

基于线图与 PSO 的网络 重叠社区发现

黄发良^{1,2} 肖南峰¹

摘要 从优化模块度的角度出发, 引入线图理论, 给出线图的硬划分与原图的重叠划分相对应的理论证明, 提出了一种基于线图与粒子群优化技术的网络重叠社区发现算法 (Communities discovery based on line graph and particle swarm optimization, LGPSO), 该方法通过粒子群优化 (Particle swarm optimization, PSO) 算法寻找网络对应线图的最优划分来发现网络重叠社区, 实验结果显示, 该方法能够在无先验信息的条件下快速有效地揭示网络的重叠社区结构。

关键词 社区发现, 线图, 粒子群优化, 复杂网络

DOI 10.3724/SP.J.1004.2011.01140

Discovering Overlapping Communities Based on Line Graph and PSO

HUANG Fa-Liang^{1,2} XIAO Nan-Feng¹

Abstract From the perspective of optimizing modularity, an overlapping community discovery algorithm, LGPSO, is proposed based on line graph and PSO. The property that a partition of a line graph corresponds to a cover of the corresponding original graph is proved. LGPSO discovers overlapping communities in original graph using PSO to optimize partition of line graph. The experiments on some real-world networks show that the algorithm can fast and effectively discover the intrinsic overlapping communities in networks without any domain information.

Key words Community discovery, line graph, particle swarm optimization (PSO), complex network

近年来有关社区发现的各种方法不断涌现^[1], 其中绝大多数方法都有着这样两个假设: 1) 假定网络中存在的社区数是已知的; 2) 假定个体只属于某个社区, 然而在现实网络中, 这些假设很难甚至是无法满足, 一方面, 人为给定社区数目很困难且具有很强的主观随意性, 另一方面网络社区经常会出现彼此重叠互相关联的情况, 一个网络节点可以同时隶属于多个不同的社区. 例如, 根据文章主题将 Web 文档划分到不同的主题分组, 同一个文档可能属于多个不同的主题分组. 社区模块度函数的提出一定程度上缓解了假设 1) 隐藏的问题, 与假设 2) 密切相关的重叠社区发现研究已经引起广泛的注意. Palla 等^[2] 首先提出 CPM 方法, 通过 K 完全图的滚动来发现重叠社区. PEACOCK 算法^[3] 运用顶点复制法挖掘网络重叠社区. 文献 [4] 在对边介数与点介数同时考虑的基础上引入带复制的删除操作来发现重叠社区. EAGLE 算法^[5] 利用基于完全图的层次聚类法生成社区谱系

树, 然后根据 Cover 的质量来划分网络. HLC 算法^[6] 借助边对的 Jaccard 相似度来进行边聚类, 进而实现重叠社区发现. Wang 等^[7] 提出社区强度的局部定义, 并以此将硬划分的社区结构改造成重叠社区结构. 本文引入线图理论, 给出并证明了线图的两个性质定理, 提出了一种新的基于线图与粒子群优化 (Particle swarm optimization, PSO)^[8-11] 的网络重叠社区发现算法 LGPSO (Communities discovery based on line graph and PSO, LGPSO), 该方法无需用户指定社区个数等算法参数, 能够有效揭示网络内在的重叠社区结构, 具有相对较好的算法性能。

1 定义与性质

定义 1. 令 V 为给定非空集合, $P = (P_1, P_2, \dots, P_m)$, 其中, $P_i \subseteq V, P_i \neq \emptyset (i = 1, 2, \dots, m)$ 且 $\cup_{i=1}^m P_i = V$, 集合 P 称作集合 V 的覆盖。

定义 2. 若集合 V 的覆盖 P 满足 $P_i \cap P_j = \emptyset (i \neq j)$, 则集合 P 可称作集合 V 的划分。

定义 3. 对于集合 V 的任意两个覆盖分别为令 $P = (P_1, P_2, \dots, P_m)$ 与 $P' = \{P'_1, P'_2, \dots, P'_n\}$, 若对每一个 P_i , 均有 P'_j 使得 $P_i \subseteq P'_j$, 则称 P 为 P' 的加细。

定义 4. 给定节点集合 V 的覆盖 P , 其中任意两个分块 P_i 与 P_j 的重叠程度称为分块重叠率 OR , $OR(P_i, P_j) = \frac{|P_i \cap P_j|}{\min(|P_i|, |P_j|)}$ 。

定义 5. 给定无向图 $G = (V, E)$, 其对应的线图 $L(G)$ 是满足如下条件的图结构: $L(G)$ 是以 $E(G)$ 为顶点集, 且只要 E 中边 e_1 与 e_2 在 G 中是相邻边, 在 $L(G)$ 中就是相邻顶点。

性质 1. 若图 G 连通, 则线图 $L(G)$ 也连通。

性质 2. 线图 $L(G)$ 节点集合的一个划分对应于原图节点集合的一个覆盖。

证明. 令 $P = (P_1, P_2, \dots, P_m)$ 为线图 $L(G)$ 的一个划分, 分块 P_i 所关联的原图的顶点集合为 $V_i = (V_{i1}, V_{i2}, \dots, V_{im})$, 由线图定义知, $\cup_{i=1}^m V_i = V, V_i \neq \emptyset (i = 1, 2, \dots, m)$, 由此可得要证划分 P 对应原图节点集合的一个覆盖, 只需证 $\exists V_i, V_j$ 使得 $V_i \cap V_j \neq \emptyset$ 成立, 证明如下: 由 $L(G)$ 的连通性可知, 必存在这样的两个分块 P_i 与 P_j : P_i 中的节点与 P_j 中的节点之间至少存在一条边, 不失一般性, 令 $e = \langle v_i, v_j \rangle$ 为 P_i 与 P_j 之间的一条边, 又由线图定义知, 线图中的一个节点对应于原图中一条边, 也就是说, 线图中的一个节点对应原图中的两个节点, 从而有 $e = \langle (v_{i1}, v_{i2}), (v_{j1}, v_{j2}) \rangle$, 另有线图中的一条边对应原图的一个顶点, 故 4 个命题 $V_{i1} = V_{j1}, V_{i2} = V_{j2}, V_{i1} = V_{j1}, V_{i2} = V_{j2}$ 有且仅有一个为真, 即 P_i 与 P_j 对应的原图中的两个节点集合之间交集不为空, $V_i \cap V_j \neq \emptyset$ 成立. 综上所述, 线图 $L(G)$ 节点集合的一个划分对应于原图节点集合的一个覆盖。 □

2 LGPSO 算法

2.1 粒子编码

粒子编码采用基于节点邻居有序表的编码方法, 其基本思想是: 首先对网络节点进行编号, 然后对每个节点根据其编号进行排序形成邻居有序表, 在初始化或粒子位置更新阶段生成新粒子时, 确保该粒子的合法性. 以图 1(a) 中的网络为例, 首先建立各节点的邻居有序表 (图 1(d)), 根据此表可以对划分 (图 1(b)) 进行粒子编码, 结果见图 1(c). 该编码方式有如下三个优势: 1) 避免非法粒子的产生; 2) 自动确定社区数; 3) 避免基于二值编码的迭代二划分策略^[13] 所遭遇的

收稿日期 2010-11-01 录用日期 2011-03-03
Manuscript received November 1, 2010; accepted March 3, 2011
国家自然科学基金 (60776816, 61171141), 广东省自然科学基金重点项目 (8251064101000005) 资助
Supported by National Natural Science Foundation of China (60776816, 61171141) and Nature Science Foundation of Guangdong Province (8251064101000005)
1. 华南理工大学计算机科学与工程学院 广州 510006 2. 福建师范大学软件学院 福州 350108
1. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006 2. Faculty of Software, Fujian Normal University, Fuzhou 350108

容易陷入局部最优划分的处境.

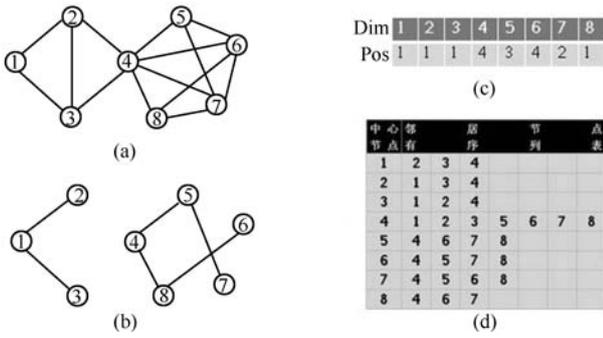


图 1 网络社区的粒子编码方案
 ((a) 网络线图; (b) 社区结构; (c) 粒子编码; (d) 邻居有序表)
 Fig. 1 Particle encoding scheme of network communities
 ((a) Line graph of network; (b) Community structure;
 (c) Encoding particle; (d) Ordered neighbor list)

2.2 粒子更新策略

LGPSO 分别采用式 (1) 与式 (2) 对粒子速度与位置进行更新:

$$y_i(t+1) = wy_i(t) + c_1 \times r_1 \times (P_i - x_i(t)) + c_2 \times r_2 \times (P_g - x_i(t)) \quad (1)$$

$$x_i(t+1) = \begin{cases} k, & \text{若 } \rho < sig(v_{ij}(t+1)) \text{ 且 } deg(i) > 1 \\ x_i(t), & \text{否则} \end{cases} \quad (2)$$

$x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ 表示粒子 i 的位置, $y_i = (y_{i1}, y_{i2}, \dots, y_{id})$ 表示粒子 i 的速度, $P_i = (P_{i1}, P_{i2}, \dots, P_{id})$ 表示粒子 i 的具有最佳适应度值的历史最优位置, $P_g = (P_{g1}, P_{g2}, \dots, P_{gd})$ 表示整个粒子群的历史最优位置. $sig(y_{ij}) = \frac{1 - \exp(-y_{ij})}{1 + \exp(-y_{ij})}$, ρ 为预定阈值, k 为除当前连接邻居外的任一随机的邻居节点, 即 $k = ceil(rand \times deg(v_i))$, $k \neq x_i(t)$, $ceil$ 为上取整函数, $deg(v_i)$ 表示节点 v_i 的度, t 为进化代数, w 为惯性系数, c_1 与 c_2 称为学习因子, 可取常数也可根据算法需要进行动态修正, r_1 与 r_2 为均匀分布在 $[0, 1]$ 之间的随机数.

2.3 粒子适应度

由于每个粒子对应网络线图的一种社区划分方案, 故粒子适应度就是社区划分的质量. 社区划分质量的评价函数有很多种, 不同的社区定义对应着不同的社区划分质量评价函数. 本文采用模块度 Q 值函数^[14] 来评价线图划分的质量. 该函数可简述如下:

假设粒子形成图 G 的 k 划分, 则可定义 $k \times k$ 的矩阵 R , 其对角元 R_{ii} 表示第 i 个社区中的节点内部度在图中所有节点度和的比例, 而非对角元素 R_{ij} 表示社区 i 的节点到其他社区 j 的边数在图中所有边数的比例. 据此给出 Q 值函数

$$Q(y) = \sum_i (R_{ii} - a_i^2) = tr(R) - \|R\|^2 \quad (3)$$

其中, $a_i = \sum_j R_{ij}$, 表示与第 i 个社区中的节点相连的边在所有边中所占的比例.

2.4 优化重叠社区结构

在实验的过程中, 我们发现线图的一个最优划分未必对应于原图的一个最优覆盖, 而更多的是对应于原图的一个次

优覆盖, 这个次优覆盖是最优覆盖的加细. 那么如何对此次优覆盖进行求精得到最优的重叠社区结构. 关于如何将模块度 Q 值函数进行拓展来测度有重叠的社区结构这方面的研究成果不多, 借鉴文献 [5] 中的 Q_{ov} 函数, 我们设计了一个社区重叠社区的后续优化过程, 其基本思想是根据社区重叠率进行重叠社区的合并, 进而依据 Q_{ov} 值来确定最优的重叠社区.

假设 P 为某个粒子所对应的原图的一个覆盖, m 是原图中的边数目, 原图中节点 i 的度为 k_i , 原图的邻接矩阵式 A , 原图的节点 i 所属社区的数目为 O_i , 则适用于重叠社区结构的评价函数可公式化如下:

$$Q_{ov} = \frac{1}{2m} \sum_{c \in P} \sum_{i,j} \frac{1}{O_i O_j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{ic} \delta_{jc} \quad (4)$$

其中函数 δ_{ic} 的取值为 1, 若节点 i 属于社区 C ; 反之, 取值为 0.

2.5 算法描述及复杂度分析

LGPSO 算法步骤可划分为三个部分: 1) 步骤 1~3, 主要负责初始化算法, 建立所需要的相关参数与数据结构; 2) 步骤 4~9, 主要通过 PSO 实现线图的最优划分; 3) 步骤 10~14, 在层次合并的过程中发现最优重叠社区并输出. 该算法基本流程描述如下:

- 步骤 1. 将原始网络图转变成对应的线图, 建立各节点的邻居有序表;
- 步骤 2. 设置粒子位置和速度的范围, 以及粒子群惯性因子 w , 根据线图节点的数量设置粒子的位置向量和速度向量的维度;
- 步骤 3. 初始化粒子群, 运用粒子位置修正策略确保粒子合法;
- 步骤 4. 复制粒子的当前位置向量到经验位置, 并将每个粒子当前位置的适应度复制到经验适应度;
- 步骤 5. 选出适应度最高的粒子, 并将其经验位置向量和经验适应度分别作为群最优位置和群最优经验;
- 步骤 6. 根据式 (1) 与式 (2) 对每个粒子更新自身的位置和速度, 运用粒子位置修正策略确保粒子合法;
- 步骤 7. 根据式 (3) 计算每个粒子的适应度, 并与其经验适应度相比较, 如果优于其经验, 则更新该粒子的经验位置及其适应度;
- 步骤 8. 计算出群体中最优粒子, 与当前群最优适应度相比较, 如果优于当前群最优粒子, 则更新群最优位置和群最优适应度;
- 步骤 9. 如果停止条件不满足, 则转到步骤 4;
- 步骤 10. 根据线图与其原图的对应关系将线图划分转变成原图的覆盖;
- 步骤 11. 计算当前重叠社区结构的 Q_{ov} 值与每一对社区的重叠率;
- 步骤 12. 选择具有最大重叠率的社区对进行合并, 对合并后的社区进行包含检测, 将被其他社区包含的社区删除;
- 步骤 13. 重复步骤 11~13, 直到社区数为 1;
- 步骤 14. 输出具有最大 Q_{ov} 的重叠社区.

假定原网络的节点数为 n , 边数为 m , 迭代次数为 t , 粒子数为 k , 则算法的复杂度可以估计如下: 首先分析第 1 部分: 对于步骤 1, 令线图节点的平均度数为 d , 又由于真实网络的 d 往往是一个很小的常量, 将原始网络图转变成对应线图的复杂度为 $O(2 \times d \times m) \approx O(m)$, 建立各节点的邻居有

序表的复杂度为 $O(m \times d \times \log d)$, 故有复杂度 $O(m)$; 步骤 2 与步骤 3 都是粒子的初始化, 则其复杂度为 $O(m \times k)$. 故第 1 部分的复杂度为 $O(m \times k)$. 然后分析第 2 部分: 步骤 4 是粒子经验的复制, 其计算复杂度为 $O(m \times k)$; 步骤 5 是获取最优粒子, 其复杂度为 $O(k)$; 步骤 6 是粒子经验的更新, 其计算复杂度为 $O(m \times k)$; 步骤 7 是粒子适应度的计算, 其复杂度的估算为: 根据粒子构造社区结构需要 $O(m)$ 的时间, 若令一次迭代产生 r 个社区, 社区的平均大小为 $[m/r]$, 则社区内链接度的计算复杂度为 $O(r \times (m/r)^2) = O(m^2/r)$, 社区间链接度的计算复杂度为值的复杂度为 $O(r \times ((m/r) \times (m/r) \times (r-1))) \approx O(m^2/r)$. 故步骤 7 的复杂度约为 $O(m^2/r)$; 步骤 8 是计算最优粒子, 其复杂度为 $O(k)$; 步骤 9 是停止条件 (迭代次数) 判定, 为常量复杂度 $O(1)$. 故此块的复杂度为 $O(t \times m^2/r)$. 最后是第 3 部分: 步骤 10 是将线图划分转变原网络的覆盖, 其复杂度为 $O(m)$; 步骤 11~13 是一个执行社区合并运算的循环, 假定算法第 2 部分产生的最优社区结构中含有 s 个社区, $O(m^2/2 + m^2/3 + \dots + m^2/s) \approx O(m^2 \times (\ln(s+1) - 1 + 0.577218)) \approx O(m^2)$; 步骤 14 输出结果, 复杂度为 $O(1)$. 故算法第 3 部分的复杂度为 $O(m^2)$. 综合上述分析可知, 算法的复杂度为 $O(t \times m^2/r)$. 由复杂度分析可以看出, LGPSO 算法在处理大规模网络的速度比较慢的, 这是由于大规模网络的边数 m 是很大, 尽管此时 r 值也会很大, 但是在维数约为 $m \times d/2$ 的高维空间中进行寻优, 这势必需要较大迭代次数 t 才能发现较优的社区结构.

3 仿真实验

3.1 数据集

三个真实网络分别是: 1) Karate 网络, 该网络是美国一所大学中的空手道俱乐部成员间的相互社会关系, 共有 34 个节点, 78 条连接边; 2) Dolphins 网络, Lusseau 等构造的一个宽吻海豚群体关系网, 共有 62 个节点, 159 条边; 3) HLM 网络, 我们从名著《红楼梦》中选取 77 个主要人物, 以 5 个家族 (宁国府、荣国府、王府、史府和薛府) 为依据生成的社会网络, 其中人物关系非常复杂, 在此我们主要考虑血缘关系与典型的媒介关系, 如图 4(a) 所示, 圆圈表示史府人员 (7 人), 上三角表示薛府人员 (12 人), 方形表示王府人员 (15 人), 下三角表示荣国府人员 (28 人), 田方形表示宁国府人员 (15 人), 该网络包含 77 个节点, 121 条边.

3.2 算法有效性

为了验证算法的有效性, 我们采用规范化互信息 (Normalized mutual information, NMI)^[15] 来度量网络的真实社区结构与算法计算所得社区结构的相似性, 首先采用 LGPSO 对 Karate 网络进行社区分析, 图 2(a) 对应的是没有经过优化的重叠社区结构, 包含三个有重叠的社区, 有节点 3, 9, 10 与 31 为社区共享节点, 其他节点均为单社区节点, 该初始社区结构的 $Q_{ov} = 0.222$, 与真实社区结构相比较得 $NMI = 0.823$, 对此初始社区结构进行基于社区重叠率的优化合并后得到最终的社区结构, 如图 2(b) 所示, 只有节点 3 为两个社区的节点, 其他节点都准确地归属于其真实社区中, 该最终社区的 $Q_{ov} = 0.2313$, $NMI = 0.9028$.

接着分析 Dolphins 网络, 图 3(a) 对应的是没有经过优化的重叠社区结构, 包含 4 个有重叠的社区, 其中有节点 15, 30 与 38 为区域 1 表示的社区与区域 2 表示的社区共享, 节点 40, 8 与 31 为区域 2 表示的社区与区域 3 表示的社区共享, 节点 10, 14 与 18 为区域 3 表示的社区与区域 4 表示

的社区共享, 其他节点均为单社区节点, 该初始社区结构的 $Q_{ov} = 0.156$, 与真实社区结构相比较得 $NMI = 0.47$, 显然此时社区结构的质量还有提升, 我们进一步对此初始社区结构进行基于社区重叠率的优化合并后得到最终的社区结构, 如图 3(b) 所示, 只有节点 8, 31 与 40 为两个社区共享, 其他节点都准确地归属于其真实社区中, 此时社区结构的 $Q_{ov} = 0.318$, $NMI = 0.824$. 最后分析 HLM, 从网络的最终社区 (图 4(b)) 可以发现: 社区史府与荣国府的共享节点“史侯”, 宁国府社区与荣国府社区有共享节点“尤二姐”、“贾源”与“贾演”, 王府社区与荣国府社区共享节点“刘姥姥”、“王熙凤”与“王夫人”, 薛府社区与薛府社区、荣国府社区分别共享节点“薛姨妈”与“邢岫烟”. 从这些共享节点可以看出, 红楼梦中的四大家族主要是以婚姻裙带关系盘根错节而形成的难以分割的一个集团: “尤二姐”嫁给“贾琏”, “王熙凤”嫁给“贾琏”, “王夫人”嫁给“贾政”, “薛姨妈”嫁给“王公之子”, “邢岫烟”嫁给“薛蝌”, 还有史侯的女儿“贾母”嫁给“贾代善”, 当然还有兄弟关系“贾演”与“贾源”, 有趣的是“刘姥姥”成为一个共享节点. 此时社区结构的 $Q_{ov} = 0.517$, $NMI = 0.835$.

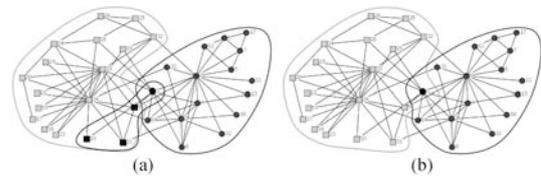


图 2 Karate 网络 ((a) 初始社区结构; (b) 最终的社区结构)
Fig. 2 Karate network ((a) Initial community; (b) Final community)

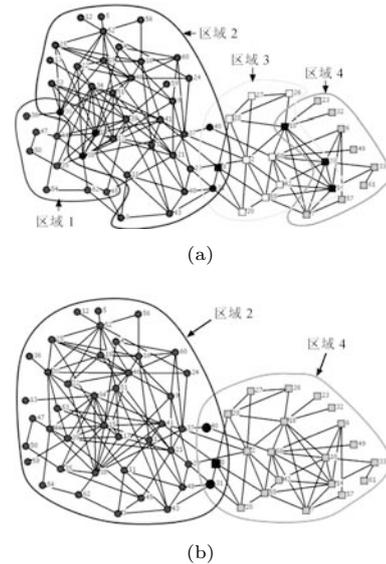


图 3 Dolphins 网络 ((a) 初始社区结构; (b) 最终的社区结构)
Fig. 3 Dolphins network ((a) Initial community; (b) Final community)

为了更好地评价算法的有效性, 我们将 LGPSO 算法在三个数据集上分别运行 50 次, 取 NMI 的平均值, 并与算法 CPM^[2]、HLC^[6]、PEACOCK^[3] 进行比较, 见表 1 (HLC 中的 t 为边相似度阈值). 由该表知, 与 CPM、HLC 和 PEACOCK 相比较, 只有在 HLM 网络中, PEACOCK 获得与

LGPSO 相当的效果, 在其他情况下, LGPSO 所得到的社区结构与真实社区相一致的程度远远高于前两者.

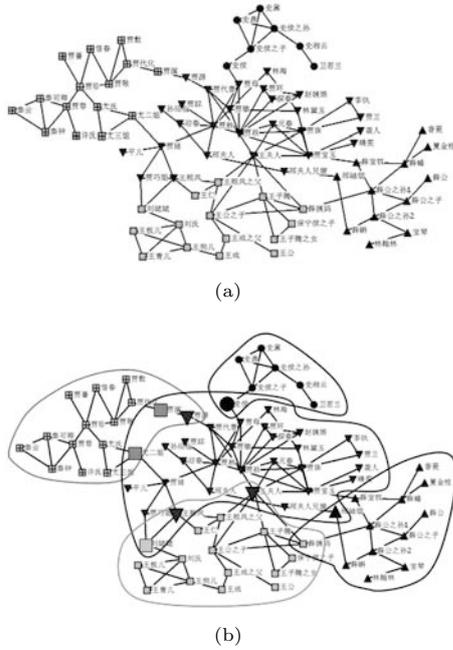


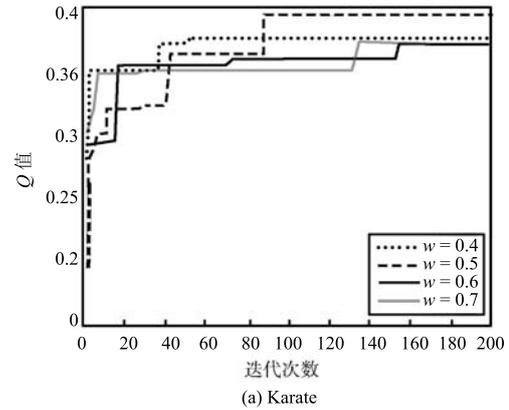
图4 HLM 网络 ((a) 初始社区结构; (b) 最终的社区结构)
Fig.4 HLM network ((a) Initial community; (b) Final community)

表1 LGPSO 与 HLC、PEACOCK、CPM 的有效性 NMI 比较
Table 1 Comparison of NMI effectiveness of LGPSO, CPM, HLC, and PEACOCK

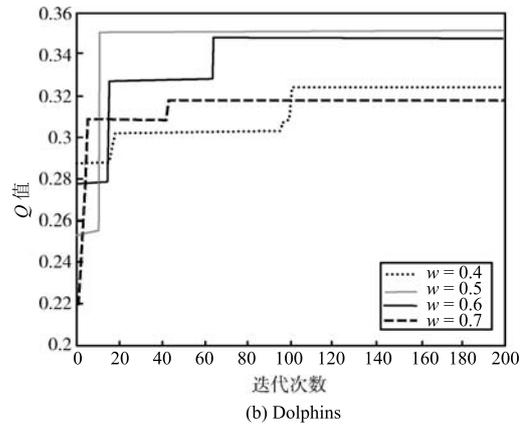
	CPM ($k = 3$)	HLC	PEACOCK	LGPSO
Zachary	0.335	0.231 ($t = 0.3$)	0.558	0.905
Dolphins	0.461	0.257 ($t = 0.25$)	0.746	0.822
HLM	0.211	0.432 ($t = 0.15$)	0.839	0.836

3.3 算法收敛性

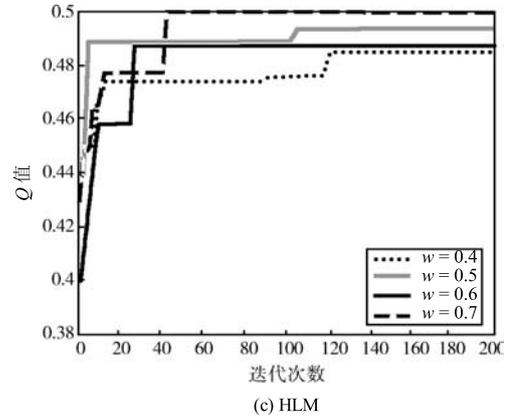
一般地, 惯性因子与粒子个数这两个参数对 PSO 算法的收敛性有着较大影响, 因此, 本文从这两个参数的不同设置来对算法的收敛性进行分析, 下面是 LGPSO 在 3 个真实网络 50 次寻优的实验结果分析. 首先来看惯性因子的影响, 从图 5 可以看出, 惯性因子在不同的网络中对算法收敛性的影响作用不同: 对于 Karate 网络, 惯性因子大小为 0.6 时, 算法收敛较快且都能收敛到最优 Q 值; 对于 Dolphins 网络, 惯性因子大小为 0.5 时, 算法收敛性最好; 对于 HLM 网络, 惯性因子为 0.7 时算法收敛性最好. 比较分析可知, 惯性因子的影响作用与具体问题相关, 对于不同问题, 选取合适的惯性因子可以加快算法收敛速度和精度. 接下来讨论粒子个数的影响, 从图 6 中可以看出, 对于 Karate 网络, 粒子数为 20 ~ 40 时, 算法收敛较快且都能收敛到最优 Q 值, 其中粒子数为 40 时收敛最快, 而粒子数为 10 时算法陷入局部最优; 对于 Dolphins 网络, 粒子数为 30 ~ 40 时, 算法能较快的收敛到最大 Q 值, 但当粒子数为 30 时需要更多的迭代次数, 而对于种群大小为 20 与 10 时, 算法都陷入到局部最优值; 对于



(a) Karate



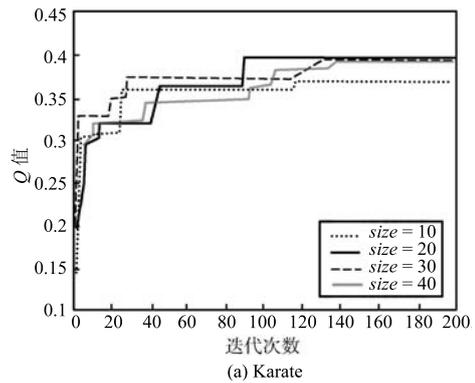
(b) Dolphins



(c) HLM

图5 惯性因子对算法收敛性能的影响

Fig.5 Inertia factor's effect on algorithm convergence



(a) Karate

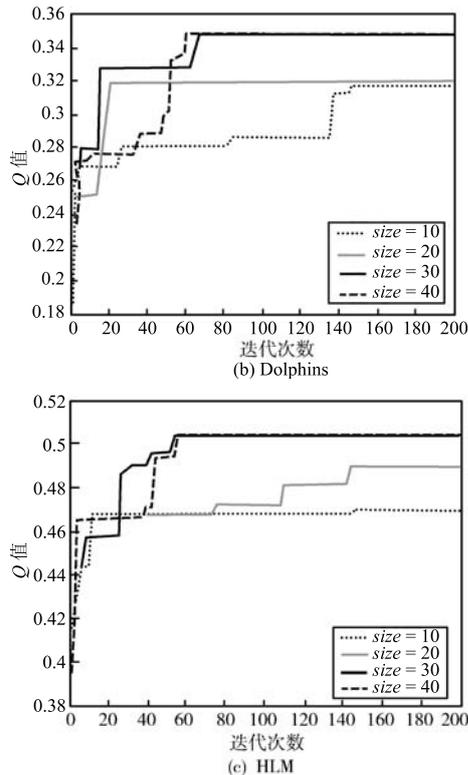


图 6 粒子数目对算法收敛性能的影响

Fig. 6 Number of particles' effect on algorithm convergence

HLM 网络, 其情形与 Dolphins 网络类似, 但当种群大小为 10 时, 其 Q 值的最优性很差. 从上面的分析可知, 粒子数目对算法性能的影响程度是与网络规模 (即原网络的边数) 相关的.

4 结束语

社区模式挖掘是复杂网络研究中的一个重要课题, 常用的社区发现算法都存在着网络内含社区数已知的假设和/或网络社区之间无重叠的假设, 不能有效地发现网络中隐含的真实社区结构. 本文从图论的角度考察网络, 将网络图转变成其对应的线图, 并给出线图的硬划分与原图的有重叠划分 (覆盖) 相对应的理论证明, 运用 PSO 算法寻找线图的最优划分, 将线图社区结构转变到其原图的重叠社区结构, 最后依据社区重叠率与加强模块度 Q_{ov} 对重叠社区模块结构进一步优化. 采用真实网络测试本文方法的有效性与收敛性, 实验结果显示, 该方法无需用户指定社区个数等算法参数, 能够有效地揭示网络的重叠社区结构, 且具有相对较好的收敛性能.

References

- Huang Fa-Liang. Studies on community detection and its application in information network. *Complex Systems and Complexity Science*, 2010, **7**(1): 64–74 (黄发良. 信息网络的社区发现及其应用研究. *复杂系统与复杂性科学*, 2010, **7**(1): 64–74)
- Palla G, Derenyi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005, **435**(7043): 814–818

- Gregory S. Finding overlapping communities using disjoint community detection algorithms. In: *Proceedings of the Complex Networks: Complex Net 2009*. Berlin, Germany: Springer-Verlag, 2009. 47–61
- Pinney J W, Westhead D R. Betweenness-based decomposition methods for social and biological networks. In: *Proceedings of the International Conference Interdisciplinary Statistics and Bioinformatics*. Leeds, UK: Leeds University Press, 2006. 87–90
- Shen H, Cheng X, Cai K, Hu M B. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and Its Applications*, 2008, **388**(8): 1706–1712
- Ahn Y Y, Bagrow J P, Lehmann S. Communities and hierarchical organization of links in complex networks. 2009, arXiv: 0903.3178
- Wang X, Jiao L, Wu J. Adjusting from disjoint to overlapping community detection of complex networks. *Physica A: Statistical Mechanics and Its Applications*, 2009, **388**(24): 5045–5056
- Kennedy J, Eberhart R. Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks*. Perth, Australia: IEEE, 1995. 1942–1948
- Liao C J, Tseng C T, Luarn P. A discrete version of particle swarm optimization for flowshop scheduling problems. *Computers and Operations Research*, 2007, **34**(10): 3099–3111
- Pan Q K, Tasgetiren M F, Liang Y C. A discrete particle swarm optimization algorithm for the no-wait flowshop scheduling problem. *Computers and Operations Research*, 2008, **35**(9): 2807–2839
- Jin Y X, Cheng H Z, Yan J Y, Zhang L. New discrete method for particle swarm optimization and its application in transmission network expansion planning. *Electric Power Systems Research*, 2007, **77**(3–4): 227–233
- Kennedy J, Eberhart R C, Shi Y. *Swarm Intelligence*. San Francisco: Morgan Kaufmann, 2001
- Duan Xiao-Dong, Wang Cun-Rui, Liu Xiang-Dong, Lin Yan-Ping. Web community detection model using particle swarm optimization. *Computer Science*, 2008, **35**(3): 18–21 (段晓东, 王存睿, 刘向东, 林延平. 基于粒子群算法的 Web 社区发现. *计算机科学*, 2008, **35**(3): 18–21)
- Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 2001, **99**(12): 7821–7826
- Nicosia V, Mangioni G, Carchiolo V, Malgeri M. Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, **2009**(3): 1–22

黄发良 博士研究生. 主要研究方向为数据挖掘. 本文通信作者.

E-mail: faliang.huang@gmail.com

(HUANG Fa-Liang Ph.D. candidate. His main research interest is data mining. Corresponding author of this paper.)

肖南峰 博士, 教授. 主要研究方向为智能机器人.

E-mail: xiaonf@scut.edu.cn

(XIAO Nan-Feng Ph.D., professor. His main research interest is intelligent robotics.)