doi:10.11918/j.issn.0367-6234.2016.11.008

稳定标签传播的社区发现方法

张 鑫, 刘秉权, 王晓龙

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘 要: 为提高标签传播算法的稳定性,解决标签传播算法随机性导致社区发现结果相差较大的问题,对标签初始化、随机队列设置和标签传播中随机选择过程进行了改进,提出一种稳定的标签传播社区发现方法. 该方法首先通过寻找不重叠三角形进行标签初始化,然后以节点标签的熵确定节点队列并分段随机排序,最后考虑邻接点的邻接点标签分布情况进行标签选择. 实验结果表明,在Zachary's Karate Club、Dolphin Social Network 和 American College Football 3个社会网络上,本文方法的稳定指标和质量指标结果均高于其他方法. 稳定标签传播的社区发现方法保持了标签传播算法优点的同时,提高了社区发现结果的质量和稳定性.

关键词: 社区发现;标签传播;随机性;标签的熵;稳定性

中图分类号: TP301.6 **文献标志码**: A **文章编号**: 0367-6234(2016)11-0047-06

Community discovery method based on stable label propagation

ZHANG Xin, LIU Bingquan, WANG Xiaolong

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: In order to improve the stability of label propagation algorithm and reduce the randomness which causes difference in the results of community discovery, labels initialization, random nodes queues setting and labels random selection are improved respectively, and a stable label propagation method for community discovery is proposed. This method first initializes labels by searching for non-overlapping triangles in the networks, and then forms nodes queues based on labels entropy and random sorted nodes in the sub queues. At last, this method chooses labels for each node by the distribution of adjacent nodes labels. Experimental results shows that, stability indexes and quality indexes of our method are higher than other methods' on three social networks—Zachary's Karate club, dolphin social network and American College football. Community discovery based on stable label propagation method not only maintains the advantages of label propagation algorithm, but also improves the quality and stability of community discovery results.

Keywords: community discovery; label propagation; randomness; entropy of labels; stability

网络聚簇结构是复杂网络的重要特征之一,网络聚簇结构特征表明社区结构存在于复杂网络中. 社区,即其内部节点之间关系相对紧密、内部节点与外部节点关系相对稀疏的节点集合. 通过分析复杂网络的结构特征,挖掘复杂网络中的社区结构,这个过程就是社区发现. 起初,研究人员利用图论和概率统计相关理论挖掘网络的本质和特点. 随着互联网的信息爆炸和人们沟通方式的转变,复杂网络的数据规模越来越大,快速、有效的社区发现方法成为多领域研究的热点问题之一[1]. 研究复杂网络社区发现方法对分析复杂网络的拓扑结构和层次结构、理解社区的形成过程、预测复杂网络的动态变化、发现复杂网络中蕴含的规律特征具有重要意义,在众多领域有广泛的应用前景[2-5].

1970年,Kernighan 和 Lin 基于贪婪算法提出了 Kernighan-Lin 方法[6],用于将网络划分为两个规模确定的社区. 该方法需要预先设定社区规模等较多先验知识,在实际网络中应用有限. GN 算法[7]是社区发现经典方法之一,由 Girvan 和 Newman 于2001年提出. 该方法核心思想为社区内的边介数应小于社区间的边介数. GN 算法时间复杂度较高,为 $O(m^2n)$,其中,m 表示网络中边的数量,n 表示网络中的节点数. Newman 等提出模块度[8]作为衡量社区发现结果质量优劣的标准. Palla 等[4,9-10]首次针

收稿日期: 2015-10-26

基金项目: 国家自然科学基金青年科学基金(61300114);国家自然科学基金面上项目(61272383);国家自然科学基金(61572151)

作者简介: 张 鑫(1984—),男,博士研究生;
王晓龙(1955—),男,教授,博士生导师

通信作者: 刘秉权, xzhang@insun.hit.edu.cn

对重叠社区提出了极大团过滤社区发现方法. 该方法需要事先确定参数,不同的参数值使得社区发现结果差异较大.

针对上述传统方法参数难以确定、算法复杂度高的不足,Zhu 等[11]提出了标签传播算法 LPA(label propagation algorithm). 由于 LPA 方法时间复杂度低且效果好,研究人员对其进行了大量深入研究. Raghavan 等[12]首次将 LPA 方法用于复杂网络中的社区发现,提出了 RAK 方法. 该方法首先将网络中每个节点赋予一个唯一的标签,然后根据当前节点的邻接点标签分布情况更新当前节点的标签,重复上述过程直到每个节点的标签都与其邻接点最多的标签相同,标签相同的节点划分为同一个社区. RAK 方法节点初始化标签时间复杂度为 $O(n)$, 每次标签传播的时间复杂度为 $O(m)$. 此外,RAK 方法无需社区数量、社区规模等先验知识,仅根据网络自身结构发现社区,因此 RAK 方法对网络结构有很好的适应性.

为了提高 RAK 方法社区发现的性能,许多研究人员做出很多尝试[13-23]. Barber 等[13]提出了一种模块化标签传播算法,定义目标函数H,将社区发现映射到最优化目标函数 H, 避免整个网络仅为一个社区的情况;Cordasco 等[14]提出了一种基于半同步标签传播过程的方法,标签传播过程是并行的,从而提高了 RAK 方法社区发现的计算速度;Leung 等[15]提出了一种扩展的 RAK 方法用于实时社区监测,通过设定参数,使得算法具有扩展性,提高了 RAK 方法的计算速度. 为降低 RAK 方法的随机性,Zhao 等[16]提出了基于标签的熵的标签传播方法 LPA-E(label propagation in entropic order),将节点按照标签的熵从小到大排序进行标签传播;康旭彬等[17]提出了基于节点相似度的标签传播算法;Sun 等[18]提出了利用邻接点影响力确定标签传播顺序的方法. 尽管上述方法一定程度上提高了标签传播社区发现方法的性能和稳定性,但都是仅从一个方面进行改进,且改进的方面完全消除了随机性,不能体现 RAK 方法的仅依据网络自身结构发现社区的特点.

本文提出一种稳定的标签传播社区发现方法,既保留了 RAK 方法无需先验知识等优点,又提高了标签传播社区发现结果的质量和稳定性. 首先通过网络中不重叠三角形进行标签初始化,然后根据节点标签计算得到的熵确定随机队列,最后考虑邻接点的邻接点标签分布情况确定传播标签.

1 稳定标签传播的社区发现方法

为提高标签传播社区发现方法的稳定性,本文在 RAK 方法的标签初始化、随机队列设置和标签传播过程分别进行了改进. 在标签初始化中,发现网络中所有不重叠三角形,给予三角形三个节点相同的标签,每个三角形标签各不相同,剩余节点赋予其他不同标签;在随机队列设置上,先将节点标签计算得到的熵从小到大对节点排序,再在排序基础上分三段随机排序;针对标签传播的随机选择,考虑被传播节点邻接点的邻接点标签与邻接点标签相同的概率,选择概率大的标签确定传播标签选择.

1.1 不重叠三角形标签初始化

RAK 方法中,每个节点的初始化标签是各不相同的,本文提出了一种无重叠三角形标签初始化方法,用来减少初始标签数量. 网络中,有很多联系紧密具有团体性的节点簇,如极大团,节点簇往往属于同一社区,且易成为社区的核心部分. CPM 算法中[4],极大团往往作为社区的核心部分,进行社区发现. 社区核心部分的确定,社区发现的结果也将更加稳定.

基于网络和社区的这个特点,在标签初始化前,首先找出网络中所有极大团,赋予每个极大团内节点相同的标签. 然而,发现网路中所有极大团是 NP 完全问题,算法的时间复杂度较高,发现所有极大团所耗时间远远超过标签传播整个算法的时间. 因此,本文提出了采用发现网络中没有节点重叠的不重叠三角形的方法,赋予发现到的三角形节点相同的标签,进行网络节点标签初始化,如算法 1 所示.

算法 1 不重叠三角形标签初始化方法伪代码

输入:邻接矩阵 AdjacentMatrix,节点个数 VerticeNum,节点邻居集合 Neighbor.

输出:标签数组 Community.

```
for i ← to VerticeNum Do
    isVisited [i] ← False;
c = 0;
for i ← 0 to VerticeNum Do
    for j ← 0 to Neighbor [i]. size Do
        for k ← 0 to Neighbor [j]. size Do
            if AdjacentMatrix [k][i] = 1 and isVisited [i\j\k] = False then
                Community [i\j\k] ← c;
                isVisited [i\j\k] ← True;
                c ++;
for i ← to VerticeNum Do
    if isVisited [i] ← False;
        Community [i] ← c;
        isVisited [i] ← True;
return Community;
```

如图 1 所示，节点 v_1，v_2 和 v_3 组成一个三角形，初始化标签均为 l_1，其他不能组成三角形的节点分别赋予不同的标签.

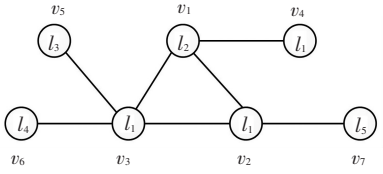

图 1　不重叠三角形标签初始化

Fig.1　Label initialization on non-overlapping triangles

不重叠三角形标签方法的时间复杂度为 $O(n^2)$，比 RAK 方法的近似线性时间复杂度有所提升，但减少了初始标签的数量. 其原因是 RAK 方法初始标签数量等于节点数量，而不重叠三角形的 3 个节点被赋予相同的标签，因此，初始标签数量要少于节点数量，即少于 RAK 方法的初始标签数量.

1.2　基于节点标签的熵的随机队列

分析随机队列对社区发现结果稳定性的影响，考虑图 2 所示网络，6 个节点 v_1，v_2，v_3，v_4，v_5，v_6，初始化标签分别为 l_1，l_2，l_3，l_4，l_5，l_6. 从直观角度看，节点 v_1，v_2 和 v_3 构成一个社区，节点 v_4，v_5 和 v_6 构成另一个社区. 若 v_1 最先进行标签更新，选择的标签可能为 l_2 或 l_3. 接下来无论 v_2 还是 v_3 先更新标签，v_1，v_2，v_3 都能被划分到同一个社区. 若 v_5 或 v_6 先进行标签更新，或 v_4 标签更新的时候不选择 l_3，则 v_4，v_5，v_6 将被划分到另外一个社区. 如果 v_3 最先更新标签，则可能为 l_1，l_2 或 l_4. 若为 l_1 或 l_2，则结果与上述分析一样；若为 l_4，则 6 个节点可能划分为一个社区. 因此，降低随机队列的随机性将提高社区发现结果的稳定性.

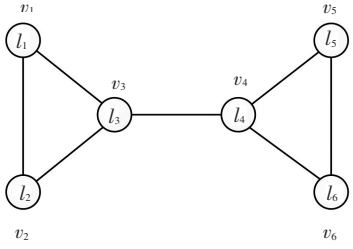

图 2　6 个节点的网络

Fig.2　Network with 6 nodes

为降低随机队列的随机性，本文首先采用了文献[16]的方法，利用节点标签计算得到的熵的大小对节点进行先后排序. 节点标签计算得到的熵公式为

$$H_v = -\sum_{l \in L(v,N(v))} p(l)\log p(l).$$

式中：$L(v, N(v))$ 为节点 v 和其邻居节点的标签集合；$p(l)$ 为标签 l 在 $L(v, N(v))$ 中的概率，即在节点 v 和 $N(v)$ 中，标签为 l 的节点数与节点 v 及其邻接点 $N(v)$ 节点数的比. 具体算法如算法 2 所示.

算法 2　基于节点标签的熵的随机队列方法伪代码

输入：邻接矩阵 AdjacentMatrix，节点个数 VerticeNum，标签数组 Community.

输出：基于节点标签的熵的随机队列 Ssort.

```
for i ← 0 to VerticeNum Do
    FindNeighbor (i, AdjacentMatrix, NeighBor)
    for j ← 0 to Neighbor.size Do
        labelNum[Community[Neighbor[j]]] ++;
    for k ← 0 to Neighbor.size Do
        pl ← labelNum[k] / (Neighbor.size() + 1.0);
        Ssort[i].S += -pl * log(pl);
qsort(Ssort)
RandomSort(Ssort, VerticeNum/3);
RandomSort(Ssort + VerticeNum/3, VerticeNum/3 * 2);
RandomSort(Ssort + (VerticeNum/3) * 2, VerticeNum);
return Ssort;
```

这种排序方法消除了传播节点队列的随机性，使结果变得确定，标签传播方法的适应性大幅度降低. 为了保证标签传播算法的适应性，本文提出将这种方法排序好的队列平均分成三个部分，每个部分内节点进行随机排列. 这样既降低了算法的随机性，又未彻底消除算法随机性，保持了标签传播算法仅依靠网络本身连接结构进行社区发现的初衷.

1.3　基于邻接点的邻接点标签分布的标签选择

标签传播过程中，当遇到最多数量的相同标签不唯一时，RAK 方法采用随机的方式进行选择，这使得最终社区发现结果随机性较大. 为降低标签传播过程中的随机性，本文提出根据被传播节点邻接点的邻接点集标签分布情况，进行标签选择的方法，如算法 3 所示.

v 为当前被传播节点，K 为 $N(v)$ 中相同标签数量最多的节点集合，$k_l \subseteq K$，k_l 为标签是 l 的节点集合. 考虑 $N(k_l)$ 中标签与标签 l 相同的节点所占比例，选择比例最大的那个标签，作为节点 v 新的标签. 如果比例相同，则随机选择一个.

算法 3　基于邻接点的邻接点标签分布的标签传播方法伪代码

输入：邻接矩阵 AdjacentMatrix，节点个数 VerticeNum.

输出：传播后的标签数组 Community.

```
for i ← 0 to VerticeNum Do
```

```
VectorFrequency(Neighbor[i], label);
if label.size() = 1 then
  Community[i] ← label[0];
else then
  for j ← 0 to label.size Do
    LabelFrequency(label[j], freqmax);
  if freqmax.size = 1 then
    Community[i] ← freqmax[0].label;
  else then
    Community[i] ← freqmax[random].label;
return Community;
```

考虑邻接点的邻接点标签分布情况,相当于给邻接点标签加上了一个权重. 如果权重值高,则说明该邻接点的标签背后有更多的支撑,邻接点的标签具有更强的影响力,应该选择该权重值高的邻接点标签. 这使得原来的标签随机性选择变成确定性选择,从而提高了社区发现结果的稳定性. 同时,在权重值相同的情况下,保留了标签选择的随机性,保持了 RAK 方法的适应性.

2 实验及分析

2.1 实验数据

选择了 Zachary's Karate Club[24]、Dolphin Social Network[25]和 American College Football [7](简称 Karate、Dolphins 和 Football 网络)这 3 个被广泛使用的社会网络进行测试,网络具体数据如表 1 所示.

表 1 实验网络的基本数据

Tab.1 Basic data of test networks

网络	节点数	边数
Karate	34	78
Dolphins	62	159
Football	115	616

实验环境为 intel(R) Core(TM) i5 CPU M 430 @2.27GHz,2.27GHz,4GB,Windows 7 操作系统.

2.2 实验评测方法

采用文献[12]提出的 f_{same} 函数和 $J_{j\text{accard's index}}$ 函数作为衡量不同社区相似度标准,将本文方法与 RAK 方法和 LPA-E 方法进行比较. f_{same} 函数用于比较两个社区发现结果的相似度,计算公式为

$$f_{\text{same}} = \frac{1}{2}\left(\sum_i \max_j\{M_{ij}\} + \sum_j \max_i\{M_{ij}\}\right)\frac{100}{n}.$$

式中 M_{ij} 表示在一个社区发现结果中社区 i 和在另一个社区发现结果中社区 j 相同节点的个数. f_{same} 函数对在一个社区发现结果中几个小的社区在另一个社区发现结果中合并成一个大的社区这种情况不是很敏感. 因此,还用到了 $J_{j\text{accard's index}}$ 函数,计算公式为

$$J_{\text{jaccard's index}} = \frac{a}{a+b+c}.$$

式中: a 是在两次发现结果中都在同一个社区的节点对数量,b 是第一次在同一个社区而第二次在不同社区的节点对数量,c 是第一次在不同社区而第二次在同一社区的节点对数量. $J_{j\text{accard's index}}$ 函数值越大,表明两种社区发现结果越相近.

为评测社区发现结果的质量,采用了 Newman 等提出的模块度[8]作为评价标准. Newman 等认为,复杂网络社区最优发现结果并不代表社区间的边数在绝对数量最少,而是比期望边数少. 模块度定义为社区内的边数减去随机生成图中的期望边数,形式化定义如下: 网络划分为 k 个社区,$k*k$ 的矩阵 $\boldsymbol{E}=(e_{ij})$,e_{ij} 表示网络中社区 i 与社区 j 之间的边数占所有边数的比例; 矩阵的迹 $\text{Tr}(\boldsymbol{E}) = \sum_i e_{ij}$,表示网络中社区内部的边数占所有边数的比例;矩阵中第 i 行的和 $a_i = \sum_j e_{ij}$,表示与社区 i 中的点相连边数占所有边数的比例;如果不考虑社区,假定节点间随机连接,那么 $e_{ij} = a_i a_j$. 模块度可以定义为

$$Q = \sum_i (e_{ij} - a_i^2) = \text{Tr}(\boldsymbol{E}) - \|\boldsymbol{E}^2\|,$$

式中 $\|\boldsymbol{X}\|$ 为所有 x 元素之和.

2.3 实验结果与分析

在 Karate、Dolphins 和 Football 网络上对本文方法、RAK 方法和 LPA-E 方法进行测试,选择 5 个社区发现结果进行两两比较. 实验结果如表 2 ~ 4 所示,表中右上半部分为 f_{same} 函数值,左下半部分为 $J_{\text{jaccard's index}}$ 函数值.

表 2 RAK 方法、LPA-E 方法和本文方法在 Karate 网络上社区发现结果比较

Tab.2 Comparison of the results in Karate network among RAK, LPA-E and our method

结果编号	RAK 方法					LPA-E 方法					本文方法				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		83.823	86.764	97.058	92.647		94.234	95.308	96.132	100.00		92.298	95.308	93.443	100.00
2	0.554		91.176	86.764	91.176	0.893		91.041	93.443	94.234	0.881		91.041	94.346	92.298
3	0.627	0.708		89.705	94.117	0.902	0.881		95.308	95.308	0.902	0.912		96.132	95.308
4	0.881	0.609	0.695		89.705	0.923	0.912	0.932		96.132	0.923	0.893	0.923		93.443
5	0.818	0.710	0.790	0.727		1.000	0.893	0.902	0.923		1.000	0.881	0.902	0.923	

表 3　RAK 方法、LPA-E 方法和本文方法在 Dolphins 网络上社区发现结果比较

Tab.3　Comparison of the results in Dolphins network among RAK, LPA-E and our method

结果编号	RAK 方法					LPA-E 方法					本文方法				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		89.405	93.536	86.543	91.034		92.540	91.034	89.405	93.536		93.536	90.405	88.506	94.619
2	0.650		86.543	90.405	84.305	0.704		86.543	85.983	88.506	0.720		89.405	86.543	85.983
3	0.720	0.790		88.506	85.983	0.740	0.783		89.405	91.034	0.734	0.778		90.405	92.540
4	0.690	0.668	0.803		83.504	0.832	0.790	0.734		90.405	0.778	0.813	0.720		91.034
5	0.753	0.679	0.778	0.765		0.778	0.803	0.813	0.753		0.884	0.832	0.720	0.740	

表 4　RAK 方法、LPA-E 方法和本文方法在 Football 网络上社区发现结果比较

Tab.4　Comparison of the results in Football network among RAK, LPA-E and our method

结果编号	RAK 方法					LPA-E 方法					本文方法				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		91.304	92.173	90.000	92.647		91.424	94.442	93.142	95.336		92.243	95.635	92.654	95.423
2	0.729		89.565	84.782	91.176	0.831		86.425	91.424	89.453	0.854		90.312	94.324	90.359
3	0.697	0.622		83.919	94.117	0.803	0.754		88.423	86.425	0.840	0.831		87.423	89.553
4	0.657	0.590	0.521		89.705	0.840	0.772	0.793		90.443	0.829	0.803	0.829		94.649
5	0.758	0.678	0.587	0.847		0.831	0.803	0.813	0.829		0.793	0.789	0.840	0.854	

从表 2~4 实验结果看，LPA-E 方法和本文方法的 f_{same} 函数值和 $J_{\text{jaccard's index}}$ 函数值均高于 RAK 方法，表明两种方法都提升了社区发现稳定性. 在规模较小的 Karate 网络中，随着算法随机性的下降，经常会出现社区发现结果完全一致的情况，如表 2 中 LPA-E 方法和本文方法左下角数值为 1.000.

为从整体上比较三种方法的稳定性，对 Karate、Dolphins 和 Football 网络分别用 RAK 方法、LPA-E 方法和本文方法进行 100 次社区发现，计算两两结果 $J_{\text{jaccard's index}}$ 函数值的平均值，如表 5 所示.

表 5　100 次社区发现结果相互之间的 jaccard's index 函数平均值

Tab.5　Average value of jaccard's index with 100 trails

网 络	RAK 方法	LPA-E 方法	本文方法
Karate	0.609	0.903	0.893
Dophins	0.634	0.784	0.798
Football	0.650	0.818	0.831

在 Dophins 和 Football 网络中，本文方法的 $J_{\text{jaccard's index}}$ 函数值平均值最高；在 Karate 网络中，LPA-E 方法的 $J_{\text{jaccard's index}}$ 函数值平均值最高. 为了分析造成这种结果的原因，统计了初始化时不重叠三角形个数 F_1 和标签传播时遇到邻接点最多数量标签不唯一的次数 F_2，用来分析本文 1.1 节中改进方法和 1.3 节中改进方法在不同网络中的影响力，如表 6 所示.

表 6　100 次社区发现中不重叠三角形个数和标签传播时邻接点最多数量标签不唯一次数的平均值

Tab.6　Average value of non-overlapping triangles and number of maximum label not unique with 100 trials

网 络	F_1	F_2
Karate	4	6
Dophins	11	21
Football	31	11

从表 6 数据可知，Karate 网络中的不重叠三角形个数和标签传播时遇到邻接点最多数量标签不唯一的次数都要少于 Dophins 和 Football 网络中的个数和次数，这表明本文 1.1 节中改进方法和 1.3 节中改进方法在 Karate 网络中的影响力低于在 Dophins 和 Football 网络中的影响力. 表 5 Karate 结果中，本文方法结果低于 LPA-E 方法结果，是由于本文 1.1 节中改进方法和 1.3 节中改进方法所提高的稳定性不足以抵消 1.2 节中改进方法里保留的随机性，这主要是由网络规模决定的，网络规模越大，网络中的不重叠三角形数量越多，发生标签传播时邻接点最多数量标签不唯一的情况越多. 虽然本文方法的 $J_{\text{jaccard's index}}$ 函数平均值略低于 LPA-E 方法，但远高于 RAK 方法，说明本文方法较好地提高了社区发现结果稳定性，同时也验证了本文方法没有完全消除 RAK 方法的随机性，保留了 RAK 方法对网络本身结构适应性的优点.

对 Karate、Dolphins 和 Football 网络分别用 RAK 方法、LPA-E 方法和本文方法进行 100 次社区发现，并计算社区发现结果的 Q 函数平均值，结果如表 7 所示.

表 7　100 次社区发现结果的 Q 函数平均值

Tab.7　Average value of Q function of community discovery with 100 trails

网 络	RAK 方法	LPA-E 方法	本文方法
Karate	0.367	0.375	0.384
Dophins	0.425	0.445	0.449
Football	0.460	0.478	0.482

Q 函数表示的是社区内边数与随机生成图中期望边数的差，反映了社区内部紧密程度，Q 函数值越大，表明发现的社区结构越紧密，越符合社区的定义，社区发现结果质量越好. 实验结果表明，本文方法的 Q 函数平均值高于 RAK 方法和 LPA-E 方法，提升了社区发现质量.

实验表明,本文算法较好地提升了标签传播算法的稳定性和社区发现结果质量.

3 结 论

1)改进了RAK方法标签初始化过程,通过挖掘网络中的不重叠三角形,赋予三角形节点相同的标签,减少了初始化标签数量;

2)降低且未完全消除方法的随机性,采用节点标签计算得到的熵和邻接点的邻接点标签分布情况进行标签传播选择,提高了社区发现结果的稳定性,同时保持了标签传播方法的无需先验知识,仅依靠网络结构本身的特点.

3)本文改进的算法较好地提高了标签传播社区发现结果的质量和稳定性.

参考文献

[1] ADAMIC L A, HUBERMAN B A, BARABÁSI A L, et al. Power-law distribution of the world wide web[J]. Science, 2000, 287(287):2115a-2115a. doi: 10.1126/science.287.5461.2115a.

[2] SIDIROPOULOS A, PALLIS G, KATSAROS D, et al. Prefetching in content distribution networks via web communities identification and outsourcing[J]. World Wide Web-internet & Web Information Systems, 2008, 11(1):39-70. doi:10.1007/s11280-007-0027-8.

[3] WANG Zhi, ZHANG Jianzhi. In search of the biological significance of modular structures in protein networks[J]. Plos Computational Biology, 2007, 3(6): 1011-1021. doi: 10.1371/journal.pcbi.0030107.

[4] PALLA G, DERENYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435(7043):814-818. doi:10.1038/nature03607.

[5] LI Xin, LIU Bing, YU P S. Discovering overlapping communities of named entities[J]. Lecture Notes in Computer Science, 2006, 4213:593-600. doi: 10.1007/11871637_60.

[6] KERNIGHAN B W, LIN S. An efficient heuristic procedure for partitioning graphs[J]. Bell System Technical Journal, 1970, 49(2): 291-307. doi: 10.1002/j.1538-7305.1970.tb01770.x.

[7] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12):7821-7826. doi: 10.1073/pnas.122653799.

[8] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004, 69(2 Pt 2):026113-026113. doi: 10.1103/PhysRevE.69.026113.

[9] DERENYI I, PALLA G, VICSEK T. Clique percolation in random networks[J]. Physical Review Letters, 2005, 94(16): 160202-160202. doi: 10.1103/PhysRevLett.94.160202.

[10] PALLA G, FARKAS I J, POLLNER P, et al. Directed network modules[J]. New Journal of Physics, 2007, 9(26):186-206. doi: 10.1088/1367-2630/9/6/186.

[11] ZHU X, GHAHRAMANI Z. Learning from labeled and unlabeled data with label propagation: Technical Report CMUCALD-02-107[R]. Pittsburgh PA: Carnegie Mellon University, 2002.

[12] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(3 Pt 2). doi: 10.1103/PhysRevE.76.036106.

[13] BARBER M J, CLARK J W. Detecting network communities by propagating labels under constraints.[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2009, 80(2 Pt 2):283-289. doi: 10.1103/PhysRevE.80.026129.

[14] CORDASCO G, GARGANO L. Community detection via semi-synchronous label propagation algori-thms[C]// 2010 IEEE International Workshop on Business Applications of Social Network Analysis (BASNA). Bangalore: IEEE, 2010:1-8.

[15] Leung I X Y, PAN Hui, LIO P, et al. Towards real-time community detection in large networks[J]. Physical Review E, 2009, 79(6):853-857. doi: 10.1103/PhysRevE.79.066107.

[16] ZHAO Yuxin, LI Shenghong, CHEN Xiuzhen. Community detection using label propagation in entropic order[C]// 2012 IEEE 12th International Conference on Computer and Information Technology (CIT). Si Chuan: IEEE, 2012: 18-24.

[17] 康旭彬, 贾彩燕. 一种改进的标签传播快速社区发现方法[J]. 合肥工业大学学报(自然科学版), 2013,36(1):43-47. doi:10.3969/j.issn.1003-5060.2013.01.010.
KANG XUBIN, JIA CAIYAN. An improved fast community detection algorithm based on label propagation[J]. Journal of HeFei University of technology, 2013,36(1): 43-47. doi:10.3969/j.issn.1003-5060.2013.01.010.

[18] SUN Heli, HUANG Jianbin, ZHONG Xiang, et al. Label propagation with α-degree neighborhood impact for network community detection[J]. Computational Intelligence & Neuroscience, 2014 (2014): 130689-130689. doi:10.1155/2014/130689.

[19] XING Yan, MENG Fanrong, ZHOU Yong, et al. A node influence based label propagation algorithm for community detection in networks[J]. Scientific World Journal, 2014, 2014(3): 627581-627581. doi: 10.1155/2014/627581.

[20] SUN Heli, LIU Jiao, HUANG Jianbin, et al. CenLP: a centrality-based label propagation algorithm for community detection in networks[J]. Physica A Statistical Mechanics & Its Applications, 2015, 436:767-780. doi:10.1016/j.physa.2015.05.080.

[21] HOSSEINI R, AZMI R. Memory-based label propagation algorithm for community detection in social networks[C]// 2015 International Symposium on Artificial Intelligence and Signal Processing (AISP). Mashhad : IEEE, 2015:256-260.

[22] DICKINSON B, HU Wei. The effects of centrality ordering in label propagation for community detection[J]. Social Networking, 2015, 4(4):103-111. doi: 10.4236/sn.2015.44012.

[23] ZHANG Xiankun, TIAN Xue, LI Yanan, et al. Label propagation algorithm based on edge clustering coefficient for community detection in complex networks[J]. International Journal of Modern Physics B, 2014, 28(30): 1450216-1450216. doi: 10.1142/S0217979214502166.

[24] ZACHARY W W. An Information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452-473.

[25] LUSSEAU D, SCHNEIDER K, BOISSEAU O J, et al. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations[J]. Behavioral Ecology & Sociobiology, 2003, 54(4):396-405. doi: 10.1007/s00265-003-0651-y.

(编辑 王小唯 苗秀芝)