

# 基于角色划分的动态社区挖掘算法研究

马瑞新<sup>1</sup> 邓贵仕<sup>2</sup>

(大连理工大学软件学院 大连 116621)<sup>1</sup> (大连理工大学管理与经济学部 大连 116621)<sup>2</sup>

**摘要** 传统社区挖掘算法根据静态的网络拓扑结构进行分析,忽视了个体能动性对网络的影响。针对社会网络中的特殊节点进行研究,引入社区种子和联系者的概念,从个体主义和结构主义两个方面分析社会网络的形成与演化机制,提出了一种基于角色划分的动态社区挖掘算法。在人工网络和真实世界网络上进行了多次测试,并与 G-N、快速 G-N、Polish 等算法进行了比较,结果表明,该算法明显优于 G-N 算法,且其挖掘到的社区都是强连通社区,具有较好的适应性和可扩展性。

**关键词** 个体能动性,社区种子,联系者,角色划分,动态挖掘,强连通社区

**中图分类号** TP393 **文献标识码** A

## Research of Dynamic Community Discovery Based on Role Assorted Thoughts

MA Rui-xin<sup>1</sup> DENG Gui-shi<sup>2</sup>

(School of Software, Dalian University of Technology, Dalian 116621, China)<sup>1</sup>

(Faculty of Management and Economics, Dalian University of Technology, Dalian 116621, China)<sup>2</sup>

**Abstract** Traditional community discovery algorithms focus on the analysis of static topology structure of networks while ignoring the influence of individual activity on the formation of networks. This paper introduced the concept of community seed and liaison, and aiming at the special nodes, researched and analyzed the formation and evolution mechanism of social network from both individualism and structuralism perspectives, proposed a role assorted community discovery algorithm. This paper tested the performance of this algorithm both on artificial network and real-world networks and compared the results with G-N, fast G-N and Polish. Experimental results show that the results of role assorted algorithm are much better than G-N algorithm, with great suitability and expandability. Besides, the discovery communities are all strong connected communities.

**Keywords** Individual activity, Community seed, Liaisons, Role assorted thoughts, Dynamic discovery, Strong connected communities

## 1 引言

网络作为许多大型复杂系统的数学表现形式,吸引了来自物理、生物、万维网、社会学和数学等众多领域的研究学者的关注<sup>[1-20]</sup>。任何复杂系统都可以从实际问题出发,抽象成由相互作用的个体构成的网络。近来的研究表明,复杂网络具有明显的社区结构。社区结构是网络模块化与异质性的反映,表示真实网络由许多不同类型节点组合而成<sup>[6]</sup>。

传统的社会网络研究专注于静态的社会结构,而忽视了具有能动性的个体,因此,社区挖掘只是针对社会网络在某一时刻的静态拓扑结构进行分析与研究。然而,个体之间的交流和互动决定了其所在社会网络的结构形成,只要有个体的行为规则就能对社会网络中社区结构的形成演化机制进行有效的预测和判断。大量研究表明,真实网络受到两个定律的控制:增长和优先情节<sup>[6]</sup>。增长意味着早期出现的节点比晚

期出现的节点有更多的机会积累链接,因此,增长特性使先存在的节点拥有无可比拟的优越性,成为链接最为丰富的节点。优先情节的内涵在于:每个节点吸引链接的能力都与其当前所拥有的链接数量成正比。

本文受增长和优先情节定律的启发,根据社会网络静态的拓扑结构进行逆向扩展,通过分析节点的度数推导其出现的时间,进而以时间为轴,模拟社会网络的形成及演化。在将节点加入到社会网络的同时进行社区划分,注重个体的影响力,提出了一种动态的角色划分(Role Assorted, R-A)的社区挖掘算法。

## 2 常用社区挖掘算法

复杂网络中社区挖掘的研究起源于社会学的研究工作者 Girvan 和 Newman 及其他相关学者的研究成果。在现有已知的社区发现算法中,以 Newman 提出的基于边中介性的

到稿日期:2011-10-08 返修日期:2011-12-30 本文受国家自然科学基金项目(60803074),中央高校基本科研业务费专项资金资助。

马瑞新(1975—),男,博士,讲师,主要研究领域为电子商务、社会计算、群智能,E-mail:teacher\_mrx@126.com;邓贵仕(1945—),男,博士,教授,主要研究领域为复杂系统分析、电子商务、经济管理。

GN算法<sup>[7]</sup>影响最为广泛。然而,GN算法对计算的需求太高,仅仅能够支持对一万个节点以下的社区进行分析,后来提出的快速G-N虽然能够以较低的时间复杂度完成对社区的寻找和划分,却以牺牲准确率为代价;Kernighan算法是一种基于贪婪算法原理<sup>[8]</sup>,在已知网络的确切规模下将网络分割为两个大小已知的社区的二分法,该算法需要较多的先验信息并且不支持对存在奇数个数的社区的挖掘,因此在实际网络中难以得到较好的应用;Radichii等人<sup>[9]</sup>在GN算法的基础上提出基于网络中三角环数量的快速分裂算法,但是如果Web网络中存在的三角形数目很少,那么相应的边聚集系数就会很小,算法就无法正确地搜索网络中的社区,特别是针对树形结构的网络。

近年来,我国学者在社区挖掘算法研究方面取得了不错的研究成果。郭崇慧、张娜<sup>[10]</sup>提出利用共邻矩阵对社会网络进行结构分析和划分,降低了算法的时间复杂度,却也同时降低了算法的挖掘精度;何晓东等<sup>[11]</sup>根据聚类融合的遗传算法对复杂网络中的社区结构进行划分,以在保证低开销的前提下提高社区划分的模块度。但是,该算法仅能分析某一时刻的网络拓扑结构,并且算法复杂度很高。

因此,尽管人们对于复杂网络的社区发现问题已进行了大量的研究,但是仍然存在一些目前无法解决的基本问题,如社区的概念虽然使用广泛,但却缺少严格的数学定义;多数算法都是针对静态网络,不能动态地分析网络结构变化;此外,现有的社区挖掘算法仅仅根据网络的拓扑图进行结构挖掘,而忽略了个体对社区形成的作用。因此,复杂网络中社区发现的研究还未成体系,很多工作有待完善。

### 3 基于角色划分的动态网络社区挖掘分析与研究

#### 3.1 基本概念

##### (1)关系矩阵

假设所要研究的网络包括  $n$  个节点,用  $A$  来表示网络的邻接矩阵,用  $T$  来表示网络的关系矩阵。本文根据关系相似性判断节点之间的密切程度,如果对角线元素为零会影响节点之间的相似性度量,造成一定的关系误差,因此假设所有的节点与自身都存在邻接关系,即邻接矩阵对角线上的元素都设置为 1;同时,这样表示的邻接矩阵可以直接用作无权网络的关系矩阵。在加权图中,边上的数值代表相邻两个节点之间的关系密切程度。

以图 1 为例,该小型社会网络的关系矩阵如下所示。

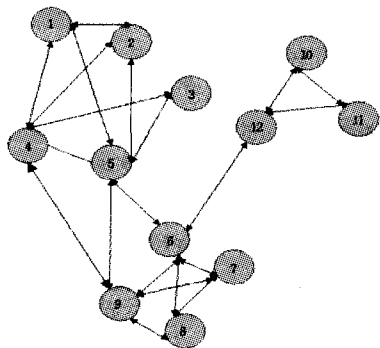


图 1 小型社会网络

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

对网络的关系矩阵进行分析得知,矩阵 Matrix 的第  $i$  行表示节点  $i$  的社会关系向量,则节点  $i$  的度数可以根据式(1)计算:

$$Degree_i = \sum_{j=1}^n A_{i,j} - 1 \quad (1)$$

通过图 1 中的网络可以看出,该网络具有明显的社区结构。“六度空间”<sup>[12]</sup>基本理论告诉我们,具有相同朋友的两者是朋友关系的概率要远远大于从社会中随机选取两个人是朋友关系的概率。据此得出定理 1。

**定理 1** 在网络  $N$  中我们知道:  $A$  和  $B$  是朋友,  $B$  和  $C$  是朋友,  $D, E$  是从  $N$  中随机选择的两个人,则  $\text{similarity}(A, C) \gg \text{similarity}(D, E)$ , 即  $A$  和  $C$  之间同属一个社区的概率要远大于  $D$  和  $E$  同属一个社区的概率。

图论常常被用于解决网络中出现的问题,要证明定理 1,只需要证明  $A, B, C$  形成三角形的概率要远远高于  $B, D, E$  形成三角形的概率。

证明:(1)构建如图 2 所示的无向图;

(2)设置  $B$  为三角形的一个固定顶点;

(3) $A$  和  $C$  是  $B$  的最近邻居,故存在边  $E_{A,B}$  和  $E_{C,B}$ ,则  $A, C$  之间存在边的概率为 50%,即  $\text{triangle}(A, B, C) = 50\%$ 。

(4)随机从网络中选取节点  $D$  和  $E$ ,则  $\text{Probability}(B, D) = \text{Probability}(B, E) = \text{Probability}(D, E) = 50\%$ ,即  $\text{triangle}(B, D, E) = 50\% * 50\% * 50\% = 12.5\%$ 。

(5)因此,  $A, C$  之间的关系相似性要远远高于  $D, E$  之间的关系相似性。

证毕。

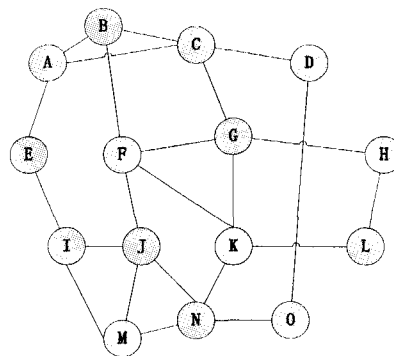


图 2 无向图示例

##### (2)社区向量

在本文中,使用向量空间模型<sup>[21]</sup>表示网络中节点的社会

关系,为了对粒子与已存在的社区进行相似性度量,我们创新性地提出,使用社区特征向量表示社区内的社会关系。社区的特征向量由社区内粒子的关系向量共同组成,表示社区中存在的社会关系。

### ①二进制向量表示法

二进制社区特征向量表示法是指用 0,1 表示社区的关系特征向量。

以图 1 为例,该社会网络的关系矩阵为  $A$ ,节点  $i$  的关系向量为  $A_i$ 。社区  $SN$  的社区向量用  $V_{SN}$  表示,其长度为 12 (节点数量),初始值为零向量。则二进制社区向量计算方法如下所示。

```
foreach node i in SN
 $V_{SN} = V_{SN} \cup A_i$  (2)
```

若  $V_{SN}(i) = 1$ ,说明社区  $SN$  中存在到节点  $i$  的边,否则说明社区中的节点均没有到  $i$  的边。该方法仅仅能够表示社区中节点是否存在到节点  $i$  的边,却不能衡量社区中有多少条到节点  $i$  的边,因此,在此基础上提出了一种关系加权社区向量表示法。

### ②关系加权社区特征向量表示法

关系加权社区特征向量表示法能够统计社区中节点到自由节点  $i$  的边的数量,在一定程度上阐明了社区中节点的吸引力的大小。社区  $SN_i(a)$  的值越大,说明社区  $SN_i$  中到节点  $a$  的边的数量越多, $a$  的吸引力也就越大。关系加权社区特征向量的计算方法如下所示:

```
foreach node i in SN
 $V_{SN} = V_{SN} + A_i$  (3)
```

在进行运算时,向量的加法运算法则不变。仍以图 1 为例,从图中可以看出,节点 4 和节点 5 属于同一社区  $SN_1$ 。如果采用二进制社区向量表示方法,则有:

$$V_{SN_1} = V_4 \cup V_5 = \langle 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0 \rangle$$

该向量仅说明社区存在到节点 1, 2, 3, 4, 5, 6, 9 的关系(边),但这些关系没有权重,不能准确表示社区对节点的吸引力的大小。若采用关系加权社区特征向量表示法,则有:

$$V_{SN_1} = V_4 + V_5 = \langle 2, 2, 2, 2, 2, 1, 0, 0, 2, 0, 0, 0 \rangle$$

该向量说明社区  $SN_1$  中存在两条边指向节点 1, 2, 3, 4, 5, 9, 存在一条边指向节点 11。

### (3)强联通社区

文献[22]提出,在一个社区中, $k_i^{in}(V)$  表示社区  $V$  中节点  $i$  的对内链接, $k_i^{out}(V)$  表示社区  $V$  中  $i$  的对外链接。若每个节点的  $k_i^{in}(V)$  都大于  $k_i^{out}(V)$ ,即存在  $\forall i \in V, k_i^{in}(V) > k_i^{out}(V)$ ,就称每一个节点对于该社区都是忠诚的,该社区是一个强连通社区。假若一个社区  $V$  中所有节点的  $k_i^{in}(V)$  之和大于  $k_i^{out}(V)$  之和,即  $\forall i \in V, \sum_{i \in V} k_i^{in}(V) > \sum_{i \in V} k_i^{out}(V)$ ,则称该社区为弱连通社区。

## 3.2 算法思想

角色划分的思想来自于方守兴的“个别人物法则”<sup>[23]</sup>。他提出,有 3 类人在网站的传播过程中起到关键作用:内行、联系者和推荐者。

内行:在某些领域有丰富经验知识的人,也被称为“权威者”,他们不仅具有较多的最近邻居(距离为 1 的用户),还与多数邻居保持密切的联系。

联系者:社区之间信息交互的桥梁,具有社交天赋的人

群。网络中的大部分用户都通过这些人与整个网络保持联系,他们是社区的边界,其人际关系可能涉及到好几个领域。

推销者:说服用户接受信息的人群。在社会网络中,推销者具有模糊不确定性,他们可能是目标用户的最近邻居,也可能是商务网站的销售人员。由于每个目标用户都具有不同的推销者且推销者对社会网络安全性的研究没有实用价值,因此,真实社会网络分析往往忽略对推销者的研究。

这 3 类人被称为“意见领袖”,在口碑传播的时代,意见领袖的话往往是最容易被人接受的,他们为人们提供建议和心得,以乐于助人的姿态为其他用户提供服务,很难使人抗拒。

增长和优先情节定律告诉我们,由于新节点偏向于和链接多的节点建立连接,因此,早出现的节点会比晚到的链接少的节点增长得快,而且,每个节点吸引的链接都和其当前的链接数量成正比。正是由于增长和优先情节的存在导致了真实网络存在的幂率。根据优先情节和增长的定律,假设度数最多的节点出现的时间最早,度数最少的节点出现的时间最晚。在该假设下,以时间为轴,根据现有的静态社会网络拓扑结构逆向模拟该社会网络的生成及演化机制。

## 3.3 算法步骤及实现

社区种子是社区的核心所在,它限定社区形成的主题和目标,并且引导社区内部成员的行为。受到真实社会中团体组织形式的启发,本文同时提出了社区种子和候选社区种子的概念,候选社区种子是社区中的第二号成员,与社区种子的关系十分密切,两者的邻接关系十分相似。

联系者负责不同社区之间的消息传播和互动,如果没有联系者,信息就不能像病毒一样席卷整个网络。

在优先情节和增长定律启发下,本文提出:节点存在的时间与自身的度数成正比。因此,根据节点的度数分布对社会网络的形成演化机制进行模拟与分析,在不断地将节点加入到网络中的同时进行社区划分,过程如下:

将所有粒子按照度数的降序排列组成一张列表  $L_{ji}$ ,社区种子集合  $S$ 、社区联系者集合  $L$  初始化为空;将列表中的第一个节点取出,作为第一个社区的种子;从第二个节点开始对列表从上往下进行检查,如果列表中的自由节点  $i$  的关系向量与已发现社区  $SN$  的特征向量之间的相似性小于最小相似性阈值  $\delta$ ,则计算该节点的平均关系密度  $C_i$ ,若  $C_i > \epsilon$  (最小关系密度阈值),则该粒子成为新的社区种子,并将节点  $i$  加入种子集合  $S$  中,否则, $i$  被加入社区联系者集合  $L$  中;若自由节点仅仅与一个已存社区的余弦相似性大于最小相似性阈值  $\delta$  且该自由节点是社区中的第二个成员,则标识为候选社区种子,否则,判断其仅作为成员加入社区中;若节点与多个社区的余弦相似性均大于最小相似性阈值,则将节点加入到与之关系最相近的社区中,并将其标识为社区联系者。然后,将  $L$  集合中的联系者取出,并与已存社区进行相似性计算,将其加入到与之关系最密切的社区中。最后,计算每个社区的联通性质,若最后得到的每个社区都是强联通社区,则算法结束,否则,重新划分满足  $k_i^{in}(V) < k_i^{out}(V)$  的节点的所属社区。

从该过程中可以看出,度数最高的节点首先出现在社会网络中,新加入的每一个节点都会与已存在的节点进行相似性度量,判断自己是不是和已存在的节点属于一类,如果是,就加入到已存在的节点团体中,否则,自成一派。在真实的社会网络中,每时每刻都有新的节点出现,虽然我们不断定其

所属类别,但是可以根据每个节点的行为特征和属性,对其进行引导,促进社会网络的收敛。

80/20 规则同样适用于社会网络,即 20% 的节点拥有 80% 的链接,而 80% 的节点仅仅拥有 20% 的链接。因此,本文提出,度数分布排名前 20% 的用户成为网络中的核心节点。核心节点的关系链接众多,因此,在进行最小相似性阈值判断时,要区别对待。本文设定,  $\delta(core) = 0.5$ ,  $\delta(normal) = 1/n$ ,  $n$  为社会网络的规模。

对于无权网络,节点的度数是衡量节点核心程度的唯一指标,而对于加权关系网,节点的度数和平均关系密切程度共同对节点所在位置的核心程度进行度量,因此,我们构造适应度函数,如式(5)所示。 $C_i$  表示节点  $i$  与网络中其他节点的平均关系密切程度,利用式(4)进行计算,对于无权网络,  $C_i = 1$ 。式(5)中,  $T$  代表当前研究的网络的加权矩阵,  $C_i$  表示节点  $i$  的所有边的平均加权,即其平均关系密度。在真实社会网络分析中,  $C_i$  在一定程度上反映了节点  $i$  的信任度,  $\alpha$  和  $\beta$  分别表示节点链接数量和节点信任度的权重。

$$C_i = \frac{\sum_{j=1}^n T_{i,j}}{\sum_{j=1}^n A_{i,j}} \quad (4)$$

$$F_i = \alpha \times \sum_{j=1}^n A_{i,j} + \beta \times C_i \quad (5)$$

以社区  $S$  和粒子  $i$  为例,粒子与社区的余弦相似性根据式(6)进行计算:

$$\text{Sim}(S, i) = \cos(S, i) = \frac{V_S \cdot V_i}{|V_S| \cdot |V_i|} \quad (6)$$

最小关系密度根据网络中所有边的平均值进行设定,如式(7)所示。因加权矩阵是对称矩阵,故要除边数的两倍。

$$\epsilon = \frac{\sum_{i=1}^n \sum_{j=1}^n T_{i,j}}{2 \times |E|} \quad (7)$$

对上述社区挖掘过程进行分析可知,该过程模拟了社会网络的形成机制,最先加入社区的个体成为社区中的中心节点,随着时间的推移,又有新的节点陆续加入,新节点总是偏向于与链接数量多的节点形成连接,导致早出现的节点出现较快的链接增长,而后加入的节点则保持较少的链接数目,长此以往,最终形成稳定的社区结构。

### 3.4 实验结果分析与对比

#### (1) Zachary Club<sup>[24]</sup>

Zachary 社会关系网是复杂网络与社会网分析领域中常用的一个小型测试网络。该网络有 34 个节点、78 条边,所有的节点分为 2 个社区(见图 4)。每个节点表示一个俱乐部成员,节点间的连接表示两个成员在俱乐部内部和外部的社会关系。该网络作为一个真实的小型社会关系网,常常被用于测试社区发现方法的有效性。

在试验过程中我们发现,由于二进制社区向量表示法不能准确地衡量社区内部的社会关系,导致计算结果不准确,因此,本文中的试验均使用关系加权社区向量。

本文使用 Karate Club 数据集作为无权网络的代表进行分角色社区挖掘算法测试,  $L_{fit} = \{34, 1, 33, 3, 2, 4, 32, 9, 24, 6, 7, 8, 14, 15, 28, 30, 31, 5, 11, 20, 25, 26, 29, 10, 13, 16, 17, 18, 19, 21, 22, 23, 27, 12\}$ 。●代表社区种子,■代表候选社区种子,◆代表联系者,▲代表普通用户,社区划分结果如图 4 所示。图 3 中虚线代表社区划分的阈值线,若节点与社区之间的相似度大于阈值线,则属于相应社区。

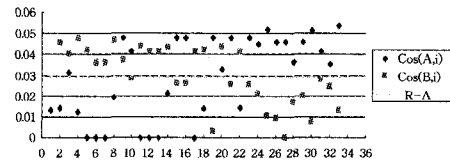


图 3 Zachary Club 社会网络中节点划分图

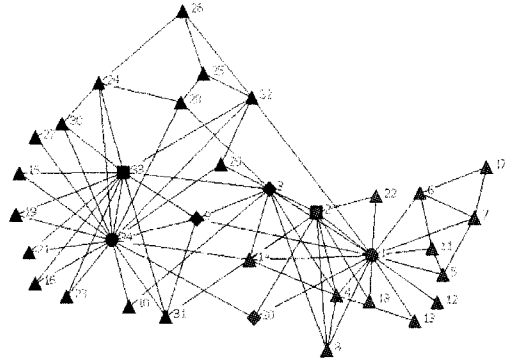


图 4 分角色社区挖掘对 Zachary 的结构划分结果

#### (2) Les Misereables<sup>[25]</sup>

D. E. Knuth 根据 Victor Hugo(维克多·雨果)的小说《悲惨世界》(Les Misérables)进行人物关系整理,构造了一个真实的社会网络数据集。网络中的节点表示小说中的角色,边表示两个角色同时出现在一幕或多幕中。网络中有 5 个主要人物,即主人公 Jean Valjean(节点 11)、探长 Javert(节点 27)、神父 Bishop Myriel(节点 0)、女工 Fantine(节点 23)及其女儿 Cosette(节点 26)。研究人际网络中关键的边(即人物之间的联系)对网络整体性能的影响,发现网络内关键的边将对舆情和疾病等的传播具有非常重要的意义。本文采用关系加权的社区向量法,使用该数据集对分角色社区挖掘算法进行实验验证,得到图 5 所示的结果。

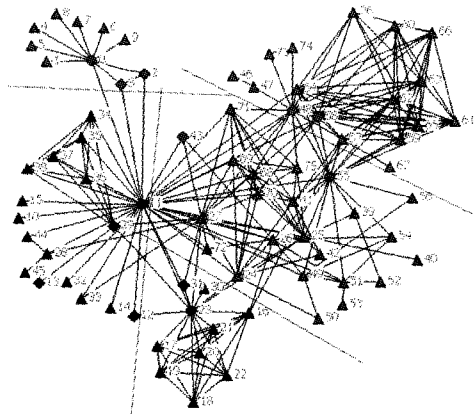


图 5 关系加权的分角色挖掘对 Les Miserables 的结构划分

以节点 14 为例,如果使用二进制社区向量表示法,则有:

$$\text{Sim}(14, 11) = \frac{1+1}{\sqrt{2} \times \sqrt{36}} = \frac{1}{3\sqrt{2}} > \zeta \quad (8)$$

$$\text{Sim}(14, 23) = \frac{1+1}{\sqrt{2} \times \sqrt{18}} = \frac{1}{3} > \zeta \quad (9)$$

根据计算结果,节点 14 属于社区 B 和 C 之间的联系者,而实际情况如图 5 所示,节点 14 严格属于社区 B。若使用关系加权社区向量表示法,则不会出现这种错误,因此,本文中所有的试验采用关系加权社区向量表示法。

(下转第 70 页)

- [11] Monks J P, Bharghavan V, Hwu W M W. A power controlled multiple access protocol for wireless packet networks[C]//Ieee Infocom 2001: The Conference on Computer Communications. Vols 1-3, 2001; 219-228
- [12] Sheth A, Han R. SHUSH: Reactive transmit power control for wireless MAC protocols[C]//First International Conference on Wireless Internet. 2005; 18-25
- [13] Koksal C, Kassab H, Balakrishnan H. An analysis of short-term fairness in wireless media access protocols[C]//Proc. of ACM SIGMETRICS'00. ACM, 2000; 118-119
- [14] 李方敏, 刘新华, 旷海兰. 基于最优连通功率的无线传感器网络稳定成簇算法[J]. 通信学报, 2008, 30(3): 75-83
- [15] Zawodniok M, Jagannathan S. A distributed power control MAC protocol for wireless ad hoc networks[C]//2004 Ieee Wireless Communications and Networking Conference. Vols 1-4, 2004; 1915-1920

(上接第 63 页)

### 3.5 模块度对比

表 1 对不同社区挖掘算法的模块度进行了对比,从中可以看出,与其他经典社区挖掘算法相比,分角色的挖掘取得了最高的模块度。特别对于悲惨世界数据集, RA-CDA 的模块度是 G-N 的 5 倍, 是 Polish 算法的 4 倍。此外, RA-CDA 根据自由节点与社区特征向量的邻接相似性进行判断和规划, 社区中每一个节点的对内连接都多于对外连接, 因此, 最终得到的都是强连通社区。

表 1 社区挖掘结果的模块度对比

Name	Karate Club	Dolphins	Les Miserables
GN	0.395	0.381	0.082
Fast G-N	0.360	0.379	0.074
Polish	0.358	0.226	0.107
RACDA	0.401	0.383	0.486

**结束语** 基于角色划分的动态社区挖掘算法受到社会分化和“特殊人物法则”的启发, 根据网络中节点度数分布进行角色划分, 分析精英个体在社会网络形成中的领导作用, 进而模拟社会网络的形成演化过程, 动态地实现节点的规划, 以便人们深入理解复杂社会网络的内在机理和演化机制。

### 参 考 文 献

- [1] Newman M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Rerview E, 2006, 74(3): 1-22
- [2] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: Structure and dynamics[J]. Physics Reports, 2006(424): 175-308
- [3] Watts D J, Strogatz S H. Collective dynamics of ‘small- world’ networks[J]. Nature, 1998, 393(6638): 440-442
- [4] 段晓东, 王存瑞, 刘向东, 等. 基于粒子群算法的社区发现[J]. 计算机科学, 2008, 35(3): 18-22
- [5] Albert R, Jeong H, Barabasi AL. The internet’s Achilles heel: error and attack tolerance of complex networks[J]. Nature, 2000, 406(2115): 378-382
- [6] Barabasi A L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512
- [7] Guimera R, Amaral L A N. Functional Cartography of Complex Metabolic Networks[J]. Nature, 2005, 433(7028): 895-900
- [8] Krapivsky P L, Redner S, Leyvraz F. Connectivity of Growing Random Networks[R]. Physical Review Letters, 2000, 85: 4629-4632
- [9] He Dong-xiao, Zhou Xu, Wang Zuo, et al. Community Mining in Complex Networks—Cluster Genetic Algorithm[J]. Acta Automatica Sinica, 2010, 36(8): 1160-1170
- [10] 郭崇慧, 张娜. 基于共邻矩阵的复杂网络社区结构划分方法[J]. 系统工程理论与实践, 2010, 30(6): 1077-1084
- [11] 何晓东, 周栩, 王佐. 复杂网络社区挖掘—基于聚类融合的遗传算法[J]. 自动化学报, 2010, 36(8): 1160-1170
- [12] 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
- [13] 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
- [14] Barber M J, Clark J W. Detecting network communities by propagating labels under constraints[J]. Physical Review E, 2009, 80(2): 026129
- [15] Leung I X Y, Hui P, Lio P, et al. Towards real-time community detection in large networks[J]. Physical Review E, 2009, 79(6): 066107
- [16] Zhang Y Z, Wang J Y, Wang Y, et al. Parallel community detection on large networks with propinquity dynamics[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France: ACM, 2009: 997-1006
- [17] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 2004, 69(6): 066133
- [18] Kernighan B W, Lin S. An efficient heuristic procedure for partitioning graphs[J]. The Dell System Technical Journal, 1970, 49(1): 291-307
- [19] Liu X, Li D Y, Wang S L, et al. Effective algorithm for detecting community structure in complex networks based on GA and clustering[C]//Proceedings of the 7th International Conference on Computational Science. Beijing, China: Springer, 2007: 657-664
- [20] Gog A, Dumitrescu D, Hirsbrunner B. Community detection in complex networks using collaborative evolutionary algorithms [C]//Proceedings of the 9th European Conference on Artificial Life. Lisbon, Portugal: Springer, 2007: 886-894
- [21] 高琪, 张永平. PageRank 算法中主题漂移的研究[J]. 网络与通信, 2010, 3(3): 117-119
- [22] 胡健, 董跃华, 杨炳儒. 大型复杂网络中社区结构发现算法[J]. 计算机工程, 2008, 34(19): 92-93
- [23] 王娟, 谢弛, 荣雪, 范雯. SNS 网站运营的现状和未来趋势研究 [N/OL]. [http://media. people. com. cn/GB/22114/119489/140165/8454258. html](http://media.people.com.cn/GB/22114/119489/140165/8454258.html), 2008-12-03
- [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] Knuth D E. The Stanford Graphbase: “A Platform for Combinatorial Computing[M]. Addison-Wesley, 1993