

文章编号: 1007-757X(2021)06-0001-04

# 基于多阶局部度数峰值点的局部社区发现算法

王得翊<sup>1</sup>, 焦澳琛<sup>1</sup>, 陈音拿<sup>1</sup>, 安静<sup>2\*</sup>, 康琦<sup>1</sup>, 汪镭<sup>1</sup>

(1. 同济大学 控制科学与工程系, 上海 201804;

2. 上海应用技术大学 电气与电子工程学院, 上海 201418)

**摘要:** 复杂网络研究面临网络规模过大难以获取全局信息的困境, 局部社区研究及其发现算法在其中发挥重要的作用。针对复杂网络的局部社区发现问题, 提出了“多阶局部度数峰值点”的新概念, 并以此构建了种子节点选取方法和局部社区检测框架。实验结果表明, 基于多阶局部度数峰值点的局部社区检测方法显著提高了局部社区发现的精确度。

**关键词:** 复杂网络; 社区检测; 局部扩张优化;  $n$  阶局部度数峰值点

**中图分类号:** TP181

**文献标志码:** A

## Local Community Detection Based on $n$ -order Local Degree Central Nodes

WANG Deyi<sup>1</sup>, JIAO Aochen<sup>1</sup>, CHEN Yinna<sup>1</sup>, AN Jing<sup>2\*</sup>, KANG Qi<sup>1</sup>, WANG Lei<sup>1</sup>

(1. Department of Control Science and Engineering, Tongji University, Shanghai 201804, China; 2. School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai 201418, China)

**Abstract:** Complex network research is focused on the problem that the network scale is too large to obtain global information. A local community research and its discovery algorithm play an important role in complex network research. A new concept of “multi-order local degree peak point” is proposed to solve the problem of local community discovery in complex networks, and a new seed node selection method and a local community detection framework are constructed. The experimental results show that the local community detection method based on the multi-order local degree peak point significantly improves the accuracy of local community detection.

**Key words:** complex network; community detection; local extension optimization;  $n$ -order local degree peak point

## 0 引言

随着科学研究对象规模的不断扩大和结构的进一步复杂, 针对复杂网络的研究已经成为生物学、社会学、经济学、流行病学、统计物理学等诸多学科前沿领域的研究热点之一<sup>[1]</sup>。复杂网络一般具有社区结构, 即复杂网络往往可以被自然地分成一系列节点组, 同一组的节点一般更具有相连相关的倾向<sup>[2]</sup>。社区发现能够检测复杂网络中的社区结构, 是分析复杂网络、挖掘网络信息的有力方法, 例如, 通过研究蛋白质中氨基酸的连接关系发现功能模块和预测, 根据节点信息熵相似性和社区信息熵稳定性进行洗钱社区发现, 因而自 Kernighan-Lin 算法<sup>[3]</sup>被完善并被用于图分割始, 社区发现算法有了长足的发展。

传统的社区发现算法已知全局网络信息对网络中所有节点进行社区划分, 分为非重叠社区发现算法和重叠社区发

现算法。非重叠社区发现算法的结果中每个节点只能属于一个社区, 例如, 基于边介数的 GN 算法<sup>[4]</sup>、基于模块度概念 FastUnfolding 算法<sup>[5]</sup>和基于 LeaderRank 排序的标签传播算法<sup>[6]</sup>等。现实中复杂网络社区结构并非一个节点仅属于一个社区, 例如, 微博用户关注内容广泛难以被分入一个社区<sup>[7]</sup>。重叠社区发现算法可以允许节点同时属于多个社区, 例如, 基于完全子图的派系过滤算法<sup>[8]</sup>等。传统的社区发现算法需要事先知道全局网络的信息, 网络规模巨大, 不易获知全局网络信息。

局部算法<sup>[9]</sup>选择种子节点并以一定的方式扩张形成社区。在种子节点选取问题上, LFM 算法<sup>[10]</sup>主张以随机选择的方式确定种子节点; 基于图遍历的局部社区发现算法<sup>[11]</sup>将度数较低的节点作为种子节点; 提出局部度数中心点的概念, 主张以局部度数最高的节点作为种子节点<sup>[12]</sup>, 但由于种子节点选择标准严苛, 算法不能较为全面地感知各个社区的

基金项目: 国家自然科学基金项目(61703279); 上海市大学生创新训练项目(S201910003)

作者简介: 王得翊(1999-), 男, 本科生, 研究方向: 机器学习算法。

焦澳琛(1999-), 男, 本科生, 研究方向: 网络数据分类算法。

陈音拿(1998-), 男, 本科生, 研究方向: 基于大数据的分类算法。

康琦(1980-), 男, 博士, 教授, 研究方向: 智能自动化、机器学习。

汪镭(1970-), 男, 博士, 教授, 研究方向: 智能自动化、系统工程。

通信作者: 安静(1978-), 女, 博士, 副教授, 研究方向: 智能信息处理。

关键节点,发现社区的稳定性略有欠缺。

本文的主要贡献:(1)在局部度数中心点的基础上,提出了多阶局部度数峰值点的概念,并将其作为种子节点的选择依据。(2)提出了基于多阶局部度数峰值点的种子节点选择方法和局部社区检测框架。

1 相关工作

1.1 局部社区扩张算法

Clauset<sup>[13]</sup>提出了衡量局部节点组社区结构显著度的标准,即局部模块度。算法从一个给定节点作为初始节点组开始,不断将节点组的邻居节点加入节点组并更新局部模块度,直到加入任何邻居节点都不能使模块度值增大,得到扩张的局部社区。

罗峰等<sup>[14]</sup>提出了基于模块度的扩张算法。先将能使子图模块度增大的相邻节点加入子图,再将子图中能使模块度值进一步增大的节点删去,直至任何节点的加入和删去都不能使子图的模块度值增大为止,得到扩张社区。

Bagrow 等定义了与给定初始节点间最短路径等于  $L$  的节点集合  $L$ -shell<sup>[15]</sup>及其总出度  $K_L$ 。 $L$ -shell 不断扩张,直至  $K_L$  增长速度小于给定阈值  $\alpha$ ,扩张停止, $L$ -shell 已覆盖节点构成扩张社区。

1.2 局部度数中心点框架

对于网络  $G$  中的节点  $v$ ,如果节点  $v$  的度数不小于它的任何一个邻居节点,则称节点  $v$  为“局部度数中心点”。从初始节点开始找到与初始节点最近的“局部度数中心点”,并以此作为新的初始节点扩张得到局部社区。

2 基于多阶局部度数峰值点的局部社区发现算法

网络中度数最高的节点往往具有与其他节点紧密的联系,以它为关键节点进行扩张能达到较好的扩张社区效果,而一个社区内的关键节点度数可能不大。规模大、节点间联系紧密的社区内往往囊括了全网络大部分度数较高的节点,局部度数中心点是基于此建立的关键节点概念。

局部度数中心点框架下容易出现:某节点是其所在社区的关键节点,但与它的所有邻居节点相比,它并非度数最高的节点,这是因为度数高于它的节点属于其他的社区,如图 1 所示。

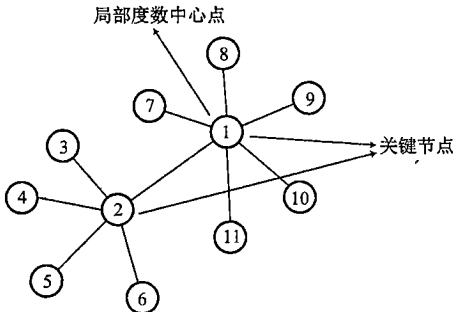


图 1 局部度数中心点概念失效

图 1 中,有 {1,7,8,9,10,11} 和 {2,3,4,5,6} 两个社区,节点 1 和 2 分别是两个社区的关键节点。按照局部度数中心点的概念,节点 2 并非关键节点,这是因为它的邻居节点 1 的度数比它高,而节点 1 属于另外的社区。

要避免上述情况,首先要放宽标准,即在寻找关键节点时,允许其邻居节点中有一些度数高于它,称这样的节点为准峰值点。

定义 1:对于网络  $G$  中的节点  $v$ ,如果节点  $v$  的邻居节点中有  $t$  个节点的度数高于它,则称节点  $v$  为“ $t$  阶准峰值点”。

一般地,处于同一社区内的节点间联系相对紧密,其内部的同阶准峰值点之间更有可能相互连接,不大可能成为社区的关键节点;相对孤立的准峰值点更有可能成为社区的关键节点,如图 2 所示。

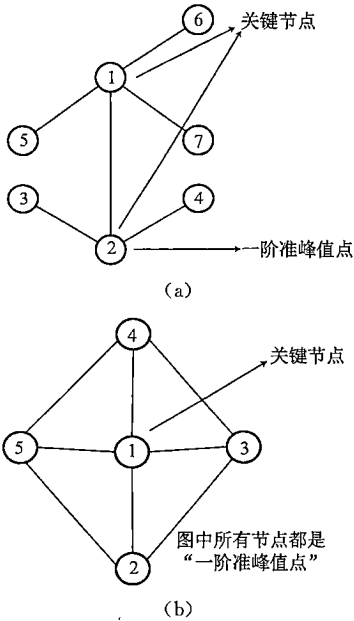


图 2 “准峰值点”为关键节点(a)和“准峰值点”不是关键节点(b)

据此给出关键节点的定义。

定义 2:对于网络  $G$  中的节点  $v$ ,如果节点  $v$  是“ $t$  阶准峰值点”,且它的邻居节点中没有其他“ $t$  阶准峰值点”,则称节点  $v$  为“ $t$  阶局部度数峰值点”。

算法从给定节点出发,不断检测邻居节点,直到发现第一个符合要求的多阶局部度数中心点,则以该节点为种子节点,按照既定方法扩张得到局部社区。若发现多个符合要求的节点,则任意选取一个阶数最低的作为种子节点。

给定一个阶数灵敏度系数  $T$ ,只有被发现的峰值点的阶数不大于  $T$  才称为找到种子节点。一般地,灵敏度系数越小,算法精确度越高,但稳定性越低。

3 实验分析

采用 LFR 基准网络数据集<sup>[16]</sup>作为研究对象。LFR 基准网络生成后节点的社区分布已知,故使用 F 值作为对给定节点社区划分效果评价指标。

3.1 峰值点阶数灵敏度系数  $T$  的选定

网络的规模越大,各节点的度数普遍越高,LFR 基准网络的混合参数  $\mu$  越大,对应现实中网络的社区越不明显,探究不同网络规模与不同混合参数对灵敏度系数的要求是很有必要的。

(1) 网络规模与灵敏度  $T$  的关系

固定 LFR 社区的混合参数为  $\mu=0.1$ ,生成社区规模为 1 000、2 000、3 000 个节点的网络。网络参数如表 1 所示。

表 1 网络规模与灵敏度  $T$  关系实验参数

节点总数(N)	平均度数(k)	度数最大值(k max)	混合参数( $\mu$ )	最小社区规模(min c)	最大社区规模(max c)
1 000	30	50	0.1	20	50
2 000	60	100		40	100
3 000	90	150		60	150

灵敏度系数由 0 开始逐渐增长,建立全局社区划分精确度-灵敏度系数关系,如图 3 所示。

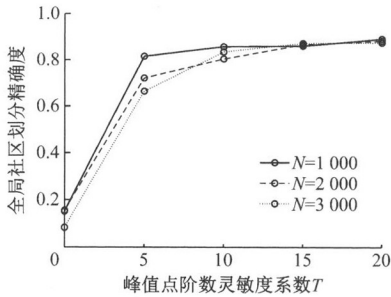


图 3 不同规模网络全局社区划分精确度与峰值点阶数灵敏度系数的关系

在网络规模相同的情况下,加入多阶局部度数峰值点较之只有局部度数中心点的情况,划分精度得到了明显改善,选取的灵敏度系数  $T$  越大,划分精度越高。

(2) 网络混合参数与灵敏度  $T$  的关系

将 LFR 社区的社区规模固定在 1 000 个节点,生成混合参数为  $\mu=0.1、0.2、\cdots、0.5$  的五个网络。网络具体参数如表 2 所示。

表 2 网络混合参数与灵敏度  $T$  关系实验参数

节点总数(N)	平均度数(k)	度数最大值(k max)	混合参数( $\mu$ )	最小社区规模(min c)	最大社区规模(max c)
1 000	30	50	0.1	20	50
			0.2		
			0.3		
			0.4		
			0.5		

对于不同的灵敏度建立全局社区划分精确度-灵敏度系数关系,如图 4 所示。

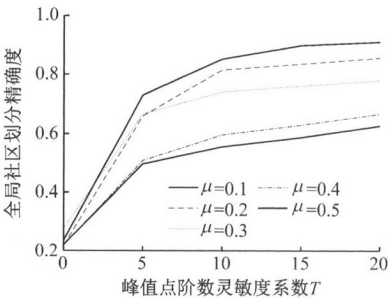


图 4 精确度与峰值点阶数灵敏度系数的关系

混合参数越大,社区结构越不明显的网络,其社区划分的精度越低。随着灵敏度系数的提高,不同混合参数的网络之间的差异并未被消除。网络混合参数相较于规模对社区划分质量影响更大。

3.2 算法分析

为说明本算法框架的优势,将其与局部度数中心点框架进行对比实验。实验网络具体参数如表 3 所示。

表 3 对比实验网络参数

节点总数	平均度数	度数最大值	混合参数	最小社区规模	最大社区规模
1 000	30	50	0.2	20	50

对网络中的每个节点运行算法,并将所得局部社区与真实社区进行对比,得到节点在指定算法下的  $F$  值。 $F$ -节点关系如图 5 所示。

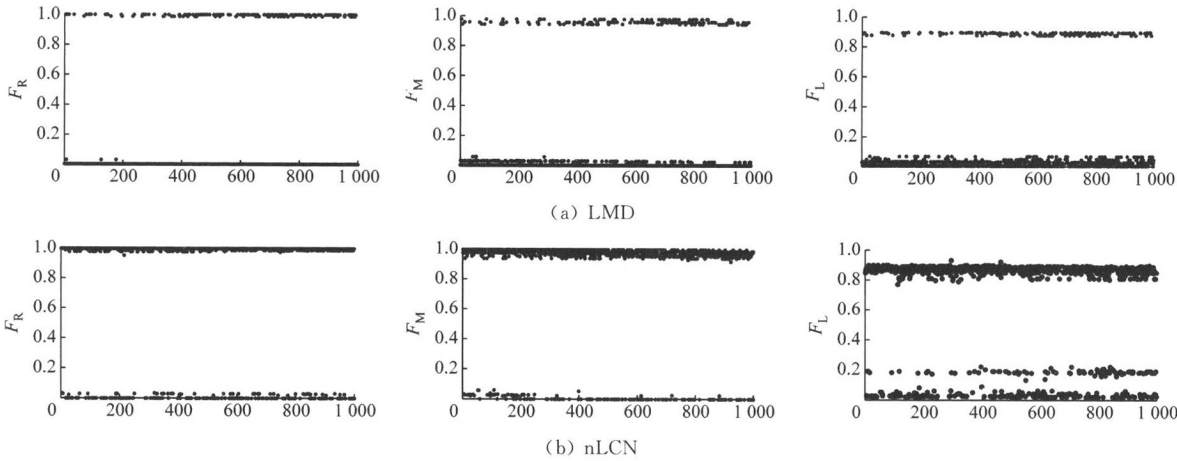


图 5 三种局部社区扩张算法基于局部度数中心点(a)和多阶局部度数峰值点(b)进行局部社区检测结果

三种扩张算法在两种框架下的 F 值对比如表 4 所示。

表 4 两种框架下各节点 F 值平均值对比

	R	M	L
LMD	0.159	0.156	0.154
nLCN	0.795	0.823	0.675

由表 4 可知,基于多阶峰值点选取种子节点的算法效果较之局部度数中心点有明显的提高。

4 总结

本文基于社区内关键节点度数较高和关键节点间联系松散对社区结构的认识,提出了多阶局部度数峰值点的概念及社区发现算法,精确度更高,稳定性更强。在一定范围内,选取的阶数灵敏度系数越大,算法精确度越高。然而,随着阶数的增大,其峰值特性逐渐弱化,带动算法精确度提高的边际效益递减,不稳定性增加。当阶数足够大时,算法退化为以既定节点为扩张初始节点的普通算法。

本文为解决局部社区发现种子节点选取问题,提供了寻找多阶局部度数峰值点的框架。以多阶局部度数峰值点为种子节点进行扩张能在一定程度上改善原局部扩张算法的性能。该算法无须得知全局网络的信息,仅通过初始给定节点便能自主探索网络,有着较低的计算复杂度,有助于复杂网络局部结构的研究。在同一网络中研究多个节点的社区归属时,本算法不限制各节点归属社区的数量,有助于探索复杂网络中的重叠社区。

参考文献

[1] 吴联仁,李瑾頔,齐佳音.一种基于分支过程的信息流行度动力学模型[J].物理学报,2019,68(7):296-301.

[2] 程学旗,沈华伟.复杂网络的社区结构[J].复杂系统与复杂性科学,2011,8(1):57-70.

[3] Karypis G, Kumar V. Multilevel  $k$ -way partitioning Scheme for irregular graphs[J]. Journal of Parallel and Distributed Computing, 1998, 48(1):96-129.

[4] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2):026113.

[5] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008.

[6] He Y S, Wang J, Xie J L, et al. Research on mining key nodes of complex web-based communities based

on mining algorithm[J]. International Journal of Web Based Communities, 2020, 16(2):202.

[7] 马慧芳,陈海波,赵卫中,等.融合标签平均划分距离和结构关系的微博用户可重叠社区发现[J].电子学报,2018,46(11):2612-2618.

[8] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435(7043): 814-818.

[9] Shang C X, Feng S Z, Zhao Z Y, et al. Efficiently detecting overlapping communities using seeding and semi-supervised learning. International[J]. International Journal of Machine Learning and Cybernetics, 2017, 8(2): 455-468.

[10] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3): 033015.

[11] 吴建,王梓权,易亿,等.基于图遍历的局部社区发现算法[J].计算机应用研究,2019,36(9): 2636-2638.

[12] 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 and Its Applications, 2013, 392(3): 529-537.

[13] Clauset A. Finding local community structure in networks[J]. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 2005, 72(2 Pt 2): 026132.

[14] Luo F, Wang J Z, Promislow E. Exploring local community structures in large networks[J]. Web Intelligence and Agent Systems: an International Journal, 2008, 6(4):387-400.

[15] Bagrow J P, Bollt E M. Local method for detecting communities[J]. Physical Review E, 2005, 72(4 Pt 2):046108.

[16] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms [J]. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 2008, 78(4 Pt 2):046110.

(收稿日期:2020.09.27)