C M Y K E:\发完的文件\15年6期文件\1171-1175黄佳鑫.PS 2015/05/14 09:03:46

小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems

2015 年 6 月 第 6 期
Vol. 36 No. 6 2015

融入节点重要性和标签影响力的标签传播社区发现算法

黄佳鑫,郭 昆,郭 红

(福州大学 数学与计算机学院,福州 350108)
E-mail:jiaxin_huang_miss@163.com

摘 要: 近年来,高质量社区的挖掘和发现已经成为社会网络研究一个热点.其中,基于标签传播的社区挖掘算法(Label Propagation Algorithm,简称 LPA)由于具有近似线性时间复杂度且无须预先定义目标函数和社区数量等优点而得到广泛关注.但是,LPA 算法的标签传播过程存在不确定性和随机性,影响了社区发现的准确性和稳定性.提出一种新的基于标签传播的社区发现算法 LPA_SI(Label Propagation Algorithm based on Significance and Influence).首先,采用新的节点重要性度量方法对节点进行排序;其次,提出一种新的标签影响力计算方法更新每个节点的标签;最后,在真实数据集和人工数据集上的实验表明,LPA_SI 在复杂度相近的情况下能够显著提高社区发现的质量,并具有较好的稳定性.

关 键 词: 社会网络;社区发现;标签传播;标签影响力;节点重要性
中图分类号: TP393 **文献标识码:** A **文 章 编 号:** 1000-1220(2015)06-1171-05

Label Propagation Algorithm for Community Detection Based on Vertex Significance and Label Influence

HUANG Jia-xin,GUO Kun,GUO Hong
(College of Mathematics and Computer Science,Fuzhou University,Fuzhou 350108,China)

Abstract: In recent years,mining and detecting of high quality communities has become a hot orientation in social network research. Among them,the community detection algorithms based on label propagation (LPA) receive broad attention for the advantages of near-linear complexity and no prerequisite for any object function or cluster number. However,the propagation of labels contains uncertainty and randomness,which affects the accuracy and stability of the LPA algorithms. In this paper,a novel community detection algorithm based on LPA called LPA_SI (LPA based on Significance and Influence) is proposed. Firstly,the vertices are sorted according to a new vertex importance measure. Then,the label of each vertex is updated according to a new label influence measure. The experiments on both the artificial datasets and the real-world datasets demonstrate that the LPA_SI algorithm significantly improves the community quality while the detection stability is preserved.

Key words: social network;community detection;label propagation;label influence;node significance

1 引 言

社会网络是指个体成员之间由于发生交互而形成一种相对稳定的关系体系.社会网络通常被抽象成一个图,图中的节点表示个人,而边(或链接)表示人与人之间的关系[1].社会网络中的社区结构挖掘对于社会网络分析具有重要意义.近 10 年来,已有很多社会网络社区挖掘方法被提出,依据采用的求解策略不同,主要可以分为基于优化的社区发现方法和基于启发式的社区发现方法[2].基于优化的方法通过设置目标函数并迭代逼近函数最优值实现社区发现,具有代表性的方法包括谱方法和模块性最大化方法[3-6].其中,谱方法[3]将社区识别问题转换为放松的二次型优化问题,通过求 Laplacian 矩阵的特征向量得到网络的近似最优划分.模块性最大化方法,是对网络模块性函数(modularity,又称 Q 函数)求最大值,代表性算法如 FN 算法[4]、GA 算法[5]和 EO 算法[6].基于启发式策略的方法通过设置启发规则来寻找最优社区划分,代表性的算法如 GN 算法[7]和 WH 算法[8].

此外,还有其他一些有效的社区发现方法,例如:基于层次聚类的社区发现方法将社会网络看成由多层社区组成,利用传统层次聚类算法实现社区发现;基于进化计算的社区发现方法将社区发现过程看作目标函数的最优化,利用进化计算方法寻求问题的近似最优解.这些算法一般具有较高的时间和空间复杂度,不适用于大规模网络挖掘.2007 年,Raghavan 等人[9]提出了一种基于标签传播思想的快速社区发现算法 LPA(Label Propagation Algorithm),该算法具有线性的时间复杂度,在处理大规模的网络时具有很好的时间效率,并且不需要优化预定义的目标函数,也不需要关于社区的数量和规模等先验信息,因此逐渐得到学者的广泛关注.但是,LPA

收稿日期:2014-05-30 收修改稿日期:2014-06-28 基金项目:国家自然科学基金项目(61103175,61300104)资助;教育部科学技术研究重点项目(212086)资助;福建省科技创新平台建设项目(2009J1007)资助;福建省自然科学基金项目(2013J01230)资助;福建省高校杰出青年科学基金项目(JA12016)资助;福建省高等学校新世纪优秀人才支持计划项目(JA13021)资助. 作者简介:黄佳鑫,女,1989 年生,硕士研究生,研究方向为社交网络社区挖掘、社交圈子识别;郭 昆(通信作者),男,1979 年生,博士,讲师,研究方向为灰色系统理论的数据挖掘和复杂网络(特别是社交网络)上的数据挖掘;郭 红,女,1965 年生,硕士,副教授,研究方向为人工智能和数据库技术.

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

C M Y K E:\发完的文件\15年6期文件\1171-1175黄佳鑫.PS 2015/05/14 09:03:46

算法在迭代更新节点标签的过程中存在不确定性和随机性，导致其结果准确性和稳定性常常不能达到预期. 文献[10-12]分别从不同角度对 LPA 算法进行改进，但是这些算法仅根据标签的个数评判标签的影响力，没有考虑节点信息和节点间紧密度对标签选择的影响. 文献[13]综合考虑节点度和边的权重对标签选择的影响，但没有考虑邻近节点的信息和邻近节点的集聚程度对标签选择的影响. 本文提出一种节点重要性度量方法，并在此基础上设计一种新的标签影响值力计算公式，应用于标签迭代传播. 实验表明：新的方法能够显著提高标签选择的准确性，从而提高社区发现的准确性和稳定性.

本文其余部分组织如下：第 2 节介绍标签传播算法的基本思想和实现步骤；第 3 节介绍本文算法的基本思想和具体步骤，并进行复杂度分析；第 4 节通过在人工数据集和真实数据集上的实验对提出的算法进行检验；第 5 节总结本文的工作并展望下一步的研究方向.

2 标签传播算法

LPA 算法是一类启发式算法，其启发规则为：不断在节点及其近邻间传递标签信息，经过多次迭代后，属于同一个社区的节点的标签将趋于一致. 其核心思想是采用邻近节点的标签数量最多的标签作为每个节点自身的标签，以确定节点所在的社区. 标准 LPA 算法的具体步骤如下[2]：

输入： 网络 $G=(V,E)$，V 为节点集，E 为边集，最大迭代次数 $maxIter$

输出： 社区集 $Setc=\{C_1,\cdots,C_k\}$，k 为社区数

算法步骤：

1) 初始化每个节点 $v\in V$ 的标签；

2) 迭代次数 $t=1$；

3) 随机对节点排序，生成有序列表 $V'=\{v_1,v_s,\cdots,v_n\}$；

4) 对任意 $v_i\in V'$，根据下式(1)或(2)进行节点标签迭代更新，在迭代过程中，每个节点采用大多数邻近节点的标签更新自身的标签，如果有多个标签的数量同为最大值，则随机选取一个标签作为该节点的标签. 标签更新分为同步更新和异步更新. 同步更新即节点 x 在第 t 次迭代时的标签根据其邻近节点第 $t-1$ 次迭代时的标签来更新，公式如下：

$$C_x(t)=f(C_{x_1}(t-1),\cdots,C_{x_k}(t-1)),x_i\in\alpha(x) \tag{1}$$

其中 $C_x(t)$ 为节点 x 在第 t 次迭代时的标签，$\alpha(x)$ 为节点 x 的邻近节点集. 同步更新方式在具有二分结构的图中可能会出现震荡现象，而异步更新能很好的解决这个问题. 异步更新即节点 x 在第 t 次迭代时的标签由其邻近节点中第 t 次迭代和第 $t-1$ 次迭代时的标签共同决定. 其公式如下：

$$C_x(t)=f(C_{x_{i1}}(t-1),\cdots,C_{x_{im}}(t-1),C_{x_{i(m+1)}}(t),\cdots,C_{x_{ik}}(t)),x_{im}\in\alpha(x) \tag{2}$$

5) 若迭代次数 $t==maxIter$ 或每个节点的标签与其大多数邻近节点的标签相同，将具有相同标签的节点归入相同社区，算法结束；否则 $t=t+1$，返回步骤 3)；

由上述步骤可知，标准 LPA 算法的不确定性和随机性主要源于步骤 4)中对节点标签的随机选择，这影响了 LPA 算法的的准确性和稳定性.

近年来，不同学者对标准 LPA 算法进行了改进. Steve 等[10]对 LPA 算法进行扩充，提出一种挖掘重叠社区结构的算法 COPRA，每个节点保留若干个社区标签，从而能够检测重叠社区. 文献 Barber 等[11]提出一种带约束的标签传播算法 LPAm，将 LPA 等价为一个优化问题，并给出对应目标函数 H，从而解决 LPA 的聚类性能问题. Leung 等[12]经过实验发现 LPA 经过五次迭代后，95% 的节点已正确的聚集，后面的迭代主要是对社区内节点的更新，因此，改进了 LPA 的更新准则和迭代规则. 但这些算法均未考虑到节点的重要性和不同标签的影响力. 在社会网络中，不同用户节点对社区的重要性可能存在很大差别，而不同标签对不同用户也可能存在完全不同的影响. 赵卓翔等[13]提出一种基于标签影响力的标签传播算法 LIB，在更新节点标签时选择影响力最大的标签作为节点的新标签. 但 LIB 算法没有考虑不同节点的重要性对社区生成的影响.

3 LPA_SI 算法

多数标签传播算法在标签更新阶段，随机对节点进行排序，无考虑节点的重要性对节点更新的影响，可能导致重要性较小的节点反过来影响一些重要性较大的节点，使传播过程产生“逆流”现象，并且在标签选择阶段，单纯采用标签个数度量标签的影响力，无考虑节点信息、节点间紧密度和邻近节点信息等对标签选择的影响. 本节提出一种新的节点重要性度量公式，基于该公式设计一种标签影响力计算公式，并进一步提出一种基于节点重要性和标签影响力的标签传播算法.

3.1 节点重要性度量

节点重要性用于度量节点在整个网络中的影响力. 文献[14]提出一种基于节点点权的节点重要性度量方法，但没有考虑邻近节点集聚系数对节点重要性的影响. 文献[15]提出一种基于度和集聚系数的节点重要性度量方法，但没有考虑点权对节点重要性影响，借鉴文献[14]和文献[15]的思想，综合考虑点权和集聚系数，提出一种新的节点重要性度量方法如下：

$$S(x)=\frac{W(x)}{\sqrt{\sum_{y\in\alpha(x)}W(y)^2}}+\frac{C(x)}{\sqrt{\sum_{y\in\alpha(x)}C(y)^2}} \tag{3}$$

其中，$S(x)$ 为节点 x 的重要性，$W(x)$ 表示节点 x 自身的点权和其邻居节点的点权之和，$\alpha(x)$ 为节点 x 的邻近节点集. $W(x)$ 根据下式计算：

$$W(x)=\sum_{y\in\beta(x)}w(x,y)+\sum_{z\in\beta(x)}\sum_{y\in\beta(z)}w(z,y) \tag{4}$$

其中，$w(x,y)$ 表示节点 x 与其邻近节点 y 的紧密度[14]，$\beta(x)$ 为节点 x 的邻近节点集. $C(x)$ 为节点 x 的集聚系数，定义为：

$$C(x)=\frac{2\times e(x)}{k(x)\times(k(x)-1)} \tag{5}$$

其中，$e(x)$ 表示节点 x 与其任意两个邻近节点所形成的三角形的个数，$k(x)$ 表示节点 x 的度数. 公式(3)中的 $W(x)$ 和 $C(x)$ 都采用同趋化函数 $u(x)=x/\sqrt{x^2}$ 进行归一化.

3.2 标签影响力计算

节点重要性越大，对其他节点的影响力越大，节点的标签越容易被传播，并且节点间紧密度越高，邻近节点的标签对目标节点的影响越大，因此，综合考虑节点重要性和节点间紧密度，提出一种新的标签影响力计算公式：

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

C M Y K E:\发完的文件\15年6期文件\1171-1175黄佳鑫.PS 2015/05/14 09:03:46

$$Influence(l) = \frac{\sum w(l_i)}{\sqrt{\sum_{y\in L} w(y)^2}} + \frac{\sum S(l_i)}{\sqrt{\sum_{y\in L} S(y)^2}} \quad (6)$$

其中,$Influence(l)$ 表示标签 l 的影响力,$l \in L$,L 为 x 在所有邻近节点上的标签组成的集合,$w(l_i)$ 表示每一个标签为 l_i 节点与目标节点的紧密度,$S(l_i)$ 表示每一个标签为 l_i 的节点的重要性.

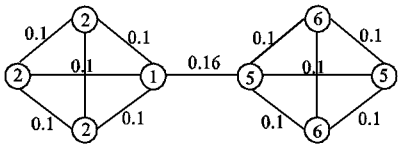

图 1　标签选择过程

Fig. 1　Label selection process

下面给出一个模拟示例,说明采用新的节点重要性度量方法和标签影响力后的标签更新过程. 如图 1 所示,标签 1 对应的节点共有标签 2 和标签 5 可以选择,采用公式(6)计算两者的影响力值,其中标签 2 的值为 1.82,而标签 5 的值为 0.81,因此,选择标签 2 作为节点 x 的标签.

3.3　算法实现

以上述节点重要性度量方法和标签影响力计算公式为基础,设计新的标签传播的社区发现算法 LPA_SI(Label Propagation Algorithm based on Significance and Influence),具体步骤如下:

输入: 网络 $G=(V,E)$,V 为节点集,E 为边集,最大迭代次数 $maxIter$

输出: 社区集 $Setc=\{C_1,\cdots,C_k\}$,k 为社区数

算法步骤:

1) 初始化每个节点 $v \in V$ 的标签;

2) 根据公式(1)计算每个节点的重要性,并按节点重要性从高到低对节点排序,生成有序列表 $V'=\{v_1,v_s,\cdots,v_n\}$,其中 $S(v_1) \geqslant S(v_s) \geqslant \cdots \geqslant S(v_n)$;

3) 迭代次数 $t=1$;

4) 对任意 $v_i \in V'$,根据下式将 v_i 的标签更新为其近邻节点标签集中影响值最大的标签:

$$Label(v_i) = argmax\{Influence(l_i)\}, l_i \in L(v_i) \quad (7)$$

其中,$Label(v_i)$ 为节点 v_i 的新标签,$L(v_i)$ 为节点 v_i 的近邻标签集.

5) 若迭代次数 $t==maxIter$ 或每个节点的标签为影响值最大的标签,将具有相同标签的节点归入相同社区,算法结束;否则 $t=t+1$,返回步骤(4);

LPA_SI 算法依据节点的重要性对节点进行排序更新,使得重要性较大的节点影响重要性较小的节点,有利于减少"逆流"现象的产生,且综合考虑了节点间的紧密度以及节点的重要性对标签选择的影响,使得节点的标签选择更准确,从而提高了算法的稳定性和准确性.

3.4　复杂度分析

设 $n=|V|$,$m=|E|$,首先分析LPA_SI算法的时间复杂度. 步骤 1 需要的时间为 $O(n)$,步骤 2 中,计算节点重要性需要的时间为 $O(m)$,采用计数排序对节点重要性排序时间为 $O(n)$,总时间复杂度为 $O(n+m)$. 步骤4 中,一次节点标签更新时间为 $O(m)$,$maxIter$ 次迭代更新时间为 $O(m*maxIter)$. 步骤5 中生成社区的时间为 $O(n)$. 因此,LPA_SI 算法总的时间复杂度为 $O(n+m*maxIter)$,与标准 LPA 算法同样具有接近线性的时间复杂度.

接着分析 LPA_SI 算法的空间复杂度. 步骤 1 用邻接表存储节点和边信息,需要的空间为 $O(n+m)$. 步骤 2 采用计数排序对节点重要性排序需要的空间为 $O(n)$. 步骤5 社区生成需要的空间为 $O(n)$. 因此,LPA_SI 算法总的空间复杂度为 $O(n+m)$,与标准 LPA 算法相同.

4　实验结果与分析

为了验证本文提出的 LPA_SI 算法的性能. 分别选择人工生成的数据集和真实的数据集进行实验. 人工数据集利用 Lancichinetti[16] 等提出的 LFR 基准程序生成. LFR 基准网络能够根据输入的参数生成不同数据量、不同度分布及不同簇数的仿真数据集. 真实数据集采用斯坦福大学的大规模网络数据集 SNAP 中的 ego-Facebook 数据集以及 Karate、Dolphins、Polbooks、Football 和 HepPh 数据集. 实验数据集的详细信息如表 1 所示.

表 1　实验数据集

Table 1　Information of real networks

数据集	参　数	
人工数据集	$N=10000 \sim 100000$　$\mu=0.1 \sim 0.6$ $k=5 \sim 55$　$kmax=60$ $minc=10 \sim 20$　$maxc=50 \sim 100$	
真实数据集		
Karate	$N=34$	$E=78$
dolphins	$N=62$	$E=159$
polbooks	$N=105$	$E=441$
Football	$N=115$	$E=613$
3437	$N=534$	$E=4813$
1912	$N=747$	$E=30025$
107	$N=1034$	$E=26749$
netscience	$N=1461$	$E=2742$
HepPh	$N=34546$	$E=420877$

表 1 中,参数 N、E、k、$kmax$、$minc$、$maxc$ 和 μ 分别表示数据集节点个数,边的数量、节点的平均度,节点的最大度,最小社区包含的节点个数,最大社区包含的节点的数量和混合参数. 所有算法均采用 Java 语言实现,并在硬件配置为 Inter(R) Core(TM) i3-2120 CPU @ 3.30GHz,3GB RAM,软件配置为 Microsoft Windows 7,JDK 7.0 的平台上进行测试.

实验比较了 LPA_SI 算法与 LPA 算法和 LIB 算法在不同节点数和不同混合参数时的模块度 Q、标准互信息 NMI 和运行时间等指标的值. 为了克服随机性对算法测试的影响,所有实验结果均取 50 次运行结果的平均值. 同时,还统计了标准互信息和模块度运行结果的标准差. 算法的最大迭代次数均设置为 100.

4.1　评价指标

4.1.1　模块度 Q

模块度是 Newman 和 Girvan 在文献[17]中提出的用于评价社区质量的指标,对于一个没有重叠社区的数据集,模块

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

C M Y K E:\发完的文件\15年6期文件\1171-1175黄佳鑫.PS 2015/05/14 09:03:46

度定义为：

$$Q = \sum_{c=1}^{n_c}\left[\frac{l_c}{m} - \left(\frac{d_c}{2m}\right)^2\right] \quad (8)$$

其中，n_c 表示社区数量，m 为网络中的总边数，l_c 表示社区 c 的内部边的数量的总和，d_c 表示社区 c 中节点度数的总和.

4.1.2 标准互信息 NMI

对于 LFR 基准网络，由于网络社区结构的真值是已知的，因此，采用文献[18]中给出的规范化互信息(normalized mutual information，简称 NMI)作为社区发现的评价指标，可以有效评价社区发现算法的准确性. NMI 的定义为：

$$NMI(X|Y) = 1 - [H(X|Y) + H(Y|X)]/2 \quad (9)$$

其中，X 为所有真实社区的集合，Y 为所有预测社区的集合，$H(X|Y)$ 为 X 在 Y 上的规范化条件熵.

4.2 人工数据集上的实验结果

4.2.1 不同节点数的实验结果

不同算法的运行时间、NMI 值和 NMI 标准差值随着节点数变化的实验结果分别列于图 2 至图 4.

从图2可以看出，LPA_SI和LPA、LIB算法一样，都具有较高的，在“运行效率”与“算法之间”的运行时间与网络中节点个数呈近似线性关系. 其中，LPA_SI 算法在初始化阶段，需要计算节点的重要性并对其进行排序，在标签更新阶段，需要计算标签的影响值，导致算法的时间效率略低于 LPA 和 LIB 算法. 尽管 LPA 和 LIB 的时间效率略优于 LPA_SI 算法，但是从图 3 和图 4 可以看出，它们在社区发现质量和算法稳定性方面不如于 LPA_SI 算法. 这主要是由于：LPA_SI 算法在标签更新阶段，根据节点重要性由高到低的顺序对节点进行标签更新，避免重要性较小的节点影响重要性较大的节点，减少了“逆流“想象的产生，并且在标签选择中，综合考虑节点间的紧密度，以及节点的重要性对标签的影响，提高了标签选择的准确性，从而提高了算法的稳定性和准确性.

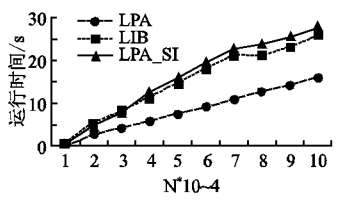

图 2 算法运行时间比较

Fig. 2 Time efficiency comparison of various algorithms

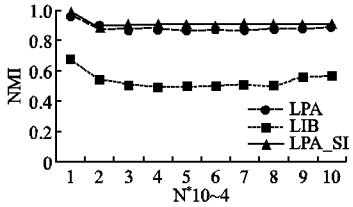

图 3 算法 NMI 值比较

Fig. 3 NMI comparison of various algorithms

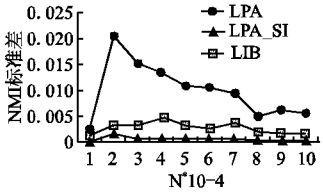

图 4 算法 NMI 标准差比较

Fig. 4 NMI variance comparison of various algorithms

4.2.2 不同混合参数的实验结果

不同算法在小社区网络和大社区网络中，NMI 值随着混合参数 μ 的变化的实验结果分别列于图 5 和图 6.

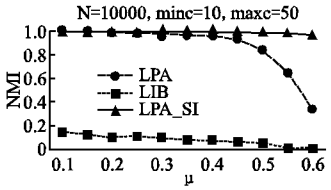

图 5 小社区网络算法 NMI 值比较

Fig. 5 NMI on big networks with small communities

由图 5 和图 6 可以看出，当社区结构比较明显时，LPA_SI 和 LPA 算法的 NMI 都较高，但是随着混合系数 μ 增大，社区结构变得越来越模糊，社区发现的难度也逐步增大. 此时，LPA_SI、LPA 和 LIB 算法的 NMI 值都呈现下降的趋势. 但是，通过结合考虑节点重要性和标签的影响力，LPA_SI 算法能够在社区结构不显著的条件下仍保持高于 LPA 和 LIB 算法的准确率.

4.2.3 真实数据集上的实验结果

表 2 显示了 3 种算法分别在 9 个真实网络上的 Q 值(AvgQ)及对应的 Q 标准差(StdQ). 表 2 中加粗的数字分别表示 AvgQ 和 StdQ 的最优值，AvgQ 值越大，社区识别的结果越准确；StdQ 值越小，算法的社区识别结果越稳定.

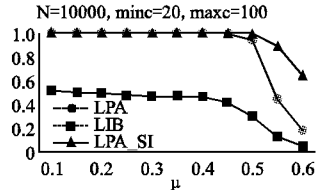

图 6 大社区网络算法 NMI 值比较

Fig. 6 NMI on big networks with big communities

由表 2 可以看出，LPA_SI 算法在多数网络上得到最高的 Q 值. 此外，LPA_SI 算法在所有的测试数据集上都具有最小的 Q 值标准差，甚至在其中 6 个数据集上的标准差为 0. 这表明 LPA_SI 算法在真实数据集上的准确性和稳定性显著优于 LPA 算法和 LIB 算法.

表 2 真实网络的社区发现结果

Table 2 Community detection results on real neworks

ID	LPA		LIB		LPA_SI	
	AvgQ	*Std*	AvgQ	*Std*	AvgQ	*Std*
Karate	0.353	0.08	0.379	0	0.395	0
dolphins	0.441	0.178	0.462	0	0.512	0
polbooks	0.49	0.04	0.314	0.007	0.521	0
Football	0.583	0.06	0.322	0.027	0.612	0
3437	0.664	0.032	0.405	0.031	0.673	0.002
1912	0.526	0.071	0.416	0.033	0.522	0
107	0.501	0.014	0.386	0.015	0.534	0
Netscience	0.925	0.016	0.751	0.012	0.919	0.003
HepPh	0.5988	0.095	0.411	0.015	0.6569	0.005

综上所述，一方面，LPA_SI 算法在标签更新阶段考虑了节点的重要性对标签更新顺序的影响，使节点重要性较高的节点

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

C M Y K E:\发完的文件\15年6期文件\1171-1175黄佳鑫.PS 2015/05/14 09:03:46

影响节点重要性较低的节点,减少“逆流”现象的产生;另一方面,在标签选择过程中,LPA_SI 算法综合考虑了节点间紧密度和节点重要性对标签选择的影响,提高标签选择的准确性,从而在不改算法复杂度的同时提高了社区发现的质量和稳定性.

5 总 结

本文在分析了现有标签传播算法的特点和不足的基础上,提出在传播标签时需要综合考虑节点重要性和标签影响力,设计了新的节点重要性度量方法和新的标签影响力计算方法,将其融入节点标签的迭代传播.通过理论分析及在人工和真实数据集上的实验表明,提出的算法能够显著提高社区发现的准确性和稳定性,且仍然保持近似线性的时间与空间复杂度.

在接下来的工作中,将进一步考虑更多的节点重要性度量方法和标签影响力度量方法,比较不同方法对标签传播算法的影响,并考虑引入 MapReduce、MPI 等并行计算框架,实现算法的并行化,使算法具有更高的实用价值.

References:

[1] Lin You-fang, Wang Tian-yu, Tang Rui, et al. An effective model and algorithm for community detection in social networks[J]. Journal of Computer Research and Development, 2012, 49(2): 337-345.

[2] Liu Da-you, Jin Di, He Dong-xiao, et al. Community mining in complex networks[J]. Journal of Computer Research and Development, 2013, 50(10): 2140-2154.

[3] Shiga M, Takigawa I, Mamitsuka H. A spectral approach to clustering numerical vectors as nodes in a networks[J]. Pattern Recognition, 2011, 44(2): 236-251.

[4] Guimera R, Sales-Pardo M, Amaral L A N. Modularity from fluctuations in random graphs and complex networks[J]. Physical Review E, 2004, 70(2): 025101.

[5] Guimera R. Functional cartography of complex metabolic networks [J]. Nature, 2005, 433(7028): 895-900.

[6] Duch J, Arenas A. Community detection in complex networks using extremal optimization [J]. Physical Review E, 2005, 72 (2): 027104.

[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.

[8] Wu F, Huberman B A. Finding communities in linear time: a physics approach[J]. The European Physical Journal B-Condensed Matter, 2004, 38(2): 331-338.

[9] Raghavan U N, Albert R, Kumara S. Near linear-time algorithm to detect community structures in large-scale networks[J]. Physical Review E, 2007, 76(3): 036106.

[10] Steve G. Finding overlapping communities in networks by label propagation[J]. New Journal of Physics, 2010, 12(10): 103018.

[11] Barber M J, Clark J W. Detecting network communities by propagating labels under constraints[J]. Physical Review E, 2009, 80(2): 026129.

[12] Leung I X Y, Hui P, Liò P, et al. Towards real time community detection in large networks [J]. Physical Review E, 2009, 79 (6): 066107.

[13] Zhao Zhuo-xiang, Wang Yi-tong, Tian Jia-tang, et al. A novel algorithm for community discovery in social networks based on label propagation[J]. Journal of Computer Research and Development, 2011, 48(z2): 8-15.

[14] Yi Xiu-shuang, Han Ye-ting, Wang Xing-wei. Algorithm based on vertex influence for detecting local community structure[J]. Journal of Chinese Computer Systems, 2013, 34(9): 1975-1979.

[15] Ren Zhuo-ming, Shao Feng, Liu Jian-guo, et al. Node importance measurement based on the degree and clustering coefficient information[J]. Acta Physica Sinica, 2013, 62(12): 128901.

[16] Lancichinetti A, Fortunato S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities[J]. Physical Review E, 2009, 80(1): 1-8.

[17] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026i13.

[18] 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.

附中文参考文献:

[1] 林友芳,王天宇,唐 锐,等. 一种有效的社会网络社区发现模型和算法[J]. 计算机研究与发展,2012,49(2): 337-345.

[2] 刘大有,金 弟,何东晓,等. 复杂网络社区挖掘综述[J]. 计算机研究与发展,2013,50(10): 2140-2154.

[13] 赵卓翔,王轶彤,田家堂,等. 社会网络中基于标签传播的社区发现新算法[J]. 计算机研究与发展,2011,48(z2): 8-15.

[14] 易秀双,韩业挺,王兴伟. 一种基于节点影响力的局部社区发现算法[J]. 小型微型计算机系统,2013,34(9): 1975-1979.

[15] 任卓明,邵 凤,刘建国,等. 基于度与集聚系数的网络节点重要性度量方法研究[J]. 物理学报,2013,62(12): 128901.

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn