

一种基于边界节点识别的复杂网络局部社区发现算法

刘 阳 季新生 刘彩霞

(国家数字交换系统工程技术研究中心 郑州 450002)

摘 要: 在网络日益巨大化和复杂化的背景下, 挖掘全局网络的社区结构代价较高。因此, 基于给定节点的局部社区发现对研究复杂网络社区结构有重要的应用意义。现有算法往往存在着稳定性和准确性不高, 预设阈值难以获取等问题。该文提出一种基于边界节点识别的复杂网络局部社区发现算法, 全面比较待合并节点的连接相似性进行节点聚类; 并通过边界节点识别控制局部社区的规模和范围, 从而获取给定节点所属社区的完整信息。在计算机生成网络和真实网络上的实验和分析证明, 该算法能够自主挖掘给定节点所属的局部社区结构, 有效地提升局部社区发现稳定性和准确率。

关键词: 复杂网络; 社区发现; 局部社区; 边界节点识别

中图分类号: TP311

文献标识码: A

文章编号: 1009-5896(2014)12-2809-07

DOI: 10.3724/SP.J.1146.2013.01955

Detecting Local Community Structure Based on the Identification of Boundary Nodes in Complex Networks

Liu Yang Ji Xin-sheng Liu Cai-xia

(National Digital Switching System Engineering & Technological R & D Center, Zhengzhou 450002, China)

Abstract: In the context that social network becomes more and more complicated and huge, it is extremely difficult and complex to mine the global community structures of large networks. Therefore, the local community detection has important application significance for studying and understanding the community structure of complex networks. The existing algorithms often have some defects, such as low accuracy and stability, the preset thresholds difficult to obtain, etc.. In this paper, a local community detecting algorithm is proposed based on boundary nodes identification, and a comprehensive consideration of the external and internal link similarity of neighborhood nodes for community clustering is given. Meanwhile, the method can effectively control the scale and scope of the local community based on the boundary node identification, so as the complete structure information of the local community is obtained. Through the experiments on both computer-generated and real-world networks, the results show that the proposed algorithm can automatically mine local community structure from the given node without predefined parameters, and improve the performance of local community detection in stability and accuracy.

Key words: Complex networks; Community detection; Local community; Identification of boundary nodes

1 引言

近年来复杂网络的研究蓬勃发展, 已成为人们深入了解和分析真实世界网络的重要理论和方法之一。其中, 社区结构作为复杂网络中广泛存在的重要属性, 是认识复杂网络的功能与特征的基础, 对研究网络的模块化、异质性等特征, 分析网络内部相互作用、网络演化都具有重要意义。而在许多实际应用中, 由于条件限制往往只能获取局部网络信息, 或者只需分析某特定节点、特定群体的社区结构, 这就从网络拓扑信息和复杂度等方面对社区发

现提出了限制和要求, 需要根据局部网络信息进行局部社区发现。

局部社区发现属于复杂网络社区发现领域的子课题, 一般定义为: 对于已知复杂网络 $G(V, E)$, V 为节点集, E 为边集, 给定任意节点 s , 挖掘节点 s 在网络 G 中所属的完整社区结构。局部社区发现致力于挖掘给定节点所属社区的局部结构和聚集特性, 与全局网络社区发现的区别主要有两点: (1) 局部社区发现结果反映的是局部网络背景下的社区结构, 而不是网络的整体社区结构, 因而也不要求网络的完整拓扑信息; (2) 局部社区发现算法的复杂度一般较低, 但由于局部网络信息不全, 难以确定社区边界, 导致易引入异质节点, 需要设定有效的聚类停止规则。在网络信息海量、复杂化背景下, 利用局部网络信息快速准确地发现给定节点所属的

2013-12-17 收到, 2014-04-24 改回

国家 863 计划项目 (2011AA010604) 和国家重大科技专项 (2012ZX03006002) 资助课题

*通信作者: 刘阳 liuyang198610@163.com

局部社区具有重要的应用意义。

近年来,复杂网络中局部社区发现问题的研究开始引起人们的关注。文献[1]提出了一种局部社区结构评价指标,以最大化该指标作为聚类规则选取给定节点的邻近节点构建局部社区。文献[2]定义了类似于“弱社区”概念的局部社区模块度,以极大化局部社区模块度为目标,迭代合并满足聚类条件的邻接节点以实现局部社区发现。文献[3]认为局部中心节点(local central node)作为局部网络的核心节点,对局部社区影响巨大,因而围绕局部中心节点而非给定节点来扩展局部社区更具有实际意义,能够改善局部社区发现的精确度。文献[4]定义了外部节点与局部社区间的连接相似度(link similarity)指标,并通过聚类、优化、剪枝三重步骤循环迭代;直至不能通过聚类外部节点使局部模块度 M 增大时则停止迭代,以期发现最优的局部社区。文献[5]从构建局部边社区的角度出发,通过排序选择位于高密度区域的种子边,最大化边适应度函数以膨胀种子边得到局部边社区,并将该方法扩展至整个网络中从而获取所有的边社区结构。文献[6]基于树图中父节点与子节点间连接密度降幅越大,子节点越有可能形成独立社区的思想,构建了网络树形图,根据最大流最小割定理自适应地围绕给定节点进行网络切割,从而实现局部社区发现。上述方法通过构建局部社区结构评价指标,在局部社区由给定节点向外部扩展的过程中,仅根据待合并的邻居节点与局部社区的内部度(internal degree)带来的局部社区评价指标增益,从而判断节点社区归属,没能同时兼顾邻居节点与外部邻接子图的连接状况。这就使得局部社区发现算法受给定节点位置和评价指标的影响较大,稳定性和准确性难以得到保障。

基于此,本文提出了一种基于边界节点识别的局部社区发现算法,主要贡献如下:(1)在局部社区聚合的过程中,综合考虑并比较待合并节点与局部社区及其外部节点的连接相似性,避免直接从给定节点进行外部扩张而引入异质节点;(2)通过边界节点识别以控制社区聚类,从而确定局部社区的范围和大小,无需设置人工参数进行干预。通过计算机生成网络和真实网络实验发现,本文算法能够基于给定节点自主有效地挖掘并识别其真实边界,从而发现接近真实完整的局部社区结构,具有相对较高的稳定性和准确率。

本文第 2 节分析现有局部社区发现方法的问题与局限;第 3 节介绍基于边界节点识别的局部社区发现算法;第 4 节介绍了局部社区发现的评价准则与方法,并基于计算机生成网络和真实网络给出实验和分析结果;第 5 节对全文进行总结。

2 问题分析

现有局部社区发现方法将社区看作是网络中具有某种相似特性的节点集合,根据社区内部节点联系紧密而社区间联系相对稀疏的特点,通过构建类似于 N-G 模块度^[7]的局部社区结构评价指标,通过计算节点聚类带来的指标增益判断其社区归属,从而逐步扩展地完成局部社区发现。目前比较典型的局部社区评价指标有 R 值, M 值, f 值, $\Delta K'_j$ 值等。这些算法计算复杂度较低,适用于社区结构明显且规则的局部社区,但往往也存在如下问题:(1)只考虑邻接节点与局部社区的连接状况,而没有考虑该节点的外部连接状态,可能导致社区归属误判而引入异质节点。(2)社区发现结果对于给定节点位置敏感。基于同一社区的不同给定节点,发现的局部社区往往差异较大。(3)难以确定聚类停止规则。即使通过设定社区结构评价指标阈值来控制聚类过程,但始终无法有效地确定局部社区的真正边界。

如图 1 所示,在局部社区“由内而外”的扩展过程中,沿用了全局网络社区发现中最大化模块度的思想,将满足或者能够改善社区内聚指标的节点加入到该社区中,使得局部社区倾向于优先引入大度数节点。而大度数节点的引入往往又会引起连锁反应,改变了局部社区扩展的方向和范围,使其邻近节点容易被划分到给定节点所属的局部社区,引入原本不属于该局部社区的异质节点,影响了局部社区发现的性能。

基于上述分析,本文提出了一种基于局部社区聚合和边界节点识别的两阶段局部社区发现算法,一方面通过全面分析局部社区邻接子图范围内连接相似性,从而获取给定节点所在社区的局部结构特征并进行节点聚类;同时在社区聚类过程中,基于邻接节点内外的连接状况反复识别其边界节点,从而控制社区聚类的规模和范围以完成局部社区发现。该算法采用了有别于最大化局部社区结构指标的思路和方法,以节点相似性模块度作为局部社区发现的衡量标准,通过边界节点识别解决已有方法存在的“何时停止聚类迭代”和给定节点位置敏感等问题。

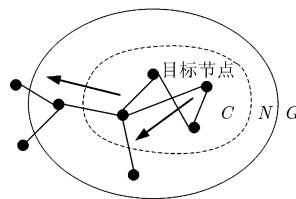


图 1 “由内而外”的局部社区发现模型

3 基于边界节点识别的局部社区发现算法

3.1 基本定义

定义 1 (邻接子图) 在复杂网络 $G(V, E)$ 中, 邻接子图是指与某节点(或某节点集、某社区)有直接连接的所有邻接节点共同组成节点集合。节点 i 的邻接子图可表示为 $N(i) = \{v \in V \mid \exists e_{iv} \in E\}$, V 为节点集, E 为边集。

在社区发现过程中, 将整个局部社区 C 视为一个聚合后的普通节点。如图 1 所示, $N(C)$ 表示局部社区 C 的邻接子图。 $U = V - N(C)$ 表示网络中局部社区 C 和邻接子图 $N(C)$ 之外的部分。

定义 2 (连接相似度) 在网络 $G(V, E)$ 中, 定义任意两邻接节点 i 与 j 的连接相似度为

$$L(i, j) = \text{Sim}_{ij}^{\cos} = \frac{\sum_{k \in N(i) \cap N(j)} \omega(i, k) \cdot \omega(j, k)}{\sqrt{\sum_{k \in N(i)} \omega^2(i, k)} \cdot \sqrt{\sum_{k \in N(j)} \omega^2(i, k)}} \quad (1)$$

其中, $\omega(i, k)$ 表示连接节点 i, k 的边 e_{ik} 的权重。若 G 为非权重网络, 则认为 $\omega(i, k) = 1$ 。

连接相似度借鉴了余弦相似性^[8]作为衡量节点相似性的标准, 表示两节点的共有连接在两节点的所有直接连接中权重越大, 说明二者联系越紧密, 节点连接的相似性越强, 越有可能属于同一社区。

定义 3 (边界节点) 对于节点 $j \notin C$, 如果该节点与局部社区 C 存在连接, 且满足条件 $L(j, C) < L(j, k), k \notin C$, 则定义节点 j 为局部社区 C 的边界节点。局部社区 C 所有的边界节点构成了其边界集合 $B(C)$ 。

如图 2 所示, 由于网络结构具有社区内部连接紧密而社区间连接稀疏, 呈现出“内紧外松”的结构特征, 使得距离社区核心较远的边缘节点构成了所属社区的天然边界。在局部社区发现过程中, 随着社区聚类的扩展, 必然逐步到达社区的边界。边界节点大多处于多个网络社区的交界处, 并往往同时与多个社区建立了连接。因此, 如何基于节点及其邻接子图的连接关系与拓扑结构判断其所属社区, 是局部社区发现的一个重要难点。而通过边界节点识别, 能够帮助确定局部社区的边界, 从而有效地精确控制局部社区发现的范围和规模, 确保社区发现的稳定性和有效性, 使得从同一社区不同节点出发所获得局部社区趋于一致。

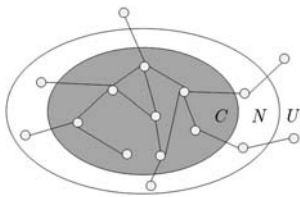


图 2 网络组成结构示意图

3.2 算法描述

本文提出了基于边界识别的局部社区发现算法, 主要包含两个环节: 社区节点聚类阶段及社区边界节点识别阶段。该算法基于网络节点的局部结构信息来衡量节点间的连接相似度, 从给定节点开始进行节点聚类; 并基于节点相似性模块度判别边界节点, 以优化局部社区发现。通过两阶段的反复迭代, 最终得到具有稳定边界的局部社区。

在社区节点聚类阶段, 将局部社区 C 视作一个普通节点。对于局部社区邻接子图 $N(C)$ 范围内任一节点 i , 遍历计算该节点与其邻接节点的连接相似度。若 $L(i, C)$ 为节点 i 在其邻接子图范围内的最大值, 则认为该节点在邻接子图范围内与局部社区具有最大的局部相似性, 将该节点 C 并入局部社区 C (如图 3(a)所示); 若 $L(i, C)$ 不满足该条件, 则将节点 i 暂时标记为边界节点, 并等待进一步的边界识别(如图 3(b)所示)。

在社区边界节点识别阶段, 遍历边界节点集 $B(C)$ 中的所有节点。对于任一节点 $i \in B(C)$, 若 p 是与 i 有最大的连接相似度的邻接节点, 将节点 p, i 视为局部社区 C 的一个邻居子图 M 。这里引入一种基于节点相似性的模块度函数 Q_S ^[9]:

$$Q_S = \sum_{i=1}^k \left[\frac{\text{IS}_i}{\text{TS}} - \left(\frac{\text{DS}_i}{\text{TS}} \right)^2 \right] \quad (2)$$

其中, k 表示网络 G 中存在的社区个数, $\text{IS}_i = \sum_{i, j \in C_i} L(i, j)$ 表示社区 C_i 内部节点的相似性累积之和, $\text{DS}_i = \sum_{i \in C_i, j \notin C_i} L(i, j)$ 表示社区 C_i 内部节点与外部节点的相似性累积之和, $\text{TS} = \sum_{i, j \in V} L(i, j)$ 表示网络 G 中所有节点的相似性累积之和。 Q_S 是对 N-G 模块度^[7]的有效扩展, 能够基于节点的连接相似性衡量网络社区划分的质量。因此, 可以根据边界节点聚类对 Q_S 的影响来判断其局部社区的归属。计算将网络子图 M 并入 C 后获得的模块度增益 ΔQ_S :

$$\Delta Q_S = Q_S^{C \cup M} - Q_S^C - Q_S^M = \frac{2\text{IS}_{CM}}{\text{TS}} - \frac{2\text{DS}_C \cdot \text{DS}_M}{\text{TS}^2} \quad (3)$$

其中, $\text{IS}_{CM} = \sum_{i \in C, j \in M} L(i, j)$ 表示模块 C, M 间节点相似性的累积之和。若 $\Delta Q_S > 0$, 则将该子图并入局部社区; 反之, 则确认节点 i 为局部社区 C 的边界节点。需要指出的是, 节点相似性模块度有助于准确区分聚类过程中的异质节点。根据 ΔQ_S 增益来判定局部社区边界, 不是为了获取具有最大 Q_S 的局部社区, 而是为了优化局部社区的发现结果。

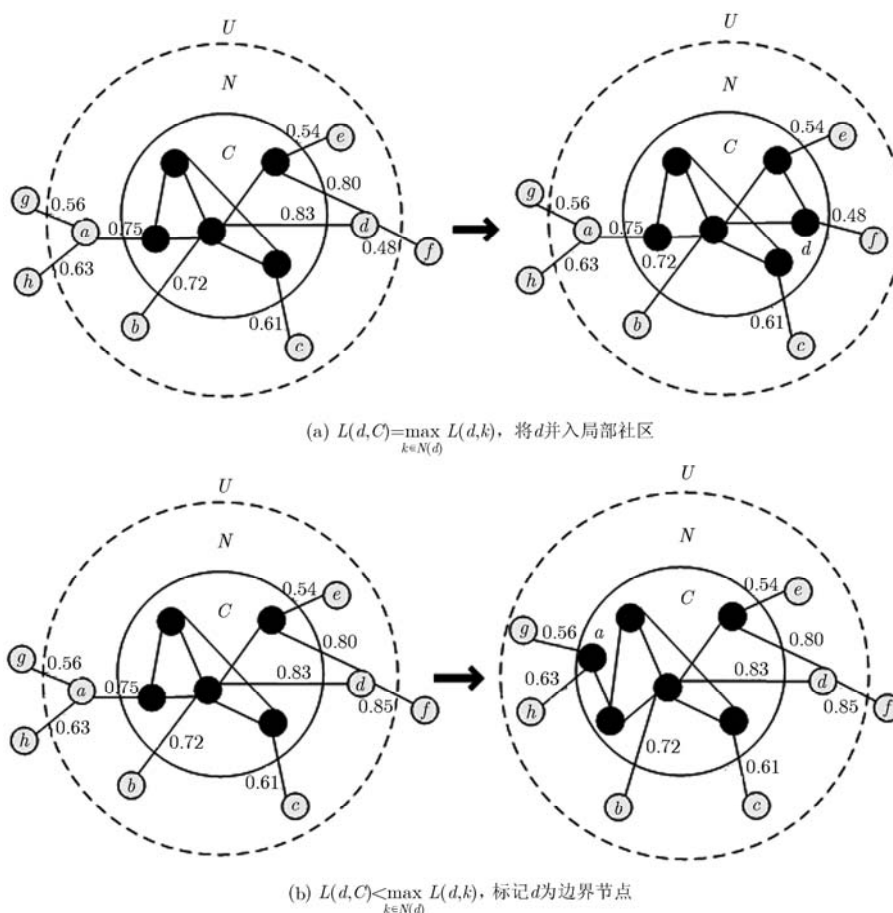


图 3 局部社区节点聚类示意图

当局部社区 C 的节点构成发生变化时, 需要及时更新邻接子图 $N(C)$ 和边界节点集 $B(C)$, 重复迭代上述两阶段的局部社区发现过程。直至局部社区的邻接子图 $N(C)$ 所有节点都被确认为边界节点, 没有邻接节点满足局部社区扩展聚类的条件, 则认为局部社区 C 发现完成。基于以上描述, 设计局部社区发现算法的具体步骤如表 1 所示。

4 实验

4.1 评价准则与方法

局部社区发现的评价准则相比全局社区发现, 除了评价社区发现精确度, 还需要评价节点划分的正确性。因而选择的评价准则也必须能够从这两方面来考察该模型的性能。本文实验部分采用了复杂网络局部社区发现领域常用的典型指标: precision^[10](正确率), recall^[10](查全率)及调和指标 F-measure 评价局部社区发现算法的性能。

若已知网络真实社区结构的标准数据集 F_i , 将局部社区发现的结果 X 映射到该标准数据集, 则 precision 表示局部社区中被正确划分的节点数在整

个局部社区所占的比例, recall 表示局部社区中被正确划分的节点数在标准数据集所占的比例, F-measure 是前两种指标的加权调和指标, 下文三者分别用 p , r 和 F 指代。

$$p = \frac{|X \cap F_i|}{|X|} \quad (4)$$

$$r = \frac{|X \cap F_i|}{|F_i|} \quad (5)$$

$$F = 2 \times \frac{p \times r}{p + r} \quad (6)$$

由于计算机生成网络和部分真实网络社区结构已知, 其给定节点所属社区的标准数据集易于获取; 对于未知社区结构的真实网络, 本文采用当前公认准确度较高的 Louvain^[11]社区发现算法挖掘给定节点的所属社区结构作为标准数据集。为了分析基于边界节点识别的局部社区发现算法的性能, 实验同时基于计算机生成网络和真实网络进行局部社区发现, 并选取文献[1], 文献[3]和文献[12]3种常见的典型算法与之进行性能对比。

表 1 局部社区发现算法

基于边界节点识别的局部社区发现算法	
输入:	$G(V, E)$, 给定节点 s
输出:	节点 s 所属的局部社区 C
(1)	Initialize $N(C)$, U , $B = \emptyset$;
(2)	while true do
(3)	//Phase 1: 局部社区聚合阶段
(4)	for each $i \in N(C)$ do
(5)	Calculate $L(i, C)$ by Eq. 1;
(6)	if $L(i, C) > \max_{k \in N(i)} L(i, k)$
(7)	$C \leftarrow C \cup \{i\}$; //merge i into C
(8)	Update $N(C)$, U , $B(C)$;
(9)	else
(10)	$B(C) \leftarrow B(C) \cup \{i\}$; //set i to be a boundary node
(11)	end if
(12)	end for
(13)	//Phase 2: 局部社区边界识别阶段
(14)	for $i \in B$ do
(15)	$p \leftarrow \arg \max_{k \in N(i)} L(i, k)$; // $p \notin C \wedge p \in N(i)$
(16)	$M \leftarrow \{p\} \cup \{i\}$; //merge i , p into a module node
(17)	$\Delta Q_S(C) \leftarrow Q_S(C \cup M) - Q_S(C)$;
(18)	if $\Delta Q_S(C) > 0$
(19)	$C \leftarrow C \cup M$; //merge M into C
(20)	update $N(C)$, U ;
(21)	end if
(22)	end for
(23)	end while
(24)	return C

4.2 计算机生成网络

计算机生成网络采用 LFR(Lancichinetti Fortunato Radicchi)程序^[13]生成的基准网络数据集。LFR 在社区发现测试方面应用广泛,且生成网络的节点数、平均节点度、 μ 等相关参数都可以灵活设置,能够生成较高质量的网络测试数据。其中,参数 μ 表示网络中社区间的连接数占所有边总和的

比例, μ 越大,计算机生成网络的社区结构越不明显,社区发现也就越困难。本文生成了一组 μ 间隔为 0.1,编号为 Network1 到 Network9 的 LFR 网络测试数据集,具体配置如表 2 所示。

从上述 9 个计算机生成网络中分别随机选择 100 个给定节点进行局部社区发现,与给定节点所属的真实社区进行对比,计算 3 种度量指标的均值。其中, r 比 p 大,说明社区发现的规模比标准社区小;反之,说明社区发现的规模比标准社区大。基于给定节点扩展局部社区的过程中,局部社区越大,越有可能完整覆盖实际社区, r 值越高;但同时引入的异质节点可能性越大,导致 p 值下降。好的局部社区发现算法必须同时具有较高的 p 值和 r 值,使得综合指标 F 值也处于较高水平。

如图 4 所示,局部社区发现受 LFR 网络的社区结构强度(μ 参数)影响较大,随着 μ 的递增,4 种算法在 3 种评价指标上都呈现出逐步下降的趋势。当 $\mu \leq 0.5$ 时,4 种算法在 3 种指标上都显示局部社区发现的准确率较高;当 $\mu > 0.5$ 时,4 种算法在 3 种指标上都出现了不同程度的下降。这是因为 LFR 网络社区结构较弱时,导致局部社区发现过程中引入异质节点的可能性增加,局部社区发现算法准确率也随之下降。

分析局部社区发现的整体性能发现,4 种算法都呈现出 r 值较高, p 值相对较低的趋势,且 p 、 F 值下降相对平缓,而 r 值间的下降较快。本文算法由于通过边界节点识别,所发现局部社区的位置和规模较为准确,因此 p 值在 4 种算法表现最好,均值为 72.07%;而在 r 指标的均值都维持在 77.86%,性能与文献[3]算法较为接近。综合起来,使得 F 均值也维持在 74.79%。对比 4 种算法的 3 项性能指标发现:基于边界节点识别的局部社区发现算法在 p 和 r 表现比较均衡且稳定,使得 F 指标值最高,说明其社区发现结果与标准社区结构更为匹配。

表 2 LFR 网络配置参数

LFR datasets	节点数	平均边数	平均节点度	最大节点度	社区规模	μ
Network 1-9	10000	74876	8	20	20-120	0.1-0.9

4.3 真实网络

真实网络相比计算机生成网络更不规则,因而其社区结构往往也更加复杂。这里采用 5 个具有不同规模的典型真实网络数据集:2000 赛季大学生美式足球联赛网络(Football)^[14],2004 年美国政治博客网络(Polblogs)^[15]、美国航空网络(US Airport)^[16]、美

国电力网络(US power grid)^[17]、2005 年 Internet 拓扑网络(Internet)^[18]来进一步测试算法性能,表 3 描述了这些真实网络的基本属性。

从上述 5 个真实网络中分别随机选择 100 个给定节点进行局部社区发现,与 Louvain 算法发现的给定节点所属社区结构进行对比,得到其 r 、 p 、 F

的均值。以美式足球联赛网和美国政治博客网两个真实网络为例，图 5 给出了各算法的各项性能表现，综合比较发现：文献[1]和文献[12]算法同时在 p 值上较低，而 r 值较高，使得两项指标相差悬殊；文献[3]在 r 和 p 上较为均衡，但 r 值对比其它算法性能较差。本文算法同时具有较高的 r 值和 p 值，使得 F 值也维持在较高水平，证明了其局部社区发现的准确性和有效性。

表 3 实验采用的真实网络的基本属性

网络数据集	节点数	边数	连接类型	网络结构
Football ^[14]	115	616	无向	已知
Poldlogs ^[15]	1490	19025	有向	已知
US Airport ^[16]	1574	28236	有向	未知
US Power grid ^[17]	4941	6594	无向	未知
Internet ^[18]	34761	171403	无向	未知

在真实网络中，4 种局部社区发现算法的调和指标如表 4 所示。分析可知：由于真实网络的社区结构彼此存在差异，给定节点的选取也具有一定的随机性，因而 4 种算法的调和评价指标在不同真实网络中也各不相同。对于 Football, US power grid, 本文算法对局部社区发现性能改善有限；对于 Polblog, US Airport, Internet 网络，本文算法的社

区发现性能优势更加明显。相比真实网络条件下的其它 3 种算法，基于边界节点识别的局部社区发现算法都表现出了相对较高的社区发现精确度；在给定节点随机选取的条件下，本文算法的综合调和指标均值达到了 86.07%，体现了较强的抗干扰能力，说明本文算法稳定性更优，并且与标准社区结构的匹配效果较好。

综合计算机生成网络和真实网络的实验可以得出结论：在不同网络环境、不同给定节点条件下，基于边界节点识别的局部社区发现算法在正确率、查全率方面相比其它算法都有不同程度的改善。尽管并不能在正确率、查全率两项指标上同时达到最优，但在这两项评价指标上较为均衡且处于较高水平，使其性能具有一定的整体优势。该结论在表 4 所示的综合调和指标上也得到了充分验证。

表 4 调和指标均值：4 种算法在真实网络中基于随机给定节点的局部社区发现性能

网络数据集	文献[1]	文献[12]	文献[3]	本文算法
Football ^[14]	0.6267	0.7623	0.7633	0.7815
Poldlogs ^[15]	0.5433	0.8111	0.8025	0.8603
US Airport ^[16]	0.4950	0.7833	0.8257	0.9145
US Power grid ^[17]	0.6882	0.7452	0.7584	0.8160
Internet ^[18]	0.6319	0.8573	0.8726	0.9314

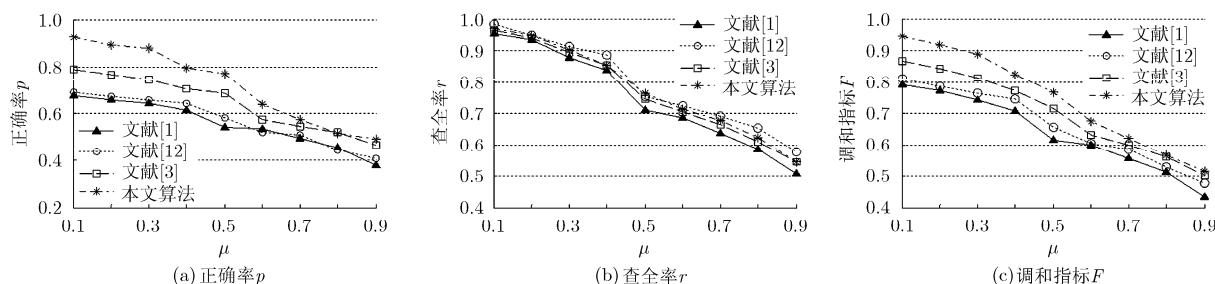


图 4 LFR 网络下 4 种算法正确率，查全率，调和指标的性能对比

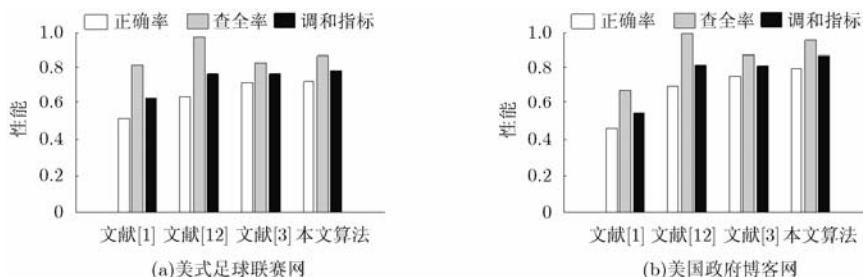


图 5 4 种算法基于真实网络的正确率、查全率、调和指标的性能对比

5 结束语

本文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，采用有别于最大化局部社区结构指标的思路和方法，通过扩展余弦相似性作为局部社区发现的衡量标准，以边界节点识别控制社区聚类的规模和范围以完成局部社区发现，从而解决已有方法稳定性不高，预设阈值难以获得等问题。通过实验和分析证明：对于不同的给定节点，基于边界节点识别的局部社区发现算法具有较高的准确率和稳定性，能够较真实地反映给定节点所属的局部社区结构。在保证准确度和稳定性的前提下，如何将局部社区发现进一步应用于大规模网络局部结构分析，以及动态时序网络中局部社区结构的演化分析等实际问题，是我们下一步研究工作的重点。

参 考 文 献

- [1] Clauset A. Finding local community structure in networks[J]. *Physical Review E*, 2005, 72(2): 026132..
 - [2] Radicchi F, Castellano C, Cecconi F, *et al.*. Defining and identifying communities in networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(9): 2658-2663.
 - [3] Chen Q, Wu T T, and Fang M. Detecting local community structures in complex networks based on local degree central nodes[J]. *Physica A*, 2013, 392(3): 529-537.
 - [4] Wu Y J, Huang H, Hao Z F, *et al.*. Local community detection using link similarity[J]. *Journal of Computer Science and Technology*, 2012, 27(6): 1261-1268.
 - [5] 潘磊, 金杰, 王崇骏, 等. 社会网络中基于局部信息的边社区挖掘[J]. *电子学报*, 2012, 40(11): 2255-2263.
Pan Lei, Jin Jie, Wang Chong-jun, *et al.*. Detecting link communities based on local information in social networks[J]. *Acta Electronica Sinica*, 2012, 40(11): 2255-2263.
 - [6] Qi X, Tang W, Wu Y, *et al.*. Optimal local community detection in social networks based on density drop of subgraphs[J]. *Pattern Recognition Letters*, 2014, 36(1): 46-53.
 - [7] Newman M E J. Modularity and community structure in networks[J]. *Proceedings of National Academy of Sciences of the United States of America*, 2006, 103(23): 8577-8582.
 - [8] Fouss F, Pirotte A, Renders J M, *et al.*. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(3): 355-369.
 - [9] Feng Z, Xu X, Yuruk N, *et al.*. A Novel Similarity-based Modularity Function for Graph Partitioning[M]. Berlin Heidelberg Springer, Data Warehousing and Knowledge Discovery, 2007: 385-396.
 - [10] Zhang Aidong. Protein Interaction Networks[M]. New York: Cambridge University Press, 2009: 44-47.
 - [11] 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, doi: 10.1088/1742-5468/2008/10/P10008.
 - [12] Luo F, Wang J Z, and Promislow E. Exploring local community structures in large networks[J]. *Web Intelligence and Agent Systems*, 2008, 6(4): 387-400.
 - [13] Lancichinetti A, Fortunato S, and Radicchi F. Benchmark graphs for testing community detection algorithms[J]. *Physical Review E*, 2008, 78(4): 046110.
 - [14] Girvan M and Newman M. Community structure in social and biological networks[J]. *Proceedings of National Academy of Science*, 2002, 9(12): 7821-7826.
 - [15] Adamic L A and Glance N. The political blogosphere and the 2004 US election: divided they blog[C]. Proceedings of the 3rd International Workshop on the Blogging Ecosystem, New York, USA, ACM, 2005: 36-43.
 - [16] US power grid network dataset-KONECT[OL]. <http://konect.uni-koblenz.de/networks/opsahl-powergrid>, November 2013.
 - [17] US airports network dataset-KONECT[OL]. <http://konect.uni-koblenz.de/networks/opsahl-usairport>, November 2013.
 - [18] Zhang Beichuan, Liu Raymond, Massey Daniel, *et al.*. Collecting the Internet AS-level topology[J]. *SIGCOMM Computer Communication Review*, 2005, 35(1): 53-61.
- 刘 阳：男，1986 年生，博士生，研究方向为社会网络分析、数据挖掘。
- 季新生：男，1969 年生，教授，博士生导师，主要研究领域为电信网安全、移动通信安全。
- 刘彩霞：女，1974 年生，副教授，硕士生导师，主要研究领域为移动通信安全、社会网络分析。