

采用离散粒子群算法的复杂网络重叠社团检测

陈国强^{1,2}, 王宇平¹

(1. 西安电子科技大学计算机学院, 710071, 西安; 2. 河南大学计算机与信息工程学院, 475004, 河南开封)

摘要: 针对复杂网络重叠社团检测的问题, 建立了衡量重叠社团划分优劣的评判函数, 并基于该函数提出了一种采用混沌量子粒子群优化的复杂网络重叠社团检测算法。该算法采用量子编码, 利用 Logistic 映射初始化粒子种群, 并以粒子群速度改变方式更新转角的大小, 以概率为 1 收敛。实验结果表明, 所提算法具有较高的检测正确率, 能更好地反映实际社团划分情况。

关键词: 复杂网络; 社团检测; 重叠社团; 离散粒子群算法

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2013)01-0107-07

Overlapping Community Detection of Complex Networks Based on Discrete Particle Swarm Algorithm

CHEN Guoqiang^{1,2}, WANG Yuping¹

(1. School of Computer Science and Technology, Xidian University, Xi'an 710071, China;

2. School of Computer and Information Engineering, Henan University, Kaifeng, Henan 475004, China)

Abstract: An evaluation function of overlapping community is proposed to solve overlapping community detection of complex networks. A chaos quantum discrete particle swarm algorithm is presented to detect overlapping community structure based on the proposed function. The quantum coding and the logistic mapping are used to generate an initial particle population, and the particle swarm speed changing method is used to update the size of the corner. It is shown that the algorithm converges with probability 1. Experimental results show that the algorithm has higher detection accuracy and can better reflect the real situation of the community divided.

Keywords: complex networks; community detection; overlapping community; discrete particle swarm algorithm

社会、生物、科技等领域中的许多复杂系统都可用网络的形式来表示^[1-2]。复杂网络的关键属性之一是社团结构, 即顶点聚集在不同的簇或组之中, 同簇中顶点之间的连接比较稠密, 不同簇之间的顶点连接比较稀疏^[3]。在单个顶点层次, 由于网络太大, 以致于无法把握网络的结构, 所以需采用网络社团的自动识别来揭示粗粒度层次的网络结构。

目前, 已提出许多社团检测算法^[4-6], 但大部分是检测非重叠社团的, 不适用于检测重叠社团。针对重叠社团检测问题, 文献[7]首次提出了重叠社团

概念及基于派系渗透的重叠社团检测算法, 但采用该算法时某些顶点可能被排除在任何社团之外。另外, 文献[8]的衡量社团划分优劣的模块度测度已被很多研究者所采用, 文献[9]基于最大派系将模块度进行了扩展, 文献[10]基于属于系数将模块度扩展为具有重叠社团结构的有向图, 但该模块度存在分辨极限的问题^[11-13]。为了克服分辨极限的问题, 文献[14]提出采用模块密度来量化社团结构, 但它仅考虑非重叠社团结构, 不能直接检测重叠社团。

基于上述, 本文将模块密度进行了扩展, 提出了

收稿日期: 2012-03-21。 作者简介: 陈国强(1977—), 男, 博士生, 副教授; 王宇平(通信作者), 男, 教授, 博士生导师。
基金项目: 国家自然科学基金资助项目(61272119, 61203372)。

网络出版时间: 2012-10-31

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20121031.1106.009.html>

量化重叠社团的新测度,构造了解决复杂网络重叠社团的新算法,并将计算机生成的数据与现实数据进行了仿真对比。

1 重叠社团量化测度

定义1 模块密度。给定一个网络 $G=(V,E)$, A 是网络 G 的邻接矩阵, V 是顶点集, E 是边集。设 $V_1 \subseteq V$, $V_2 \subseteq V$ 且 $V_1 \cap V_2 = \emptyset$, 则 $L(V_1, V_2) = \sum_{i \in V_1, j \in V_2} A_{ij}$, $L(V_1, V_1) = \sum_{i \in V_1, j \in V_1} A_{ij}$, 且 $L(V_1, \bar{V}_1) = \sum_{i \in V_1, j \in \bar{V}_1} A_{ij}$ ($\bar{V}_1 = V - V_1$)。给定 G 的一个划分 $P = \{G_1(V_1, E_1), \dots, G_m(V_m, E_m)\}$, V_i 和 E_i 分别是 $G_i (i=1, \dots, m)$ 的顶点集和边集, 则衡量不可重叠社团划分优劣的模块密度为^[14]

$$D = \sum_{i=1}^m \frac{L(V_i, V_i) - L(V_i, \bar{V}_i)}{|V_i|} \quad (1)$$

设分块

$$c \in \{V_1, V_2, \dots, V_m\}$$

$$\sigma_{ic} = \begin{cases} 1, & i \in c \\ 0, & i \notin c \end{cases}$$

$|c|$ 为分块 $c (c \in P)$ 中的顶点数, 则模块密度又可表示为

$$D = \frac{\sum_{c \in \{V_1, V_2, \dots, V_m\}} \frac{\sum_{i \in c, j \in c} A_{ij} - \sum_{i \in c, j \notin c} A_{ij}}{|c|}}{\sum_{c \in \{V_1, V_2, \dots, V_m\}} \frac{\sum_{i \in V, j \in V} \sigma_{ic} \sigma_{jc} A_{ij} - \sum_{i \in V, j \in V} \sigma_{ic} (1 - \sigma_{jc}) A_{ij}}{|c|}} \quad (2)$$

如果顶点 i 属于分块 c , 则 σ_{ic} 取值为 1, 否则 σ_{ic} 取值为 0。对于重叠社团, 一个顶点可以属于多个分块, 因此在衡量重叠社团划分优劣时还需考虑顶点属于社团的程度。

定义2 覆盖模块密度 D_c 。设 G 的一个覆盖

$$C = \{c_1, c_2, \dots, c_m\}, \quad \bigcup_{i=1}^n c_i = V$$

顶点 i 属于分块 $c (c \in C)$ 的属于系数为 α_{ic} , 则衡量可重叠社团划分优劣的度量标准

$$D_c = \sum_{c \in C} \left(\sum_{i \in V, j \in V} \alpha_{ic} \alpha_{jc} \sigma_{ic} \sigma_{jc} A_{ij} - \sum_{i \in V, j \in V} \alpha_{ic} (1 - \alpha_{jc}) \sigma_{ic} (1 - \sigma_{jc}) A_{ij} \right) / \sum_{k \in V} \alpha_{kc} \quad (3)$$

定义3 属于系数 α_{ic} 。连接顶点的边数属于社团的边数占总边数的比例

$$\alpha_{ic} = \frac{1}{\alpha_i} \frac{\sum_{j \in c} A_{ij}}{d_i} \quad (4)$$

$$\alpha_i = \sum_{c \in C} \frac{\sigma_{ic} \sum_{j \in c} A_{ij}}{d_i} \quad (5)$$

$$\sum_{c \in C} \alpha_{ic} = 1 \quad (6)$$

根据式(3)和式(4)可以获得网络覆盖划分的测度。值得注意的是, 如果任意顶点只属于唯一的分块, 即为非重叠社团, 则 $\alpha_{ic} = \sigma_{ic}$, 此时 $D_c = D$ 。也就是说, 本文定义的测度即适用于重叠社团度量, 也适用于非重叠社团度量。基于 D_c , 重叠社团检测可以看作是找到最大化 D_c 的一种覆盖划分。该问题是一个典型的 NP 难问题, 如果网络顶点较多, 则很难在合理的时间范围内获得全局最优解。

本文基于量子粒子群算法^[15], 利用混沌初始化种群, 来增加初始种群的遍历性, 由此提出一种采用混沌量子离散粒子群的重叠社团检测算法, 简记为 CQDPSO-CD, 算法中采用了泛化后的 D_c 作为优化目标。

2 CQDPSO-CD

2.1 编码

设 n 表示 G 中的顶点数, m 表示指定的社团数, $l = nm$ 表示粒子的维数, s 表示群体规模(群体数)。本文算法中社团数无需指定, 实现时先将 m 设为最大值 n , 再由算法进化自动确定。如果已知网络社团数, 则指定社团数可以加快算法收敛。

定义4 粒子的位置编码。设矩阵 $M = (\sigma_{uv}) (u = 1, 2, \dots, n; v = 1, 2, \dots, m)$ 为顶点与社团的关系矩阵, σ_{uv} 取值为 0 或 1。如果顶点 u 属于社团 v , 则 $\sigma_{uv} = 1$, 否则 $\sigma_{uv} = 0$ 。第 i 个粒子的位置可用 $l = nm$ 维向量 $p_i = (p_{i1}, p_{i2}, \dots, p_{il}) = (\sigma_{11}, \sigma_{12}, \dots, \sigma_{1m}, \sigma_{21}, \sigma_{22}, \dots, \sigma_{2m}, \dots, \sigma_{n1}, \sigma_{n2}, \dots, \sigma_{nm})$ 来表示, 它代表着网络的一种可重叠社团划分的方案。

如图1所示, 如果将网络分成2个社团 $\{1, 2, 3, 4, 5\}$ 和 $\{5, 6, 7, 8\}$, 则顶点5是重叠顶点, 该顶点与社团的关系矩阵为

$$M = [1, 0; 1, 0; 1, 0; 1, 0; 1, 1; 0, 1; 0, 1; 0, 1]$$

粒子编码为

$$p = (1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1)$$

定义5 量子粒子编码。量子粒子的每一位采用量子比特的方式进行编码, 即用一对实数定义一个量子比特位。一个具有 l 个量子比特位的量子粒子

$$q_i = \begin{bmatrix} \alpha_{i1} & \alpha_{i2} & \cdots & \alpha_{il} \\ \beta_{i1} & \beta_{i2} & \cdots & \beta_{il} \end{bmatrix}$$

且满足 $|\alpha_{ij}|^2 + |\beta_{ij}|^2 = 1 (j=1, 2, \dots, l)$, 若用相位表

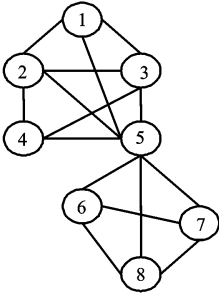


图1 具有重叠社团结构的示例网络

示,则

$$\mathbf{q}_i = \begin{bmatrix} \cos(\theta_{i1}) & \cos(\theta_{i2}) & \cdots & \cos(\theta_{il}) \\ \sin(\theta_{i1}) & \sin(\theta_{i2}) & \cdots & \sin(\theta_{il}) \end{bmatrix}$$

$$\theta_{ij} = \arctan \frac{\alpha_{ij}}{\beta_{ij}}$$

2.2 算法基本流程

本文算法的基本流程如图2所示。

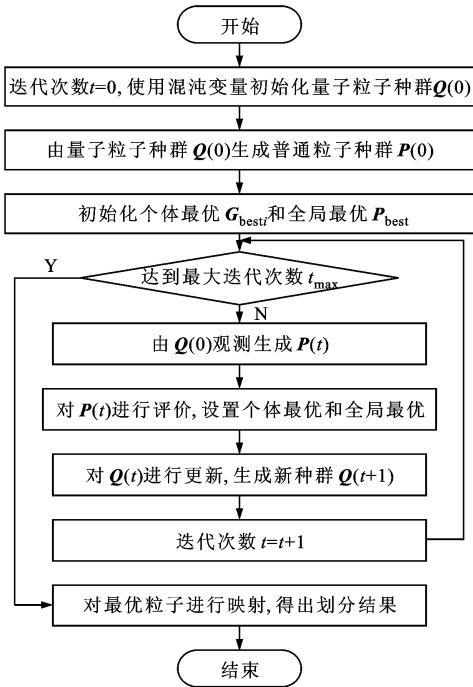


图2 算法的基本流程

2.3 CQDPSO-CD 步骤

步骤1 初始化量子粒子种群 $\mathbf{Q}(0)$ 。迭代次数为 $t=0$, 规模为 s , 粒子编码长度为 $l=nm$, 则初始化量子粒子种群 $\mathbf{Q}(t) = \{\mathbf{q}_1^t, \mathbf{q}_2^t, \dots, \mathbf{q}_s^t\}$, 第 i 个量子粒子

$$\mathbf{q}_i^t = \begin{bmatrix} \alpha_{i1}^t & \alpha_{i2}^t & \cdots & \alpha_{il}^t \\ \beta_{i1}^t & \beta_{i2}^t & \cdots & \beta_{il}^t \end{bmatrix}, \quad i = 1, 2, \dots, s$$

且满足 $|\alpha_{ij}^t|^2 + |\beta_{ij}^t|^2 = 1, j = 1, 2, \dots, l$ 。将第 i 个量子粒子用相位来表示, 则

$$\mathbf{q}_i^t = \begin{bmatrix} \cos(\theta_{i1}^t) & \cos(\theta_{i2}^t) & \cdots & \cos(\theta_{il}^t) \\ \sin(\theta_{i1}^t) & \sin(\theta_{i2}^t) & \cdots & \sin(\theta_{il}^t) \end{bmatrix}$$

$$\theta_{ij}^t = \arctan \frac{\alpha_{ij}^t}{\beta_{ij}^t}$$

为了遍历所有粒子空间, 避免可能出现的冗余, 保证粒子的随机性, 本文对初始量子粒子种群使用了 Logistic 映射, 即将

$$x_{i+1}^j = \gamma x_i^j (1 - x_i^j)$$

$$i = 1, 2, \dots, s; \quad j = 1, 2, \dots, l; \quad 0 < x_i^j < 1$$

作为混沌变量。当混沌变量控制系数 $\gamma=4$ 时, 系统处于完全混沌状态, 此时状态空间为 $(0, 1)$ 。利用 $\gamma=4$ 时的混沌特性, 针对不同的 i , 分别给定不同的混沌变量初始值, 利用 Logistic 映射生成 l 个混沌变量 $x_i^j, j=1, 2, \dots, l$, 用 l 个混沌变量分别初始化量子粒子种群中第 i 个量子粒子上的量子位, 将 $\alpha_{ij}^0, \beta_{ij}^0 (j=1, 2, \dots, l)$ 初始化为 $\cos(2x_i^j \pi), \sin(2x_i^j \pi)$, 得

$$\mathbf{q}_i^0 = \begin{bmatrix} \cos(\theta_{i1}^0) & \cos(\theta_{i2}^0) & \cdots & \cos(\theta_{il}^0) \\ \sin(\theta_{i1}^0) & \sin(\theta_{i2}^0) & \cdots & \sin(\theta_{il}^0) \end{bmatrix}$$

$$\theta_{ij}^0 = \arctan \frac{\alpha_{ij}^0}{\beta_{ij}^0} = \arctan \frac{\cos(2x_i^j \pi)}{\sin(2x_i^j \pi)}$$

步骤2 由量子粒子种群 $\mathbf{Q}(0)$ 生成普通粒子种群 $\mathbf{P}(0)$ 。通过观察 $\mathbf{Q}(0)$ 的状态, 生成一组普通粒子 $\mathbf{P}(0) = \{\mathbf{p}_1^0, \mathbf{p}_2^0, \dots, \mathbf{p}_s^0\}$, 每个普通粒子 $\mathbf{p}_i^0 (i=1, 2, \dots, s)$ 是长度为 l 的二进制串 $\mathbf{p}_i^0 = (p_{i1}^0, p_{i2}^0, \dots, p_{il}^0)$, 而 p_{ij}^0 通过观察概率幅值 $|\alpha_{ij}^0|^2, |\beta_{ij}^0|^2 (j=1, 2, \dots, l)$ 来获得。观察方法为: 随机产生一个 $[0, 1]$ 之间的数 r , 如果 $r > |\alpha_{ij}^0|^2$, 则 $p_{ij}^0 = 1$, 否则 $p_{ij}^0 = 0$ 。观察生成的每个粒子 $\mathbf{p}_i^0 (i=1, 2, \dots, s)$ 代表了一种可能的社团划分方案。

步骤3 初始化个体最优 $\mathbf{P}_{best i}$ 和全局最优 $\mathbf{G}_{best}$ 。设 $\mathbf{P}_{best i}$ 表示普通粒子 $\mathbf{p}_i$ 迄今为止搜索到的最优位置, $\mathbf{G}_{best}$ 表示整个粒子群迄今为止搜索到的最优位置, $\mathbf{Q}_{p_{best i}}$ 表示 $\mathbf{P}_{best i}$ 对应的量子粒子, $\mathbf{Q}_{G_{best}}$ 表示 $\mathbf{G}_{best}$ 对应的量子粒子。根据式(2), 通过覆盖模块密度函数计算可以获得每个粒子的覆盖模块密度 $D_c(\mathbf{p}_i^0)$ 。计算时 $\mathbf{P}_{best i} = \mathbf{p}_i^0, \mathbf{G}_{best} = \max\{D_c(\mathbf{p}_1^0), D_c(\mathbf{p}_2^0), \dots, D_c(\mathbf{p}_s^0)\}$, 同时根据 $\mathbf{P}_{best i}$ 和 $\mathbf{G}_{best}$ 设置对应的 $\mathbf{Q}_{p_{best i}}$ 和 $\mathbf{Q}_{G_{best}}$ 。

步骤4 判断算法终止条件。算法达到最大迭代次数 t_{max} 时迭代终止, 输出最优值, 否则下转步骤5。

步骤5 更新 $\mathbf{Q}(t)$, 生成新种群 $\mathbf{Q}(t+1)$ 。本文中量子粒子种群 $\mathbf{Q}(t)$ 的更新是通过量子旋转门来改变量子粒子的相位实现的。量子粒子的相移量根据种群中最优 $\mathbf{Q}_{G_{best}}$ 和 $\mathbf{Q}_{p_{best i}}$ 的相位差确定。相移量

$$\Delta\theta_{ij}^{t+1} = \omega\Delta\theta_{ij}^t + \lambda_1 r_1 (\theta_{q_{\text{best}}} - \theta_{ij}^t) + \lambda_2 r_2 (\theta_{p_{\text{best } i}} - \theta_{ij}^t) \quad (7)$$

$$\theta_{ij}^{t+1} = \theta_{ij}^t + \Delta\theta_{ij}^{t+1} \quad (8)$$

式中: $\Delta\theta_{ij}^{t+1}$ 表示第 $t+1$ 代第 i 个粒子的第 j 维的相移量; θ_{ij}^{t+1} 表示当前相位; ω 表示权重系数; λ_1, λ_2 表示加速因子; r_1, r_2 为 $[0, 1]$ 之间的随机数, 用来保持群体的多样性。设 $Q(t+1) = \{q_1^{t+1}, q_2^{t+1}, \dots, q_s^{t+1}\}$, 根据式(7)和式(8), 应用量子旋转门更新后的粒子

$$q_i^{t+1} = \begin{bmatrix} \alpha_{i1}^{t+1} & \alpha_{i2}^{t+1} & \cdots & \alpha_{il}^{t+1} \\ \beta_{i1}^{t+1} & \beta_{i2}^{t+1} & \cdots & \beta_{il}^{t+1} \end{bmatrix} = \begin{bmatrix} \cos(\theta_{i1}^{t+1}) & \cos(\theta_{i2}^{t+1}) & \cdots & \cos(\theta_{il}^{t+1}) \\ \sin(\theta_{i1}^{t+1}) & \sin(\theta_{i2}^{t+1}) & \cdots & \sin(\theta_{il}^{t+1}) \end{bmatrix} \quad i = 1, 2, \dots, s \quad (9)$$

步骤6 由量子粒子种群 $Q(t)$ 生成普通粒子种群 $P(t)$ 。通过观察 $Q(t)$ 的状态, 生成一组普通解 $P(t) = \{p_1^t, p_2^t, \dots, p_s^t\}$ 。每个 $p_i^t (i=1, 2, \dots, s)$ 是长度为 l 的二进制串, 由概率幅值 $|\alpha_{ij}^t|^2, |\beta_{ij}^t|^2 (j=1, 2, \dots, l)$ 观察得到。观察方法同步骤2。

步骤7 评估适应度函数, 更新个体最优和全局最优解。由式(2)计算每个粒子的覆盖模块密度 $D_c(p_i^t)$: 如果 $D_c(p_i^t) > D_c(p_{\text{best } i})$, 则置 $p_{\text{best } i} = p_i^t$; 如果 $G_{\text{best}} < \max\{D_c(p_1^t), D_c(p_2^t), \dots, D_c(p_s^t)\}$, 则置 $G_{\text{best}} = \max\{D_c(p_1^t), D_c(p_2^t), \dots, D_c(p_s^t)\}$, 并根据 $p_{\text{best } i}$ 和 G_{best} 更新对应的 $Q_{p_{\text{best } i}}$ 和 $Q_{p_{\text{best } i}}$ 。

步骤8 执行 $t=t+1$, 上转步骤4。

2.4 算法收敛性分析

定理1 CQDPSO-CD 以概率1收敛到问题的全局最优解。

证明1 在 CQDPSO-CD 算法中, 对于任意2个粒子 $p_a = (p_{a1}, p_{a2}, \dots, p_{al})$ 、 $p_b = (p_{b1}, p_{b2}, \dots, p_{bl})$, p_a 是由 p_b 通过粒子进化可达的。

任意 $p_a = (p_{a1}, p_{a2}, \dots, p_{al})$ 存在任意量子粒子

$$q_a = \begin{bmatrix} \cos(\theta_{a1}) & \cos(\theta_{a2}) & \cdots & \cos(\theta_{al}) \\ \sin(\theta_{a1}) & \sin(\theta_{a2}) & \cdots & \sin(\theta_{al}) \end{bmatrix} \quad \cos\theta_{ai} \neq 1; \quad \cos\theta_{ai} \neq 0 \quad i = 1, 2, \dots, l$$

而 q_a 由观察方法生成 p_a 的概率为

$$\Psi\{Q_{p(q_a)} = p_a\} = |p_{a1} - \cos^2(\theta_{a1})| \cdot |p_{a2} - \cos^2(\theta_{a2})| \cdots |p_{al} - \cos^2(\theta_{al})| > 0$$

式中: $Q_{p(q_a)}$ 表示由量子粒子 q_a 生成的普通粒子。

量子粒子 q_a 更新后生成的新的量子粒子

$$U_{p(q_a)} = \begin{bmatrix} \cos(\theta_{b1}) & \cos(\theta_{b2}) & \cdots & \cos(\theta_{bl}) \\ \sin(\theta_{b1}) & \sin(\theta_{b2}) & \cdots & \sin(\theta_{bl}) \end{bmatrix}$$

$$\theta_{bi} = \theta_{ai} + \Delta\theta_{bi}$$

$$\Delta\theta_{bi} = \omega\Delta\theta_{bi} + c_1 r_1 (\theta_{q_{\text{best}}} - \theta_{bi}) + c_2 r_2 (\theta_{p_{\text{best } i}} - \theta_{bi})$$

因为 r_1, r_2 为 $[0, 1]$ 之间的随机数, 所以 $\cos\theta_{bi} \in [-1, 1]$ 。设

$$q_b = \begin{bmatrix} \cos(\theta_{b1}) & \cos(\theta_{b2}) & \cdots & \cos(\theta_{bl}) \\ \sin(\theta_{b1}) & \sin(\theta_{b2}) & \cdots & \sin(\theta_{bl}) \end{bmatrix} \quad \cos\theta_{bi} \neq 1; \quad \cos\theta_{bi} \neq 0$$

$$i = 1, 2, \dots, l$$

则由量子粒子 q_a 更新为量子粒子 q_b 的概率

$$\Psi\{U_{p(q_a)} = q_b\} > 0$$

同理, 由 q_b 观察生成 p_b 的概率

$$\Psi\{Q_{p(q_b)} = p_b\} = |p_{b1} - \cos^2(\theta_{b1})| \cdot |p_{b2} - \cos^2(\theta_{b2})| \cdots |p_{bl} - \cos^2(\theta_{bl})| > 0$$

综上, 对于任意粒子 $p_a = (p_{a1}, p_{a2}, \dots, p_{al})$, 存在任意量子粒子

$$q_a = \begin{bmatrix} \cos(\theta_{a1}) & \cos(\theta_{a2}) & \cdots & \cos(\theta_{al}) \\ \sin(\theta_{a1}) & \sin(\theta_{a2}) & \cdots & \sin(\theta_{al}) \end{bmatrix} \quad \cos\theta_{ai} \neq 1; \quad \cos\theta_{ai} \neq 0 \quad i = 1, 2, \dots, l$$

q_a 通过观察方法生成粒子 p_a 的概率 $\Psi\{Q_{p(q_a)} = p_a\} > 0$; q_a 更新生成新的量子粒子

$$q_b = \begin{bmatrix} \cos(\theta_{b1}) & \cos(\theta_{b2}) & \cdots & \cos(\theta_{bl}) \\ \sin(\theta_{b1}) & \sin(\theta_{b2}) & \cdots & \sin(\theta_{bl}) \end{bmatrix} \quad \cos\theta_{bi} \neq 1; \quad \cos\theta_{bi} \neq 0 \quad i = 1, 2, \dots, l$$

的概率 $\Psi\{U_{p(q_a)} = q_b\} > 0$; 量子粒子 q_b 通过观察方法生成 p_b 的概率 $\Psi\{Q_{p(q_b)} = p_b\} > 0$ 。所以, 对于任意2个粒子 p_a 和 p_b , p_a 进化为 p_b 的概率

$$\Psi\{E_{G(p_a)} = p_b\} \geq \Psi\{U_{p(q_a)} = q_b\} \cdot \Psi\{Q_{p(q_b)} = p_b\} > 0$$

即对于可行域中的任意2个粒子 p_a 和 p_b , p_b 是由 p_a 通过粒子进化可达的。

证明2 设种群序列为 $X(t) = P(t) \cup G_{\text{best}}, t=0, 1, 2, \dots$, 由 CQDPSO-CD 可知, 粒子更新时会在每一代中把全局最优个体记录、保留下来, 所以种群序列 $X(0), X(1), \dots, X(t), \dots$ 显然是关于目标函数单调的, 即满足 $\forall t$, 且有

$$\max\{D_c(x) | x \in X(t)\} \leq$$

$$\max\{D_c(x) | x \in X(t+1)\}$$

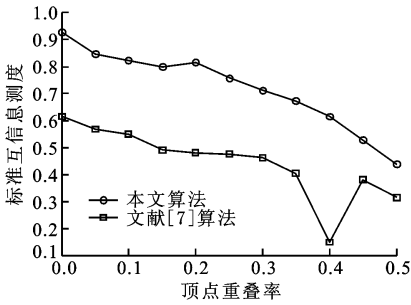
根据文献[16]的原理, 综合证明1和证明2可知, CQDPSO-CD 可以概率1收敛到社团分解的全局最优解。

3 仿真实验

3.1 计算机生成的网络

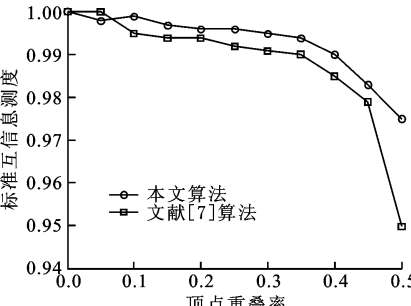
采用文献[17]的网络生成器生成标准测试集^[18],测试集应具有实际网络的典型属性,如度的异构分布性、社团大小的异构分布性等。网络生成器生成的网络允许社团之间重叠,限制重叠的顶点属于固定个数的社团。该生成器包含了多个参数:网络的顶点数 n ,平均度 k ,混合系数 μ ,最大度 $k_{\max}$,顶点度的幂律分布指数 t_1 ,社团大小的幂律分布指数 t_2 ,最小社团大小 $c_{\min}$,最大社团大小 $c_{\max}$,重叠顶点数 n_o ,重叠顶点所属社团数 m_o 。本实验中: $n=1\,000,5\,000$; $c_{\min}=10, c_{\max}=50$ 和 $c_{\min}=20, c_{\max}=100$; $\mu=0.1, 0.3$ 。这样共生成 8 组不同的网络,每组有 10 个典型网络,每组网络运行 20 次。算法参数设置如下: $t_{\max}=1\,000$; $s=50$; $\lambda_1=\lambda_2=2.0$; $\omega=0.4$ 。

采用文献[19]的标准互信息测度来衡量本文算法的性能。在 8 种不同配置的网络中,互信息结果如图 3 和图 4 所示。从图 3 和图 4 可以看出,本文算法的互信息明显优于文献[7],即检测正确率大于文献[7],可见本算法不仅对较小的社团有效,也适用于大规模网络。

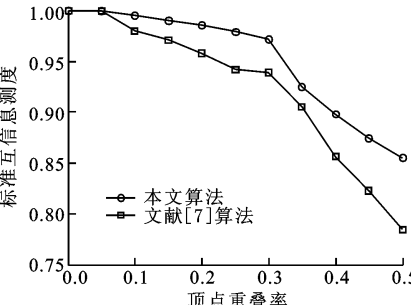


(d) $c_{\min}=20, c_{\max}=100, \mu=0.3$

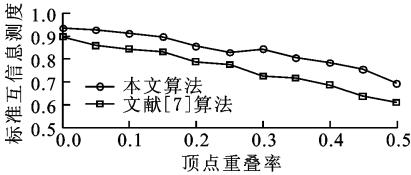
图 3 $n=1\,000, \langle k \rangle=20$ 时算法的标准互信息测度比较



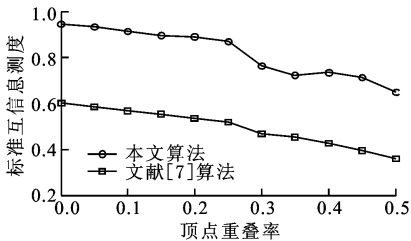
(a) $c_{\min}=10, c_{\max}=50, \mu=0.1$



(b) $c_{\min}=10, c_{\max}=50, \mu=0.3$

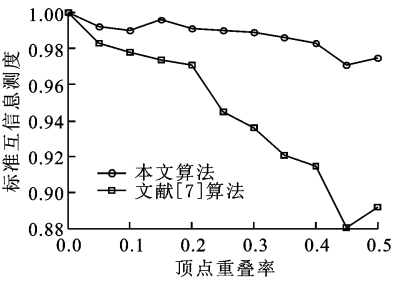


(c) $c_{\min}=20, c_{\max}=100, \mu=0.1$

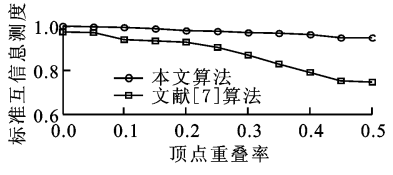


(d) $c_{\min}=20, c_{\max}=100, \mu=0.3$

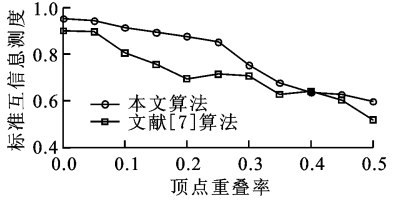
图 4 $n=5\,000, \langle k \rangle=20$ 时算法的标准互信息测度比较



(a) $c_{\min}=10, c_{\max}=50, \mu=0.1$



(b) $c_{\min}=10, c_{\max}=50, \mu=0.3$

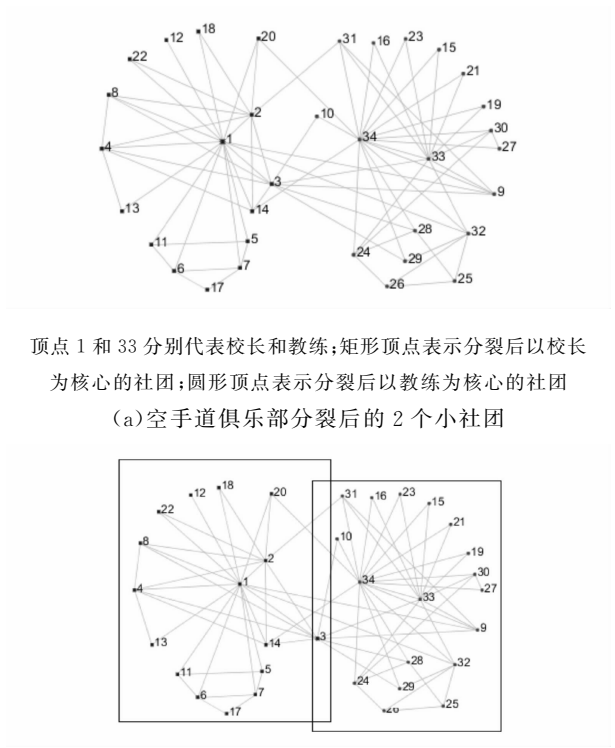


(c) $c_{\min}=20, c_{\max}=100, \mu=0.1$

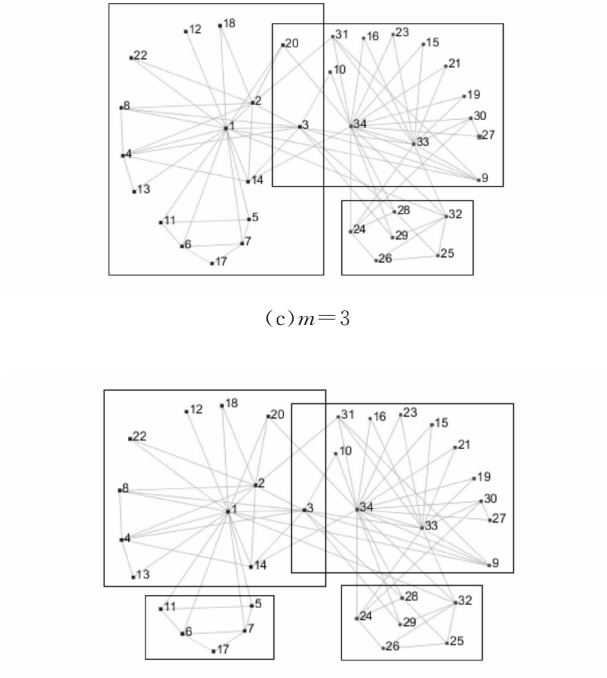
3.2 实际的网络

Zachary 空手道俱乐部网络是研究社团检测的一个经典网络^[20],其共有 34 个顶点、78 条边,该俱乐部因校长和教练意见分歧而分裂成 2 部分,形成了分别以校长和教练为核心的小社团,如图 5 所示。

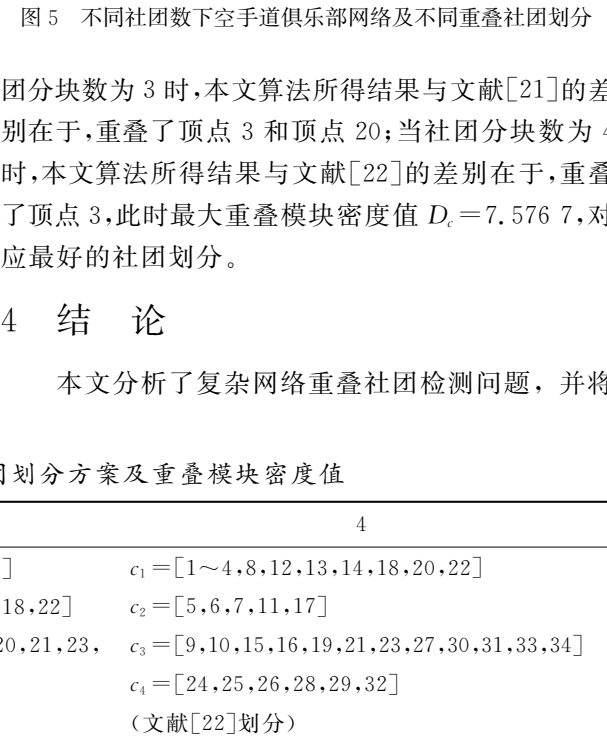
重叠模块密度如表 1 所示。当社团分块数为 2 时,本文算法检测到的 2 个社团重叠了顶点 3;当社



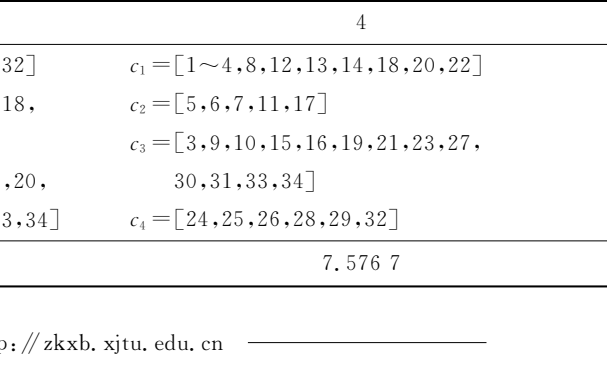
(a)空手道俱乐部分裂后的 2 个小社团



(b) $m=2$



(c) $m=3$



(d) $m=4$

图 5 不同社团数下空手道俱乐部网络及不同重叠社团划分

团分块数为 3 时,本文算法所得结果与文献[21]的差别在于,重叠了顶点 3 和顶点 20;当社团分块数为 4 时,本文算法所得结果与文献[22]的差别在于,重叠了顶点 3,此时最大重叠模块密度值 $D_c=7.576\ 7$,对应最好的社团划分。

4 结 论

本文分析了复杂网络重叠社团检测问题,并将

表 1 空手道俱乐部网络不同社团划分方案及重叠模块密度值

m	2	3	4
无重叠社团划分	$c_1=[1\sim 8,11\sim 14,17,18,20,22]$ $c_2=[9,10,15,16,19,21,23\sim 34]$ (真实划分)	$c_1=[24\sim 26,28,29,32]$ $c_2=[1\sim 8,11\sim 14,17,18,22]$ $c_3=[9,10,15,16,19,20,21,23,27,30,31,33,34]$ (文献[21]划分)	$c_1=[1\sim 4,8,12,13,14,18,20,22]$ $c_2=[5,6,7,11,17]$ $c_3=[9,10,15,16,19,21,23,27,30,31,33,34]$ $c_4=[24,25,26,28,29,32]$ (文献[22]划分)
D_c	6.833 3	6.297 4	7.509 1
m	2	3	4
有重叠社团划分	$c_1=[1\sim 8,11\sim 14,17,18,20,22]$ $c_2=[3,9,10,15,16,19,21,23\sim 34]$	$c_1=[24,25,26,28,29,32]$ $c_2=[1\sim 8,11\sim 14,17,18,20,22]$ $c_3=[3,9,10,15,16,19,20,21,23,27,30,31,33,34]$	$c_1=[1\sim 4,8,12,13,14,18,20,22]$ $c_2=[5,6,7,11,17]$ $c_3=[3,9,10,15,16,19,21,23,27,30,31,33,34]$ $c_4=[24,25,26,28,29,32]$
D_c	7.100 3	6.629 3	7.576 7

模块密度进行了改进,同时使用混沌量子粒子群算法求解问题,以适合度量重叠社团。设计的算法充分利用了量子计算的高效性、混沌算子的遍历性及粒子群计算的简便性,并在量子旋转门中采用粒子群进化来改变转角的大小。对计算机生成的数据和实际数据进行了仿真实验,并与经典算法进行了比较,结果表明本文算法所得结果更符合实际,具有较高的检测正确率。

参考文献:

- [1] SHI Chuan, YAN Zhenyu, CAI Yanan, et al. Multi-objective community detection in complex networks [J]. *Applied Soft Computing*, 2012, 12(2): 850-859.
- [2] WU Jianshe, JIAO Licheng, JIN Chao, et al. Overlapping community detection via network dynamics [J]. *Phys Rev: E*, 2012, 85: 016115.
- [3] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. *Proc Natl Acad Sci USA*, 2002, 99(12): 7821-7826.
- [4] NGUYEN N P, XUAN Ying, THAI M T. On detection of community structure in dynamic social networks [J]. *Mathematics and Statistics*, 2012, 58(3): 307-347.
- [5] 沈毅, 徐焕良. 加权网络权重自相似评判函数及其社团结构检测 [J]. *物理学报*, 2010, 59(9): 6022-6028.
SHEN Yi, XU Huanliang. The evaluation function of weight similarity and its application in community detection in weighted networks [J]. *Acta Physica Sinica*, 2012, 59(9): 6022-6028.
- [6] 黄健斌, 孙鹤立, BORTNER D, 等. 从链接密度遍历序列中挖掘网络社团的层次结构 [J]. *软件学报*, 2011, 22(5): 951-961.
HUANG Jianbin, SUN Heli, BORTNER D, et al. Mining hierarchical community structure within networks from density-connected traveling orders [J]. *Journal of Software*, 2011, 22(5): 951-961.
- [7] PALLA G, DERENYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society [J]. *Nature*, 2005, 435(9): 814-818.
- [8] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks [J]. *Phys Rev: E*, 2004, 69: 026113.
- [9] SHEN Huawei, CHENG Xueqi, GUO Jiafeng. Quantifying and identifying the overlapping community structure in networks [J/OL]. *J Stat Mech*, 2009. <http://dx.doi.org/10.1088/1742-5468/2009/07/P07042>.
- [10] NICOSIA V, MANGIONI G, CARCHIOLO V, et al. Extending the definition of modularity to directed graphs with overlapping communities [J/OL]. *J Stat Mech*, 2009. <http://dx.doi.org/10.1088/1742-4468/2009/03/P03024>.
- [11] FORTUNATO S, BARTHELEMY M. Resolution limit in community detection [J]. *Proc Natl Acad Sci USA*, 2007, 104(1): 36-41.
- [12] ROSVAL L M, BERGSTROM C T. An information-theoretic framework for resolving community structure in complex networks [J]. *Proc Natl Acad Sci USA*, 2007, 104(18): 7327-7331.
- [13] MUFF S, RAO F, CAHISCH A. Local modularity measure for network clusterizations [J]. *Phys Rev: E*, 2005, 72(5): 056107.
- [14] LI Z P, ZHANG S H, WANG R S, et al. Quantitative function for community detection [J]. *Phys Rev: E*, 2008, 77(3): 036106.
- [15] WANG Yan, FENG Xiaoyue, HUANG Yanxin, et al. A novel quantum swarm evolutionary algorithm and its applications [J]. *Neurocomputing*, 2007, 70: 633-640.
- [16] BACK T. *Evolutionary algorithm in theory and practice* [M]. Oxford, UK: Oxford University Press, 1996: 129.
- [17] LANCICHINETTI A, FORTUNATO S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities [J]. *Phys Rev: E*, 2009, 80(1): 016118.
- [18] GREGORY S. Finding overlapping communities in networks by label propagation [J]. *New Journal of Physics*, 2010, 12(10): 103018.
- [19] LANCICHINETTI A, FORTUNATO S, KERTESZ J. Detecting the overlapping and hierarchical community structure of complex networks [J]. *New Journal of Physics*, 2009, 11(3): 033015.
- [20] ZACHARY W W. An information flow model for conflict and fission in small groups [J]. *Journal of Anthropological Research*, 1977, 33(4): 452-473.
- [21] MAZUMDER R, HASTIE T, TIBSHIRANI R. Spectral regularization algorithms for learning large incomplete matrices [J]. *J Mach Learn Res*, 2010, 11(3): 2287-2322.
- [22] MA X K, GAO L, YONG X R, et al. Semi-supervised clustering algorithm for community structure detection in complex networks [J]. *Physica: A*, 2010, 389(1): 187-197.

(编辑 苗凌 武红江)