

一种基于邻域跟随关系的增量社区发现算法

陈羽中 施 松 朱伟平 於志勇 郭 昆

(福州大学数学与计算机科学学院 福州 350108)

(福建省网络计算与智能信息处理重点实验室(福州大学) 福州 350108)

摘 要 社区发现能够揭示真实社会网络的拓扑结构和动态特性. 目前的社区发现算法多针对静态社会网络所设计, 而绝大多数真实社会网络的社区结构是动态变化的. 针对动态社区发现, 现有算法通常基于社区结构平稳变化的假设, 无法处理演化过程中可能出现的大量社区消亡或涌现等突发事件. 为解决有效并高效地发现大规模动态社会网络的社区结构的问题, 提出了一种基于邻域跟随关系的社区表示模型 Follow-Community, 模型刻画的社区由不同角色的节点以及节点间的跟随关系组成, 通过发现节点间存在的直接或间接的跟随关系, 可将跟随同一个节点的节点所构成的集合归为一个社区. 基于该模型提出了一种具有接近线性时间复杂度的邻域跟随算法 NFA (Neighborhood Following Algorithm), 遍历网络节点一次即可得到静态社会网络的社区结构. 进一步扩展得到增量邻域跟随算法 iNFA (incremental Neighborhood Following Algorithm). 通过更新网络演化过程中相关节点的邻域跟随关系, iNFA 可发现动态社会网络的社区结构及社区演化. 实验结果验证了算法在大规模动态社会网络社区发现方面具有精度、效率以及稳定性的优势.

关键词 动态社会网络; 社区发现; 增量分析; 邻域跟随; 社区演化

中图法分类号 TP391

DOI号 10.11897/SP.J.1016.2017.00570

An Incremental Community Discovery Algorithm Based on Neighborhood Following Relationship

CHEN Yu-Zhong SHI Song ZHU Wei-Ping YU Zhi-Yong GUO Kun

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108)

(Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing (Fuzhou University), Fuzhou 350108)

Abstract Community discovery can help reveal the topological structures and the dynamic characteristics of the real social networks. However, most community discovery algorithms are designed for static social networks, while in most real social networks, the community structures of networks change dynamically. For community discovery in dynamic social networks, existing algorithms are based on the assumption that the community structures of dynamic networks evolve smoothly, and therefore cannot deal with the sudden events of emergence and extinction of communities during the evolution of dynamic social networks. To address the aforementioned issue of discovering community structures effectively and efficiently in large-scale dynamic social networks, this paper presents a novel community representation model called Follow-Community model, which represents the following relationships among neighbors. In the Follow-Community

收稿日期: 2016-01-18; 在线出版日期: 2016-05-08. 本课题得到国家自然科学基金(61300104, 61300103)、福建省教育厅科技重点项目(JK2012003)、福建省科技厅产学研重大项目(2014H6014)、福建省自然科学基金(2013J01230)、福建省科技创新平台项目(2014H2005)和福建省科技平台建设项目(2009J1007)资助. 陈羽中, 男, 1979年生, 博士, 教授, 中国计算机学会(CCF)会员, 主要研究方向为复杂网络、数据挖掘、计算智能. E-mail: yzchen@fzu.edu.cn. 施 松, 男, 1990年生, 硕士, 主要研究方向为复杂网络、数据挖掘. 朱伟平, 男, 1991年生, 硕士研究生, 主要研究方向为复杂网络、数据挖掘. 於志勇(通信作者), 男, 1982年生, 博士, 副教授, 中国计算机学会(CCF)会员, 主要研究方向为普适计算、移动社会网络. E-mail: yuzhiy@gmail.com. 郭 昆, 男, 1979年出生, 博士, 副教授, 中国计算机学会(CCF)会员, 主要研究方向为复杂网络、数据挖掘.

model, the community is represented by nodes with different roles and the following relationships among the nodes. By finding the direct or indirect following relationships among nodes, the collection of nodes that follow the same leader can be partitioned into one community. Based on the Follow-Community model, a Neighborhood Following Algorithm (NFA) with nearly linear time complexity is proposed to discover community structures in static social networks by just traversing the node set only once. Furthermore, an extended algorithm of NFA, named incremental Neighborhood Following Algorithm (iNFA), is also proposed. By updating the neighborhood following relationships of the relevant nodes involved in network evolution over time, iNFA can discover the community structures and community evolution in dynamic social networks. Finally, extensive experiments are conducted to validate the advantage of the proposed algorithms in accuracy, efficiency and stability for community discovery in large-scale dynamic social networks.

Keywords dynamic social networks; community discovery; incremental analysis; neighborhood following; community evolution

1 引言

社区是社会网络最重要的组成单元结构. 深入研究网络的社区结构有助于揭示真实网络的组织原则、拓扑结构和动力学特性, 具有十分重要的意义^[1]. 由于真实社会网络通常是动态变化的, 作为探究社会网络动态特性的前提, 动态社区发现问题是目前社会网络领域的重要研究方向之一.

早期的动态社区发现研究通常将动态网络看成时间轴上的网络快照序列, 利用静态社区发现算法得到各个孤立网络快照的社区结构, 再通过社区间相似度进行社区匹配实现网络演化过程中的社区跟踪^[2-7]. Leskovec 等人^[3]从网络的基本特征如平均最短路径等入手, 揭示了动态网络在演化过程中所呈现出的特性. Takaffoli 等人^[4]定义了社区新增、消亡、存活、分裂、合并等 5 种社区演化事件, 用于刻画网络的社区演化特性, 并对社区的演化进行跟踪. Backstrom 等人^[5]通过研究动态网络中社区的演化特征, 发现个体加入社区的行为与社区结构有着密切的关联. Palla 等人^[6]采用 CPM(Clique Percolation Method)算法发现网络各个快照的社区结构, 并分析不同时刻社区间的演化关系. Falkowski^[7]采用 GN(Girvan and Newman)算法得到快照的社区结构. 该类方法不但复杂度高, 而且容易导致相邻时刻网络快照的社区结构相差过大, 违背了社会网络在演化过程中大部分节点的社区归属的变化较平稳的特性, 不利于对社区进行演化分析.

针对早期研究存在的问题, 一些学者考虑将网

络的历史社区结构作为对当前时刻网络快照进行社区发现的先验知识, 并将社区发现问题转换为在社区结构平滑约束条件下的优化问题, 从而得到更为平滑的动态社区结构^[8-10]. FacetNet^[8] (a Framework for Analyzing Communities and EvoluTions in dynamic NETworks)算法通过 K-L 距离定义 Snapshot cost 和 Temporal cost 两个代价函数, 分别表示得到的社区结构与当前网络快照相符合的程度以及当前的社区结构和前一个时刻的社区结构的相似程度. FacetNet 的优化目标是最小化二者的加权和. Kawadia 和 Sreenivasan^[9]将当前网络社区的模块度作为优化目标, 通过自定义的 E 公式量化相邻时刻网络结构的距离, 最后利用 LPA 求解原优化问题的对偶问题. Chen 等人^[10]同样以模块度为优化目标, 以任意节点对在两个相邻时刻的社区归属的相似程度的变化量之和作为约束条件, 采用梯度下降方法求解原凸规划问题. 上述方法能够得到更为平滑的社区结构. 但存在复杂度较高的问题, 不适用于大规模动态网络的社区发现.

针对大规模动态网络的社区发现问题, 近几年出现了增量社区发现算法. 增量社区发现算法首先挖掘动态网络中发生变化的节点, 然后仅对这些节点及其局部区域内节点的社区归属进行更新^[11-16]. Nguyen 等人^[11]将相邻两个网络快照的变化分成 4 种情况: 新增节点、删除节点、新增边、删除边, 并以优化模块度为目标更新相关节点的社区归属. Xie 等人^[12]提出 LabelRankT 增量社区发现算法, 首先挖掘网络中的增量相关的节点集合, 利用基于标签传播的 LabelRank 算法更新集合中节点的标签. 上

述方法得到的相邻时刻的社区结构差别较大,易受网络演化过程中突发事件的影响。iLCD^[13]通过遍历新增的边集,判断新增社区或合并相似的社区,但未考虑节点或者边消失的情况,无法跟踪社区的演化事件。Coscia 等人^[14]提出 DEMON 算法,节点只需识别自己所在的可视局部社区(ego-community),而社区则是通过不断合并各个局部社区来得到。DEMON 仅需对网络节点做一次遍历,且能够不断处理网络新增节点,但是对社区合并参数较敏感,容易产生过多高度重复的社区。上述增量社区发现算法多基于社区结构变化较平稳的假设,能够获得较为平滑的动态社区结构,但是无法应对演化过程中可能出现的大量社区消亡或涌现等突发事件。

针对现有社区发现方法存在的问题,需要寻找一种高效的增量社区发现算法,既能够满足社会网络在演化过程中变化较平稳的特性,又能够及时发现演化过程中出现的突发事件。通过对真实社会网络的分析发现,在庞大而复杂的社会网络中,用户可能无法明确自己归属于哪个社区,但是用户了解其最有可能和某个好友处于同一圈子,本文称之为社区归属跟随关系。本文算法试图快速地从大规模社会网络中获取节点间的社区归属跟随关系链,并由某个领导者节点代表社区,其余节点则通过直接或间接地跟随社区领导者来构成社区。本文的主要工作如下:

(1) 提出一种基于邻域跟随关系的社区表示模型 Follow-Community Model,模型能够描述各个社区内部节点间的社区归属跟随关系。

(2) 提出邻域跟随算法(Neighborhood Following Algorithm, NFA), NFA 只需遍历一次网络节点即可得到基于 Follow-Community 模型的网络社区结构。算法运行效率优于 LPA (Label Propagation Algorithm)^[17]等具有接近线性时间复杂度的社区发现算法。

(3) 扩展 NFA 得到增量邻域跟随算法(Incremental Neighborhood Following Algorithm, iNFA), iNFA 能高效地发现动态网络中的社区结构。

2 问题描述

动态网络可表示为时间轴上各个时间片的网络快照的集合 $\{G_1, G_2, \dots, G_{T-1}, G_T\}$, 其中 $G_t = (V_t, E_t)$ 表示 t 时刻的网络快照。定义 t 时刻网络快照 G_t

的一个社区划分 $P_t = \{C_{t,1}, C_{t,2}, \dots, C_{t,k}\}$, 其中 $C_{t,i}$ 为 t 时刻网络社区集合中的第 i 个社区, k 为社区数目, 则动态网络社区发现问题定义如下。

定义 1. 动态网络社区发现问题. 给定 $t-1$ 时刻的网络快照 G_{t-1} 和对应的网络社区划分 P_{t-1} 以及 t 时刻的网络快照 G_t , 求 t 时刻的网络快照 G_t 的一个社区划分 P_t 。

增量社区发现算法需要获得相邻时刻发生变化的节点以及边, 定义 t 时刻增量相关的网络部分为 $\Delta G_t = (\Delta V_t, \Delta E_t)$, 其中 $\Delta V_t = (V_{t-1} - V_t) \cup (V_t - V_{t-1})$ 为 t 与 $t-1$ 时刻发生变化的节点集合, $\Delta E_t = (E_{t-1} - E_t) \cup (E_t - E_{t-1})$ 为 t 与 $t-1$ 时刻发生变化的边集合, 则 t 时刻的增量相关的节点集合定义为 $IV_t = \{i \mid \exists j, j \in V_t \wedge (i, j) \in \Delta E_t\}$, 则增量社区发现问题定义如下。

定义 2. 增量社区发现问题. 给定 $t-1$ 时刻的网络快照 G_{t-1} 和对应的网络社区划分 P_{t-1} , 以及 t 时刻的网络快照 G_t 与增量相关节点集合 IV_t , 求 t 时刻的网络快照 G_t 的一个社区划分 P_t 。

3 邻域跟随社区发现算法

3.1 邻域跟随社区表示模型

在传统的社区发现算法中, 通常将网络社区结构表示为节点与社区间的映射关系的集合, 如 LPA^[17] 用一个标签号作为社区 ID, 根据标签传播过程结束后节点所拥有的标签确定节点所属的社区; 基于局部优化拓展的 GCE 算法^[18] 则将被同一种子扩展的节点视为相同社区, 节点的社区归属同样依赖于节点与社区间直接的映射关系。

这些方法求解问题的出发点是, 社区通常表示为内部联系紧密的节点的集合, 社区发现问题归结为如何优化当前网络的整体社区质量。但在节点关系复杂的社会网络中, 节点(用户)本身可能无法知道潜在的社区结构, 而往往只能了解其邻域的拓扑结构, 因此可以将社区结构表示为由节点及节点间的跟随关系组成。通过发现节点间存在的直接或间接的跟随关系, 可将跟随同一个节点(领导者)的节点所构成的集合归纳为一个社区。下面给出社区表示模型的相关定义。

不失一般性, 假定所讨论的动态网络中的网络快照均为简单无向图。模型刻画的社区由不同角色的节点以及节点间的跟随关系组成。社区内

需要判断被跟随的节点是否有可能成为一个社区的领导者(LEADER). 因此定义节点领袖质量, 用于表示节点成为社区领导者的可能性, 如下所示.

定义 9. 节点领袖质量 (Leadership Quality, LQ). 用于衡量节点成为一个社区领导的可能性.

$$LQ_x = \frac{fc(x)}{fc(following(x))} \quad (7)$$

其中 LQ_x 表示节点 x 的领袖质量, 定义为节点 x 的跟随节点数与跟随目标的跟随节点数的比值, 反映了节点独立于其跟随目标的能力.

定义 10. 领导者(LEADER)判别规则:

$$isLeader(x) = \begin{cases} \text{true}, & LQ_x \geq r \\ \text{false}, & \text{else} \end{cases} \quad (8)$$

判定节点 x 是否为 LEADER 的规则是: 给定参数 $r(0 < r < 1)$, 当节点 x 为 FOLLOWER 时, 若 $LQ_x \geq r$, 则将节点 x 标记为 LEADER; 若 x 为 LEADER 但 $LQ_x < r$, 则重新标记为 FOLLOWER.

在得到所有节点的跟随关系之后, NFA 通过后处理过程, 将邻域跟随社区结构转换成社区. 该过程为每个节点寻找领导者, 然后将跟随同一领导者的节点以及领导者归为同一社区.

定义 11. 环链跟随策略. 网络中可能存在跟随环链, 即节点与自身存在一个间接跟随关系: $x \mapsto x$. 如节点 A, B, C 构成一个跟随环链: $A \rightarrow B \rightarrow C \rightarrow A$. 若环链上存在领导者节点, 则环链上的节点与该领导者属于同一个社区; 否则随机指定一个节点作为领导者, 确保算法能将所有节点划分到相应社区.

需要注意的是, 上述模型和算法基于简单无向图的假设, 但是根据模型中直接邻域跟随、间接邻域跟随的定义可发现, 有向图中节点边的有向性对上述定义没有影响. 对于二分图网络, 根据算法中的最信赖邻域跟随策略, 节点是从其邻域中按照该策略选择某个节点作为自身的社区跟随目标, 从而构成节点间的社区关系的, 所产生的包含多个节点的社区同样是二分图, 而对于社区发现过程中可能出现的孤立节点, 可以在后处理中通过孤立节点修复机制进行社区归并, 因此论文中的模型和算法经过简单扩展后即可用于有向图、二分图等特殊网络.

3.3 邻域跟随算法描述

NFA 主要包含两个步骤: 首先一次遍历网络的所有节点, 得到跟随关系和各节点的角色; 之后通过后处理将直接或间接跟随同一个 LEADER 的节点划分到同一个社区. 具体算法实现如算法 1 所示.

算法 1. NFA(邻域跟随社区发现算法).

输入: 顶点集合 V , 参数 r

输出: 社区划分结果 C

// 遍历一次网络中的所有节点

FOR $x \in V$

{

$Follow(x)$; // 获得节点 x 的跟随关系

} // END FOR

// 后处理, 根据跟随关系得到网络社区, 将直接或间接

// 跟随同一个 LEADER 的节点划分到同一个社区

$C \leftarrow PostProc()$;

return C ;

过程 1. $Follow$.

输入: 顶点 x , 参数 r

输出: 节点 x 的跟随关系

// 获取节点 x 邻居中跟随者最多的节点

$k \leftarrow \arg \max_{j \in N(x) \cup x} fc(j)$;

$following(x) \leftarrow k$; // 设置节点 x 的跟随者

$fc(k)++$; // 更新节点 k 的跟随者数目

// $status(x)$ 表示节点 x 的角色

IF ($status(x) == \text{NULL}$) { // 节点 x 的角色未确定,

 // 标志节点为跟随者角色

$status(x) \leftarrow \text{FOLLOWER}$;

} // END IF

ELSE IF ($status(k) == \text{FOLLOWER} \parallel status(k) == \text{LEADER}$) {

 // 试探节点 k 是否可以成为一个 leader

$leader \leftarrow following(k)$;

 IF ($fc(k)/fc(leader) \geq r$)

$status(k) = \text{LEADER}$;

 ELSE

$status(k) = \text{FOLLOWER}$;

 } // END ELSE-IF

过程 2. $PostProc$.

输入: 节点角色 $status$, 节点跟随信息 $following$

输出: 社区集合 C

$C \leftarrow \{\}$ // 设置 C 为空集合

FOR $x \in V$

{

 // 寻找跟随的社区领导

$cur \leftarrow x$;

 WHILE ($status(cur) == \text{FOLLOWER}$)

 {

$cur \leftarrow following(cur)$;

 } IF ($cur == x$) {

$status(cur) = \text{LEADER}$; // 避免产生跟随环路

 break;

 } // END IF

} // END WHILE

```

tmp ← C.get(cur);
tmp ← tmp ∪ {x};
//将节点添加到跟随同一领导的集合
C.put(cur, tmp);
} //END FOR
return C;

```

3.4 增量邻域跟随算法

网络演化过程本质上是节点和边的新增和删除,节点和边的变化引起网络内部分社区内聚程度和社区间的耦合程度加强或减弱,从而引起社区结构的变化.从 Follow-Community 模型的角度出发,网络的演化过程表现为部分节点的跟随目标发生变化,最终导致社区领导者发生变化.在这个过程中,邻域不变的节点占多数,这些节点在网络演化过程中无需更改其跟随关系,它们的社区归属的变化通过其与新的领导者之间的间接跟随关系得到更新.因此增量邻域跟随过程中只需保存邻域不变的节点的跟随者、跟随数目以及节点角色,并更新增量相关节点的跟随情况,从而达到增量社区发现的目的.

基于上述思想,扩展 NFA 算法得到增量社区发现算法 iNFA,包含 3 个步骤:首先用上一时刻的网络节点的跟随信息初始化新到网络节点的跟随信息;之后采用最可信赖邻域跟随策略更新增量相关节点的跟随信息;最后通过后处理过程将得到的 Follow-Community 结构转换成社区结构.具体实现如算法 2 所示,其中 $status_{t-1}$ 和 fc_{t-1} 分别存储各个节点在上一个时刻的角色和跟随者数量.

算法 2. iNFA(增量邻域跟随社区发现算法).

输入: $following_{t-1}, status_{t-1}, fc_{t-1}$

输出: t 时刻的社区划分结果 C

```

//用  $t-1$  时刻的节点跟随信息,初始化  $t$  时刻的网络
initFollow(); //节点的跟随对象、跟随数目、节点角色
//遍历增量相关节点
FOR  $x \in IV_t$ 
{

```

```

    iFollow(x); //更新  $t$  时刻的节点跟随关系

```

```

} //END FOR

```

```

//后处理,根据跟随关系得到  $t$  时刻的社区划分

```

```

 $C \leftarrow postProc()$ ;

```

```

return C;

```

过程 3. $iFollow$.

输入: 节点 x , 参数 r

输出: 更新后的 t 时刻的节点 x 的跟随关系

```

//寻找节点  $x$  邻域中跟随节点数最多的节点

```

```

 $k \leftarrow \arg \max_{j \in N(x) \cup x} fc(j)$ ;

```

```

oldLeader ← following(x);

```

```

IF (oldLeader ≠ NULL) {

```

```

    fc(oldLeader) --;
} //END IF
following(x) ← k; //设置节点  $x$  的跟随者
fc(k) ++; //更新节点  $k$  的跟随者数目
IF (status(x) == NULL) { //节点  $x$  的角色未确定
    //标志节点为跟随者角色
    status(x) ← FOLLOWER;
}
ELSE IF (status(k) == FOLLOWER || status(k) == LEADER) {
    //试探节点  $k$  是否可以成为一个 leader
    leader ← following(k);
    IF (fc(k) / fc(leader) ≥ r)
        status(k) = LEADER;
    ELSE
        status(k) = FOLLOWER;
} //END ELSE-IF

```

3.5 算法复杂度

首先分析 NFA 的时间复杂度:

步骤 1. 每个节点在选择跟随目标时需要访问其邻域一次,所需时间为 $O(k)$,其中 k 为网络平均度大小 m/n , n, m 分别为网络的节点数和边数.因此,该步骤的时间复杂度为 $O(m)$.

步骤 2. 后处理阶段,假设每个节点到跟随的领导者间的平均最短路径长度为 f ,由于 f 是一个很小的整数,则时间复杂度为 $O(fn)$.

综上,NFA 的时间复杂度为 $O(m + fn)$,和 LPA 相近,但由于 NFA 无需额外的迭代,实际运行时间大大缩短.

考虑 NFA 的空间复杂度,NFA 在计算节点间的邻域跟随关系的过程中,需记录每个节点的跟随者数量、角色、跟随目标的信息,在根据邻域跟随关系得到网络社区的后处理过程中,则需要存储每个社区的节点集合信息,所需的内存为 $O(n)$;此外存储网络所需的内存为 $O(m)$,因此 NFA 的空间复杂度为 $O(m + n)$.

iNFA 的时间复杂度如下步骤:

(1) 初始化 t 时刻的网络快照 G_t 的节点跟随信息需遍历一次 G_t 的所有节点,所需时间为 $O(n)$.

(2) 假设增量相关节点集合为 IV_t ,集合大小为 $|IV_t|$,更新一个节点的跟随目标所需时间为 $O(k)$,更新所有增量相关节点的跟随目标需要 $O(m|IV_t|/n)$,通常 $|IV_t| \ll n$.

(3) iNFA 和 NFA 的后处理过程相同,需要耗时 $O(fn)$.

综上,当网络较为平稳时,iNFA 主要耗时操作

为初始化节点跟随信息;当网络变化情况较大时,iNFA 主要的耗时操作为更新节点的跟随目标,两种操作的时间复杂度均小于 $O(m)$,因此 iNFA 在最坏情况下的时间复杂度为 $O(m+fn)$.

考虑 iNFA 的空间复杂度,iNFA 在增量社区发现过程中需要相邻时刻网络快照以及网络中节点的跟随信息,运行所需的内存为 $O(n)$;此外存储社会网络所需的内存为 $O(m)$,因此 iNFA 算法的空间复杂度为 $O(m+n)$.

4 实 验

实验首先采用静态的人工网络与真实网络,并与其他典型的静态社区发现算法进行对比,验证 NFA 的社区发现效果以及运行效率;然后基于真实动态网络,分析 iNFA 挖掘动态社区结构的效果,实验的硬件环境是:3.00 GHz 的 Pentium(R) Quad-Core CPU,16 GB 内存,Windows 7 操作系统,算法采用 Java 语言实现.

对于人工网络数据集,采用普遍使用的规范化互信息(Normalized Mutual Information,NMI)^[19]作为社区划分结果的评价指标,定义如下:

$$NMI(A,B)=\frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B}C_{ij}\log\left(\frac{C_{ij}N}{C_iC_j}\right)}{\sum_{i=1}^{C_A}C_i\log\left(\frac{C_i}{N}\right)+\sum_{j=1}^{C_B}C_j\log\left(\frac{C_j}{N}\right)} \quad (9)$$

其中: A 是网络的真实社区结构; B 是算法得到的社区划分结果; C_A 是真实划分中社区的数量; C_B 是算法得到的社区数量; C 是混淆矩阵,矩阵 C 的元素 C_{ij} 表示属于 A 的第 i 个社区的节点出现在 B 的第 j 个社区中的数量, C_i 表示矩阵 C 中第 i 行元素的和, C_j 表示矩阵 C 中第 j 列元素的和. NMI 比较的是真实社区结构和算法发现的社区结构的相似程度, NMI 的取值范围为 $[0,1]$,如果算法获得的社区划分和真实划分是一致的,则 $NMI(A,B)=1$,如果 A 和 B 完全不同,则 $NMI(A,B)=0$. 因此 NMI 值越大,意味着算法的社区划分与网络的真实社区结构相吻合的程度越高,社区划分效果越好.

对于真实网络,通常无法得到网络的真实社区结构,采用普遍使用的模块度(Modularity)指标^[20]作为社区划分结果的评价指标,定义如下:

$$Q=\frac{1}{2m}\sum_{i,j\in V}\left[A_{ij}-\frac{k_ik_j}{2m}\right]\delta(c_i,c_j) \quad (10)$$

其中: m 是网络的边数; A_{ij} 表示网络的邻接矩阵,如

果节点 i 和 j 有边相连则 $A_{ij}=1$,否则 $A_{ij}=0$. k_i 与 k_j 分别表示节点 i 和 j 的度,对于给定的一种社区划分: c_i 表示节点 i 所属的社区; c_j 表示节点 j 所属的社区,节点 i 和节点 j 属于同一个社区时 $\delta(c_i,c_j)=1$,否则为 0. 模块度的实质是比较处于同一社区的边数与随机网络的情况,随机网络不具备社区结构,网络的一个社区内部的连边数量与其在相应随机网络中的连边数量之差越大,说明算法获得的社区结构越明显,社区划分结果越好,因此,一般认为模块度值越大,获得的社区划分的质量越好,其取值范围为 $[0,1]$.

4.1 NFA 算法分析

4.1.1 LFR 网络

LFR 网络是 Lancichinetti 等人^[21]提出的基准网络,通过设定不同的参数可得到不同特性的网络,其网络生成参数如表 1 所示:其中 $minc$ 和 $maxc$ 分别为一个社区所包含的节点数的最小值与最大值, $maxk$ 为节点度的最大值, mu 为混合参数,表示节点与社区外部节点链接的概率, mu 越大,社区发现的难度越大, τ_1 为节点度的幂率分布指数,取值为 2, τ_2 为社区规模的幂率分布指数,取值为 1. LFR 网络在生成过程中,每个节点的度与每个社区的规模分别根据指数为 τ_1 和 τ_2 的幂率分布获得.

表 1 LFR 网络参数

LFR 网络参数	参数说明,默认值
N	网络节点数,5000~100 000
$minc$	最小社区节点数,10
$maxc$	最大社区节点数,50
$\langle k \rangle$	平均节点度,20
$maxk$	最大节点度,50
mu	混合参数,取值 0~1 范围
τ_1	节点度的幂律分布指数,取 2
τ_2	社区规模的幂律分布指数,取 1

本节基于 LFR 网络对比分析各个算法的运行效率,并对参数 r 进行讨论,所选取的对比算法如表 2 所示,均为较典型的具有接近线性时间复杂度的社区发现算法.

表 2 静态社区发现算法

算法	时间复杂度	备注
NFA	$O(m+fn)$	f : 节点到跟随的领导者间的平均最短路径长度
LPA ^[17]	$O(Tm+n)$	T : 标签迭代次数.
COPRA ^[22]	$O(Tvm\log(vm/n))$	v : 节点携带标签的最大数量,取 1
Demon ^[14]	$O(nK^{3-\alpha})$	K : 网络最大度数. 网络度. 大小分布满足 $p_k=k^{-\alpha}$
Infomap ^[23]	$O(m)$	—
LFM ^[24]	$O(n_c s^2)$	n_c : 社区数目. s : 社区平均大小.
GCE ^[18]	$O(mh)$	h : 为初始选取的种子数量.

为了验证 NFA 的运行效率,通过调整生成参数 N (范围为 5000~100 000)得到不同规模的网络,运行时间结果如图 2 所示.虽然各算法的运行耗时和网络的边数大致都呈线性关系,但当网络边数达到 500 000 时,NFA 的运行时间仍不超过 1 s,优于其他算法.

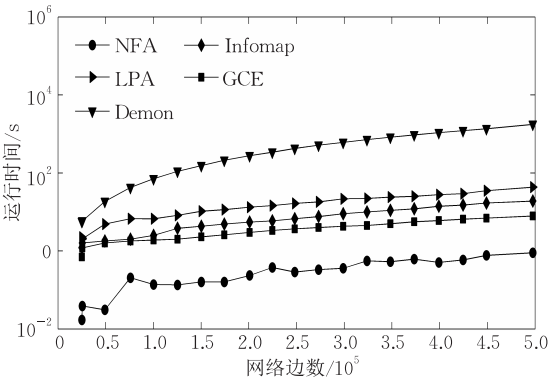


图 2 算法运行效率对比

NFA 需要根据阈值参数 r 判断节点是否能够成为领导者. r 值越大节点越不容易成为一个领导者,容易将一些规模较小的社区合并到规模更大的社区.为了讨论参数 r 对算法的影响,实验采用 3 个不同度幂律分布的 LFR 网络,其他参数设置为表 1 的默认参数,实验中 r 的取值分别为 0.1、0.5、0.9.

实验结果如图 3 所示,可以发现得到的社区的大小分布和真实分布接近,且 r 取值越高,规模较大的社区也越多.此外发现社区发现结果对参数 r 不敏感,如图 4 所示,对于 3 个服从不同的度幂律分布指数 τ_1 的 LFR 网络, r 取值的变化对 NMI 的影响不大,在不同 r 取值情况下所得到的 NMI 指标均在 0.9 以上,说明算法对参数 r 不敏感,一般情况下, r 取中间值 0.5 即可以达到较好的社区发现结果;当度幂律分布指数 $\tau_1=1$ 时,网络的节点度差异较大,NFA 能够取得更高的 NMI ,说明 NFA 能够

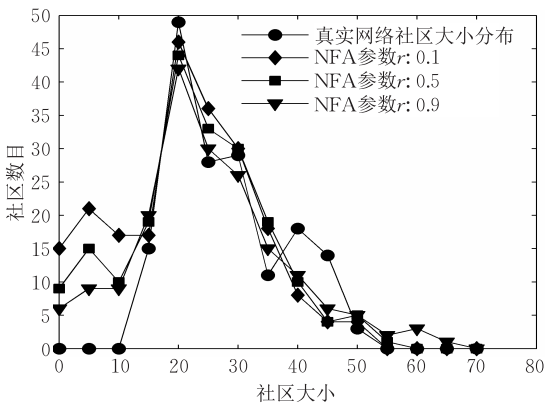


图 3 参数 r 对社区大小分布的影响

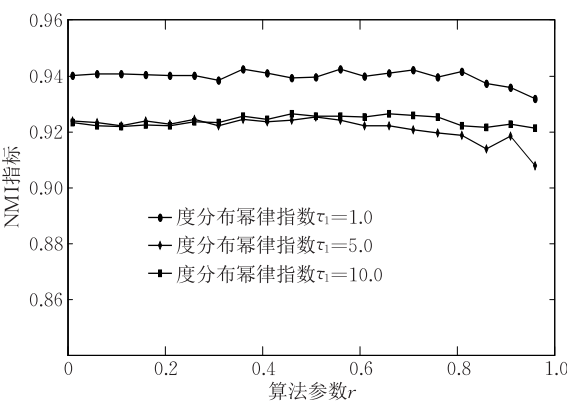


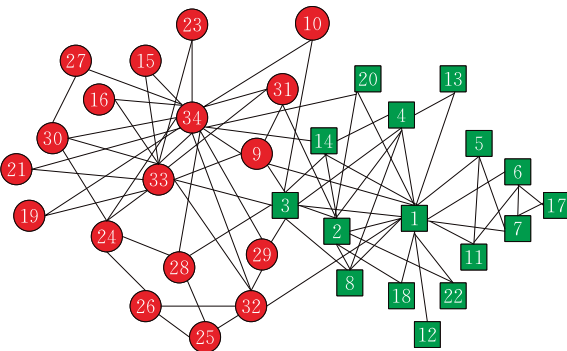
图 4 参数 r 对不同度幂律分布网络的影响

很好地适应度分布差异明显的网络.

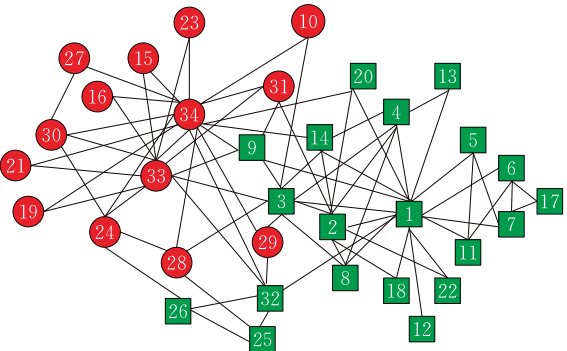
4.1.2 真实静态网络

为了验证 NFA 在真实网络中的效果,实验基于 14 个真实网络数据集对比各算法得到的模块度 Q .对比算法包括 LPA、COPRA 和 Infomap.

为分析 NFA 社区发现结果的准确性,首先以社区结构已知的空手道俱乐部网络(karate 网络)为例,对 NFA 所发现的社区结构进行分析.图 5(a)显示了空手道俱乐部网络的真实社区结构,因俱乐部的管理者(节点 34)和教练(节点 1)发生争执,俱乐部分成了以节点 34 和节点 1 为核心的两大社区.图 5(b)为 NFA 算法得到的社区发现结果,NFA 也



(a) karate网络真实社区结构



(b) NFA所发现的karate网络社区结构

图 5 karate 网络社区结构

将该网络划分成以节点 1 和节点 34 为核心的两个社区,仅有节点 9,25,26,32 的社区归属和实际情况不同;节点 9 选择了节点 1 作为跟随目标,而节点 25 和节点 26 跟随节点 32,从而间接地跟随节点 1. 原因在于节点 1 的跟随者数量大于节点 34 的跟随者数量,按照最可信赖邻域跟随原则,节点 9 和节点 32 选择了节点 1 作为跟随目标;而节点 25 和节点 26 随机选择一个节点(节点 32)作为跟随目标,从而间接地跟随节点 1. 比较图 5(a)与图 5(b)可以发现,NFA 所得到的社区划分结果中仅有节点 9,25,

26,32 的社区归属和真实情况不同.

14 个真实网络的实验结果如表 3 所示. 从表 3 可以发现,NFA 在 karate,jazz,net sciences 等 8 个网络中获得了最佳的模块度值,在 protein-protein、cond-mat-2003、PGP、soc-Pokec 网络中获得的模块度值也十分接近最佳值,但是在 football,email 网络中获得的模块度较低,这是由于一些网络对于节点的初始跟随目标的选择较敏感,错误选择的初始跟随目标节点会造成部分社区被划分为多个小社区,造成模块度的下降.

表 3 各算法在真实网络中的社区发现结果对比

网络	节点数	边数	模块度指标 Q				运行时间/s			
			Infomap	LPA	COPRA	NFA	Infomap	LPA	COPRA	NFA
karate ^[25]	34	78	0.398	0.296	0.371	0.404	0.00	0.00	0.00	< 0.01
football ^[26]	212	1226	0.600	0.589	0.579	0.490	0.00	0.01	0.01	< 0.01
jazz ^[27]	198	2742	0.442	0.440	0.442	0.442	0.10	0.19	0.25	< 0.01
email ^[28]	1133	5451	0.532	0.557	0.464	0.282	0.76	0.22	0.41	< 0.01
netscience ^[29]	2555	5484	0.908	0.901	0.904	0.908	1.02	0.11	0.15	< 0.01
protein-protein ^[30]	2445	6265	0.669	0.504	0.661	0.670	0.89	0.30	0.45	< 0.01
blogs ^[31]	3982	6803	0.802	0.756	0.798	0.804	1.23	0.61	0.60	0.01
DBLP ^[32]	2386	18930	0.265	0.544	0.553	0.560	1.65	0.92	0.93	0.02
PGP ^[33]	10680	24361	0.081	0.754	0.800	0.775	2.56	2.55	2.99	0.02
word_association ^[30]	7205	31784	0.450	0.365	0.447	0.456	2.69	2.46	3.48	0.02
blogs2 ^[34]	30557	82301	0.517	0.601	0.592	0.492	13.89	12.10	13.20	0.04
cond-mat-2003 ^[35]	27519	116181	0.662	0.616	0.652	0.622	10.56	11.44	14.45	0.03
Amazon ^[36]	510235	3356824	0.612	0.580	0.599	0.623	26.89	31.33	42.38	1.36
soc-Pokec ^①	1632803	30622564	0.282	0.421	0.451	0.429	2436.21	2420.12	2442.87	11.84

此外,NFA 的重要优势在于运行效率,NFA 与 LPA、Infomap 等算法具有相近的时间复杂度,但由于 NFA 无需额外迭代,实际运行时间大大缩短. 如 NFA 在边数少于 6803 的网络上的运行时间小于 0.01 s,随着网络规模的扩大时,NFA 的运行时间平稳增长,在节点规模为 10^7 的 soc-Pokec 网络中,NFA 的运行时间也只有 11.84 s,在运行速度上具有明显的优势.

综合表 3 和图 5 的实验结果可发现,NFA 可以取得较优的模块度,所发现的社区结构与真实社区结构相近,此外在运行速度方面较其他对比算法具有明显优势.

4.2 iNFA 算法分析

4.2.1 DBLP 合著者网络

该数据集包含了 43 个国际会议和期刊在 1990~2010 年间的论文合著关系网络^[25],包含 9 个网络快照,每个快照中的合著关系的时间跨度为 5 年,只选取至少在相应的时间跨度内发表超过 10 篇论文的作者,共得到 2723 位作者. 图 6 显示了该网络的演化情况,可以发现,网络规模平缓增大,在第 7 个时

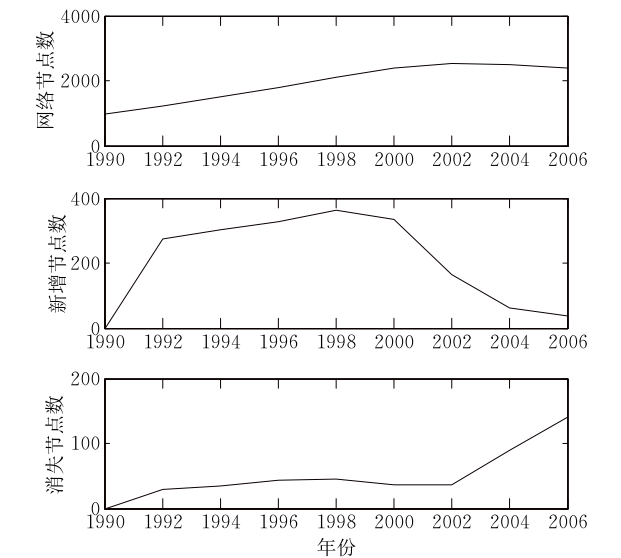


图 6 DBLP 合著者网络的变化情况

间片(2002 年)后消失的节点数明显增多.

为验证 iNFA 的社区发现效果,选择 FacetNet 和 LabelRankT 进行对比. FacetNet 需要社区数目

① <http://snap.stanford.edu/data/soc-pokec.html>

作为算法的先验参数, 设置为 iNFA 发现的平均社区数目, 历史权重参数 λ 设置为 0.3. LabelRankT 采用默认参数值. 对比结果如图 7 所示: 在前 8 个时间片, 8 个算法得到的模块度较接近; 在第 9 个时间片, 大量节点消失导致多个社区的内聚性减弱, iNFA 和 FacetNet 取得的模块度较 LabelRankT 更为平滑. 图 8 为各算法得到的社区数, 可以发现 LabelRankT 得到的社区数出现骤变. 这是由于增量更新节点标签时, LabelRankT 无法为部分非增量相关节点重新分配标签而产生了碎片社区, 同时分配新标签的增量相关节点大多被其他社区归并, 而一些较小规模社区的所有节点都属于增量相关节点集合, 导致社区被归并, 造成模块度下降. FacetNet 由于采用固定社区数, 难以适应网络的演化特性, 取得的模块度低于 iNFA.

iNFA 得到的邻域跟随社区社区结构, 还可以

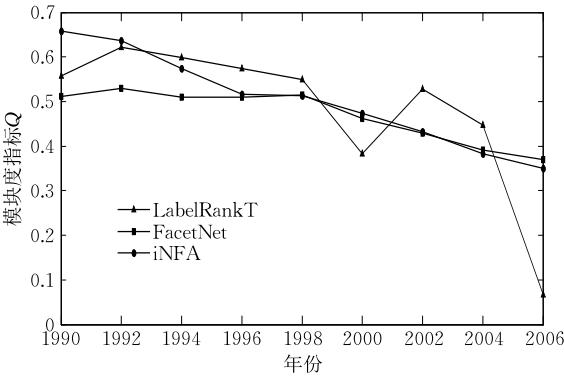


图 7 增量社区发现算法对比

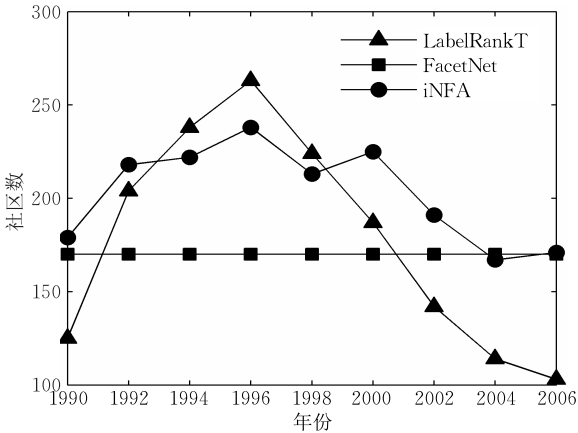


图 8 社区数目变化情况

清晰地刻画节点在社区中的重要性. 我们将利用跟随关系分析 DBLP 网络中数据挖掘方向研究群体的演变过程. 近年来, 以 Han 为代表的数据挖掘领域研究群体对该领域做出了重要的贡献. 图 9 刻画了在该著网络在演化过程中, 以 Han 为代表的研究群体的演变过程. 由图 9 可知, 在第 1 个时间片 (1990~1994), 出现以 Han 为“领导者”的社区, 且社区规模较小. 在其后几个时间片社区不断扩张. 到了第 8 个时间片 (2004~2008) 以 Yu 为首的小团体开始壮大, 但还只是 Han 为代表的社区的一部分. 将近 2010 年, Yu 势头更加旺盛, 成为数据挖掘学术界代表人物之一, 在图中表示为 Han 以 Yu 为社区跟随目标节点. 这与近年的科研现状相符, 说明 Following-Community 模型可以揭示节点对某一社区的重要程度, 具有现实意义.

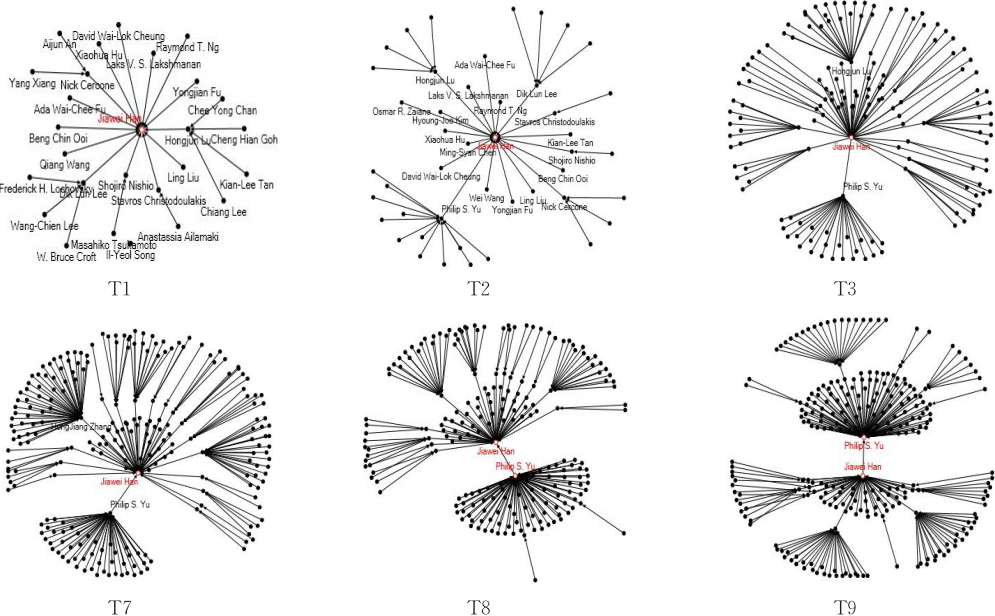


图 9 数据挖掘研究领域研究团体演化

4. 2. 2 AS-Internet Routers Graph

该网络来自互联网上边界网关日志文件^[3],记录了从1997年11月到2000年2月共733天的服务器间的通信网络快照. 各个网络快照中节点数目最多为6477(边数为13233). 网络的节点和边会随时间的推移新增或删除. 网络较为稀疏,存在较多边界节点,节点的平均度较小,社区结构不明显.

实验采用文献[4]提出的基于社区事件分析方

法来分析该网络的动态特性;其中 Form 事件表示社区在当前时刻之前没有出现过;Dissolve 事件表示社区在后续时刻中不再出现;Survive 事件表示社区在上一个时刻存在;Split 事件表示社区在当前时刻分裂成其他多个社区;Merge 事件表示社区由上一时刻的多个社区合并而来. 由图 10 可知网络在第420,640,710时刻附近出现突变情况,有大量的旧社区消失,之后又新增多个社区.

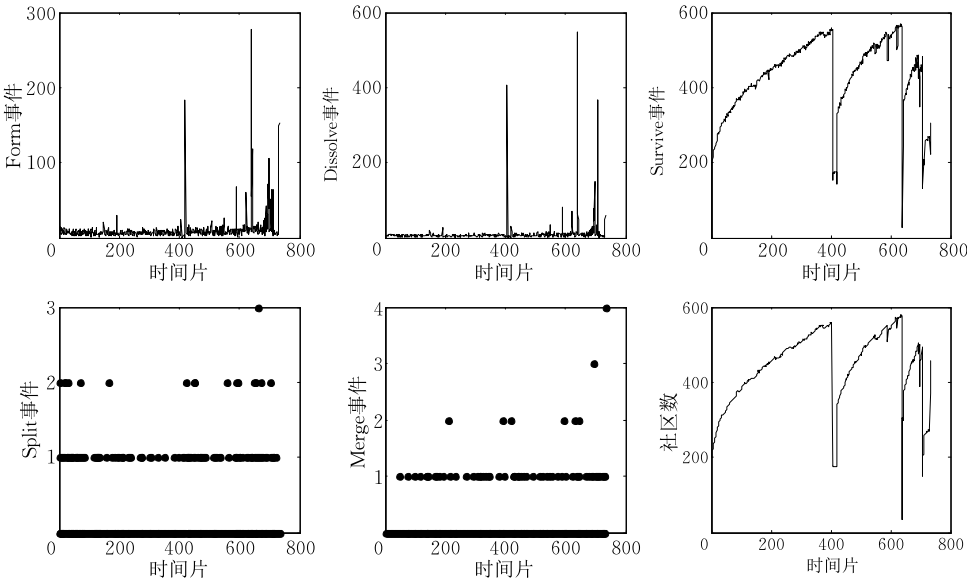


图 10 社区演化事件

AS-Internet 网络具有时间跨度长和网络不稳定的特点,有利于验证动态社区发现算法在周期较长的动态网络中的社区发现能力. 基于该网络,实验对比了 iNFA、FacetNet 和 LabelRankT 三个动态社区发现算法. 其中 iNFA 设定参数 r 为 0.1. 由图 11 可知,在第 420 时间片之前,3 个算法取得的模块度指标相近,但 LabelRankT 的波动较为明显. 在第 420 时间片网络发生突发事件,增量相关节点数目明显增多,社区结构遭到一定的破坏. 但 LabelRankT 在增量更新社区结构时,每个增量相关节点选择某个邻居的编号作为初始化节点的新标签,而节点的邻居构成已经发生较大的变化(邻居可能大多属于上一时刻的其他社区),且由于标签传播机制本身存在的随机性,导致社区结构更进一步遭到破坏. 因此 LabelRankT 在第 420 时间片得到的模块度指标急剧下降. FacetNet 以社区结构的变化为约束条件,有利于保护网络原有的社区结构,模块度下降程度比 LabelRankT 小,而 iNFA 在增量过程开始时保留了各个节点的邻域跟随信息,增量过程调整部分节点的跟随目标,由节点到原有的社区领导者间的

间接跟随关系不会轻易破坏,有利于局部密集区域的节点群体在新的时刻保持同样的社区归属,从而得到更平稳的社区结构.

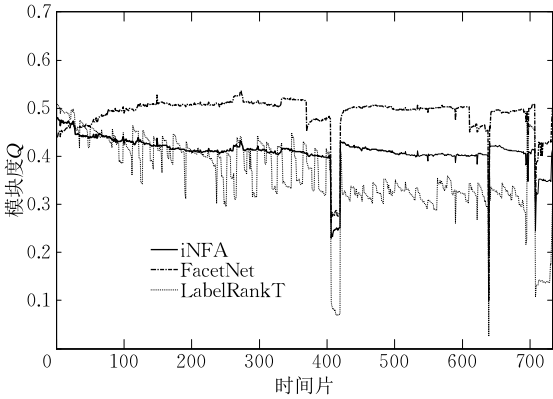


图 11 AS-Internet 网络模块度指标对比

如图 12,比较 3 个算法的运行时间可以发现,利用 iNFA 发现该动态网络的 733 个时间片的网络快照中的社区总共只花费 38.2 s,每个时间片的社区发现过程所需的时间远小于 1 s. FacetNet 总耗费时间为 19465 s,超过 5 h. LabelRankT 则需要耗费

接近 1h. 实验说明 iNFA 能够满足长周期、大规模的动态网络对社区发现效果和效率的要求.

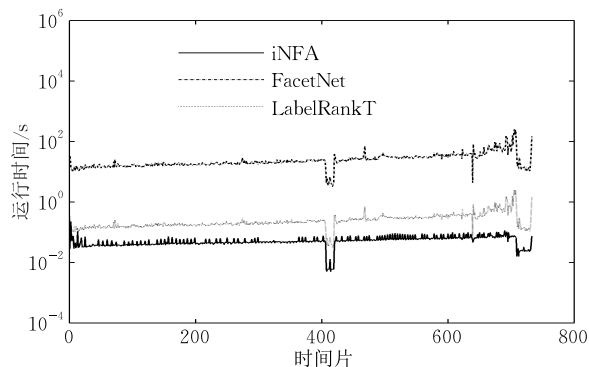


图 12 AS-Internet 运行时间指标对比

5 总 结

本文提出的邻域跟随社区模型刻画了节点间的社区跟随情况,适用于网络演化过程中通过增量更新节点跟随信息挖掘网络的社区结构,iNFA 采用邻域最可信赖跟随策略选择跟随目标,得到网络的邻域跟随社区结构.iNFA 在动态网络社区发现过程只需更新增量相关节点的跟随关系,即可得到新的网络社区结构.算法具有线性时间复杂度,运行效率优于目前典型的社区发现算法,且其社区发现能力不受网络突发事件的影响.今后将改进领导者节点选择策略,进一步提高社区发现精度,并对邻域跟随社区模型进行扩展,使之能够表示网络的重叠社区结构.

参 考 文 献

- [1] Cheng Xue-Qi, Shen Hua-Wei. Community structure of complex networks. *Complex Systems and Complexity Science*, 2011, 8(1): 57-70(in Chinese)
(程学旗, 沈华伟. 复杂网络的社区结构. *复杂系统与复杂性科学*, 2011, 8(1): 57-70)
- [2] Wang Li, Cheng Xue-Qi. Dynamic community in online social networks. *Chinese Journal of Computers*, 2015, 38(2): 219-237(in Chinese)
(王莉, 程学旗. 在线社会网络的动态社区发现及演化. *计算机学报*, 2015, 38(2): 219-237)
- [3] Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: Densification laws, shrinking diameters and possible explanations//*Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*. Chicago, USA, 2005: 177-187
- [4] Takaffoli M, Sangi F, Fagnan J, et al. Community evolution mining in dynamic social networks. *Procedia-Social and Behavioral Sciences*, 2011, 22: 49-58
- [5] Backstrom L, Huttenlocher D, Kleinberg J, et al. Group formation in large social networks: Membership, growth, and evolution//*Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Philadelphia, USA, 2006: 44-54
- [6] Palla G, Barabasi A L, Vicsek T. Quantifying social group evolution. *Nature*, 2007, 446(7136): 664-667
- [7] Falkowski T. Community Analysis in Dynamic Social networks. Sierke [Ph. D. dissertation]. University Magdeburg, Magdeburg, Germany, 2009
- [8] Lin Y R, Chi Y, Zhu S, et al. FacetNet: A framework for analyzing communities and their evolutions in dynamic networks//*Proceedings of the 17th International Conference on World Wide Web*. Beijing, China, 2008: 685-694
- [9] Kawadia V, Sreenivasan S. Online detection of temporal communities in evolving networks by estrangement confinement. *arXiv preprint arXiv: 1203.5126*, 2012
- [10] Chen Y, Kawadia V, Urgaonkar R. Detecting overlapping temporal community structure in time-evolving networks. *arXiv preprint arXiv:1303.7226*, 2013
- [11] Nguyen N P, Dinh T N, Xuan Y, et al. Adaptive algorithms for detecting community structure in dynamic social networks//*Proceedings of the 30th IEEE International Conference on Computer Communications*. Shanghai, China, 2011: 2282-2290
- [12] Xie J, Chen M, Szymanski B K. LabelRankT: Incremental community detection in dynamic networks via label propagation //*Proceedings of the 1st SIGMOD/PODS Workshop on Dynamic Networks Management and Mining*. New York, USA, 2013: 25-32
- [13] Cazabet R, Amblard F, Hanachi C. Detection of overlapping communities in dynamical social networks//*Proceedings of the 2nd International Conference on Social Computing*. Minneapolis, USA, 2010: 309-314
- [14] Coscia M, Rossetti G, Giannotti F, et al. Demon: A local-first discovery method for overlapping communities//*Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, China, 2012: 615-623
- [15] Shan Bo, Jiang Shou-Xu, Zhang Shuo, et al. IC: Incremental algorithm for community identification in dynamic social networks. *Journal of Software*, 2009, 20(Supplement): 184-192(in Chinese)
(单波, 姜守旭, 张硕等. IC: 动态社会关系网络社区结构的增量识别算法. *软件学报*, 2009, 20(增刊): 184-192)
- [16] Wang Yu, Gao Lin. Discovering communities in dynamic networks: A bridgeness incremental clustering algorithm. *Journal of Xidian University*, 2013, 40(1): 30-35(in Chinese)
(王玓, 高琳. 动态网络桥系数增量聚类算法. *西安电子科技大学学报*, 2013, 40(1): 30-35)
- [17] Raghavan U N, Albert R, Kumara S. Near linear time algorithm

- to discover community structures in large-scale networks. *Physical Review E*, 2007, 76(3): 036106
- [18] Lee C, Reid F, Mcdaid A. Detecting highly overlapping community structure by greedy clique expansion. *arXiv preprint arXiv: 1002.1827*, 2010
- [19] Danon L, Diaz-Guilera A, Duch J, et al. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005, 2005(9): P09008
- [20] Newman M E J, Girvan M. Finding and evaluating community structure in networks. *Physical Review E*, 2004, 69(2): 026113
- [21] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 2008, 78(4): 046110
- [22] Gregory S. Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 2010, 12(10): 103018
- [23] Rosvall M, Bergstrom C T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 2008, 105(4): 1118-1123
- [24] Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 2009, 11(3): 033015
- [25] Wayne W, Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 1977, 33: 452-473
- [26] Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826
- [27] Gleiser P M, Danon L. Community structure in jazz. *Advances in Complex Systems*, 2003, 6(4): 565-573
- [28] Guimera R, Danon L, Diaz-Guilera A, et al. Self-similar community structure in a network of human interactions. *Physical Review E*, 2003, 68(6): 065103
- [29] Newman M E J. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 2006, 74(3): 036104
- [30] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005, 435(7043): 814-818
- [31] Adamic L A, Glance N. The political blogosphere and the 2004 US election: Divided they blog//*Proceedings of the 3rd International Workshop on Link Discovery*. New York, USA, 2005: 36-43
- [32] Desmier E, Plantevit M, Robardet C, et al. Cohesive co-evolution patterns in dynamic attributed graphs//*Ganascia J G, Lenca P, Petit J M eds. Volume 7569 of Lecture Notes in Computer Science: Discovery Science*. Berlin: Springer Berlin Heidelberg 2012: 110-124
- [33] Boguñá M, Pastor-Satorras R, Diaz-Guilera A, et al. Models of social networks based on social distance attachment. *Physical Review E*, 2004, 70(5): 056122
- [34] Gregory S. An algorithm to find overlapping community structure in networks//*Proceedings of the Knowledge Discovery in Databases: PKDD 2007, European Conference on Principles and Practice of Knowledge Discovery in Databases*. Warsaw, Poland, 2007: 91-102
- [35] Newman M E J. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 2001, 98(2): 404-409
- [36] Leskovec J, Adamic L A, Huberman B A. The dynamics of viral marketing. *ACM Transactions on the Web*, 2007, 1(1): 5



CHEN Yu-Zhong, born in 1979, Ph.D., professor. His research interests include complex networks, data mining and computational intelligence.

SHI Song, born in 1990, M.S. His research interests include complex networks and data mining.

Background

Community discovery is an important research direction in social network analysis, which can help reveal the organizational principles of real-world networks, including their topological structures and dynamic characteristics. Informally, communities are group of nodes that connected densely inside

ZHU Wei-Ping, born in 1991, M.S. candidate. His research interests include complex networks and data mining.

YU Zhi-Yong, born in 1982, Ph.D., associate professor. His research interests include ubiquitous computing, mobile social Networking.

GUO Kun, born in 1979, Ph.D., associate professor. His research interests including complex networks and data mining.

the group while sparsely connected with the rest of the network.

In most real-world social networks especially the large-scale online social networks, the networks and the communities evolve over times. Recently, some community discovery algorithms have been proposed to meet the requirement of

discovering community structures in large scale dynamic networks. Generally, most community discovery algorithms excavate the nodes involved in network evolution, and then update community affiliation of these nodes and their neighbor nodes. These algorithms can perform well with the assumption that community structure usually evolves smoothly and steadily, but cannot deal with some critical events during the evolution, such as the possible extinction or the emergence of a large number of communities.

In a large-scale social network, it is difficult to achieve a global view of the community structures efficiently. However, a node can discover its local topology and its neighborhood in an easy way. Based on this idea, this paper presents a novel community representation model called Follow-Community Model to address the deficiencies of current incremental community discovery algorithms. This model represents the following relationship among neighbor nodes, and then based on the Follow-Community model, a Neighborhood Following Algorithm(NFA) with nearly linear time complexity is

proposed to discover community structures in static networks by just visiting each node only once. Furthermore, an extended algorithm of NFA named incremental Neighborhood Following Algorithm (iNFA) is proposed to find communities in dynamic networks. Extensive experimental results and analysis are conducted to validate the superiority of the proposed algorithms in accuracy, efficiency and stability for community discovery in dynamic networks.

This work is partly supported by the National Natural Science Foundation of China under Grant Nos. 61300104 and 61300103, the Key Project of Fujian Education Committee under Grant No. JK2012003, the Natural Science Foundation of Fujian Province under Grant No. 2013J01230, the Key Project of industry-academic cooperation of Fujian Province under Grant No. 2014H6014 and the Project of Technology Innovation Platform of Fujian Province under Grant No. 2014H2005. In recent years, our group concentrated on the research of social network analysis, and several papers have been published in international journals and conferences.