基于边标签传播的复杂网络社区识别方法

张健沛[1],邓 琨[1],杨 静[1],刘星妍[2]
(1.哈尔滨工程大学计算机科学与技术学院,黑龙江哈尔滨 150001;
2.黑龙江省电子信息产品监督检验院,黑龙江哈尔滨 150090)

摘 要: 针对传统基于标签传播的复杂网络重叠社区识别算法难以准确识别重叠节点的缺陷,本文通过分析边与其邻居边的关系,提出用来评估边归属社区的归属密度函数及归属倾向性函数,并在此基础上设计一种基于边标签传播的重叠社区识别方法(OLLP).该方法首先以每条边连接2个节点中度高的节点标签作为该边的标签;然后通过分析边的归属密度与归属倾向性迭代更新边标签,最终标签相同的边属于同一社区.在基准网络与真实网络数据集上进行测试,并与多个具有代表性的算法进行比较,实验结果表明了OLLP算法的有效性和可行性.

关键词: 复杂网络; 重叠社区识别; 标签传播
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2015)06-1113-06
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2015.06.012

Community Detection in Complex Networks Based on Link Label Propagation

ZHANG Jian-pei[1], DENG Kun[1], YANG Jing[1], LIU Xing-yan[2]
(1. *College of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang* 150001, *China*;
2. *Heilongjiang Province Electronic and Information Products Supervision Inspection Institute, Harbin, Heilongjiang* 150090, *China*)

Abstract: Since traditional overlapping community detection methods in complex networks based on label propagation that could not detect overlapping nodes accurately, this paper presented link attribution density and link attribution orientation functions through analyzing the relationship between each link and its neighbor links to assess the attribution community of each link. On this basis, overlapping community detection method based on link label propagation (OLLP) was designed. Firstly, OLLP used the label of every link to the node label which possesses the higher degree when connected by the link, and then updated the label repeatedly through analyzing attribution density and attribution orientation of the link. Finally, identical label links were attributed to the same community. By testing on both synthetic and real-world networks, and comparing with multiple representative algorithms, the experimental results verify the validity and feasibility of OLLP.

Key words: complex networks; overlapping community detection; label propagation

1 引言

复杂网络社区识别已成为当今研究的热点问题,它对于复杂网络的拓扑结构分析、功能分析和行为预测具有重要的理论意义和实际意义[1],其不仅吸引了大量的学者进行研究,而且被广泛应用于蛋白质功能预测[2]、舆情分析与控制[3,4]、搜索引擎[5]等众多领域当中.

传统的社区识别算法是将网络分解为若干个互不相连的社区,每个节点必须属于唯一社区.但在真实网络中,各社区之间并不完全独立,而是相互重叠的[6],例如在社会网络中,一个教授根据研究领域的不同,可同时属于多个不同社区(如数学、物理等).因此,复杂网络中重叠社区识别通常更具实际意义.

至今,许多重叠社区识别算法已被提出,其可以粗略的分为节点型社区识别与边社区识别两种[7].经研究表明[8],边通常表示一对节点之间的独特关系,仅存在单一角色且属于唯一社区,当边所属的社区确定后,相应的重叠节点将自然归属多个社区.因此,边社区具有识别重叠节点能力较强的优势,使其在识别重叠社区方面更具潜力.

本文利用边社区识别重叠节点能力强的优势,从边的局部视角出发,分析每条边与其邻居边的关系,提出

收稿日期:2014-03-31;修回日期:2014-05-21;责任编辑:马兰英
基金项目:国家自然科学基金(No.61370083, No.61073043, No.61073041, No.61402126);高等学校博士学科点专项科研基金(No.20112304110011, No.20122304110012);哈尔滨市科技创新人才研究专项资金(优秀学科带头人)(No.2011RFXXG015)

边归属密度函数及边归属倾向性函数,并在此基础上设计一种基于边标签传播的重叠社区识别方法 OLLP(The Overlapping Community Detection based on Link Label Propagation).

2 传统基于标签传播的社区识别算法

近年来,标签传播算法凭借其思想简单而高效的特点,已获得社区识别领域的广泛关注,并被扩展至重叠社区识别领域,例如:COPRA 算法[9]首先为每个节点初始化唯一的标签并设置相应的归属系数;然后在迭代过程中,每个节点的标签通过其邻居节点相同标签的归属系数相加与归一化的结果进行更新.为得到重叠社区,COPRA 设置参数 v 限制每个节点同时归属的社区个数.在 COPRA 基础上,Wu 等提出了 BMLPA 算法[10],该算法在迭代过程中,每个节点的标签是通过其邻居节点标签的归属比率进行更新的.其中归属比率大于某一参数阈值 p 的标签将被保留,最终拥有多个标签的节点为重叠节点.为进一步提高识别效率,Xie 等提出 SLPA 算法[11],此算法首先为每个节点初始化标签及标签存储空间;然后在迭代过程中,每个节点向邻居节点发出标签,同时从邻居节点随机接收一个标签存入标签存储空间;最终在每个标签存储空间中,相同标签所占比例大于参数 r 的标签将保存下来,此时,若存储空间中标签数量大于 1,其对应的节点被确定为重叠节点.

综上可知,传统基于标签传播的重叠社区识别算法,由于难以准确找到重叠节点,所以需要预先掌握网络中重叠节点归属特性而设置相关参数.但在纷繁复杂的网络中,掌握这些先验知识是极其困难的.因此,本文所提算法与其他标签传播算法的主要区别在于,运用边社区识别的思想,提出一种在无需掌握任何先验知识的情况下,能够较好完成重叠社区识别任务的边标签传播社区识别算法.

3 基于边标签传播的重叠社区识别算法

3.1 问题分析

依据文献[12]可知,社区识别的主要目标是识别出社区内部连接紧密,社区之间连接松散的社区结构.因此,在边社区识别算法中,分析边以何种方式结合,能够形成连接紧密的社区结构就显得尤为重要.本节将重点分析这一问题.给定网络 $G=(V,E)$,V 表示节点集合,E 表示边集合.

定义 1(邻居边) 边 $e_{v,v'}$ 的邻居边 $n(e_{v,v'})$ 是指分别与节点 v 和 v' 相连,且不包括 $e_{v,v'}$ 的边集合,即:$n(e_{v,v'})=\{e_{v,i}\in E \mid i\in n(v) \text{且} i\neq v'\}\bigcup\{e_{v',j}\in E \mid j\in n(v') \text{且} j\neq v\}$,其中,$n(v)$ 表示节点 v 的邻居节点集合.

定义 2(共同邻居边) 边 $e_{v,v'}$ 的共同邻居边 $s(e_{v,v'})$ 是指节点 v 和 v' 与共同邻居节点 u 相连的边集合,即:$s(e_{v,v'})=\{e_{v,u},e_{v',u}\in E \mid u\in n(v)\cap n(v')\}$.

定义 3(非共同邻居边) 边 $e_{v,v'}$ 的非共同邻居边 $d(e_{v,v'})$ 是指 $e_{v,v'}$ 的邻居边 $n(e_{v,v'})$ 去掉共同邻居边 $s(e_{v,v'})$ 的边集合,即:$d(e_{v,v'})=n(e_{v,v'})\setminus s(e_{v,v'})$.

定义 4(边标签) 设边 $e_{v,v'}$ 归属社区 C_i,社区 C_i 的标识为 $L(C_i)$,若 $L(e_{v,v'})=L(C_i)$,则称 $L(e_{v,v'})$ 为 $e_{v,v'}$ 的边社区标签,简称边标签.

定义 5(边社区紧密度函数)[8] 设社区 C_i 中,包含 m_{c_i} 条边及 n_{c_i} 个节点,则边社区紧密度函数定义为

$$D_{c_i}=\frac{m_{c_i}-(n_{c_i}-1)}{n_{c_i}(n_{c_i}-1)/2-(n_{c_i}-1)} \tag{1}$$

其中当 $n_{c_i}=2$ 时,设 $D_{c_i}=0$.

定理 1 设 C_v 是与边 $e_{v,v'}$ 的节点 v 相连的非共同邻居边所在社区,记 m_v,n_v 为 C_v 所含的边数及节点数.此时,若 $e_{v,v'}$ 加入 C_v 后,形成新社区为 $C_{v'}$,则 $C_{v'}$ 对应的社区紧密度 $D_{v'}$ 小于等于 C_v 对应的社区紧密度 D_v.

证明 社区 C_v 的社区紧密度 D_v 可表示为

$$D_v=\frac{m_v-(n_v-1)}{n_v(n_v-1)/2-(n_v-1)} \tag{2}$$

同理,社区 $C_{v'}$ 的社区紧密度 $D_{v'}$ 可表示为

$$D_{v'}=\frac{m_{v'}-(n_{v'}-1)}{n_{v'}(n_{v'}-1)/2-(n_{v'}-1)} \tag{3}$$

由于将 $e_{v,v'}$ 加入社区 C_v,显然有 $m_{v'}=m_v+1$,$n_{v'}=n_v+1$,此时 $D_{v'}$ 可表示为

$$D_{v'}=\frac{m_v+1-n_v}{(n_v+1)n_v/2-n_v} \tag{4}$$

其中,社区 $C_{v'}$ 对于社区 C_v 的社区紧密度增量 $\Delta D_{c_{v'}-c_v}$ 可表示为

$$\Delta D_{c_{v'}-c_v}=D_{v'}-D_v \tag{5}$$

若 C_v 为树形社区结构,则存在 $m_v=n_v-1$.此时,$\Delta D_{c_{v'}-c_v}=0$,故 $D_v=D_{v'}$.

若 C_v 较稠密时,则存在 $m_v>n_v-1$,同时 $n_v>2$.此时,$\Delta D_{c_{v'}-c_v}<0$,故 $D_{v'}<D_v$.证毕

由定理 1 可知,当一条边加入其非共同邻居边所在社区时,无法提高社区紧密度.

定理 2 设边 $e_{v,v'}$ 的共同邻居边在社区 C_u 中,记 m_u,n_u 为 C_u 所含的边数及节点数.此时,若 $e_{v,v'}$ 加入 C_u,将形成新社区 $C_{u'}$,则 $C_{u'}$ 对应的社区紧密度 $D_{u'}$ 大于 C_u 对应的社区紧密度 D_u.

证明 社区 $C_{u'}$ 对于社区 C_u 的社区紧密度增量 $\Delta D_{c_{u'}-c_u}$ 可表示为

$$\Delta D_{c_{u'}-c_u}=D_{u'}-D_u \tag{6}$$

$$\Delta D_{c_{u'}-c_u}=\frac{m_{u'}-(n_{u'}-1)}{n_{u'}(n_{u'}-1)/2-(n_{u'}-1)}-\frac{m_u-(n_u-1)}{n_u(n_u-1)/2-(n_u-1)} \tag{7}$$

不难发现,将边 $e_{v,v'}$ 加入社区 $C_{u'}$,则有 $m_{u'}=m_u+1, n_{u'}=n_u$.

显然,$\Delta D_{c_{u'}-c_u}>0$,故 $D_{u'}>D_u$.证毕

此定理表明,当边加入其共同邻居边所在社区时,可以有效提高社区紧密度.

基于定理 1 与定理 2,本文给出边归属密度函数和边归属倾向性函数.

3.2　边归属密度及边归属倾向性函数

依据上述分析可知,每条边与其共同邻居边归属同一社区会增加社区紧密度,因此,将每条边与其共同邻居边归属同一社区是合理的.然而,若一条边的共同邻居边分属于不同社区,将使该边的归属社区出现分歧,因此,最可行的做法是将每条边归属与其共同邻居边连接最紧密的社区.基于此,我们给出边归属密度函数以评价每条边与其不同社区的共同邻居边连接紧密度情况.

定义 6(边归属密度函数)　边 $e_{v,v'}$ 与社区 C_u 的边归属密度可定义为

$$B_{e_{v,v'}\to C_u}=\frac{\mathrm{count}(s(e_{v,v'})\cap C_u)}{\mathrm{count}(\mathrm{node}(s(e_{v,v'})\cap C_u))} \tag{8}$$

其中,$\mathrm{count}(s(e_{v,v'})\cap C_u)$ 表示边 $e_{v,v'}$ 的共同邻居边 $s(e_{v,v'})$ 属于社区 C_u 的边数,$\mathrm{node}(s(e_{v,v'})\cap C_u)$ 表示 $s(e_{v,v'})\cap C_u$ 所覆盖的节点.

式(8)中,若 $B_{e_{v,v'}\to C_u}$ 值越高,则节点 v 和 v' 与属于社区 C_u 的共同邻居节点集合存在更多的连接边,因此,节点 v 和 v' 与社区 C_u 连接越紧密,则边 $e_{v,v'}$ 与社区 C_u 连接越紧密.综上,将边 $e_{v,v'}$ 归属于 $B_{e_{v,v'}\to C_u}$ 值高的社区是可行的.

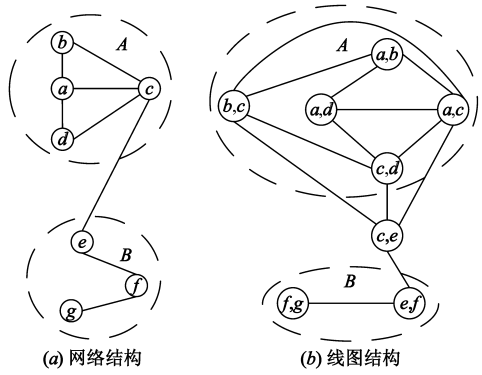
边归属密度函数针对每条边与其共同邻居边的连接紧密度进行了分析,但在真实网络中往往存在部分边没有共同邻居边,此时若仅对拥有共同邻居边的边进行处理,将使无共同邻居边的边独立成为社区,影响社区识别质量.鉴于此,给出边归属倾向性函数.

定义 7(边归属倾向性函数)　边 $e_{v,v'}$ 与社区 C_v 的边归属倾向性函数可定义为

$$T_{e_{v,v'}\to C_v}=\frac{\mathrm{count}(d(e_{v,v'})\cap C_v)}{\mathrm{count}(d(e_{v,v'}))} \tag{9}$$

其中,$d(e_{v,v'})$ 表示 $e_{v,v'}$ 的非共同邻居边.

由式(9)可知,边 $e_{v,v'}$ 的非共同邻居边 $d(e_{v,v'})$ 归属于社区 C_v 的边越多,使 $T_{e_{v,v'}\to C_v}$ 的值越高.事实上,一条边的非共同邻居边与某社区的交集数量,也反映出该边对属于某社区所具有的倾向性.从图 1 (图 1(b)是图 1(a)的网络拓扑结构对应的线图结构)可知,边 $e_{c,e}$ 与社区 A 的连接更加紧密,所以将边 $e_{c,e}$ 归属于社区 A 相对合理.此时,若采用边归属倾向性函数判断边 $e_{c,e}$ 的归属社区,则有边 $e_{c,e}$ 归属社区 A 的倾向性 $T_{e_{c,e}\to A}=0.75$,归属社区 B 的倾向性 $T_{e_{c,e}\to B}=0.25$,因此,边 $e_{c,e}$ 应归属社区 A.鉴于此,将边归属于边归属倾向性函数值高的社区是合理的.



(a) 网络结构　　(b) 线图结构

图1　网络拓扑结构与线图结构实例

综上,为有效提高社区连接紧密度,边 $e_{v,v'}$ 应属于边归属密度函数值高的社区,若边 $e_{v,v'}$ 无法通过边归属密度函数对归属社区做出判断,则边 $e_{v,v'}$ 应属于边归属倾向性函数值高的社区.本文在此基础上设计 OLLP 算法.

3.3　算法描述

在复杂网络中,度高的节点通常被称为"核心"节点,经研究表明[13],"核心"节点在局部范围内具有较强的影响力,会将其周围的节点吸收形成社区.因此,本文采用的初始化标签策略是将网络中每条边的标签初始为该边连接 2 个节点中度高的节点标签,以此形成具有树形结构的初始社区;此外,依据网络中连接节点度由高到低的顺序依次取边,并通过计算边归属密度函数值及边归属倾向性函数值,确定边归属社区并更新边标签,反复执行此操作,直到没有边标签继续更新为止,最终标签相同的边属于同一社区.基于此,本文提出 OLLP 算法,其详细描述如下.

算法 1　边标签传播算法(OLLP)

输入:复杂网络 $G=(V,E)$
输出:边社区结构 $C=\{C_1,C_2,\cdots,C_k\}$
　　//$L(e)$表示边 e 的标签
begin
(1)for each $e_{i,j}\in m$　// m 为 G 的边数
(2)　　$L(e_{i,j})\leftarrow$边 $e_{i,j}$连接 2 个节点中度高的节点标签
(3)end for
(4)labelList$\leftarrow G$ 中不重复边标签按包含边的数量降序排列的集合

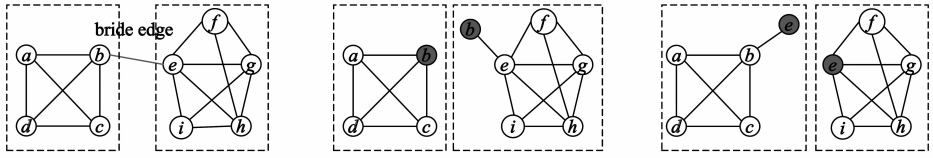
```
(5)while 无边标签需要更新
(6)   for each l∈labelList
(7)     dnEdges←从 E 中取出标签为 l 的边集合
(8)     for each e_{i,j}∈dnEdges
(9)       cList←边 e_{i,j}共同邻居边的标签集合
(10)      cGS←按 cList 中的标签,分别计算 e_{i,j}的边归属密度函数值
(11)      if length(cGS = = max(cGS)) = = 1
(12)        L(e_{i,j})←max(cGS)对应的标签
(13)      else
(14)        cnList←边 e_{i,j}非共同邻居边的标签集合
(15)        cnGS←按 cnList 中的标签,分别计算 e_{i,j}的边归属倾向性函数值
(16)        L(e_{i,j})←从 max(cnGS)对应的标签中随机取一个
(17)      end if
(18)    end for
(19)  end for
(20)end while
end
```

由于边社区识别是将边唯一归属一个社区,因此网络中若存在桥边现象将会影响社区的识别质量,例如图 2(a)中的桥边 $e_{b,e}$可能被识别为如图 2(b)或图 2(c)所示的社区结构.显然,这并不是一个好的社区结构.为此本文借鉴文献[7]提出的重叠节点微调方法解决这一问题.

定义 8(社区平均度) 设社区 C_v 中有 m_{C_v} 条边及 n_{C_v} 个节点,则社区平均度 Γ_{C_v} 可定义为

$$\Gamma_{C_v} = 2 * \frac{m_{C_v}}{n_{C_v}} \tag{10}$$

重叠节点微调方法是在社区结构中获得重叠节点集合 O,若 O 中节点 v 属于邻接社区 C_v,使 Γ_{C_v} 增大,则将节点 v 保留在 C_v 中,否则将 v 从 C_v 中删除.若节点 v 在所有邻接社区中,均无法提高社区平均度,则节点 v 仅保留在使社区平均度降低最小的社区中.



(a) 桥边实例 (b) 一种社区结构 (c) 另一种社区结构

图2 桥边实例

3.4 时间复杂度分析

设网络 G 中包含 n 个节点,m 条边,k 为节点的平均度,s 为网络 G 中"核心"节点数量.下面给出算法的时间复杂度分析.

显然,OLLP 算法的边标签初始化的时间复杂度为 $O(km)$;在网络 G 中取前 s 个"核心"节点的时间复杂度为 $O(sn\log(n))$;由于在边标签传播过程中,每条边的标签是依据邻居边的标签更新的,因此更新 1 条边标签的时间复杂度为 $O(k)$,更新全部边标签的时间复杂度为 $O(km)$,最终完成 t 次标签传播的时间复杂度为 $O(tkm)$.因此,OLLP 算法的时间复杂度为 $O(km + sn\log(n) + tkm)$.考虑到复杂网络普遍为稀疏网络,同时具有幂率分布现象,因此 s,k 为常数.所以,OLLP 算法的时间复杂度也可表示为 $O(m + n\log(n))$.

4 实验

OLLP 算法在基准网络和真实网络数据集上进行测试,并与 LC[8]、LFM[14]、COPRA[9]、CFINDER[6]、BMLPA[10] 和 SLPA[11]算法进行对比分析,以验证算法性能.

4.1 基准网络

由于 LFR benchmark[15]基准网络与真实网络的统计特性极其相似.因此,本文使用该基准网络对各算法进行测试.其共享参数设置为 $N = 200$,$k = 10$,$k_{\max} = 30$,$C_{\min} = 20$,$C_{\max} = 50$,其他参数设置如表 1 所示.

表 1 LFR benchmark 基准网络参数设置

网络	O_n	O_m	μ
$R1$	20	2	0.1 – 0.4
$R2$	100	2	0.1 – 0.4
$R3$	20	2 – 6	0.1
$R4$	20	2 – 6	0.3

在此,我们使用经典的精度评价指标,扩展统一化互信息 NMI(Normalized Mutual Information)[14]来评价各算法的性能.图 3 展示了各算法在 4 组($R1$ ~ $R4$)网络中的 NMI 比较结果.可以看出,OLLP 算法仅在少数网络中($R1$ 的 $\mu = 0.1$ 和 0.3,$R2$ 的 $\mu = 0.1$,$R3$ 的 $O_m = 4$,$R4$ 的 $O_m = 2$ 和 3)未取得最优值外,其在多数网络中所取得的 NMI 值均优于其他对比算法.此外,随着参数 μ 和 O_m 的增加,社区识别难度也逐渐增大,虽然 OLLP 算法整体呈现下降趋势,但波动幅度较小.因此也说明 OLLP 算法较为稳定.综上可知,在基准网络中,OLLP 算法能够识别出较高质量的重叠社区结构.

4.2 真实数据集

由于基准网络与真实网络的拓扑特性略有不同,因此本文通过真实网络数据集进一步测试 OLLP 算法的性能.

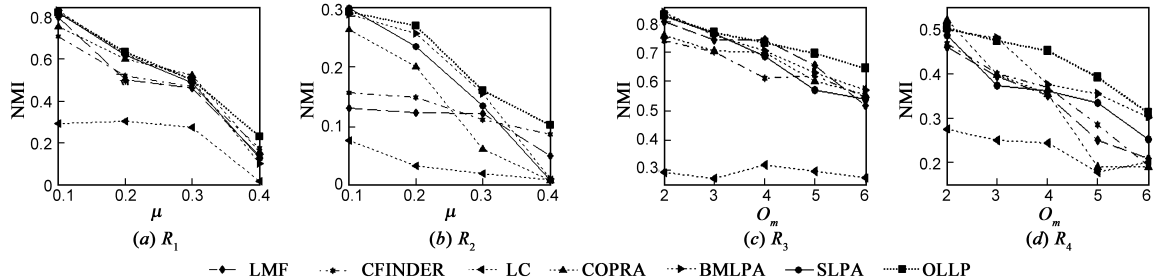


图3　各算法NMI比较结果

4.2.1　社区连接紧密度分析

本节通过扩展模块度 EQ(Extended Modularity)[16]来衡量各算法识别社区的连接紧密度.表 2 中列出了用于测试的 8 种真实网络数据集,其中包含从几十个节点的小规模网络到上万个节点的大规模网络.

表 2　真实网络数据集

网络	节点	边	描述
Karate	34	78	空手道网络[17]
Dolphins	62	159	海豚社会网络[17]
Lesmis	77	254	悲惨世界网络[17]
Polbooks	105	441	美国政治之书网络[17]
Email	1,133	5,451	电子邮件交往网络[18]
Netscience	1,588	2,742	作者合作网络[18]
Word	7,207	31,784	词汇语义网络[18]
Internet	22,963	48,436	互联网快照网络[17]

表 3 中展示了针对 8 个真实网络,执行 OLLP 算法得到的 EQ 值与其他算法进行比较的结果("\ \"表示算法识别社区失败).从中可知,OLLP 算法仅在面对 Dolphins、Lesmis 和 Email 时没有取得最优值.但在面对其他 5 个网络时,OLLP 算法取得的 EQ 值明显优于其他对比算法.因此验证了 OLLP 算法在真实网络上的社区识别能力同样优秀.

表 3　OLLP 与其他算法的 EQ 值比较

EQ	OLLP	CFINDER	LFM	COPRA	BMLPA	SLPA	LC
Karate	**0.3574**	0.1072	0.2146	0.3239	0.3478	0.3472	0.1220
Dolphins	0.4235	0.2885	0.2374	0.4206	**0.4355**	0.3879	0.1514
Lesmis	0.3236	0.1855	**0.4812**	0.4779	0.4733	0.3209	0.2484
Polbooks	**0.5124**	0.4304	0.3476	0.4586	0.5011	0.4568	0.2198
Email	0.2750	0.2641	0.1822	0.3523	**0.3583**	0.1837	0.0261
Netscience	**0.7716**	0.5905	0.2599	0.7186	0.7547	0.7126	0.7192
Word	**0.2464**	0.1649	0.1412	0.0007	0.0004	0.2032	0.0195
Internet	**0.2070**	\ \	0.1958	0.0345	0.1791	0.1161	0.0234

4.2.2　大规模网络分析

为考查各算法在大规模网络中的社区识别能力,在此选用 BrightKite (58,228 个节点,214,078 条边)[19]和 EUEmail (265,214 个节点,420,045 条边)[19]两个大型网络对各算法的社区识别能力进行评估.图 4 中展示了各算法针对以上两个网络识别社区规模的分布情况.由于 CFINDER 对以上两个网络以及 BMLPA 对 EUEmail 网络未能识别出社区结构,因此针对相应网络仅对其余算法进行了分析.从图中可知,在 OLLP 算法生成的社区结构中,规模在 3 ~ 1000 个节点的社区所占比率较高,而在真实世界网络中通常不具实际意义的社区(包含 1 ~ 2 节点)所占比率远低于其他算法.由此,从各算法识别社区规模的分布角度能够看出,OLLP 算法所识别的社区结构相对合理.

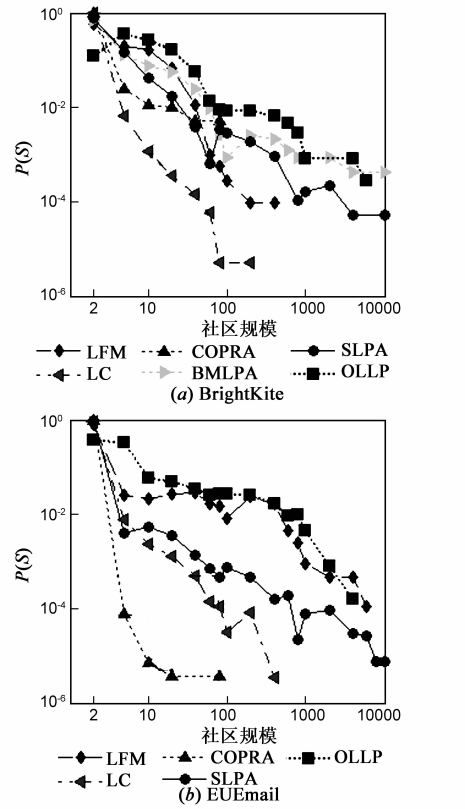


图4　各算法针对大型网络识别社区规模的分布情况

5　结论

本文通过分析每条边与其邻居边的连接紧密度,将邻居边区分为共同邻居边与非共同邻居边,提出用于评估每条边归属社区的边归属密度函数与边归属倾向性函数,最终在以上两个函数的基础上提出一种仅

考虑网络拓扑结构信息,无需进行外部参数设置的边标签传播社区识别方法(OLLP).该算法有效克服了传统标签传播方法在进行重叠社区识别时,存在重叠节点难以识别的缺陷.在基准网络和真实网络数据集上进行测试,并与多个经典算法进行对比分析,实验结果表明了 OLLP 算法是有效的、可行的.

参考文献

[1] 金弟,刘大有,杨博,等.基于局部探测的快速复杂网络聚类算法[J].电子学报,2011,11(39):2540 – 2546.
Jin Di, Liu Da-You, Yang Bo, et al. Complex network clustering algorithm using local detection[J]. Acta Electronica Sinica, 2011, 11(39): 2540 – 2546. (in Chinese)

[2] Wang Z, Zhang J. In search of the biological significance of modular structures in protein networks[J]. Plos Computational Biology, 2007, 3(6): e107.

[3] Cheng J, Cao J D, Zhang W Q, et al. Influential node control strategy for opinion evolution on social networks[J]. Abstract and Applied Analysis, 2013, 2013: 689495.

[4] Qian C, Cao J D, Lu J, et al. Adaptive bridge control strategy for opinion evolution on social networks[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2011, 21(2): 025116.

[5] Sidiropoulos A, Pallis G, Katsaros D, et al. Prefetching in content distribution networks via web communities identification and outsourcing[J]. World Wide Web, 2008, 11(1): 39 – 70.

[6] 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.

[7] Shi C, Cai Y N, Fu D, et al. link clustering based overlapping community detection algorithm[J]. Data & Knowledge Engineering, 2013, 87(9): 394 – 404.

[8] Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multi-scale complexity in networks[J]. Nature, 2010, 466(7307): 761 – 764.

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

[10] Wu Z H, Lin Y F, Steve G, et al. Balanced multi-label propagation for overlapping community detection in social networks [J]. Journal of Computer Science and Technology, 2012, 27(3): 468 – 479.

[11] Xie J R, Szymanski B K, Liu X. Slpa: uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process[A]. Proceedings of IEEE ICDM Workshop on DMCCI[C]. Vancouver: IEEE, 2011. 344 – 349.

[12] Newman M E J. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(23): 8577 – 8582.

[13] Bagrow J P, Bolt E M. Local method for detecting communities[J]. Physical Review E, 2005, 72(4): 046108.

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

[15] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4): 046110.

[16] Shen H W, Cheng X Q, Cai K, et al. Detect overlapping and hierarchical community structure in networks[J]. physical A: Statistical Mechanics and Its Applications, 2009, 388(8): 1706 – 1712.

[17] Newman M E J. Network Data[DB/OL]. http://www-personal.umich.edu/~mejn/netdata, 2014-02-01.

[18] Steve G. Network Research[DB/OL]. http://www.cs.bris.ac.uk/~steve/networks/copra, 2014-02-01.

[19] Leskovec J. SNAP[DB/OL]. http://snap.stanford.edu, 2014-05-10.

作者简介

张健沛 男,1956 年生于黑龙江哈尔滨.哈尔滨工程大学计算机科学与技术学院教授、博士生导师.主要研究方向为数据库与知识工程、数据挖掘、复杂网络分析等.
E-mail: zhangjianpei@hrbeu.edu.cn

邓 琨(通信作者) 男,1980 年生于黑龙江齐齐哈尔.哈尔滨工程大学计算机科学与技术学院博士研究生.主要研究方向为数据挖掘、复杂网络分析.
E-mail: dengkun@hrbeu.edu.cn