法性能的评价质量,是模糊重叠社区检测中有待完善的重要内容。

4.2.6 EAs 全局最优化性能问题

基于 EAs 的模糊模块度优化完全依靠进化种群个体在可行搜索空间中的探索性和开采性并行搜索,获得全局最优的模糊重叠社区划分。随着网络规模、复杂度、模糊程度及重叠程度的增加,最优化社区划分搜索对应的决策空间和目标空度维度大幅增加,极端退化现象逐渐严重,全局最优解搜索难度逐渐加大,容易导致 EAs 陷入局部最优甚至求解失败,因此 EAs 的全局最优化能力对于模糊重叠社区检测质量具有重要影响。

现有基于 EAs 的模糊重叠社区检测中采用的 EA 主要为遗传算法 GA,作为传统且典型的一类进化优化算法,GA 虽然较为通用,但其在大规模复杂最优化问题上面临收敛性能急剧下降的问题,现有算法中通常需要加入局部搜索操作,如爬山法、模拟退火等,以提升算法成功收敛到全局最优解的概率。近年来,智能优化领域中涌现出多种基于群智能的高性能 EAs,包括差分进化算法(Differential Evolution, DE)^[72]、人工蜂群算法(Artificial Bee Colony Algorithm, ABC)^[73]、引力搜索算法(Gravitational Search Algorithm, GSA)^[74]、社会蜘蛛优化算法(Social Spider Optimization, SSO)^[75]等。实验研究已证实上述算法在收敛精度、收敛速度及稳定性方面相比较于 GA 具有明显优势,因此可尝试将其应用于模糊模块度优化问题的求解,以提升全局最优模糊重叠社区划分质量。此外,模糊模块度最优化问题具有一定的特殊性,即种群个体的各维分量之间由于受网络拓扑关系制约而紧密相联。EAs 优化过程中一方面需要对交叉、变异等核心进化操作进行转化,将个体各维分量与社区信息关联起来,在保证合法性的基础上,使个体在进化过程中尽可能保留优质分量中携带的社区划分信息;另一方面,需要在 EAs 中添加 Clean-up 等操作对进化操作中产生的错误划分的节点进行修正,以提升相邻节点划分至同一社区的概率,在提升社区划分精确度的同时加快算法搜索速度。

基于上述分析可知,如何根据实际网络检测需求选择高收敛性能的 EAs,在编码、进化操作核心操作中有效利用网络拓扑及社区信息,利用其在大规模复杂最优化问题中的强大的全局收敛能力,提升模糊模块度优化质量,是基于 EAs 的模块度优化模糊重叠社区检测研究中要解决的重要问题。

5 研究趋势分析

模糊重叠社区检测通过扩展隶属度取值空间,大幅细化了对于节点与社区之间隶属关系的刻画,能够在社区划分中体现所有节点对所有社区完整且相对的隶属度信息,不仅能够更加精确、细致地反映出较为复杂且模糊的真实重叠社区结构,而且能够挖掘出重叠社区结构中隐含的深层次拓扑信息。模糊重叠社区检测强大的功能特性,使其对于精确挖掘复杂网络的重叠社区结构、掌握网络中隐含的深层次拓扑结构特征、分析社区结构的功能特性及动力学特性、预测网络行为及演化态势等具有重要意义。

真实环境中复杂网络的规模日益庞大、拓扑结构日趋复杂,模糊程度及重叠程度不断增强,用户对于社区检测的需求逐渐多样化和复杂化。为提升真实网络重叠社区检测精度、尽可能深度挖掘出社区结构信息并最大化满足用户需求,模糊重叠社区检测技术的应用需求日益凸显,同时对现有模糊重叠社区检测方法的性能提出了严峻挑战。综合分析模糊重叠社区检测现阶段已取得的研究成果及目前面临的技术难题,分析未来研究方向及发展趋势主要包含以下 4 个方面。

5.1 模糊模块度优化的扩展

从国内外相关研究可以看出,基于模糊模块度优化的检测方法由于能够获得真正的模糊重叠社区划分,且在易操作性、稳定性等方面具备一定优势,成为目前模糊重叠社区检测的主要方法。然而,模糊模块度函数仅能够评价社区划分的模块性能,无法有效测量社区划分与真实社区结构之间的相似性,以及重叠社区划分的重叠性。随着真实网络环境下用户多样性检测需求的不断提升,最优社区划分同时满足多项检测性能需求的趋势愈加明显,如同时保证较高的模块性和精确性等。

当模糊重叠社区划分的性能指标涵盖模块性、精确性、重叠性等多项评价标准时,待优化的目标函数数目将超过 2 项,单目标模糊模块度优化问题将演变为多目标优化问题(Multi-objective Optimization Problems,

MOPs)甚至是高维多目标优化问题(High-dimensional Multi-objective Optimization Problems, LMOPs)(目标函数数目大于4)。MOPs中由于需要同时使多目标函数达到相对最优,需要平衡多目标函数之间的冲突关系,而LMOPs中高维的特性进一步加剧了MOPs的目标空间复杂度和最优解搜索难度,是目前智能优化领域中最难求解的优化问题之一,需要设计高性能的多目标进化算法(Multi-Objective Evolutionary algorithms, MOEAs)才能有效求解^[76]。因此,如何设计收敛精度高、分布性好、速度快且稳定性强的MOEAs,对包含模糊模块度函数在内的高质量函数集合进行最优化,以获得多尺度、多样化的Pareto最优社区划分方案集合,充分利用目标函数之间的差异性和多样性提升模糊重叠社区划分的精确性、功能性和用户满意度,是模糊重叠社区检测未来发展的必然趋势和重点研究内容。

5.2 大规模网络数据处理中精度与效率的平衡

网络社区结构的模糊重叠现象及高精度的模糊重叠社区检测需求广泛存在于在线社交网络等大规模复杂网络环境中。一方面,大规模网络社区检测由于节点及连接数目较大,导致检测算法的时间和空间计算复杂度相对较高;另一方面,模糊重叠社区检测由于定量测定节点隶属度,且在检测结果中体现完整隶属度分布,进一步加剧了计算成本。目前现有模糊重叠社区检测方法难以在可行时间内有效处理十万节点以上大规模网络的社区检测问题。因此,如何通过有效利用大规模网络数据的稀疏性减少模糊模块度计算时间成本、EAs多种群分布式计算、全局搜索与局部搜索融合等方式,在保证模糊重叠社区检测质量的同时提升运算效率将是当前及未来模糊重叠社区检测研究中要解决的关键问题。

5.3 网络属性信息的利用与社区属性的体现

现有模糊重叠社区检测都是基于网络拓扑结构信息进行社区划分,大多未考虑网络中大量存在的属性信息,包括节点自身特征属性和链接特征属性等。一方面,真实社交网络中社区结构的形成除了显性的拓扑结构因素外,隐性的属性因素也具有重要影响,尤其重叠社区结构的产生可能更多依赖于节点和链接的属性相似性。此外,真实网络环境中可能已知部分节点的社区归属及隶属度信息,若对其加以有效利用能够有助于提升检测精度及效率。因此,模糊重叠社区检测过程中应尽可能加强对属性信息的利用,通过模糊隶属度量化并区分节点与社区之间属性相似性,从而提升对网络中重叠社区的检测能力。另一方面,对于模糊重叠社区检测所得的节点隶属度分布,除了希望其能够更加精确地反映出节点与社区之间的拓扑隶属度,也希望其能够体现节点与社区之间的属性相似性,从而进一步挖掘出重叠社区结构的属性信息,如重叠节点的属性多样性、社区属性相似性及差异性,从而更好地进行网络功能特性、动力学特性、演化特性等分析。因此,模糊重叠社区检测中网络属性信息有效利用,以及模糊重叠社区划分中属性信息的体现,都将是今后研究中有待解决的重要问题。

5.4 模糊重叠社区结构的动态演化更新

在社会网络等基于用户关联性所生成网络中,节点之间除了拓扑结构关系及属性相似性关系外,还存在频繁的信息交互,信息流动使得网络具有较强的动态特性^[77]。由于网络动态演化过程中,各个时间片上节点的复合属性可能为临时或过时状态^[78],因此节点的社区归属、节点的重叠性及重叠程度、社区的重叠程度等都可能发生变化。如何有效利用网络历史数据及新增数据,获得每个时间片上的模糊重叠社区划分,提升单个时间片上的重叠社区检测质量,并根据不同时间片上的网络快照分析重叠社区结构的动态演化过程,并对未来的演化趋势进行预测,是模糊重叠社区检测及动态网络社区检测未来的重要研究方向。

6 结语

模糊重叠现象广泛存在于社交网络等基于个体关联性构成的真实网络环境中,使网络中社区结构的拓扑关系呈现出较强的复杂性和模糊性。对大规模社交网络等真实复杂网络进行模糊重叠社区检测,能够更加深入、全面、细致地体现出节点与社区之间的隶属关系,从而更加精准地反映出复杂且模糊的真实社区结构,同时揭示出重叠社区结构中隐含的深层次拓扑信息,大幅提升对真实网络重叠社区拓扑结构的检测能力以及对拓扑关系的挖掘能力。对于发现社区结构显性及隐性的拓扑结构特征、属性特征、分析网络功能特性及动力学特性、掌握并预测网络演化规律等具有重要意义。

参考文献:

- [1] Su J H, Havens T C. Quadratic program-based modularity maximization for fuzzy community detection in social networks [J]. IEEE Transactions on Fuzzy Systems, 2015, 23(5):1356-1371.

- [2] Su J H, Havens T C. Community detection in social networks using a genetic algorithm [C]. Proc of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Beijing, China, 2014: 2039-2046.
- [3] Havens T C, Bezdek J C, Leckie C. A soft modularity function for detecting fuzzy communities in social networks [J]. IEEE Transactions on Fuzzy Systems, 2013, 21(6):1170-1175.
- [4] Xie J R, 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):1-37.
- [5] Golsefid S M M, Zarand M H F, Bastani S S. Fuzzy community detection model in social networks [J]. International Journal of Intelligent Systems, 2015, 30:1227-1244.
- [6] Gregory S. Fuzzy overlapping communities in networks [J]. Journal of Statistical Mechanics-Theory and Experiment, 2011, 2: P02017.
- [7] Kundu S, Pal S K. Fuzzy-rough community in social networks [J]. Pattern Recognition Letters, 2015, 67:145-152.
- [8] Bennett L, Kittas A, Liu S S. Community structure detection for overlapping modules through mathematical programming in protein interaction networks [J]. Plos One, 2014, 9(11):e112821-e112821.
- [9] 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, 2015, 38(47):1-17.
Liu Shichao, Zhu Fuxi, Gan Lin. A label-propagation-Probability-based algorithm for overlapping community detection [J]. Chinese Journal of Computers, 2015, 38(47):1-17.
- [10] Nicosia V, Mangioni G, Malgeri M. Extending the definition of modularity to directed graphs with overlapping communities [J]. Journal of Statistical Mechanics-Theory and Experiment, 2009(3):3166-3168.
- [11] Psorakis I, Roberts S, Ebden M, et al. Overlapping community detection using Bayesian non-negative matrix factorization [J]. Physical Review E, 2011, 83(6 Pt 2):1509-1520.
- [12] Zhao K, Zhang S W, Pan Q. Fuzzy analysis for overlapping community structure of complex network [C]. Proc CCDC, 2010: 3976-3981.
- [13] Palla G, Derényi I, Farkas I. Uncovering the overlapping community structure of complex networks in nature and society [J]. Nature, 2005, 435(7043): 814-818.
- [14] Farkas I, Abel D, Palla G, et al. Weighted network modules [J]. New J Phys, 2007, 9(6):180-198.
- [15] Palla G, Dernyi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society [J]. Nature, 2005, 435: 814-818.
- [16] Kumpula J M, Kivela M, Kaski K, et al. Sequential algorithm for fast clique percolation [J]. Phys Rev E 2008, 78(2):1815-1824.
- [17] Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks [J]. Nature, 2010, 466(7307): 761-764.
- [18] Kim Y, Jeong H. Map equation for link communities [J]. Physical Review E, 2011, 84(2):1402-1409.
- [19] 朱牧,孟凡荣,周勇. 基于链接密度聚类的重叠社区发现算法[J]. 计算机研究与发展, 2013, 50(12):2520-2530.
Zhu Mu, Meng Fanrong, Zhou Yang. Density-based link clustering algorithm for overlapping community detection [J]. Journal of Computer Research and Development, 2013, 50(12):2520-2530.
- [20] Wu Z, Lin Y, Wan H, et al. A fast and reasonable method for community detection with adjustable extent of overlapping [C]. Proc of International Conference on Intelligent Systems & Knowledge Engineering. 2010:376-379.
- [21] Fortunato S. Community detection in graphs [J]. Phys Rep, 2010, 486(3-5):75-174.
- [22] Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure of complex networks [J]. New Journal of Physics, 2008, 11(3):19-44.
- [23] Lee C, Reid F, McDaid A. Detecting highly overlapping community structure by greedy clique expansion [C]. Proceedings of the 4th International Workshop on Social Network Mining and Analysis. ACM, Washington DC, USA, 2010, 33-42.
- [24] Coscia M, Rossetti G, Giannotti F, et al. DEMON: a local first discovery method for overlapping communities [C]. proc of the 18th ACM SIGKDD Int Conf on Knowledge Discovery and Data Ming. New York: ACM, 2012:615-623.
- [25] Lancichinetti A, Radicchi F, Ramasco J J, et al. Finding statistically significant communities in networks [J]. Plos One, 2010, 6(4):336-338.
- [26] Jin D, Yang B, Baquero C, et al. A markov random walk under constraint for discovering overlapping communities in complex networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2011, 2011(5):50.
- [27] 陈俊宇,周刚,南煜,等. 一种半监督的局部扩展式重叠社区发现方法[J]. 计算机研究与发展, 2016, 53(6): 1376-1388.
Chen Junyu, Zhou Gang, Nan Yu, et al. Semi-supervised local expansion method for overlapping community detection [J]. Journal of Computer Research and Development, 2016, 53(6): 1376-1388.
- [28] Blondel V D, Guillaume J L, Lambiotte R. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics Theory & Experiment, 2015, 30(2):155-168.
- [29] Shen H W, Cheng X Q, Guo J F. Quantifying and identifying the overlapping community structure in networks [J]. Journal of Statistical Mechanics Theory & Experiment, 2009, 2009(7):07042.
- [30] Wang X, Li L, Cheng Y. An overlapping module identification method in protein-protein interaction networks [J]. BMC Bioinformatics, 2012, 13 (Suppl):343-347.
- [31] Franca F O D, Coelho G P. Identifying overlapping communities in complex networks with multimodal optimization [J]. IEEE Congress on Evolutionary Computation, Cancun, México, 2013: 269-276.
- [32] 张英杰,龚中汉,陈乾坤. 基于免疫离散差分进化算法的复杂网络社区发现[J]. 自动化学报, 2015, 41(4):749-757.

- Zhang Yingjie, Gong Zhonghan, Chen Qiankun. Community detection in complex networks using immune differential evolution algorithm [J]. *Acta Automatica Sinica*, 2015, 41(4):749-757.
- [33] 金弟, 刘杰, 杨博, 等. 局部搜索与遗传算法结合的大规模复杂网络社区探测[J]. *自动化学报*, 2011, 37(7): 873-882.
Jin Di, Liu Jie, Yang Bo, et al. Genetic algorithm with local search for community detection in large-scale complex networks [J]. *Acta Automatica Sinica*, 2011, 37(7): 873-882.
- [34] 金弟, 杨博, 刘杰, 等. 复杂网络簇结构探测——基于随机游走的蚁群算法[J]. *软件学报*, 2012, 23(3): 451-464.
Jin Di, Yang Bo, Liu Jie, et al. An colony optimization based on random walk for community detection in complex networks [J]. *Journal of Software*, 2012, 23(3): 451-464.
- [35] Zhou X, Liu Y H, Zhang J D. An ant colony based algorithm for overlapping community detection in complex networks [J]. *Physica A*, 2015, 427:289-301.
- [36] 黄发良, 肖南峰. 基于线图与 PSO 的网络重叠社区发现[J]. *自动化学报*, 2011, 37(9): 1140-1144.
Huang Faliang, Xiao Nanfeng. Discovering overlapping communities based on line graph and PSO [J]. *Acta Automatica Sinica*, 2011, 37(9): 1140-1144.
- [37] Pizzuti C. A multi-objective genetic algorithm for community detection in networks [C]. *Proceedings of the 21st IEEE International Conference on Tools with Artificial Intelligence*, 2009, 379-386.
- [38] 刘辰龙. 基于进化算法的社区检测及其在个性化推荐的应用[D]. 西安:西安电子科技大学, 2014.
Liu Chenlong. Evolutionary community detection algorithms and their applications in personalized recommendation [D]. Xi'an: Xidian University, 2014.
- [39] Wu Z H, Lin Y F, Gregory S, et al. Balanced multi-label propagation for overlapping community detection in social networks [J]. *Journal of Computer Science and Technology*, 2012, 27(3):468-479.
- [40] Qiang H, Yan G. A method of personalized recommendation based on multi-label propagation for overlapping community detection [J]. *System science, engineering design and manufacturing informatization (ICSEM)*, 2012 3rd International Conference on IEEE, 2012, 1: 360-364.
- [41] Gregory S. Finding overlapping communities in networks by label propagation [J]. *New Journal of Physics*, 2009, 12(10):2011-2024.
- [42] Xie J, Szymanski B K. Towards linear time overlapping community detection in social networks [C]. *Proc PAKDD Conf*, 2012, 25-36.
- [43] Zhao K, Zhang SW, Pan Q. Fuzzy analysis for overlapping community structure of complex network [C]. *Proc CCDC*, 2010, 3976-3981.
- [44] Psorakis I, Roberts S, Ebden M, et al. Overlapping community detection using Bayesian non-negative matrix factorization [J]. *Physical Review E*, 2011, 83(6 Pt 2):1509-1520.
- [45] 李玉翔, 李弼程, 郭志刚. 基于非负矩阵分解的网络重叠社区发现研究[J]. *系统仿真学报*, 2014, 26(3): 643-649.
Li Yuxiang, Li Bicheng, Guo Zhigang. Research on overlapping community detection in networks using non-negative matrix factorization [J]. *Journal of System Simulation*, 2014, 26(3):643-649.
- [46] Zhang S H, Wang R S, Zhang X S. Uncovering fuzzy community structure in complex networks [J]. *Physical Review E (S1550-2376)*, 2007, 76(4): 046103. 1-046103. 7.
- [47] 刘勇. 复杂网络的非重叠与重叠社区检测方法[D]. 西安:西安电子科技大学, 2014.
Liu Yong. Method for non-overlapping and overlapping communities detection in complex networks [D]. Xi'an: Xidian University, 2014.
- [48] 陈俊宇, 周刚, 熊小兵. 一种采用邻居投票机制的重叠社区发现方法[J]. *小型微型计算机系统*, 2014, 35(10):2272-2277.
Chen Junyu, Zhou Gang, Xiong Xiaobing. Detection overlapping community structure with neighbor voting [J]. *Journal of Chinese Computer Systems*, 2014, 35(10): 2272-2277.
- [49] Bennett L, Kittas A, Liu S S. Community structure detection for overlapping modules through mathematical programming in protein interaction networks [J]. *Plos One*, 2014, 9(11):e112821-e112821.
- [50] Liu J. Detecting the fuzzy clusters of complex networks [J]. *Pattern Recognit*, 2010, 43:1334-1345.
- [51] Wang W J, Liu D, Liu X, et al. Fuzzy overlapping community detection based on local random walk and multidimensional scaling [J]. *Phys A*, 2013, 392(24): 6578-6586.
- [52] 李刘强, 桂小林, 安健, 等. 采用模糊层次聚类的社会网络重叠社区检测算法[J]. *西安交通大学学报*, 2015, 49(2): 7-13.
Li Liuqiang, Gui Xiaolin, An Jian, et al. Overlapping community detection algorithm based on fuzzy hierarchical clustering in social network [J]. *Journal of Xi'an Jiaotong University*, 2015, 49(2): 7-13.
- [53] Suman K D, Sankar K P. Fuzzy-rough community in social networks [J]. *Pattern Recognition Letters*, 2015, 67, 145-152.
- [54] Nepusz T, Petroczi A, Negyessy L, et al. Fuzzy communities and the concept of bridgeness in complex networks [J]. *Phys Rev E*, 2008, 77(1):119-136.
- [55] Zhang S, Wang R, Zhang X. Identification of overlapping community structure in complex networks using fuzzy c-means clustering [J]. *Physica A Statistical Mechanics & Its Applications*, 2007, 374(1):483-490.
- [56] Liu J. Fuzzy modularity and fuzzy community structure in networks [J]. *Physics of Condensed Matter*, 2010, 77(4): 547-557.
- [57] Zhan W H, Guan J H, Chen H H, et al. Identifying overlapping communities in networks using evolutionary method [J]. *Physica A*, 2016, 442:182-192.
- [58] 刘瑶, 康晓慧, 高红, 等. 基于节点亲密度和度的社会网络社区发现方法[J]. *计算机研究与发展*, 2015, 52(10): 2363-2372.
Liu Yao, Kang Xiaohui, Gao Hong, et al. A community detection method based on the node intimacy and degree in social network [J]. *Journal*

- of Computer Research and Development, 2015, 52(10): 2363-2372.
- [59] 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; P10008.
- [60] 冷作福. 基于贪婪优化技术的网络社区发现算法研究[J]. 电子学报, 2014, 42(4): 723-729.
Leng Zuofu. Community detection in complex networks based on greedy optimization [J]. Acta Electronica Sinica, 2014, 42(4): 723-729.
- [61] Guimera R, Amaral L. Functional cartography of complex metabolic networks [J]. Nature, 2005, 433: 895-900.
- [62] Medus A, Acuna G, Dorso C. Detection of community structures in networks via global optimization [J]. Physica A: Statistical Mechanics and Its Applications, 2005, 358: 593-604.
- [63] Ruan J, Zhang W. Identifying network communities with a high resolution [J]. Physical Review E, 2008, 77: 1-12.
- [64] 陈国强, 王宇平. 基于极值优化模块密度的复杂网络社区检测[J]. 华中科技大学学报(自然科学版), 2011, 39(4): 82-85.
Chen Guoqiang, Wang Yuping. Community detection in complex networks using extremal optimization modularity density [J]. Journal of Huazhong University of Science and Technology(Natural Science Edition), 2011, 39(4): 82-85.
- [65] Aloise D, Cafieri S, Caporossi G, et al. Column generation algorithms for exact modularity maximization in networks [J]. Physical Review E, 2010, 82: 046112.
- [66] Bennett L, Liu S, Papageorgiou L G, et al. Detection of disjoint and overlapping modules in weighted complex networks [J]. Advances in Complex Systems, 2012, 15: 11500.
- [67] 黄发良, 张师超, 朱晓峰. 基于多目标优化的网络社区发现方法[J]. 软件学报, 2013, 24(9): 2062-2077.
Huang Faliang, Zhang Shichao, Zhu Xiaofeng. Discovering network community based on multi-objective optimization [J]. Journal of Software, 2013, 24(9): 2062-2077.
- [68] Tasgin M, Bingol A. Communities detection in complex networks using genetic algorithms [C]. Proc Eur Conf Complex Syst, 2006, 1-6.
- [69] Pizzuti C. Community detection in social networks with genetic algorithms [C]. Proc 10th Annu Conf Genetic Evol comput, 2008, 1137-1138.
- [70] Shen H, Cheng X, Cai K, et al. Detect overlapping and hierarchical community structure in networks [J]. Physica A: Statistical Mechanics and its Applications, 2009, 388: 1706-1712.
- [71] 周东青, 王星, 程嗣怡, 等. 离散粒子群社区检测算法[J]. 系统工程与电子技术, 2016, 38(2): 428-433.
Zhou Dongqing, Wang Xing, Cheng Siyi, et al. Community detection algorithm via discrete PSO [J]. Systems Engineering and Electronics, 2016, 38(2): 428-433.
- [72] Bi X J, Xiao J. Classification-based self-adaptive differential evolution with fast and reliable convergence performance [J]. Soft Computing, 2011, 15(8): 1581-1599.
- [73] 王艳娇, 肖婧. 基于改进人工蜂群算法的高维多目标优化[J]. 中南大学学报(自然科学版), 2015, 46(6): 2109-2117.
Wang Yanjiao, Xiao Jing. Optimization of multi-objective problems based on artificial bee colony algorithm [J]. Journal of Central South University(Science and Technology), 2015, 46(6): 2109-2117.
- [74] Esmat R, Hossein N, Saeid S. GSA: a gravitational search algorithm [J]. Information Sciences, 2009, 179, 2232-2248.
- [75] Cuevas E, Cienfuegos M, Zaldivar D, et al. A swarm optimization algorithm inspired in the behavior of the social-spider [J]. Expert Systems with Applications, 2013, 40 (16): 6374-6384.
- [76] 肖婧, 毕晓君, 王科俊. 基于全局排序的高维多目标进化算法研究[J]. 软件学报, 2015, 26(7): 1574-1583.
Xiao Jing, Bi Xiaojun, Wang Kejun. Research of global ranking based many-objective optimization [J]. Journal of Software, 2015, 26(7): 1574-1583.
- [77] 索勃, 李战怀, 陈群, 等. 基于信息流动分析的动态社区发现方法[J]. 软件学报, 2014, 25(3): 547-559.
Suo Bo, Li Zhanhuai, Chen Qun, et al. Dynamic community detection based on information flow analysis [J]. Journal of Software, 2014, 25 (3): 547-559.
- [78] 国琳, 左万利, 彭涛. 基于隶属度的社会化网络重叠社区发现及动态集群演化分析[J]. 电子学报, 2016, 44(3): 587-594.
Guo Lin, Zuo Wanli, Peng Tao. Overlapping community detection and dynamic group evolution analysis based on the degree of membership in social network [J]. Acta Electronica Sinica, 2016, 44(3): 587-594.

(责任编辑 李进)