

离散 Jaya 算法的复杂网络社区发现^①

李剑雄, 鲍志强

(华南师范大学 计算机学院, 广州 510631)

通讯作者: 李剑雄, E-mail: 1194353393@qq.com



摘要: 社区结构是复杂网络的重要特性之一, 基于模块度的复杂网络社区发现问题是一个 NP 难度的组合优化问题, 常用启发式算法求解. 最近出现的 Jaya 算法是求解连续优化问题的一种简单有效的元启发式方法. 本文在遵循 Jaya 算法按靠近最好解、远离最差解的方式更新种群个体的基础上, 针对复杂网络社区发现问题给出了 Jaya 算法离散化的策略, 提出一种复杂网络社区发现的离散 Jaya 算法. 实验表明, 在几个典型真实网络实例和一类人造网络实例上, 与几个经典算法和元启发式算法相比, 本文算法具有求解精度高、能自动确定社区数目等优点.

关键词: 复杂网络; 进化算法; 模块度; 社区发现; Jaya

引用格式: 李剑雄, 鲍志强. 离散 Jaya 算法的复杂网络社区发现. 计算机系统应用, 2020, 29(6): 146-154. <http://www.c-s-a.org.cn/1003-3254/7432.html>

Discrete Jaya Algorithm for Complex Network Community Detection

LI Jian-Xiong, BAO Zhi-Qiang

(School of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract: Community structure is one of most important characteristics of complex networks. The community detection problem based on modularity is NP-hard as a combinatorial optimization problem, which is often solved by heuristic algorithms. Jaya algorithm is a simple and effective meta-heuristic method for solving continuous optimization problems. In this study, the strategy of discretizing Jaya algorithm for complex network community discovery is given on the basis of updating the population individuals according to the way Jaya algorithm works, that is, an individual is updated close to the best solution and far away from the worst solution, and thus a discrete Jaya algorithm for complex network community discovery is proposed. Experiments show that the proposed algorithm has the advantages of high resolution and automatic determination of the number of communities compared with the classical algorithms in several real network instances and a class of artificial network instances.

Key words: complex networks; evolutionary algorithms; modularity; community detection; Jaya

许多现实世界中的复杂系统都可以使用复杂网络来描述, 例如社交网、生物网、万维网、交通网等. 复杂网络可建模为图, 其中节点表示网络中的对象, 边表示对象间的关系. 大量研究表明: 复杂网络通常具有小世界性和无标性等统计特性, 并呈现出社区结构等拓扑特性. 一个网络可以分成若干社区, 社区内部的

节点连接相对紧密, 不同社区之间的节点连接相对稀疏^[1,2]. 社区发现的目的是找出复杂网络的社区结构, 提取社区结构中的信息.

来自不同领域的研究中提出了许多不同的社区发现算法, 这些算法可大致分为: 图划分的分割算法^[3]、聚类算法^[4-6]、基于网络动力学特性的算法^[7-9]和基于目标

① 基金项目: 国家自然科学基金 (61370003)

Foundation item: National Natural Science Foundation of China (61370003)

收稿时间: 2019-10-25; 修改时间: 2019-11-20; 采用时间: 2019-12-05; csa 在线出版时间: 2020-06-10

函数的优化算法^[10-14]. 其中以模块度函数 (Modularity, Q) 的为目标函数的优化算法是目前较为普遍的一类算法. 模块度函数 Q 是 Newman 提出的评价社区结构优劣的指标^[6], 借助于该函数可将社区发现问题转化为优化问题. 如 Duch^[15]提出了基于模块度的极值优化算法, Guimera 等^[16]基于模拟退火的 GA(Genetic algorithm) 算法, 2008 年 Blondel 等^[17]提出的 BGLL 算法. 然而求最优模块度的社区结构是 NP 难问题, 常常用启发式算法求解. 许多启发式算法属于贪心法, 一般很难求得满意的社区划分结果, 且需要预先设定网络的社区数目. 智能优化元启发式算法作为社区发现问题的一种有效解决手段, 目前已被广泛应用求解^[18-20]. Talbi 等^[21]提出了一种求解图划分问题的并行遗传算法, 该算法具有超线性加速特性. Bui 等^[22]提出了基于图划分的预处理模式从而提高遗传算法的空间搜索能力. Tasgin 等^[23]基于模块度使用了遗传算法进行社区发现. Gog 等^[24]基于种群中个体的信息分享机制提出了一种协同进化算法用于社区发现. Pizzuti^[25]提出了 GA-Net 遗传算法, 利用了 locus-based 表示法和交叉操作, 该方法在操作时只考虑了该点的实际连接关系, 有效的减少了无效搜索. Gong 等^[26]提出了一种文化基因算法用于优化模块度进行社区发现. Duch 等^[15]通过对标准差分进化算法的交叉操作进行离散化, 提出了离散差分进化算法用于社区发现. Change 等^[27]采用最大最小蚁群算法, 通过信息素的挥发和路径的选择进行社区的划分. 金弟等^[13]通过分析模块度函数的局部单调性, 以及改进遗传算法的变异策略, 提出一种基于局部搜索的改进遗传算法. 张英杰等^[14]提出了一种基于免疫差分进化算法 IDDE (Immune Discrete Differential Evolution). Said 等^[28]提出了基于聚类协同系数的 CC-GA (Clustering Coefficient based Genetic Algorithm) 算法, 该算法可用于稠密网络和稀疏网络的社区发现.

Jaya 算法^[29]是由 Rao RV 提出的一种元启发式算法, 适用于求解连续型最优化问题. 与其它元启发式算法相比, Jaya 算法简单, 需要输入的参数少, 只需设置种群大小和迭代次数. 该算法通过离散化, 近年来已被应用在许多不同的组合优化问题上^[30,31]. 本文针对复杂网络社区发现问题, 遵循 Jaya 算法的基本思想, 提出了一种离散化的算法形式称作 DJaya (Discrete Jaya algorithm), 其核心思想是针对种群中个体不同的倾向

性, 采取不同的更新方式, 平衡全局探索性和局部搜索性来提高搜索的有效性. 在标准人造网络数据集 GN Benchmark 和几个典型的真实网络算例上的实验分析表明, DJaya 算法能自动确定社区的数目, 且在模块度、互信息等方面表现出良好的性能^[32].

1 相关工作

1.1 模块度函数

不同的社区发现算法在同一复杂网络上进行社区划分时结果通常是不同的. 为了统一评价标准, Newman 等提出了社区结构的模块度函数. 模块度函数易于理解且计算简单, 被广泛用来评价社区划分质量^[5]. 一个社区结构的模块度 Q 定义如下:

$$Q = \frac{1}{2m} \sum \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (1)$$

其中, k_i 表示节点 i 的度, $k_i k_j / 2m$ 表示节点 i 与节点 j 之间的连接概率, A_{ij} 表示图的邻接矩阵, 如果节点 i 与节点 j 有边相连, 则 $A_{ij} = 1$, 否则 $A_{ij} = 0$. c_i 表示节点 i 所属的社区, 如果 $c_i = c_j$, 则 $\delta(c_i, c_j) = 1$, 否则 $\delta(c_i, c_j) = 0$. 通常, Q 值的范围是 $[-0.5, 1)$, 越接近于 1, 则复杂网络的社区结构越明显. Q 值一般在 0.3-0.7 的范围内^[2], 当 $Q < 0.3$ 时表明网络几乎没有社区结构.

1.2 基于模块度的社区发现的典型算法

基于模块度优化的社区发现算法有许多, 其中包括许多以模块度为目标函数的算法, 其典型的代表如 FMM (Fast Modularity Maximization) 算法^[5]、LPA (Label Propagation Algorithm) 算法^[33]、BGLL 算法^[24]等贪心启发式算法, 以及近年提出的 ACO (Ant Colony Optimization)^[28]、LGA (Genetic Algorithm with Local search)^[14]、IDDE、CC-GA 等元启发式算法.

FMM 算法将图中每个节点初始化为一个单独的社区, 然后选择使模块度 Q 的增值最大的社区进行合并, 重复该步骤, 直到网络中所有节点属于同一个社区. 整个过程是自底向上的过程, 最终得到一个层次图. 网络中节点对应层次图的最底层的叶子节点. 层次图中每一层对应网络的某种划分, 层次图中所有层次对应的划分中模块度最大的划分为网络最终划分.

标签传播算法 LPA 是由 Raghavan 等提出的^[34], 其基本思想是让每个节点有一个自己特有的标签, 节点会选择自己邻居中出现次数最多的标签, 如果每个

标签出现次数一样多,就随机选择一个标签替换自己原始的标签,如此往复,直到每个节点标签不再发生变化,那么持有相同标签的节点就归为一个社区.标签传播算法具有简单、高效、快速的优点,常常和其它算法结合,用来初始化种群,增加种群的多样性和精度,加快算法的收敛.

BGLL 算法是一种层次性贪心算法,它是一个两阶段过程.第一阶段首先将每一个节点划分作为单独的社区,然后考虑将每个节点合并到其邻接点所在的社区中,执行能使模块度增大的合并.第二阶段将第一阶段划分出来的社区聚合成为一个点后形成新的网络,再在新的网络上重复以上过程,直到网络中的结构不再改变为止.

ACO 算法以网络的一个随机社区划分作为初始解,通过编码把初始解作为路径上的初始信息素,然后利用网络的邻接矩阵作为路径选择启发函数,采用最大最小蚁群算法不断对模块度函数进行优化.迭代一定次数后,产生的路径即解码为最终的社区划分.

LGA 算法以模块度 Q 为目标函数,采用基于社区标识的编码方式,然后利用标签传播法初始化种群,采用单路交叉、局部搜索变异 LMA 和 $\mu+\lambda$ 选择等 3 个算子来探测网络社区结构.

IDDE 算法的基本思想是先用标签传播法初始化种群;然后借鉴差分进化算法基本特征,采用随机单点变异、单路交叉和贪婪选择策略生成中间种群,最后对执行免疫克隆选择操作,生成下一代种群.该算法融合离散差分进化算法的全局搜索能力和免疫克隆选择策略的局部搜索能力,增加了算法的鲁棒性.

CC-GA 算法通过使用聚类协同系数来初始化种群,以模块度为适应度函数,利用遗传中的统一交叉和提出的变异操作进行更新种群,该算法不需要预先知道社区的数量.可用于稠密和稀疏网络的社区发现.

1.3 Jaya 算法

Jaya 算法是近年来出现的一种元启发式优化算法. Jaya 这个词在梵语中是胜利的意思.因为该算法力求达到最优解从而取得胜利,因此命名为 Jaya 算法^[30]. Jaya 算法中,第一步是随机初始化规模大小为 N_p 的种群,后续每一步不断迭代,每次迭代中要完成从一个种群更新到下一个种群的任务,直到满足最大迭代次数为止.每一代种群要选出最好解和最差解用于更新种群中的每个个体解.种群中个体的更新采用下述公式:

$$X_i(t+1) = X_i(t) + r_1 \times (X_B(t) - |X_i(t)|) - r_2 \times (X_W(t) - |X_i(t)|), \quad i = 1, 2, \dots, N_p \quad (2)$$

其中, $X_B(t)$ 和 $X_W(t)$ 分别表示第 t 代种群中最好解和最差解, i 表示种群中第 i 个个体解. r_1 和 r_2 每个个体更新时产生的 0 到 1 的随机数,用作缩放因子试图增强算法的多样性.式中 $+r_1 \times (X_B(t) - |X_i(t)|)$ 表明个体 X 有朝着最优个体靠近的趋势,式中 $-r_2 \times (X_W(t) - |X_i(t)|)$ 则表明个体 X 有远离最差个体的趋势.在每一代种群个体更新过程中,如果种群中下一代个体 $X_i(t+1)$ 的目标函数值更优,则替代上一代个体 $X_i(t)$, 否则,保持不变.

Jaya 与其它元启发式优化算法相比,简单易理解,参数较少,只需要设定种群的大小和迭代的次数. Jaya 算法的流程图如图 1.

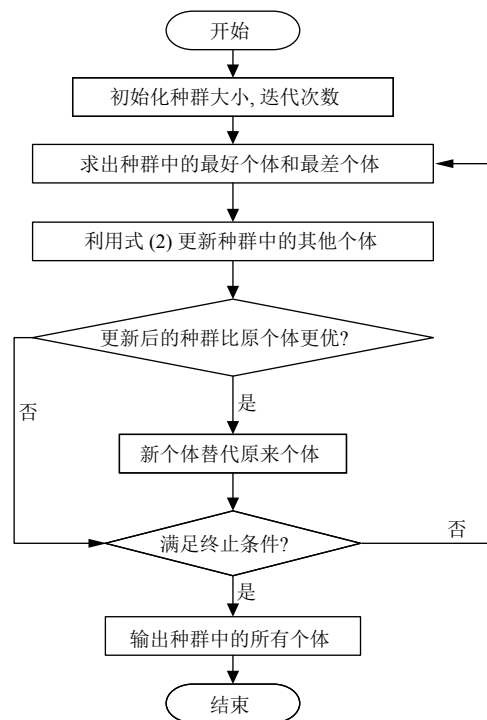


图 1 Jaya 算法流程图

2 离散 Jaya 的社区发现算法

在本节中,我们主要介绍如何对 Jaya 算法进行离散化.主要从社区的编码和初始化、离散化策略以及算法伪代码几方面展开.

2.1 社区结构的表示和初始化

社区结构对应网络的划分,其表示方式往往与算法相关且是影响算法效果的一个关键.社区发现的进化类算法中,一个社区结构对应的网络划分可以看作

是一个个体. 目前有两种经典的个体表示方法^[32], 一种是基于社区标识符表示法 LR (Label-based Representation), 另外一种是局部邻接表示 LAR (Locus-based Adjacency Representation). 本文采用局部邻接表示法, 在该表示法中, 向量的元素值代表该节点的某个邻居节点的节点标识符. 如图 2, 网络局部邻接表示如表 1, 其对应的网络划分见图 3.

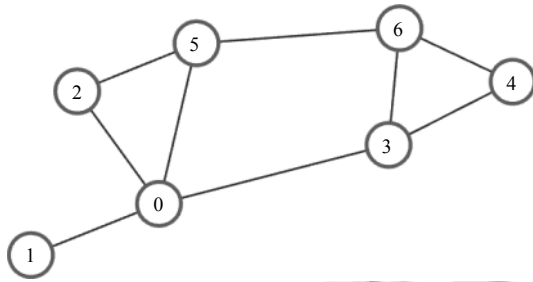


图 2 一个网络示例

表 1 网络的局部邻接表示

| 索引 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|----|---|---|---|---|---|---|---|
| 元素 | 2 | 0 | 5 | 4 | 6 | 0 | 3 |

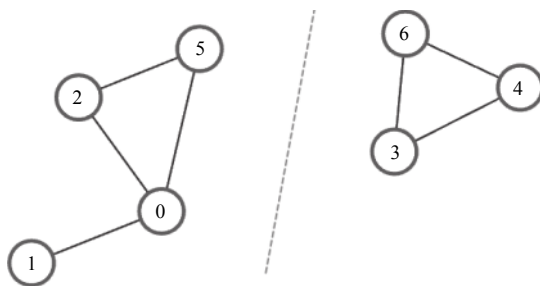


图 3 表 1 对应的局部邻接表示所导出的网络划分

根据上述社区表示结构方式, 本文使用类似标签传播思想的方法进行种群的初始化. 在初始化过程中, 首先使网络中的每个节点都随机选择其中一个邻居节点, 每一次迭代中, 如果节点 i 的邻居节点出现的次数最多, 则该邻居节点作为节点 i 的值; 如果邻居节点出现的次数一样, 则随机选择一个邻居节点.

2.2 离散化策略

Jaya 算法适用于求解连续性的最优化的问题. Jaya 算法的主要思想是通过远离最差解, 靠近最优解更新种群. 社区发现是一个离散型的组合优化问题. 当使用个体向量表示社区的一个划分时, 该向量中的每一个元素代表一个节点, 如果利用 Jaya 算法对个体向

量更新, 如对个体向量的作差运算, 但节点之间的作差运算并无意义. 为了采用 Jaya 算法的思想求解社区结构划分问题, 我们需要对 Jaya 算法做离散化处理, 为此先引入几个概念.

(1) 倾向性

为了判断种群中的个体是更靠近最优解个体还是更靠近最差解个体, 本文提出倾向性判断公式:

$$bestGap = Q(x_{best}) - Q(x) \quad (3)$$

$$worstGap = Q(x) - Q(x_{worst}) \quad (4)$$

$$Gap = bestGap - worstGap \quad (5)$$

$Q(x_{best})$, $Q(x_{worst})$, $Q(x)$ 分别表示最优个体的模块度、最差个体的模块度、当前个体的模块度, $bestGap$ 表示当前个体模块度和最优模块度的差值, $worstGap$ 表示当前个体模块度和最差模块度的差值, Gap 表示当前个体在最优个体和最差个体之间的倾向性. 当 $Gap > 0$, 则该个体倾向于最差个体, 当 $Gap < 0$, 则倾向于最优个体.

(2) 最优相似率

最优相似率即当前个体向量与最优解个体向量相同节点个数占全部元素的比率, 其计算公式如下:

$$similarRate = \sum_{d=1}^n (X_d \Theta X_d^{best}) / n \quad (6)$$

其中, X_d 表示当前个体, X_d^{best} 表示最优解个体, n 表示节点个数, d 表示该个体中元素的序数, Θ 表示同或, 如果 X_d 和 X_d^{best} 相等, 则结果为 1, 否则为 0.

(3) 模块度函数的局部更新

为了使得模块度函数最大化, 我们定义一种模块度函数的局部贪心更新方式. 通过依次遍历该节点的邻居节点, 选择模块度最大化的邻居节点进行替代. 第 $t+1$ 代的个体更新用式 (7) 表示如下:

$$X_d^{t+1} = \arg \max_j Q(X_d^j), j \in Neighbor_d \quad (7)$$

其中, X 表示一个个体向量解 (向量的元素对应图中的节点), t 表示第 t 次迭代, Q 为对应式 (1) 的模块度函数, $Neighbor_d = \{n_1, n_2, \dots, n_k\}$ 为节点 d 的邻居节点集合. 式 (7) 表示取最大模块度的邻居节点作为下一代个体的节点更新.

(4) 贪心选择

当个体更新后, 选取适应度较高的个体作为下一

代的父种群个体,即贪心选择.该操作相当于遗传中的“优胜劣汰”,同时和 Jaya 算法保持了一致的筛选策略,对比非贪心选择策略,即优化后的个体不进行比较直接作为下一代,选用贪心选择策略保证了下一代种群中每一个个体都至少不会变得更差,从而使解的平均质量更优,加快了收敛速度.

基于上述几个概念,下面给出 Jaya 算法离散化策略,离散化的 Jaya 算法称作 DJaya 算法. DJaya 算法通过式 (3)~式 (5) 来判断种群中每个个体是靠近种群的最优解还是最差解,从而对种群中的个体离散化为两类. ① 对于种群中更靠近最差解的个体,采用类似标签传播的方法依次对该个体的各个节点进行更新操作,该操作提高了较差个体的多样性和精度,达到了进一步远离最差解的效果,有利于 DJaya 算法的全局探索性. ② 对于种群中靠近最优解的个体,先对该个体向量采用式 (6) 计算出最优相似率,然后基于相似率对该个体的某些节点采用式 (7) 进行局部更新,该操作有利于加快种群的优化速度,使整个种群更加靠近最优解,从而有利于增强 DJaya 算法的局部开发性.

DJaya 算法通过判定种群中个体的倾向性来平衡各个阶段的全局搜索能力和局部搜索能力,当个体靠近最优个体解,采取类似标签传播方法进行更新,相当于重新定义初始解,有利于探索出最优解,体现了全局搜索能力;当算法靠近最有个体解,采取了基于局部函数的更新方法,进一步优化了精英解,加快了算法收敛,体现了局部搜索能力.

(5) 随机抖动

为了更进一步避免算法陷入全局较优,我们加入了随机抖动操作,相当于遗传算法中的变异操作,是一种局部开发,有利于增加搜索方向,增大搜索空间.当连续迭代几次后种群中的个体没有得到优化,则执行随机抖动,即随机选择该个体节点的某个邻居节点值替代当前个体节点值.该操作设置了一个较小的抖动率,一方面保持原始种群的较优性,另一方面增加多样性.

2.3 DJaya 算法流程

综上,通过选取模块度为适应度函数,采用 LAR 表示种群的划分和类似 LPA 策略进行初始化,然后在每代种群更新中采用离散化的 DJaya 算法进行不断的迭代优化,选取适应度高的个体作为下一代的父种群.

算法 DJaya 的描述如算法 1.

算法 1. DJaya 算法

输入: G, M, T, R // G 表示网络, M 表示种群规模, T 表示最大迭代次数, R 表示随机抖动率

输出: $Result$ // 网络社区结构

```

1. Begin
2.  $P \leftarrow LAR$  进行表示,并用 LPA 初始化父种群
3.  $p^{(new)} \leftarrow \emptyset$  //子种群初始化为空集
4. For  $i = 1: T$ 
5.   For  $j = 1: M$ 
6.      $Gap(j) \leftarrow$  利用式 (3)~(5) 判断个体倾向性
7.     If  $Gap(j) > 0$  //靠近最差个体
8.        $p^{(new)} \leftarrow LPA(j)$  //类似标签传播法更新
9.     Else //靠近最优个体
10.      If  $rand() < similarRate(j)$  //满足最优相似率
11.         $p^{(new)} \leftarrow$  利用公式 (7) 更新
12.       $p^{(new)} \leftarrow p^{(new)} \cup P$  //更优个体进入下一代
13.      If  $rand() < R$ 
14.         $p^{(new)} \leftarrow$  随机抖动
15.    End For
16.   $P \leftarrow p^{(new)}$  //生成新的种群
17. End For
18.  $Result \leftarrow$  选择  $P$  中适应度最高的个体
19. End

```

下面分析 DJaya 算法的时间复杂度.假设网络中 $G = (V, E)$ 的节点个数为 n ,种群规模为 M ,迭代次数为 T .使用局部邻接表示法初始化种群时,对于每个节点随机选择其中一个节点值赋值,设网络的节点平均度为 \bar{k} ,则该初始化种群的其时间复杂度为 $O(Mn\bar{k})$;利用 DJaya 算法离散策略对种群进行更新时,计算模块度函数的时间复杂度为 $O(Mn^2)$,如果个体靠近最差解,则利用类似标签传播法进行更新,其时间复杂度为 $O(TM \times (n\bar{k} + n^2))$;如果个体靠近最优解,则采用局部函数进行更新,其时间复杂度为 $O(TM \times (\bar{k}n^2 + n^2))$.因此, DJaya 算法的时间复杂度为 $O(TMn^2\bar{k})$.

3 实验结果与分析

实验平台为 Windows 10 系统, Intel(R) Core(TM) i5 CPU 1.8 GHz, 内存 8.0 GB.实验数据包括不同参数情况下的扩展 GN Benchmark^[33] 人工网络和 4 个经典真实网络.实验中将 DJaya 算法与 LPA, FMM, BGLL 等经典贪心算法,以及智能算法如 ACO, LGA, IDDE, CC-GA 进行比较.为了减少统计误差,每个算法在每个算例上独立运行 20 次,取平均值作为算法运行的结果.

按照惯例, 针对人工网络, 已知其真实社区结构, 实验中仅比较各算法的标准互信息值; 对于真实的网络, 我们比较不同算法的模块度值^[34].

标准互信息 (Normalized Mutual Information, *NMI*)^[35]常用来评价算法的划分结果与网络真实划分结果的吻合程度. 对于网络的两种划分 *A* 和 *B*, 它们之间的互信息定义如下:

$$NMI = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} \log(C_{ij}N/C_i C_j)}{\sum_{i=1}^{C_A} C_i \log(C_i/N) + \sum_{j=1}^{C_B} C_j \log(C_j/N)} \quad (8)$$

其中, *N* 表示节点个数. *C_A* 表示划分 *A* 中网络的社区数目, *C_B* 表示划分 *B* 中网络的社区数目, *C_i* 表示第 *i* 行元素和, *C_j* 表示第 *j* 列元素之和, *C_{i,j}* 表示划分 *A* 中的社区 *i* 和划分 *B* 中的社区 *j* 两者之间的交集中含有的节点数目. 如果 *A* = *B*, 则 *NMI* 值为 1, 如果 *A* 划分和 *B* 划分完全不同, 则 *NMI* 值为 0.

3.1 人工网络比较实验

为了分析 DJaya 算法的优化性能, 我们在扩展 GN Benchmark 上进行了实验分析. 该人工网络的社区结构已知, 它含有 128 个节点, 共划分为 4 个不同的社区, 每个社区有 32 个节点, 社区中每个节点的平均度数为 16, 如表 2. 网络中的混合参数 *u* 主要用于确定某一社区中任意一个节点与其他社区的节点之间共享边的数量, 当混 *u* < 0.5 时, 网络中的社区结构较为明显, 但随着 *u* 的增大, 网络中的社区结构将变得越来越模糊, 性能一般的算法难以检测出网络中存在的社区结构, 从而可以区分不同算法的性能, 比较结果如图 4.

表 2 人工网络的参数 (GN benchmark)

| | |
|-------|------|
| 节点数 | 128 |
| 边数 | 2048 |
| 平均节点数 | 16 |

由图 4 可见, 当 *u* < 0.35 时候, 各个算法的 *NMI* 值达到了 1, 都能准确找出社区结构; 当 *u* = 0.4, BGLL、ACO 算法的 *NMI* 开始下降; 当 *u* = 0.45, LGA、IDDE、DJaya 算法能找到和真实社区完全一样的社区结构, 相比其它算法, 它们较为稳定; 当 *u* = 0.5, 此时网络的社区结构已经非常模糊, DJaya 算法仍能准确划分接近 80% 的社区结构, 表明 DJaya 算法的一定优越性, 总体来说 DJaya 算法在人工网络上优于其它算法.

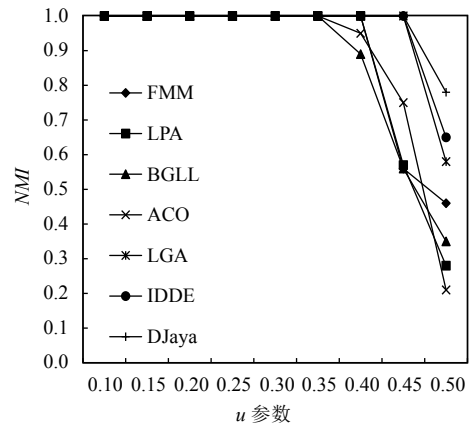


图 4 各种算法划分精度的比较

3.2 真实网络比较实验

真实的复杂网络与人工网络相比, 有不同的拓扑属性. 这里采用了 4 个被广泛使用的真实网络作为实验数据集, 其点数和边数见表 3. Karate^[36]网络是对一个美国大学空手道俱乐部进行观测而构建出的一个社会网络. Dolphins^[37]是由 Lusseau 等使用长达 7 年的时间观察新西兰 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. Polbooks^[6]是由 Amazon 上销售的美国政治相关书籍页面上建立起来的网络. Football^[1]网络是根据美国大学生足球联赛而创建的一个复杂的社会网络.

表 3 4 个真实网络数据的参数

| 网络 | 节点数 | 边数 |
|----------|-----|-----|
| Karate | 34 | 78 |
| Dolphins | 62 | 159 |
| Polbooks | 105 | 441 |
| Football | 115 | 613 |

表 3 中, football 真实社区的划分效果与 DJaya 算法划分的效果图分别如图 5 和图 6.

如图 5, football 对应的真实社区有 12 个, 图 6 采用 DJaya 算法划分的社区则有 10 个, 可以看出 DJaya 算法合并了一些较小的社区, 保持了大部分社区划分是和真实社区一致, 文献[2]认为这种划分是合理的.

表 4 给出了各种算法在 4 个标准真实网络算例上所得的模块度值. 对于算例 Karate, CC-GA、LGA、DJaya 都达到 0.4198 最优值^[38]; 在算例 Dolphins 上 DJaya 算法取得了最高的模块度值; 在算例 polbooks

上, DJaya 和 LGA 都达到了该数据集的理论最高值 0.5272^[39]; 在算例 football上, LGA、IDDE、DJaya 都表现良好, 达到了最优值^[39]。综上所述, 在 4 个算例上, DJaya 算法展现出较明显优势。

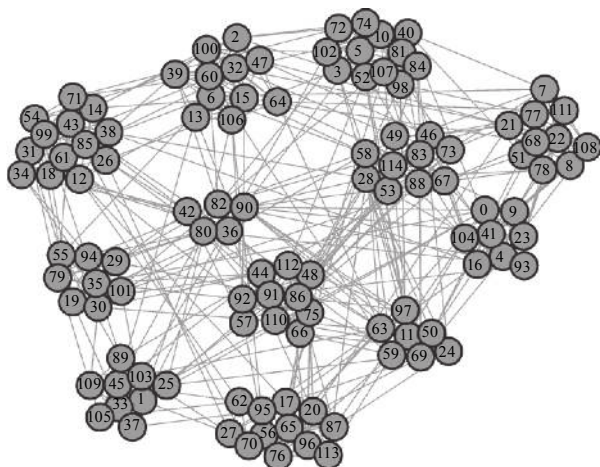


图 5 真实的 football 的社区 (12 个社区)

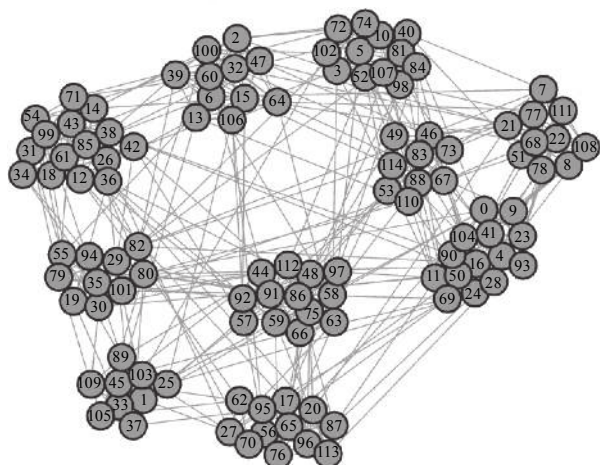


图 6 DJaya 划分的 football 社区 (10 个社区)

4 总结

本文提出的 DJaya 算法用于求解复杂网络社区发现问题, 该算法采用局部邻接表示法的编码方式对种群进行社区表示, 并用标签传播思想进行初始化, 然后借鉴连续空间 Jaya 算法的靠近最优解、远离最差解的基本思想, 先判断种群个体的倾向性, 针对靠近最差解采用标签传播思想更新种群, 靠近最优解采用局部模块度函数更新种群, 最后采用贪心选择策略生成下一代种群。DJaya 算法实现简单, 容易理解, 具有全局搜索能力和局部开发能力。通过在真实网络和人工网络上

的实验对比发现, 该算法具有较好的优化能力和鲁棒性, 达到了较高的社区发现精度且自动确定社区个数。该算法适用于非重叠、静态社区, 考虑到模块度作为优化函数的局限性^[39], 今后将改进算法的目标函数, 或建立多目标优化函数, 同时考虑扩展算法的适应范围, 增大数据规模, 应用在重叠、动态社区发现问题。

表 4 各种算法在不同网络中的模块度对比

| 类型 | 算法 | Karate | Dolphins | Polbooks | Football |
|----------|-------|---------------|---------------|---------------|---------------|
| 贪心启发式算法 | FMM | 0.4013 | 0.5194 | 0.5168 | 0.5996 |
| | LPA | 0.3536 | 0.4860 | 0.5039 | 0.5881 |
| | BGLL | 0.4005 | 0.5164 | 0.5059 | 0.6018 |
| | DJaya | 0.4198 | 0.5283 | 0.5272 | 0.6046 |
| 智能元启发式算法 | ACO | 0.4198 | 0.5271 | 0.5271 | 0.6045 |
| | IDDE | 0.4198 | 0.5282 | 0.5272 | 0.6046 |
| | CC-GA | 0.4198 | 0.5263 | 0.5252 | 0.5632 |
| | LGA | 0.4198 | 0.5280 | 0.5272 | 0.6046 |
| | DJaya | 0.4198 | 0.5283 | 0.5272 | 0.6046 |

参考文献

- Girvan M, Newman MEJ. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821–7826. [doi: 10.1073/pnas.122653799]
- Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Physical Review E*, 2004, 69(2): 026113. [doi: 10.1103/PhysRevE.69.026113]
- Kernighan BW, Lin S. An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal*, 1970, 49(2): 291–307. [doi: 10.1002/j.1538-7305.1970.tb01770.x]
- Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. *Physical Review E*, 2004, 70(6): 066111. [doi: 10.1103/PhysRevE.70.066111]
- Newman MEJ. Fast algorithm for detecting community structure in networks. *Physical Review E*, 2004, 69(5): 066133.
- Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(23): 8577–8582. [doi: 10.1073/pnas.0601602103]
- Xing Y, Meng FR, Zhou Y, et al. A node influence based

- label propagation algorithm for community detection in networks. *The Scientific World Journal*, 2014, 2014: 627581.
- 8 Van Dongen S. Graph clustering by flow simulation [Ph. D. thesis]. Utrecht: University of Utrecht, 2000. 1–173.
- 9 Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(4): 1118–1123. [doi: [10.1073/pnas.0706851105](https://doi.org/10.1073/pnas.0706851105)]
- 10 Jia GB, Cai ZX, Musolesi M, *et al.* Community detection in social and biological networks using differential evolution. *Proceedings of the 6th International Conference on Learning and Intelligent Optimization*. Paris, France. 2012. 71–85.
- 11 Cai Q, Gong MG, Ma LJ, *et al.* Greedy discrete particle swarm optimization for large-scale social network clustering. *Information Sciences*, 2015, 316: 503–516. [doi: [10.1016/j.ins.2014.09.041](https://doi.org/10.1016/j.ins.2014.09.041)]
- 12 Tasgin M, Herdagdelen A, Bingol H. Community detection in complex networks using genetic algorithms. arXiv: 0711.0491, 2006.
- 13 金弟, 刘杰, 杨博, 等. 局部搜索与遗传算法结合的大规模复杂网络社区探测. *自动化学报*, 2011, 37(7): 873–882.
- 14 张英杰, 龚中汉, 陈乾坤. 基于免疫离散差分进化算法的复杂网络社区发现. *自动化学报*, 2015, 41(4): 749–757.
- 15 Duch J, Arenas A. Community detection in complex networks using extremal optimization. *Physical Review E*, 2005, 72(2): 027104. [doi: [10.1103/PhysRevE.72.027104](https://doi.org/10.1103/PhysRevE.72.027104)]
- 16 Guimerà R, Amaral LAN. Functional cartography of complex metabolic networks. *Nature*, 2005, 433(7028): 895–900. [doi: [10.1038/nature03288](https://doi.org/10.1038/nature03288)]
- 17 Blondel VD, Guillaume JL, Lambiotte R, *et al.* Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 2008(10): 10008. [doi: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008)]
- 18 Pizzuti C. Evolutionary computation for community detection in networks: A review. *IEEE Transactions on Evolutionary Computation*, 2018, 22(3): 464–483. [doi: [10.1109/TEVC.2017.2737600](https://doi.org/10.1109/TEVC.2017.2737600)]
- 19 Cai Q, Ma LJ, Gong MG, *et al.* A survey on network community detection based on evolutionary computation. *International Journal of Bio-Inspired Computation*, 2016, 8(2): 84–98. [doi: [10.1504/IJBIC.2016.076329](https://doi.org/10.1504/IJBIC.2016.076329)]
- 20 Liu J, Abbass HA, Tan KC. Evolutionary community detection algorithms. In: Liu J, Abbass HA, Tan KC, eds. *Evolutionary Computation and Complex Networks*. Cham: Springer, 2019. 77–115.
- 21 Talbi EG. A parallel genetic algorithm for the graph partitioning problem. *Proceedings of the 5th International Conference on Supercomputing*. New York, NY, USA. 1991. 312–320.
- 22 Bui TN, Moon BR. Genetic algorithm and graph partitioning. *IEEE Transactions on Computers*, 1996, 45(7): 841–855. [doi: [10.1109/12.508322](https://doi.org/10.1109/12.508322)]
- 23 Tasgin M, Herdagdelen A, Bingol H. Community detection in complex networks using genetic algorithms. arXiv: 0711.0491, 2007.
- 24 Gog A, Dumitrescu D, Hirsbrunner B. Community detection in complex networks using collaborative evolutionary algorithms. *Proceedings of the 9th European Conference on Artificial Life*. Lisbon, Portugal. 2007. 886–894.
- 25 Pizzuti C. Ga-Net: A genetic algorithm for community detection in social networks. *Proceedings of the 10th International Conference on Parallel Problem Solving from Nature*. Dortmund, Germany. 2008. 1081–1090.
- 26 Gong MG, Fu B, Jiao LC, *et al.* Memetic algorithm for community detection in networks. *Physical Review E*, 2011, 84(5): 056101.
- 27 Chang HH, Feng ZR, Ren ZG. Community detection using ant colony optimization. *Proceedings of 2013 IEEE Congress on Evolutionary Computation*. Cancun, Mexico. 2013. 3072–3078.
- 28 Said A, Abbasi RA, Maqbool O, *et al.* CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks. *Applied Soft Computing*, 2018, 63: 59–70. [doi: [10.1016/j.asoc.2017.11.014](https://doi.org/10.1016/j.asoc.2017.11.014)]
- 29 Rao RV. Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *International Journal of Industrial Engineering Computations*, 2016, 7(1): 19–34.
- 30 Gao KZ, Yang FJ, Zhou MC, *et al.* Flexible job-shop rescheduling for new job insertion by using discrete Jaya algorithm. *IEEE Transactions on Cybernetics*, 2019, 49(5): 1944–1955. [doi: [10.1109/TCYB.2018.2817240](https://doi.org/10.1109/TCYB.2018.2817240)]
- 31 Gao KZ, Zhang YC, Sadollah A, *et al.* Jaya, harmony search and water cycle algorithms for solving large-scale real-life urban traffic light scheduling problem. *Swarm and Evolutionary Computation*, 2017, 37: 58–72. [doi: [10.1016/j.swevo.2017.05.002](https://doi.org/10.1016/j.swevo.2017.05.002)]
- 32 Hruschka ER, Campello RJGB, Freitas AA. A survey of evolutionary algorithms for clustering. *IEEE Transactions on*

- Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2009, 39(2): 133–155. [doi: [10.1109/TSMCC.2008.2007252](https://doi.org/10.1109/TSMCC.2008.2007252)]
- 33 Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 2008, 78(4): 046110. [doi: [10.1103/PhysRevE.78.046110](https://doi.org/10.1103/PhysRevE.78.046110)]
- 34 Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 2007, 76(3 Pt 2): 036106.
- 35 Danon L, Díaz-Guilera A, Duch J, *et al.* Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005, 2005(9): 09008. [doi: [10.1088/1742-5468/2005/09/P09008](https://doi.org/10.1088/1742-5468/2005/09/P09008)]
- 36 Zachary WW. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 1977, 33(4): 452–473. [doi: [10.1086/jar.33.4.3629752](https://doi.org/10.1086/jar.33.4.3629752)]
- 37 Lusseau D. The emergent properties of a dolphin social network. *Proceedings of the Royal Society B: Biological Sciences*, 2003, 270(S2): S186–S188.
- 38 Aloise D, Cafieri S, Caporossi G, *et al.* Column generation algorithms for exact modularity maximization in networks. *Physical Review E*, 2010, 82(4): 046112. [doi: [10.1103/PhysRevE.82.046112](https://doi.org/10.1103/PhysRevE.82.046112)]
- 39 Fortunato S, Barthelemy M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(1): 36–41. [doi: [10.1073/pnas.0605965104](https://doi.org/10.1073/pnas.0605965104)]