

基于势能背景信息的社团标签探测算法

宋砚秋 李桂君 李慧嘉

(中央财经大学管理科学与工程学院 北京 100081)

摘要 近年来,社团结构分析已经引起很多领域的关注,一些探测方法也陆续被提出。然而,其中大多数方法只利用了网络拓扑结构,并没有考虑内在的背景信息。基于离散势能理论,提出了一种新的半监督社团探测方法,利用标记节点产生的静电场来确定未标记节点的标签(社团标号)。首先给一定数目的节点赋予用户定义的标签;然后利用稀疏线性方程组计算余下节点的标签,其中每个节点的标签被设定为计算出的最大势能值;最后将该方法与现有算法进行比较。实验结果表明,所提算法在现实世界网络和人工基准网络上都展现了很强的探测能力,特别是在只具有模糊大规模社团结构的情况下,该算法仍然具有很高的准确性。

关键词 复杂网络,社团探测,背景信息,势能理论,标签探测

中图分类号 TP393 文献标识码 A

Community Label Detection Algorithm Based on Potential Background Information

SONG Yan-qi LI Gui-jun LI Hui-jia

(School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China)

Abstract In recent years, community structure analysis has attracted much attention in many fields, which aims to partition nodes in a graph into several clusters, in order to achieve a satisfactory state in which each cluster has a densely connected intra-cluster structure and homogeneous attribute value. Existing methods mainly assume that nodes in graphs are cooperative to optimize a given objective function, but ignore their background information in real-life contexts. Based on potential theory, this paper proposed a new semi-supervised community detection algorithm, which uses the electrostatic field generated by the tag node to determine the label of unlabeled nodes (community label). This paper firstly gave a certain number of nodes to the user-defined label, and then used the sparse linear equations to calculate the label of the remaining nodes, where each node's label was set to calculate the maximum potential value. By comparing with the existing algorithms, it is showed that the proposed algorithm has a strong detection ability in terms of the real world network and artificial benchmark network. It is also very accurate even through in the case of fuzzy large-scale community structure.

Keywords Complex networks, Community detection, Background information, Potential theory, Labels detection

1 引言

复杂网络作为新兴的跨学科研究方向,是现代科学中最活跃的研究领域之一,对于人们理解现实世界系统具有非常重要的作用^[1]。复杂网络作为复杂系统的拓扑结构,在很多情况下具有社团结构这一特征,即同一个社团内的点之间存在较多的关联边,而在不同的社团之间只有少量的关联边^[2]。社团是网络中具有特定功能的独立单元,同一社团内的个体具有较强的结构相似性和功能相似性,对于分析现实世界各个系统具有很高的价值^[3]。因此,社团探测技术在计算机科学、社会学和生命科学等传统网络分析学科中具有非常重要的地位。

到目前为止,为了更好地发现网络社团结构,科学家提出了很多算法,比如中心度算法^[2,4]、模块度优化算法^[5-8]、自旋模型^[9]和随机矩阵模型^[10]。但是其中大多数算法属于无监

督方法的范畴,它们仅仅使用了网络的拓扑信息而忽略了内在的背景信息,而这很大程度是由于缺少网络的先验信息导致其并不适合于实际应用。事实上,先验背景信息在挖掘社团结构时是非常重要的,比如在蛋白质交互网络(Protein-protein interaction network)中我们已知一些蛋白质属于特定的功能类型^[11]。因此,将先验背景信息应用于社团结构挖掘是一个值得研究的新兴领域。

最近,一些相关的社团探测算法逐渐被提出。Ma等^[12]提出了一种基于对称非负矩阵分解的半监督方法,将包含了成对约束(通过肯定相连的边来肯定不相连的边)的边的信息应用到网络社团识别中。Eaton等^[13]提出了一种基于自旋模型的半监督算法,它将个体自旋标号(已知小部分节点的社团分配)和成对约束等先验知识作为提取社团结构过程中的重要信息。Zhang^[14]提出了一种方法,它通过修改网络的邻接矩阵来隐式地编码约束,从而可看成是社团归属矩阵的去噪

本文受国家自然科学基金项目(71473285,71401194),中央高校基本科研业务费专项资金中央财经大学科研创新团队支持计划:科技金融协同创新模式与机制设计研究资助。

宋砚秋(1980—),女,副教授,主要研究方向为数据挖掘、社交网络、创新管理;李桂君(1973—),男,教授,主要研究方向为数据挖掘、城市可持续发展,E-mail:ligj@cufe.edu.cn(通信作者);李慧嘉(1985—),男,副教授,主要研究方向为数据挖掘、应急管理。

过程。总体来说,虽然目前已经有一些社团探测算法被提出,但其划分性能和适用性仍有很大的提升空间。

Biggs 于 1997 年提出了网络中的离散势能理论,其已经被广泛运用到不同的学科领域中。2006 年,Grady 等^[15]将离散势能理论用于影像分割,其在精确度和强健度上显示了非常高的性能。2013 年,Zhang 等^[16]将势能理论应用于有向网络,并提出了一种链路预测算法,它可以在局域有向网络上探测缺失关联并进行关系推荐。Wang 等^[17]提出了一种基于电荷模型的半监督方法,即先将一些数据标记为节点上的电荷,然后模拟静电场,利用已知电荷和网络关联,将其余未标记节点的电荷计算出来。目前基于电荷模型的半监督方法主要被用来进行影像分割和文本数据分类等,还没有被应用于网络社团探测问题。

本文提出了一种新的基于离散势能理论的半监督社团检测方法。为了实现社团探测,本文利用标记节点产生的静电场来确定未标记节点的标签(社团标号)。首先,给一定数目的节点赋予用户定义的标签;然后利用稀疏线性方程组计算余下节点的标签,其中每个节点的标签被设定为计算出的最大势能值。最后将其与现有算法(如 Girvan-Newman 算法、Markov 聚类算法和 Infomap 算法)做比较。实验结果表明,本文算法在现实世界网络和人工基准网络上都展现了很强的探测能力,即使是在只具有模糊社团结构的大规模网络中,该算法仍然具有很高的准确性。

2 半监督社团探测算法

2.1 问题描述和相关概念

本文首先给出相关的概念和定义。对于无权无向网络 $G=(V,E)$,其中 $V=\{v_1, v_2, \dots, v_n\}$ 表示节点的集合, $E \subseteq V \times V$, E 表示边的集合。本文不考虑多重边(加权边)和自环边。通常,一个网络可以由它的邻近矩阵 $A=(A_{ij})$ 来表示;如果节点 v_i 和 v_j 之间有边,那么邻接矩阵中 A_{ij} 的值等于 1,否则等于 0。另外,矩阵 A 对应的度矩阵 D 为对角矩阵,对角线上的元素为节点的度 $d(v_i)$ ($i=1, 2, \dots, n$),它的具体形式如式(1)所示:

$$D = \begin{pmatrix} d(v_1) & 0 & \cdots & 0 \\ 0 & d(v_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d(v_n) \end{pmatrix} \quad (1)$$

另一种重要的矩阵为拉普拉斯矩阵 L ,它的表达式为:

$$L = D - A \quad (2)$$

可以看出,拉普拉斯矩阵的每行元素之和都等于 0。假设社团数目为 k , $K=\{1, 2, \dots, k\}$ 。为了确保结果的唯一性,在每个社团中至少要有一个已经拥有标号的节点。对于节点 $V=\{v_1, v_2, \dots, v_n\}$,本文将节点分成两组: $V_L = \{(v_1, y_1), (v_2, y_2), \dots, (v_i, y_i)\}$ 为已标记的节点集合,其中 $y_i \in K$ 为节点 v_i 的标号; $V_U = \{v_{i+1}, \dots, v_n\}$ 为未标记节点集合, $V_L \cap V_U = \emptyset$ 。另外,本文重新排列了节点的顺序,使得已标记的节点排在前面,未标记的节点排在后面。

本文的主要任务是确定未标记节点的标号,这等价于解决这样一个问题:考虑一个随机游走者(Random walker),它从某个节点开始,最有可能(可能性最大)先到达 k 类节点中的哪一类。该游走者先到哪一类,该节点就属于那一类。而这种可能性与静电场中未标记节点的势能值完全一致。

2.2 探测算法

基于网络随机游走和离散势能理论之间的关联,本文可以计算出未标记节点 v_i 的势能 x_i^s ,而该未标记节点 v_i 位于由已标记节点生成的静电场中。势能可以通过解决电路理论问题得到,而电路理论问题和组合 Dirichlet 问题^[18]一致。首先为一些已标记的节点赋上势能的值,然后计算每个未标记的节点处的势能,即一个随机游走者从未标记的节点开始到最先到达具有标签 s 的那个节点。通过线性稀疏对称方程组的计算结果,这些势能也许会被计算出来。本文先回顾组合 Dirichlet 问题,并展示如何在序列中找到答案。

对于区域 Ω 和场 u ,我们可以定义 Dirichlet 积分为:

$$D[u] = \frac{1}{2} \int_{\Omega} |\nabla u|^2 d\Omega \quad (3)$$

Dirichlet 积分在热能转移、随机过程等领域中有着广泛的应用。接下来定义一个满足 Laplace 方程的协调函数:

$$\nabla^2 u = 0 \quad (4)$$

然后,Dirichlet 问题可以转化为寻找满足边界条件的协调函数的问题。考虑到拉普拉斯矩阵 L ,根据文献^[15],式(3)的组合形式为:

$$D[x] = \frac{1}{2} x^T L x \quad (5)$$

此时, x 是最小化式(3)所得到的势能值。重新对网络所有节点进行排序,并将已标记节点放在前面,本文将方程(5)分解为:

$$\begin{aligned} D[x] &= \frac{1}{2} \begin{bmatrix} x_L^T & x_U^T \end{bmatrix} \begin{bmatrix} L_L & B \\ B^T & L_U \end{bmatrix} \begin{bmatrix} x_L \\ x_U \end{bmatrix} \\ &= \frac{1}{2} (x_L^T L_L x_L + 2x_U^T B^T x_L + x_U^T L_U x_U) \end{aligned} \quad (6)$$

其中, x_L 和 x_U 分别为已标记节点和未标记节点的势能。然后设 $D[x]$ 关于 x_U 的导数等于 0,可以得到下列线性方程组:

$$L_U x_U = -B^T x_L \quad (7)$$

其中, x_U 是 $|V_U|$ 的未知数。如果每个社团中至少包含一个已标记节点,那么式(7)是非奇异的。

对于式(7),本文提前指定已标记节点的势能值,接下来只需要计算未标记节点的势能。对于每一个标号 s ,在由标号为 s 的节点生成的静电场中,本文定义 x_i^s 为在节点 v_i 上的势能。将已标记节点的标号集合定义为下列函数:

$$m(v_j) = s, v_j \in \{1, 2, \dots, k\} \quad (8)$$

其中, k 为社团的数量。然后对于节点 $v_j \in V_L$,我们可以给每个标签 s 定义一个 $|V_L| \times 1$ 的向量:

$$p_j^s = \begin{cases} 1, & \text{if } m(v_j) = s \\ 0, & \text{if } m(v_j) \neq s \end{cases} \quad (9)$$

在上述分析的基础上,给每个标号 s 建立一个线性方程系统:

$$L_U x^s = -B^T p^s \quad (10)$$

通过求解方程(10)可以得到组合 Dirichlet 问题的解。对所有的标签来说,式(10)可以转化为:

$$L_U X = -B^T P \quad (11)$$

其中, $X=[x_1, x_2, \dots, x_k]$ 代表 k 列未知数, $P=[p_1, p_2, \dots, p_k]$ 是 k 列已标记节点的势能。显然,方程(11)包括 k 个稀疏线性方程组,但是由于任意节点的势能值总和为 1,因此只有 $k-1$ 个线性稀疏方程需要求解:

$$\sum_{s=1}^k x_i^s = 1, \forall v_i \in V_U \quad (12)$$

这样便可以计算出在已标记节点产生的静电场中未标记节点的势能。然后,比较每个未标记节点的势能,如果某节点与某个已标记节点的势能一样,那么该节点就相应地属于那个已标记节点的社团。

根据离散势能理论,方程(10)的结果可以用电路理论来解释。把 z 表示为一个分支电流向量, f 为电源, b 为电压, p 为势降。根据基尔霍夫电流定律、欧姆定律、基尔霍夫电压定律 3 个电路理论,可以得到以下方程:

$$M^T z = f \text{ (Kirchhoff's Current Law)} \tag{13}$$

$$C_p = z \text{ (Ohm's Law)} \tag{14}$$

$$p = Mx + b \text{ (Kirchhoff's Voltage Law)} \tag{15}$$

其中, M 代表节点边关联矩阵; C 表示构成关系矩阵并被约减成单位矩阵 I , 其中每个元素都是 1。从图论的基本知识可以判断,这 3 个矩阵的关系是 $L = M^T C M$ 。将上面的 3 个方程组合成一个线性的系统:

$$M^T C M x + M^T C b = f \tag{16}$$

这等价于:

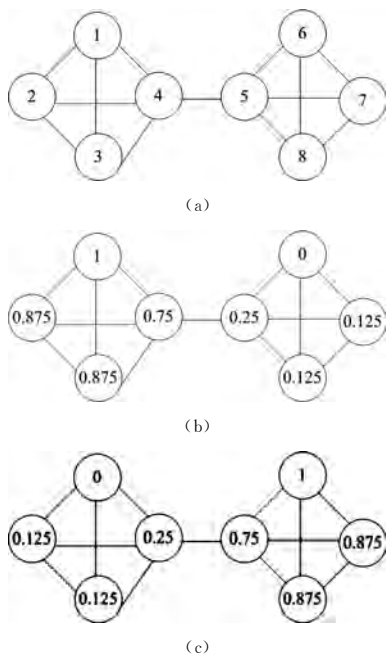
$$L x = f - M^T C b \tag{17}$$

如果让 $f=0$ (没有电流源), 方程(17)等价于方程(10), 电压源(已标记节点)发挥着类似的作用。

为了使基本模型更便于理解,图 1(a)中展示了具有 8 个节点的小型网络。很显然,该网络包括两个社团,一个由节点 v_1, v_2, v_3, v_4 组成,另一个由节点 v_5, v_6, v_7, v_8 组成。将 v_1 和 v_6 定为已标记节点。相应的两个线性方程系统为:

$$L_U x^s = -B^T p^s, s=1,2 \tag{18}$$

其中, $p_1 = [1, 0]^T, p_2 = [0, 1]^T$ 。根据本文提出的方法,得到的社团分区如图 1(b)、图 1(c)所示。



注:节点 v_1 和 v_6 标记为势能来源;图(b)和图(c)分别代表由 v_1 和 v_6 生成的静电场,每个小圆圈中的数字代表该位置相应的势能

图 1 包含两个社团的小型网络(每个社团包括 4 个节点)

2.3 复杂度分析

本节对算法进行总结并分析其时间复杂度,算法的详细步骤如算法 1 所示。

算法 1 基于离散势能论的半监督社团探测算法

输入:邻接矩阵 A , 初始化的标签矩阵 Y

输出:未标记节点的标签

1. 构造出图的拉普拉斯矩阵 L
2. 构造组合 Dirichlet 问题 $D[x]$
3. 重新排序拉普拉斯矩阵,将已标记节点放在前面,其次是未标记节点
4. 建立对称线性方程组 $L_U X = -B^T P$ (参考方程(11))
5. 利用共轭梯度算法,求解第 4 步的方程
6. 对于 $i=1; i \leq n; i++$
返回 $(v_i, y_i = \arg \max_{j \leq k} (X)_{ij})$
7. 结束

本文算法主要包含 3 部分:计算拉普拉斯矩阵、求解 x_u 的优化问题(6)以及比较势能。算法的总时间复杂度取决于其包含的 3 个部分中的最高复杂度。显然,第二部分耗费的时间是最多的,其他两个部分相对来说非常小。对于小型网络,优化问题(6)可以转换为求解稀疏线性方程(11)。而对于大型网络来说,由于拉普拉斯矩阵 L 是一个对称正定矩阵,因此一般采用共轭梯度方法对其进行求解^[19],这样不仅可以在有限的步数内找到最优解,而且需要的内存空间很少,适合大规模网络的计算。应用共轭梯度法来求解问题(4)时最多需要 n 次迭代, n 为节点数目,这相当于每个已标记节点将自己的势能提供给其他未标记节点。因此,在拥有 k 个已标记节点的网络中第二部分的时间复杂度是 $O(k_n)$ 。本文算法具有接近于线性的复杂度,非常适合在包括大规模网络在内的现实网络中应用。

3 实验

本节将所提算法应用到现实网络及人工构建的网络来验证其有效性,其中实验数据集包括 Zachary 空手道俱乐部网络、Lusseau 宽吻海豚网络和 LFR 人工基准网络。

3.1 空手道俱乐部网络

空手道俱乐部网络由 Zachary 建立^[20],他在两年内观察了俱乐部成员间的社会关系变化情况,此网络的 34 个节点与成员相对应,78 条边代表俱乐部成员间的联系。在研究过程中,由于俱乐部管理员(节点 34)和教练(节点 1)之间的分歧,网络被分成两个对立的团体。本文假设这两个团体是空手道俱乐部网络真正的社团结构(见图 2),即第一个是以教练为中心且包括 16 个节点的集合:1-8, 11-14, 17-18, 20 和 22;第二个是由管理员为中心且包括剩余 18 个节点的集合。



注:两种形状节点分别代表两个社团

图 2 Zachary 空手道俱乐部网络划分结果

从图 1 可以看出,在大多数情况下这两个社团可以被正确地分类。但由于节点 3 与同一社团的 4 个节点和另一个社团的 5 个节点相连通,因此节点 3 常常被错误分类,它是唯一

一个不稳定的节点,这与 Girvan-Newman 算法^[2]、Markov 聚类算法^[1]的结果相同。此外,节点 9,10,29,32 等边界节点很少被错误分类,而且在任何情况下社团内部节点都不会被错误划分。因此,本文拥有的背景信息中已标记节点越多,算法就越优良。

3.2 Lusseau 海豚网络

海豚网络是 Lusseau 在 1994—2001 年之间创建的,其中的 62 个节点代表宽吻海豚,159 条边代表海豚之间比随机更经常出现的关联。根据文献^[24],由于其中一只海豚(记为 SN100)的暂时消失,海豚被分成了两组。从文献^[24]中可以推断,海豚网络两个社团之间的团间边共有 6 条。

在海豚网络实验中,本文随机选择了一些已标记节点。表 1 中给出了不同标记节点的相应结果。从表 1 可以发现,在大多数情况下两组海豚都可以被正确分类,最有可能被误判的节点是 SN89。但是如果所选已标记节点不是两组之间的边界节点,那么网络总是能够被正确地划分。总之,只有选择边界节点作为已标记节点才可能导致分区偏差;此外,已标记节点的选择对社团划分精度具有重要影响,它反映了一个事实,即以社团内部的节点作为已标记节点将大大提高划分精度。

表 1 海豚网络社团划分结果

第一组	第二组	被错误分类的节点
Knite	Oscar	无
Mus	CCL	无
Jet SN4	SN89	无
Gallatin	Grin	DN63, SN89
Web	SN9	DN63, SN89
Number1	TR77	SN89
BeescratchOscar	SN100	SN89
Jet, DN63	Kringel	Double PL, SN89
Knite	Web MN60, SN96	SN89
UpbangSN9	MN60, Grin	无

3.3 人工网络

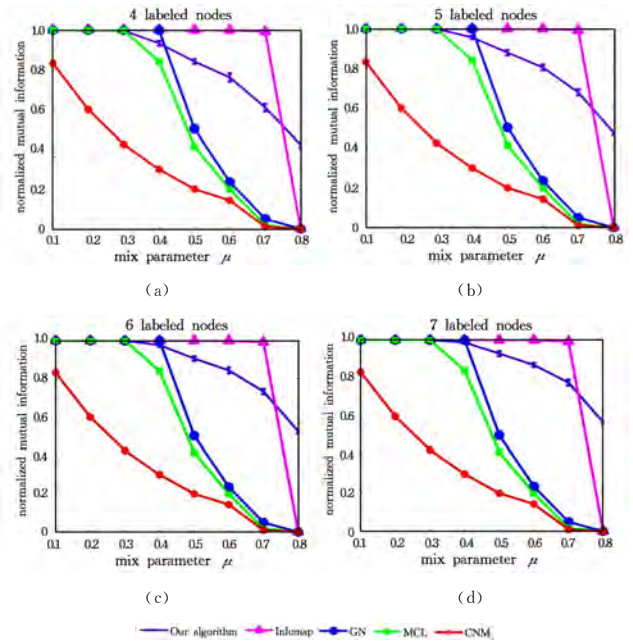
本文在 LFR 基准网络中验证算法的效率。LFR 网络模型由 Lancichinetti 等人提出^[22],其具有无标度特征的度分布和社团规模分布,因此更加接近现实中的社会和生物系统。LFR 网络模型的生成由一些参数控制:节点数目 n 、节点的平均度 $\langle k \rangle$ 、最大节点度、混合比例参数(Mixing Parameter) μ (即每个节点与其他社团的节点共享 μ 比例的边)、最小社团规模 min 和最大社团规模 max 。 μ 的变化范围为 $[0, 1]$,用来调整网络的模糊性,较大的 μ 代表更弱的社团结构。在本次实验中,统一设定网络的节点平均度 $\langle k \rangle = 10$,最大节点度 $max_i, max = 20$ 。每个社团的规模有大小之分,其由参数 min_i, min 和 max_i, max 控制,因此在网络规模固定的情况下可以调节社团的个数。

本文利用归一化互信息的值(NMI)作为比较评价的指标,其具体形式如(19)所示:

$$NMI(A, B) = \frac{-2 \sum_{i=1}^{c_A} \sum_{j=1}^{c_B} N_{ij} \log(\frac{N_{ij} N}{N_i N_j})}{\sum_{i=1}^{c_A} N_i \log(\frac{N_i}{N}) + \sum_{j=1}^{c_B} N_j \log(\frac{N_j}{N})} \quad (19)$$

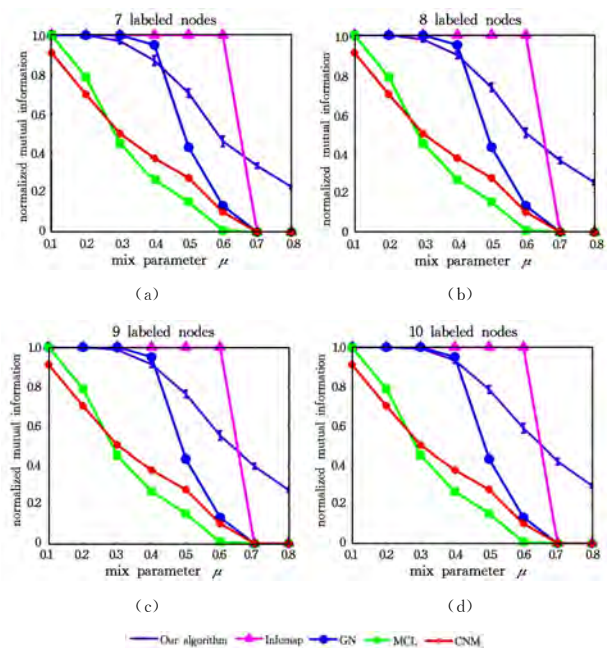
其中, c_A 是社团的真实数量, c_B 表示已发现的社团数量。矩阵 N 代表混合矩阵, N_{ij} 表示属于真实社团 i 却被划分到社团 j 中的节点数量。 N_i 和 N_j 分别是混合矩阵中行 i 和列 j 的和, N 代表网络节点的数目。如果已发现的社团和真实社团完全相同,那么 NMI 取最大值 1,并且 NMI 越接近 1,网络划分的质量越好。

图 3 展示了在小型网络上本文算法与 Girvan-Newman (GN)^[2]、马尔可夫聚类算法 (MCL)^[1]、Clusset 的算法 (CNM)^[21] 和 Infomap 算法^[23] 的实验比较结果。对小型网络来说,当混合比例参数 μ 在 0.1~0.3 时,本文算法与 GN 算法、MCL 算法和 Infomap 算法有着几乎相同的效果。但是当 μ 大于 0.4 时,除了 Infomap 算法,本文算法的效果比其他 4 个算法都好。根据图 4,在大型网络上,当 μ 从 0.4 增加到 0.6 时,本文算法除了比 Infomap 算法的性能稍差,相较于其他算法有着明显的性能优势。特别地,当 μ 处于 0.7 和 0.8 之间时,本文算法的 NMI 值大于 0.2,但是其余 4 种算法的 NMI 值几乎都是 0。也就是说,本文算法特别适用于高混合比例参数的情况,因此更适合划分模糊社团。



注:每个点为 20 次运算结果的平均值

图 3 本文算法和其他算法在小型社团基准网络上的比较结果



注:每个点为 20 次运算结果的平均值

图 4 本文算法和其他算法在大型社团基准网络上的比较结果

- [8] CHEN F, GUO K, LIN J, et al. Intra-cloud lightning: Building CDNs in the cloud[C]//Proc. of IEEE INFOCOM 2012. 2012: 433-441.
- [9] HE J, WU D, ZENG Y, et al. Toward Optimal Deployment of Cloud-Assisted Video Distribution Services[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 23(10): 1717-1728.
- [10] PACIFICI V, DAN G. Distributed algorithms for content allocation in interconnected content distribution networks[C]//Proceedings of IEEE INFOCOM 2015. 2015: 2362-2370.
- [11] 王智. 在线社交视频内容分发策略研究[D]. 北京:清华大学, 2013.
- [12] LI S, XU J, SCHAAR M, et al. Popularity-driven content caching[C]//Proceedings of INFOCOM 2016. 2016: 1-9.
- [13] 叶剑虹, 叶双. 基于混合模式的流媒体缓存调度算法[J]. 计算机科学, 2013, 40(2): 61-64.
- [14] 聂华, 张敏, 郭敬荣. 基于内容流行度差异性的 CDN-P2P 融合分发网络缓存替换机制研究[J]. 通信学报, 2015, 36(s1): 9-15.
- [15] TANG G, WU K, BRUNNER R. Rethinking CDN design with distributee time-varying traffic demands[C]//IEEE INFOCOM 2017 IEEE Conference on Computer Communications. IEEE, 2017: 1-9.
- [16] TANG J, QUEK T Q S, TAY W P. Joint resource segmentation and transmission rate adaptation in Cloud RAN with Caching as a Service[C]//IEEE, International Workshop on Signal Processing Advances in Wireless Communications. IEEE, 2016.
- [17] DAI B, YU W. Energy Efficiency of Downlink Transmission Strategies for Cloud Radio Access Networks[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(4): 1037-1050.
- [18] LIU D, YANG C. Energy Efficiency of Downlink Networks With Caching at Base Stations[J]. IEEE Journal on Selected Areas in Communications, 2015, 34(4): 907-922.

(上接第 317 页)

结束语 根据网络背景知识, 本文提出了一种高效的基于离散势能理论的半监督社团探测方法。从实验结果可以看出, 当将背景知识融入社团探测过程中时, 可以大大提高算法的性能, 特别是当网络社团结构是模糊的情况下。此外, 背景知识在本文算法中应用得越多, 社团划分的结果就会越好。

值得一提的是, 在实际的应用中社团数量这一信息必须提前给出, 这是一个非常严格的约束。未来的工作将集中于没有社团数量先验信息的情况, 使之能够准确地识别更加一般情况下的社团结构。

参 考 文 献

- [1] VAN D S M. Graph clustering by flow simulation[D]. Utrecht: University of Utrecht, 2000.
- [2] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821-7826.
- [3] FORTUNATO S. Community detection in graphs[J]. Physics Reports, 2010, 486(3): 75-174.
- [4] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [5] LI H J, DANIELS J J. Social significance of community structure: Statistical view[J]. Physical Review E, 2015, 91(1): 012801.
- [6] NEWMAN M E J. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences, 2006, 103(23): 8577-8582.
- [7] LI H J, WANG Y, WU L Y, et al. Potts model based on a Markov process computation solves the community structure problem effectively[J]. Physical Review E, 2012, 86(1): 016109.
- [8] 李慧嘉, 李慧颖, 李爱华. 多尺度的社团结构稳定性分析[J]. 计算机学报, 2015, 38(2): 301-312.
- [9] SON S W, JEONG H, NOH J D. Random field Ising model and community structure in complex networks[J]. The European Physical Journal B-Condensed Matter and Complex Systems, 2006, 50(3): 431-437.
- [10] 李慧嘉, 李爱华, 李慧颖. 社团结构迭代快速探测算法[J]. 计算机学报, 2017, 40(4): 970-984.
- [11] WESTON J, LESLIE C, IE E, et al. Semi-supervised protein classification using cluster kernels[J]. Bioinformatics, 2005, 21(15): 3241-3247.
- [12] MA X, GAO L, YONG X, et al. Semi-supervised clustering algorithm for community structure detection in complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2010, 389(1): 187-197.
- [13] EATON E, MANSBACH R. A Spin-Glass Model for Semi-Supervised Community Detection[C]//AAI. 2012.
- [14] ZHANG Z Y. Community structure detection in complex networks with partial background information[J]. Scientific Reports, 2013, 3(1): 3241.
- [15] GRADY L. Random walks for image segmentation[J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2006, 28(11): 1768-1783.
- [16] ZHANG Q M, LÜ L, WANG W Q, et al. Potential theory for directed networks[J]. PloS One, 2013, 8(2): e55437.
- [17] WANG F, ZHANG C. Semisupervised learning based on generalized point charge models[J]. IEEE Transactions on Neural Networks, 2008, 19(7): 1307-1311.
- [18] DOYLE P G, SNELL J L. Random walks and electric networks[M]. Mathematical Association of America, 1984.
- [19] GOLUB G H, VAN LOAN C F. Matrix computations[M]. JHU Press, 2012.
- [20] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452-473.
- [21] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70(2): 066111.
- [22] LANCICHINETTI A, FORTUNATO S, RADICCHI F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4): 046110.
- [23] ROSVALL M, BERGSTROM C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the National Academy of Sciences, 2008, 105(4): 1118-1123.
- [24] LUSSEAU D, NEWMAN M E J. Identifying the role that animals play in their social networks[J]. Proceedings of the Royal Society of London B: Biological Sciences, 2004, 271(Suppl 6): S477-S481.