

万仁霞,张宇红,苗夺谦.基于吸收度的三支决策社团划分算法[J].江西师范大学学报(自然科学版) 2022 46(3):314-321.  
WAN Renxia ZHANG Yuhong MIAO Duoqian.The absorbance-based community division algorithm with three-way decision [J].Journal of Jiangxi Normal University( Natural Science) 2022 46(3):314-321.

文章编号:1000-5862(2022)03-0314-08

## 基于吸收度的三支决策社团划分算法

万仁霞<sup>1</sup>,张宇红<sup>1</sup>,苗夺谦<sup>2</sup>

(1.北方民族大学数学与信息科学学院,宁夏 银川 750021;2.同济大学计算机科学与技术系,上海 201804)

**摘要:**该文针对社团划分存在的重叠区域问题引入三支决策思想,提出了一种基于吸收度的社团划分算法(3WD-PPOC).3WD-PPOC 首先根据网络结构的重要度矩阵进行社团的初始划分,再利用  $F$  吸收度来构建社团间的重叠区,即社团边界域,并得到各社团的正域,最后通过  $P$  吸收度来完成对在社团边界域中节点的再次划分和社团正域的更新.对比同类算法,3WD-PPOC 具有较低的时间复杂度.实验结果进一步表明:3WD-PPOC 能够有效地进行社团划分,相比其他社团划分算法,3WD-PPOC 表现出更好的社团划分质量,划分后的各社团结构更紧密.该算法对社团重叠节点的划分具有较好的稳定性.

**关键词:** 社团划分; 三支决策; 吸收度; 重要度矩阵; 正域; 边界域

**中图分类号:** TP 301 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2022.03.15

### 0 引言

在现实世界中存在着各种各样的网络,如城市交通网络<sup>[1]</sup>、社会网络<sup>[2]</sup>、生物网络<sup>[3]</sup>等,它们都可以抽象成复杂网络.而复杂网络又是由若干个社团构成的,社团的结构性<sup>[4]</sup>主要表现为:在同一个社团中的节点联系较为紧密,在不同社团间的节点联系较为稀疏.研究网络的社团结构可以更加准确地理解复杂网络的拓扑结构及其内在原理.因此,如何有效地进行社团划分是社团研究者们研究方向之一.

近年来,许多学者从不同角度对网络的社团划分<sup>[5]</sup>算法进行了研究.Newman 快速算法(Newman fast algorithm, NFA)<sup>[6]</sup>依靠模块度获得最佳社团结构;自包含 GN 算法<sup>[7]</sup>(self-contained GN algorithm)给出了强社团结构和弱社团结构 2 种变量的定义,为社团结构的好坏提供了一种参考标准;B.W. Kernighan 等<sup>[8]</sup>提出了基于图划分的社团划分算法,该算法需要提前预知社团划分的数目才可以实现社团划分;U.N. Raghavan 等<sup>[9]</sup>提出一种快速标号传播算

法(label propagation algorithm, LPA),该算法利用网络自身结构来判定社团结构,复杂度较低且收敛快,但缺点是比真实社团结构的精度略低.上述算法从不同角度和不同层面对复杂网络的社团划分问题进行研究,并取得了一定的研究成果.目前,社团划分算法大多数是非重叠的社团划分算法,其对于重叠部分的处理采用传统的二支决策技术,即根据已有的信息、社团对节点的归属做出接受或拒绝的判决.然而,由于信息的模糊或不充分等因素,许多网络存在重叠的社团结构<sup>[10]</sup>,所以基于二支决策技术的社团划分会导致社团划分结果的不可靠问题.因此,如何对网络中重叠部分的节点进行有效地划分,以便发现社团潜在的规律,已引起了许多学者的关注.李敏毓等<sup>[11]</sup>提出一种社团结构特征研究,旨在处理在社交网络中的重叠社团并解决现有的社团划分算法结果分辨率低的问题.LFM(largest fitness measure)是基于局部优化的适应度函数的社团发现算法<sup>[12]</sup>,该算法在发现网络中的重叠社团和有关层次结构的社团方面具有较好的效果.郭娜等<sup>[13]</sup>提出了一种基于最大生成树的重叠社团发现算法,该算法对初始

收稿日期:2022-01-05

基金项目:国家自然科学基金(61662001),中央高校基本科研业务费专项资金(FWNX04)和宁夏自然科学基金(2021AAC03203)资助项目.

作者简介:万仁霞(1975—),男,江西南昌人,教授,博士生导师,主要从事信息系统、数据挖掘知识学习和智能计算研究.E-mail:wanrx1022@nmu.edu.cn

社团划分结果进行优化,且避免了社团之间重叠的出现。

三支决策(three-way decisions, 3WD)理论<sup>[14]</sup>的思想是由决策粗糙集理论(decision-theoretic rough sets, DTRS)<sup>[15]</sup>产生的,旨在解决在现实世界中的不确定信息的决策问题,为模糊信息处理<sup>[16]</sup>提供一种新的解决思路。由于三支决策符合人类思维和认知特点,能较好地处理在实际决策过程中出现的不确定性问题,所以它一经提出便得到国内外学者的广泛关注。杨雪洁等<sup>[17]</sup>提出了一种基于子模优化的边界域处理社团发现算法,旨在用三支决策模型及子模优化思想来划分社团结构。方莲娣等<sup>[18]</sup>提出一种基于三支决策的非重叠社团划分算法,该方法将初始聚类形成的重叠社团进行2次划分以形成最终的非重叠社团。

本文通过节点的重要度来刻画节点间的关系,采用三支决策思想来解决社团节点重叠问题,提出了一种基于吸收度的三支决策社团划分算法,即根据不同节点归入不同社团域的动作参数所产生的损失函数来定义吸收度,并依据吸收度对已获取的重叠节点进行划分,不仅较好地体现了节点的真实归属,还可以获取更好地接近全局最优社团。本文采用3个真实的网络数据集对3WD-PPOC算法进行了验证,实验结果表明本文所提算法对在社团中的节点处理可行且有效。

## 1 相关概念

### 1.1 节点属性重要度

在一个无向无权网络 $G = \langle V, E \rangle$ 中,节点集合为 $V$ ,边集合为 $E$ 。在节点集合 $V$ 中一对点即对应边集合 $E$ 中的一条边。

定义1<sup>[19]</sup> 设节点集合 $V = \{e_1, e_2, \dots, e_n\}$ ,令 $(e_i, e_j)$ 表示节点 $e_i$ 与节点 $e_j$ 之间的边。若 $e_i$ 和 $e_j$ 相连,则边存在;若 $e_i$ 和 $e_j$ 不相连,则边不存在。边集 $E = \{(e_i, e_j) : e_i, e_j \in V; 1 \leq i, j \leq n\}$ ,则节点 $e_i$ 的度 $D_i$ 表示 $e_i$ 的邻居节点的数目,即与该节点连接的其他节点的数目 $D_i$ 的表达式为

$$D_i = |\{(e_i, e_j) : e_i, e_j \in V; (e_i, e_j) \in E\}|.$$

定义2<sup>[20]</sup> 设节点集为 $V = \{e_1, e_2, \dots, e_n\}$ ,那么在节点集中所有节点的平均度为 $\langle k \rangle$ , $\langle k \rangle$ 的表达式为

$$\langle k \rangle = \sum_{i=1}^n e_i / n.$$

由于 $e_i, e_j$ 属于同一网络,且 $D_i$ 表示节点 $e_i$ 的邻

居节点数目,于是可得到如下节点平均度的结论。

定理1<sup>[20]</sup> 设节点数为 $n$ ,无自环无项网络,节

点平均度值为 $\langle k \rangle = \sum_{i=1}^n e_i / n$ ,若节点 $e_i$ 的度为 $D_i$ ,则 $e_i$ 将自身重要度的 $D_i / \langle k \rangle^2$ 传播给其他相邻的每一个节点。

定义3<sup>[21]</sup> 邻接矩阵 $W_{n \times n}$ 表示在网络中节点 $e_i$ 与节点 $e_j$ 之间的连接关系,其中 $w_{ij}$ 为在邻接矩阵中对应的元素,取值为0或1, $n$ 为节点总数。即若节点 $e_i$ 和 $e_j$ 之间有连接关系,则 $w_{ij} = 1$ ;若节点 $e_i$ 和 $e_j$ 之间没有连接关系,则 $w_{ij} = 0$ 。

定义4<sup>[22]</sup> 节点重要度矩阵 $H$ 的定义为

$$H = \begin{pmatrix} 1 & w_{12}D_2/\langle k \rangle^2 & \cdots & w_{1n}D_n/\langle k \rangle^2 \\ w_{21}D_1/\langle k \rangle^2 & 1 & \cdots & w_{2n}D_n/\langle k \rangle^2 \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}D_1/\langle k \rangle^2 & w_{n2}D_2/\langle k \rangle^2 & \cdots & 1 \end{pmatrix},$$

其中将矩阵对角线上的元素全部置为1,它表示在网络中每个节点对于自身的重要度贡献比值为1。由定理1知,在网络中的节点重要度矩阵 $H$ 是网络邻接矩阵的映射。当 $i \neq j$ 时, $w_{ij}$ 映射为 $w_{ij}D_j/\langle k \rangle^2$ ;当 $i = j$ 时, $w_{ij}$ 映射为1。

### 1.2 三支决策

三支决策<sup>[23]</sup>理论对于在信息处理中不确定决策问题的解决具有高效性,尤其是对信息不精确、条件不充分的情况。其核心思想是将决策项分成3种决策规则,分别是正域决策、负域决策和边界域决策。当证据不完整、不充足时,可以采用边界域决策;当证据精准、完善时,可以采用正域决策或者负域决策。正域决策和负域决策是明确的,即三支决策主要围绕边界域的处理展开研究。

三支决策的研究主要基于决策粗糙集,整个论域被划分为3个部分,即正域(POS)、负域(NEG)和边界域(BND),分别代表接受、拒绝和不承诺3种决策结果。决策粗糙集模型理论<sup>[24]</sup>将概率粗糙集和最小风险贝叶斯决策结合起来,通过计算各类决策风险损失值,对正域(POS)、负域(NEG)和边界域(BND)进行划分。

文献[24]对三支决策给出了具体解释:假设有3种状态的集合 $\Omega = \{X_1, X_2, X_3\}$ ,对应于概率粗糙集上的正域、负域、边界域。由分类结果的3个域构造出一个决策动作集 $D = \{\alpha_P, \alpha_N, \alpha_B\}$ ,其中 $\alpha_P, \alpha_N, \alpha_B$ 分别代表将一个对象分类到概率粗糙集上的正域、负域、边界域的决策动作,且不同的决策动作代表不同的分类结果。

## 2 基于吸收度的三支决策社团划分算法

三支决策模型根据对象的正域、负域、边界域采取不同的决策规则,尤其在处理不确定性问题时,可以通过信息量的增多,做出更为准确的判决。

### 2.1 社团中的 3 个域

基于三支决策的思想,实现对网络重叠社团结构的划分。对于初始社团划分后获得的重叠社团结构的定义如下:

- (i) 正域(POS) 表示被考察社团的非重叠节点;
- (ii) 边界域(BNG) 表示重叠部分的节点;
- (iii) 负域(NEG) 表示除正域及边界域外的节点。

如图 1 所示,左右 2 个椭圆分别代表 2 个社团结构,且这 2 个社团之间存在社团重叠结构,即节点 5、节点 6 以及节点 7。网络节点集为  $V = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ , 社团  $X = \{1, 2, 3, 4, 5, 6, 7\}$ , 依据三支决策思想,正域  $Q_{\text{POS}}(X) = \{1, 2, 3, 4\}$ , 边界域  $Q_{\text{BND}}(X) = \{5, 6, 7\}$ , 负域  $Q_{\text{NEG}}(X) = \{8, 9, 10, 11\}$ 。

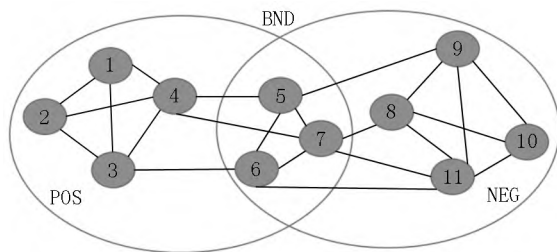


图 1 重叠社团 3 个域的划分

显然, 社团结构满足如下性质。

- 性质 1 在网络  $G = \langle V, E \rangle$  中, 对于任意社团  $X$ , 有: (i)  $Q_{\text{POS}}(X) \cup Q_{\text{BND}}(X) \cup Q_{\text{NEG}}(X) = V$ ;
- (ii)  $Q_{\text{POS}}(X)$ 、 $Q_{\text{BND}}(X)$ 、 $Q_{\text{NEG}}(X)$  两两相交为空。

这表明: 社团的正域、边界域、负域是网络节点的一个划分, 且当在社团中的 2 个域明确后, 其余 1 个域也是确定的。因此, 在本文中, 仅讨论社团的正域、边界域的构成, 不对负域赘叙。

### 2.2 吸收度

借鉴三支决策阈值<sup>[25]</sup>的概念, 本文给出社团划分的吸收度定义。

定义 5 假设一个社团可以划分为  $\Omega = \{Y_1, Y_2, Y_3\}$ , 即对应于社团的正域、负域、边界域。当在社团中的节点归到不同的社团域中时, 可以得到不同的动作参数  $L = \{\alpha_P, \alpha_N, \alpha_B\}$ , 其中  $\alpha_P, \alpha_N, \alpha_B$  分别代表节点归入不同社团域的动作参数, 且不同的动作参数代表不同的社团域划分结果, 可能出现 9 种损

失函数(见表 1)。其中第 1 列函数表示当节点归入社团  $Y_1$  动作参数为  $\alpha_P, \alpha_N, \alpha_B$  时带来的损失函数, 记为  $\lambda_{PY_1}, \lambda_{NY_1}, \lambda_{BY_1}$ ; 第 2 列函数表示当节点归入社团  $Y_2$  动作参数为  $\alpha_P, \alpha_N, \alpha_B$  时带来的损失函数, 记为  $\lambda_{PY_2}, \lambda_{NY_2}, \lambda_{BY_2}$ ; 第 3 列函数表示当节点归入社团  $Y_3$ , 动作参数为  $\alpha_P, \alpha_N, \alpha_B$  时带来的损失函数, 记为  $\lambda_{PY_3}, \lambda_{NY_3}, \lambda_{BY_3}$ 。

表 1 损失函数

参数	$Y_1$	$Y_2$	$Y_3$
$\alpha_P$	$\lambda_{PY_1} = \lambda(\alpha_P   Y_1)$	$\lambda_{PY_2} = \lambda(\alpha_P   Y_2)$	$\lambda_{PY_3} = \lambda(\alpha_P   Y_3)$
$\alpha_N$	$\lambda_{NY_1} = \lambda(\alpha_N   Y_1)$	$\lambda_{NY_2} = \lambda(\alpha_N   Y_2)$	$\lambda_{NY_3} = \lambda(\alpha_N   Y_3)$
$\alpha_B$	$\lambda_{BY_1} = \lambda(\alpha_B   Y_1)$	$\lambda_{BY_2} = \lambda(\alpha_B   Y_2)$	$\lambda_{BY_3} = \lambda(\alpha_B   Y_3)$

节点归入不同社团域可以得到不同的损失函数, 因此不同节点在不同的情况下会归入不同的域。本文将根据吸收度的不同职能, 将吸收度区分为  $F$  吸收度和  $P$  吸收度:

$$F = (\lambda_{PY_3} - \lambda_{BY_3}) / ((\lambda_{PY_3} - \lambda_{BY_3}) + (\lambda_{BY_1} - \lambda_{PY_1})), \quad (1)$$

$$P = (\lambda_{BY_3} - \lambda_{NY_3}) / ((\lambda_{BY_3} - \lambda_{NY_3}) + (\lambda_{NY_1} - \lambda_{BY_1})), \quad (2)$$

其中  $F$  吸收度用来控制社团边界域的形成,  $P$  吸收度用来控制社团边界域节点的再划分。

可以证明  $F, P$  吸收度满足如下性质。

性质 2  $0 \leq P < F \leq 1$ 。

### 2.3 PPC 算法流程

本文首先根据重要度矩阵, 给出非重叠的社团划分算法, 该算法在不引入吸收度时形成了初始社团结构, 主要实现步骤如算法 1 所示。

算法 1 PPC 算法。

输入: 无向无权网络  $G = \langle V, E \rangle$  及重要度矩阵  $H$ 。

输出: 网络社团  $A_j (j = 1, 2, \dots, k)$ 。

(i) 计算网络的节点重要度矩阵  $H$ ;

(ii) 随机选择  $k$  个节点  $z_1, z_2, \dots, z_k$  作为初始社团的中心, 令  $A_j = \{z_j\} (j = 1, 2, \dots, k)$ ;

(iii) 对于网络中的任意节点  $Z$  若  $|H_Z - H_{Z_{j_0}}| = \min_{j_i=1, 2, \dots, k} |H_Z - H_{Z_{j_i}}|$ , 则将  $Z$  归入  $Z_{j_0}$  所在社团  $A_{j_0}$ , 即  $Z \in A_{j_0}$ ;

(iv) 计算各社团中心, 若各中心未发生变化,

则转入步骤(v),否则,转向步骤(iii);

(v) 输出社团  $A_j (j = 1, 2, \dots, k)$ .

在步骤(iii)中的  $H_z$  代表在社团中所有需要被考察节点  $Z$  的重要度值.

算法1实际上是一种典型的非重叠社团划分的算法,为重叠社团划分创造了初始社团结构.不同于一般社团构建的算法,本文的算法采用节点重要度来划分节点的社团归属.

## 2.4 3WD-PPOC 算法流程

本文结合三支决策思想,引入吸收度的概念,在上述算法产生的初始社团结构的基础上进行局部再划分处理,从而达到对重叠社团节点的更精细划分,其主要实现步骤如下.

算法2 3WD-PPOC 算法.

输入: 无向无权网络  $G = \langle V, E \rangle$ 、重要度矩阵  $H$  及

$\lambda_{PY_1}, \lambda_{NY_1}, \lambda_{BY_1}, \lambda_{PY_2}, \lambda_{NY_2}, \lambda_{BY_2}, \lambda_{PY_3}, \lambda_{NY_3}, \lambda_{BY_3}$  值.

输出: 无重叠网络社团结构.

(i) 计算网络的节点重要度矩阵  $H$ ;

(ii) 随机选择  $k$  个节点  $z_1, z_2, \dots, z_k$  作为初始社团的中心,令  $A_j = \{z_j\} (j = 1, 2, \dots, k)$ ;

(iii) 对于网络中的任意节点  $Z$ ,若  $|H_Z - H_{Z_{j_0}}| =$

$\min_{j_i=1,2,\dots,k} |H_Z - H_{Z_{j_i}}|$ ,则将  $Z$  归入  $Z_{j_0}$  所在社团  $A_{j_0}$ ,即  $Z \in A_{j_0}$ ;

(iv) 计算各社团中心,若各中心未发生变化,则转入步骤(v),否则,转向步骤(iii);

(v) 通过式(1)~(2)计算  $F$  吸收度和  $P$  吸收度;

(vi) 对于任意网络节点  $Z$ ,若存在社团  $A_l, A_m$ ,有  $||H_Z - H_{z_l}| - |H_Z - H_{z_k}|| \leq F$ ,则将节点  $Z$  归入社团  $A_l, A_m$  的边界  $Q_{\text{BND}}(A_l), Q_{\text{BND}}(A_m)$ ,即  $Z$  为社团  $A_l, A_m$  的共同边界点(社团重叠节点);

(vii) 构建社团正域:  $Q_{\text{POS}}(A_i) = A_i - Q_{\text{BND}}(A_i) (i = 1, 2, \dots, k)$ ;

(viii) 计算社团正域的中心  $\{r_1, r_2, \dots, r_k\}$ ,其中  $r_i = (\sum_{t_j \in Q_{\text{POS}}(A_j)} Q_{\text{POS}}(A_i)_{t_j}) / |Q_{\text{POS}}(A_i)|$ ;

(ix) 边界节点再划分: 节点  $Z$  为  $Q_{\text{BND}}(A_{i_1}), Q_{\text{BND}}(A_{i_2}), \dots, Q_{\text{BND}}(A_{i_l}) (1 \leq i_1, i_2, \dots, i_l \leq k)$  的重叠节点,令  $B_j = ||H_Z - H_{r_j}| - |H_Z - H_{r_j}||$ ,若  $B_{j_0} = \min_{j=i_1, i_2, \dots, i_l} B_j$  且  $B_{j_0} < P$ ,则将  $Z$  归入社团正域  $Q_{\text{POS}}(A_{j_0})$ ;

(x) 更新  $Q_{\text{POS}}(A_{j_0})$  的中心  $r_{j_0}$ ;

(xi) 更新边界域:  $Q_{\text{BND}}(A_j) = Q_{\text{BND}}(A_j) - \{Z\}$

$(j = i_1, i_2, \dots, i_l)$ ;

(xii) 若  $Q_{\text{BND}}(A_l) \neq \emptyset (l = 1, 2, \dots, k)$ ,则转步骤(ix),否则,转步骤(xiii);

(xiii) 输出社团正域  $Q_{\text{POS}}(A_i) (i = 1, 2, \dots, k)$ .

算法2的步骤(i)~(iii)实际上执行的是算法1的内容,对重叠社团的划分起到初始化的作用.即算法2是算法1引入吸收度概念后的改进,从而能更好地处理重叠社团节点的划分.其中,在步骤(vi)中的  $F$  吸收度用于形成社团边界域,可以刻画社团的重叠区.步骤(ix)是对社团重叠区的节点(即边界域中的节点)进行最终的社团归属判决.  $P$  吸收度用于刻画边界域中节点的划分.  $F$  吸收度作用于边界粗社团的形成过程,即若  $F$  吸收度值越大,则所构成的社团结构就越粗糙;  $P$  吸收度作用于区分边界域中社团节点的细化过程,即若  $P$  吸收度越小,则边界社团中节点的划分就越精细.由于算法的边界域是基于多个社团的共同边界点而产生的(即步骤(vi)),不同于一般的三支决策处理类似问题(如三支聚类<sup>[26]</sup>)的结果,所以在本文算法的边界域中的节点同时为多个社团潜在的节点,这为步骤(ix)~(xii)在多社团边界域上开展更新提供了必要条件.

算法输出为社团正域,可以证明这些社团正域满足如下的性质.

性质3 在无向无权网络  $G = \langle V, E \rangle$  中,经过算法2产生的社团正域  $Q_{\text{POS}}(A_i) (i = 1, 2, \dots, k)$  满足:

(i)  $\bigcup_{i=1,2,\dots,k} Q_{\text{POS}}(A_i) = V$ ;

(ii)  $Q_{\text{POS}}(A_i) \cap Q_{\text{POS}}(A_j) = \emptyset (1 \leq i, j \leq k, \text{且 } i \neq j)$ .

即经过算法2后输出的所有社团正域是整个网络节点的一个有效划分.

## 3 算法验证及分析

### 3.1 实验数据与实验环境

为了验证本文所提算法的有效性和可行性,本文采用了3个典型的社交网络数据集作为实验数据集,分别是著名的空手道俱乐部成员网络(Zachary's karate club)、足球联盟网络(American college football)和海豚社会关系网络(Dolphins).数据集可在网上(<http://www-personal.umich.edu/~mejn/netdata>)的数据集中获取.数据集的基本信息如表2所示.

实验环境为英特尔酷睿双核 P8500 处理器,内

存 8 GB 64 位 Windows10 操作系统. 主要编程语言为 Matlab、R 语言编程工具.

表 2 实验数据集

数据集	节点数	边数	平均核度	平均聚集系数
Zachary	34	78	2.206	0.246
Football	115	613	5.330	0.202
Dolphins	62	159	5.129	0.303

### 3.2 评价指标

衡量社交网络中社团划分质量的标准主要有内部评价指标和外部评价指标<sup>[27-29]</sup>.

3.2.1 内部评价指标 在社交网络中, 模块度函数作为社团划分好坏的量化标准已经被广泛使用.

若某个网络包含  $n$  个社团, 则该网络对应的邻接矩阵  $W$  是一个  $n \times n$  的对称矩阵  $e_{ij} (1 \leq i, j \leq n)$  表示在网络中连接 2 个不同社团节点的边在所有边中所占的比例, 这 2 个节点分别位于第  $i$  个社团和第  $j$  个社团.  $T_{re} = \sum_{i=1}^n e_{ij}$  表示某社团内部的边数占该网

络中所有边的比例.  $a_i = \sum_{j=1}^n e_{ij}$  表示所有与社团  $i$  的内部节点相连接的边数占该网络中所有边的比例.

模块度函数  $Q$ <sup>[28]</sup> 定义如下:

$$Q = \sum_{i=1}^n (e_{ij} - a_i^2) = T_{re} - \|W^2\|.$$

若社团内部边的比例不大于在任意连接时的期望值, 则有  $Q = 0$ , 且  $Q$  的上限为 1. 若社团结构越明显, 则越接近 1. 在实际的网络中,  $Q$  的取值范围一般为 0.3 ~ 0.7.

3.2.2 外部评价指标  $N_{MI}$  指标<sup>[29]</sup> 可以用来估计具有已知分区的真实社团结构与社团划分结果之间的相似性.  $N_{MI}$  反映了划分的社团结构与真实社团结构非常相似, 若  $N_{MI}$  值为 1, 则 2 个社团结构完全相同, 若  $N_{MI}$  值为 0, 则 2 个社团结构完全不同. 其计算公式为

$$N_{MI}(X|Y) = 1 - (H(X|Y) + H(Y|X)) / 2,$$

其中  $X$  表示原社团结构的集合;  $Y$  表示使用本算法得到的社团结构的集合;  $H(X|Y)$  表示  $X$  在  $Y$  上的规范化条件熵;  $H(Y|X)$  表示  $Y$  在  $X$  上的规范化条件熵.

### 3.3 实验与结果分析

3.3.1 吸收度与模块度的关系 从 3WD-PPOC 的算法描述可以看出, 损失函数通过吸收度  $F$ 、 $P$  来影

响社团划分的效果, 图 2 直观展示了本文算法 (3WD-PPOC) 的吸收度  $F$ 、 $P$  对模块度  $Q$  的影响. 从图 2 可以看出, 对于 Zachary 数据集, 当损失函数为  $\lambda_{PY_1} = 0, \lambda_{PY_3} = 6.4, \lambda_{NY_1} = 13.6, \lambda_{NY_3} = 0, \lambda_{BY_1} = 4, \lambda_{BY_3} = 0.4$  时  $F$ 、 $P$  吸收度为  $(F, P) = (0.60, 0.40)$ , 此时模块度参数最优值为 0.417; 对于 Football 数据集, 当损失函数为  $\lambda_{PY_1} = 0, \lambda_{PY_3} = 5.6, \lambda_{NY_1} = 15.6, \lambda_{NY_3} = 0, \lambda_{BY_1} = 4.5, \lambda_{BY_3} = 0.1$  时  $F$ 、 $P$  吸收度为  $(F, P) = (0.550, 0.005)$ , 模块度参数最大值为 0.604; 对于 Dolphins 数据集, 当损失函数为  $\lambda_{PY_1} = 0, \lambda_{PY_3} = 5.6, \lambda_{NY_1} = 15.6, \lambda_{NY_3} = 0, \lambda_{BY_1} = 4.5, \lambda_{BY_3} = 0.1$  时  $F$ 、 $P$  吸收度为  $(F, P) = (0.550, 0.005)$ , 此时模块度参数最佳值为 0.546.

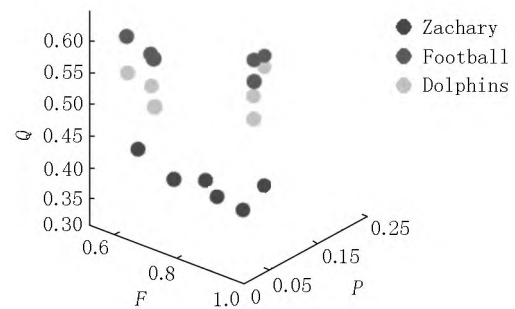


图 2 基于实验数据集吸收度与模块度的关系图

3.3.2 基于典型数据集的划分效果 (i) 基于 Zachary 空手道俱乐部网络实验. Zachary 空手道俱乐部网络<sup>[30]</sup> 是美国某大学空手道俱乐部的关系网络, 该网络包含 34 个节点及 78 条边, 其中节点表示俱乐部成员, 边表示成员之间存在的关系. Zachary 空手道俱乐部成员关系网络是复杂网络、社团发现技术等领域的典型测试网络数据集, 在网络中的人物关系因某种原因而被分成若干个小社团, 该网络的原始结构如图 3 所示.

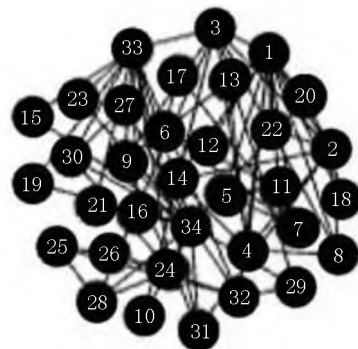


图 3 Zachary 网络初始社团结构

据上述“吸收度与模块度的关系”实验, 得到社团划分结构, 即将 Zachary 网络划分为 4 个社团, 如图 4 所示.

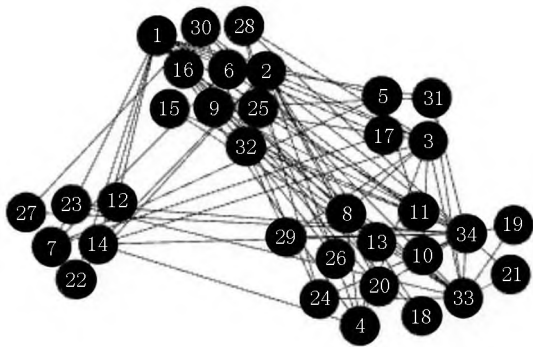


图 4 3WD-PPOC 算法对 Zachary 网络的社团划分结果

(ii) 基于 Football 足球联盟网络实验. Football 数据集是经典的社团研究数据集之一, 该网络由 115 个球队的 613 场比赛抽象而成, 如何根据不同球队之间的实力合理划分球队, 并合理安排相应的赛事是该实验关注的重点, 该网络的原始结构如图 5 所示.

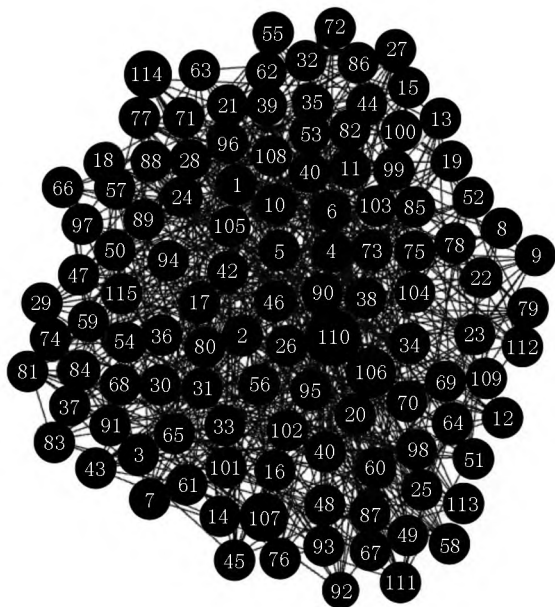


图 5 Football 网络初始社团结构

根据上述“吸收度与模块度的关系”实验, 得到社团划分结构, 将 Football 网络划分为 6 个社团, 社团划分的结果如图 6 所示.

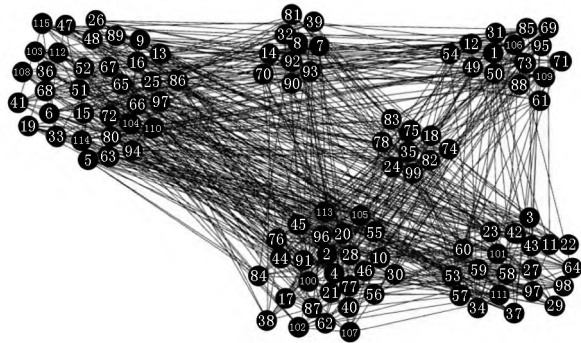


图 6 3WD-PPOC 算法对 Football 网络的社团划分结果  
(iii) 基于 Dolphins 海豚社会关系网络实验.

Dolphin 海豚数据集是 D. Lusseau 等使用长达 7 a 的时间观察新西兰 Doubtful Sound 海峡 62 只海豚群体的交流情况而得到的海豚社会关系网络. 这个网络具有 62 个节点及 159 条边, 节点表示海豚, 而边表示海豚间的接触的频率, 该网络的原始结构如图 7 所示.

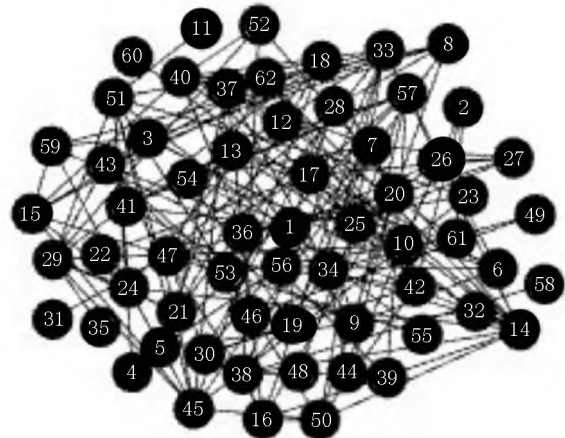


图 7 Dolphins 网络初始社团结构

根据上述“吸收度与模块度的关系”实验, 得到社团划分结构, 将 Dolphins 网络划分为 4 个社团, 社团划分的结果如图 8 所示.

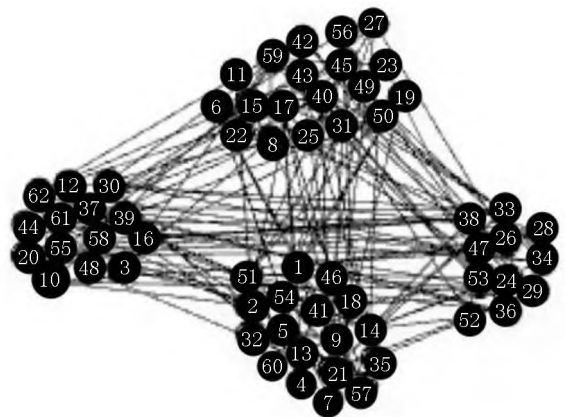


图 8 3WD-PPOC 算法对 Dolphins 网络社团划分结果

从图 4、图 6 和图 8 可以看出, 划分后 3 个数据集的社团结构比较紧密, 这说明本文的算法对边界域的节点得到了合理的划分.

3.4 模块度、 $N_{MI}$  值、时间复杂度分析

为验证本算法的有效性, 本文首先将 PPC 算法 (本文算法 1) 和 3WD-PPOC 算法进行模块度  $Q$  值对比, 结果如表 3 所示.

表 3 基于实验数据集的 PPC 算法和 3WD-PPOC 的模块度  $Q$  值对比

模块度	Zachary 网络	Football 网络	Dolphins 网络
PPC	0.395	0.601	0.501
3WD-PPOC	0.417	0.604	0.546

从表 3 可以看出,在引入吸收度后 Zachary 网络、Football 网络及 Dolphins 网络的模块度均大于没有引入吸收度的模块度,即吸收度的引入可以使延迟决策划分到边界域的重叠节点做出 2 次决策,使划分后的社团结构更加紧密。基于吸收度的决策结果,对重叠社团的划分好坏有较显著的影响,

表 4 各算法在实验数据集上的  $Q$  值对比

Network	3WD-PPOC 算法	Newman 算法	GN 算法	LMF 算法	文献[13] 算法	文献[18] 算法
Zachary 网络	0.417	0.371	0.401	0.20	0.20	0.39
Football 网络	0.604	0.546	0.601	0.22	0.29	0.59
Dolphins 网络	0.546	0.379	0.455	0.24	0.28	0.47

表 5 各算法在实验数据集上的  $N_{MI}$  值对比

Network	3WD-PPOC 算法	Newman 算法	GN 算法	LMF 算法	文献[13] 算法	文献[18] 算法
Zachary 网络	0.941	0.95	0.93	0.59	0.69	0.62
Football 网络	0.817	0.81	0.82	0.17	0.22	0.66
Dolphins 网络	0.935	0.75	0.80	0.18	0.31	0.71

表 6 算法时间复杂度对比

算法	时间复杂度
3WD-PPOC 算法	$O(n^2)$
Newman 算法	$O(n^2)$
GN 算法	$O(n^3)$
LMF 算法	$O(n^2 \log n)$
文献[13] 算法	$O(n^3)$
文献[18] 算法	$O(n^2)$

从表 4 ~ 表 6 可以看出:本文所提出的 3WD-PPOC 算法和 Newman 算法、文献[18] 算法的时间复杂度均为  $O(n^2)$ ,其他算法的时间复杂度都高于  $O(n^2)$ ,这表明 3WD-PPOC 具有良好的计算开销。在 Zachary 网络、Football 网络、Dolphins 网络中,3WD-PPOC 都获得了最高的模块度值,3WD-PPOC 的  $N_{MI}$  值在 3 个实验数据集上均在 0.8 以上,除了在 Zachary 网络数据集上的  $N_{MI}$  值略低于 Newman 算法外,在其他网络数据集上均优于其他比较算法。Newman 算法属于贪心算法的一种,它通过不断迭代更新形成新的社团结构,社团划分结果能较好地刻画社团间的关系。Newman 算法与 3WD-PPOC 的时间复杂度相同,但在模块度方面 Newman 算法低于 3WD-PPOC。在实验数据集上,3WD-PPOC 的模块度  $Q$  值比 Newman 算法、GN 算法、LMF 算法的分别提升了 12.4%、10.6%、44.0%,这表明 3WD-PPOC 比 Newman 算法在刻画社团内部节点连接稳定性方面具有更好的优势。特别是在 Dolphins 网络上,3WD-PPOC 的  $N_{MI}$  值为 0.935,远高于其他比较算法,这说明 3WD-PPOC 算法对该数据集的社团划分精度已达到相当高的程度,划分结构的质量优良。

综上所述,本文所提出的 3WD-PPOC 算法在处

边界域中的重叠节点的划分更为合理和稳定。

为了进一步验证算法性能,本文将 3WD-PPOC 算法与经典的 Newman 算法、GN 算法、重叠社团划分算法 LMF 算法、重叠社团划分最新算法(文献[13]、文献[18])进行模块度值、 $N_{MI}$  值比较,结果如表 4 和表 5 所示。各算法的时间复杂度如表 6 所示。

理社团网络划分问题上具有一定优势,在保持较好的处理时间开销下还能有效地对复杂网络节点进行社团划分,且划分出来的社团内部节点具有较好的连接稳定性。

## 4 结束语

本文将三支决策的思想应用于重叠区域的社团划分,提出了一种基于吸收度的三支决策社团划分算法。该算法根据社团的重要度矩阵和吸收度产生社团重叠区,再通过三支决策建立社团节点与边的界域、正域、负域的对应关系。三支决策思想的引入,有效提高了社团划分的质量。基于真实数据的实验结果表明:本文所提算法能够有效地进行社团划分, $F$  吸收度刻画了社团边界的细节, $P$  吸收度的引入则可以增加边界域重叠节点的归属程度,即提高了社团的模块度。对比其他社团划分算法,本文所提算法在实验网络中能取得较高的划分质量。划分后的各社团结构紧密,这表明该算法对社团重叠节点的划分具有较好的稳定性。下一步将考虑以提高社团的  $N_{MI}$  值为目标改进初始重叠社团的划分方法。

## 5 参考文献

- [1] 吴样平,郭飞,曾明华.基于复杂网络的城市综合交通网络特征分析与优化研究[J].江西师范大学学报(自然科学版),2015,39(3):326-330.
- [2] MUTAMBIK I,ALMUQRIN A,LIU Yulong,et al.Gender differentials on information sharing and privacy concerns

- on social networking sites: perspectives from users [J]. Journal of Global Information Management ,2021 ,29( 3) : 50110.
- [3] ZHAO Da ,YI Yaqiao ,HE Qi ,et al.Exploring the regulatory mechanism of Nao Tai Fang on vascular dementia's biological network based on cheminformatics and transcriptomics strategy [J].Journal of Ethnopharmacology ,2021 , 274( 1) : 114065.
- [4] LÁZÁR Z I ,PAPP I ,VARGA L ,et al.Stochastic graph Voronoi tessellation reveals community structure [J]. Physical Review ,2017 ,95( 2) : 22306.
- [5] FILIPPO R ,CLAUDIO C ,FEDERICO C ,et al.Defining and identifying communities in networks [J].Proceedings of the National Academy of Sciences of the United States of America ,2004 ,101( 9) : 2658-2663.
- [6] NEWMAN M E J.Detecting community structure in networks [J]. The European Physical Journal B ,2004 , 38( 2) : 321-330.
- [7] XU Shenmeng ,WANG Xianwen ,LIU Zeyuan ,et al.Network structural analysis of technology: a study from patent perspective [J].Journal of Science and Technology Policy in China ,2013 ,4( 3) : 214-235.
- [8] KERNIGHAN B W ,LIN S.An efficient heuristic procedure for partitioning graphs [J].Bell System Technical Journal , 1970 ,49( 2) : 291-307.
- [9] 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 and Soft Matter Physics ,2007 ,76( 3) : 36106.
- [10] LIN Hanyang ,ZHAN Yongzhao ,ZHAO Zizheng ,et al. Overlapping community detection based on attribute augmented graph [J].Entropy ,2021 ,23( 6) : 680.
- [11] 李敏毓 ,陈兴蜀 ,尹雅丽 等.面向多源社交网络的社团结构特征研究 [J].工程科学与技术 ,2017 ,49( S2) : 195-202.
- [12] LANCICHINETTI A ,FORTUNATO S ,KERTÉSZ J.Detecting the overlapping and hierarchical community structure of complex networks [J].New Journal of Physics ,2008 ,11 ( 3) : 19-44.
- [13] 郭娜 ,郑晓艳.基于最大生成树的重叠社团发现算法 [J].计算机应用研究 ,2020 ,37( S2) : 170-172.
- [14] LIU Dun ,YAO Yiyu ,LI Tianrui.Three-way investment decisions with decision-theoretic rough sets [J].International Journal of Computational Intelligence Systems ,2011 , 4( 1) : 66-74.
- [15] XING Ruikang ,LI Chenghai ,ZHANG Xin ,et al.Multigranulation decision-theoretic rough set based on incomplete interval-valued information systems [EB/OL]. [2021-11-18].<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=IPFD&filename=LRCM201808001057>.
- [16] ANDREAS M ,RUDOLF S.Vague information processing [J].HMD Praxis der Wirtschaftsinformatik ,2018 ,55( 3) : 465-466.
- [17] 杨雪洁 ,曹风云 ,陈洁 等.基于子模优化的边界域处理社团发现算法 [J].电子测量与仪器学报 ,2020 , 34( 4) : 111-117.
- [18] 方莲娣 ,张燕平 ,陈洁 等.基于三支决策的非重叠社团划分 [J].智能系统学报 ,2017 ,12( 3) : 293-300.
- [19] 聂祥林 ,张玉梅 ,吴晓军 等.基于节点依赖度和相似社团融合的社团结构发现算法 [J].计算机工程与科学 , 2017 ,39( 7) : 1273-1280.
- [20] 陈东明 ,王云开 ,黄新宇 等.基于社团密合度的复杂网络社团发现算法 [J].东北大学学报(自然科学版) , 2019 ,40( 2) : 186-191.
- [21] ANIRBAN B ,MARK R.Sharp transition of the invertibility of the adjacency matrices of sparse random graphs [J]. Probability Theory and Related Fields ,2021 ,180( 5) : 233-308.
- [22] 吴卫江 ,周静 ,李国和.一种基于节点重要度的社团划分算法 [J].中南民族大学学报(自然科学版) ,2016 , 35( 1) : 119-122.
- [23] 胡峰 ,张苗 ,于洪.基于三支决策的主动学习方法 [J].控制与决策 ,2019 ,34( 4) : 718-726.
- [24] 姚一豫 ,祁建军 ,魏玲.基于三支决策的形式概念分析、粗糙集与粒计算 [J].西北大学学报(自然科学版) , 2018 ,48( 4) : 477-487.
- [25] SUN Bingzhen ,CHEN Xiangtang ,ZHANG Liye ,et al. Three-way decision making approach to conflict analysis and resolution using probabilistic rough set over two universes [J].Information Sciences ,2020 ,507: 809-822.
- [26] YU Hong ,WANG Xincheng ,WANG Guoyin. An active three-way clustering method via low-rank matrices for multi-view data [J].Information Sciences ,2020 ,507: 823-839.
- [27] 唐益明 ,丰刚永 ,任福继 等.面向结构复杂数据集的模糊聚类有效性指标 [J].电子测量与仪器学报 ,2018 , 32( 4) : 119-127.
- [28] NEWMAN M E J ,GIRVAN M. Finding and evaluating community structure in networks [J].Physical Review E: Statistical , Nonlinear , and Soft Matter Physics ,2004 , 69( 2) : 26113.
- [29] SHAO Chenxi ,DUAN Yubing. Identifying community structure in complex networks [J].International Journal of Modern Physics B ,2015 ,29( 19) : 65-67.
- [30] ZACHARY W. An information flow model for conflict and fission in small groups [J].Journal of Anthropological Research ,1977 ,33( 4) : 452-473.

(下转第330页)



## The Galvanometer Position Tracking Based on Kalman Filter and Discrete Sliding Mode

LI Cong ,WANG Zhonghua\*

( College of Information Engineering ,Nanchang Hangkong University ,Nanchang Jiangxi 330063 ,China)

**Abstract:** In order to solve the influence of external disturbance and noise imposed on the trajectory state of galvanometer scanning mirror ,the control method combining Kalman filter and discrete sliding mode is proposed to track desired angular position. Firstly ,around the discrete mathematical model of the driving motor of the galvanometer system ,the Kalman filter is constructed to estimate the real motion change of the galvanometer system ,so as to reduce the influence of external disturbance and noise on the system. Secondly ,the tracking error is corrected by discrete sliding mode control ,which makes the position and velocity of the galvanometer system stable quickly ,and further improves the anti-interference ability and position tracking ability of the system. Finally ,the comparative simulation experiment and the verification experiment of galvanometer system platform are designed for two both instantaneous and continuous types of external disturbances. The simulation and experimental results show that under two different types of external disturbances ,the control method can not only make the scanning mirror of the galvanometer track the expected value quickly and accurately ,but also suppress the chattering problem of discrete sliding mode control effectively.

**Key words:** galvanometer system; Kalman filter; discrete sliding mode control; position tracking

( 责任编辑: 冉小晓)

( 上接第 321 页)

## The Absorbance-Based Community Division Algorithm with Three-Way Decision

WAN Renxia<sup>1</sup> ,ZHANG Yuhong<sup>1</sup> ,MIAO Duoqian<sup>2</sup>

( 1.College of Mathematics and Information Science ,North Minzu University ,Yinchuan Ningxia 750021 ,China;

2.College of Computer Science and Technology ,Tongji University ,Shanghai 201804 ,China)

**Abstract:** The idea of three-way decision is introduced to solve the overlapping problem of community division ,and the community division algorithm ( 3WD-PPOC) based on absorbance is proposed. In the 3WD-PPOC algorithm ,the initial division of communities is firstly carried out according to the importance matrix of the network structure ,and then the overlap regions between communities ,namely the boundary region of communities ,are constructed by the  $F$  absorbance and the positive region of each community is obtained. Finally ,the redistribution of the nodes in the community boundary regions and the updating of the community positive regions are completed by  $P$  absorbance. Compared with other community division algorithms ,3WD-PPOC has lower time complexity. The experimental results further show that 3WD-PPOC can effectively divide communities. Contrast to the comparison algorithms ,3WD-PPOC has better community division quality. And the community structure after division is close ,which indicates that the algorithm has good stability for dividing the overlapping nodes in community.

**Key words:** community division; three-way decisions; absorbance; importance matrix; positive region; boundary region

( 责任编辑: 冉小晓)