

# 一种基于差分进化的社团检测算法

孙韩林<sup>1,2</sup>, 马素刚<sup>1,2</sup>, 王忠民<sup>1,2</sup>

(1.西安邮电大学计算机学院, 陕西 西安 710121;

2.西安邮电大学陕西省网络数据智能处理重点实验室, 陕西 西安 710121)

**摘要:** 复杂网络的社团结构分析可抽象为一个优化问题, 用进化算法求解。进化类算法的一个基本问题是如何把问题的候选解编码到进化个体中。本文将索引局部邻接表示法用于社团检测进化算法的个体表示, 把社团结构分析转化为一个整数优化问题。在该个体表示方法的基础上, 提出了一种基于差分进化的社团检测算法。在一组合成网络和真实网络上验证了算法性能, 并与两种基于遗传算法的典型社团检测进化算法进行了对比。实验结果表明, 当网络社团结构较为清晰时, 基于差分进化的算法检测到的社团结构具有更好的质量。

**关键词:** 社团检测; 社团结构分析; 差分进化; 复杂网络

**中图分类号:** TP311 **文献标识码:** A

## A Community Detection Algorithm Using Differential Evolution

SUN Hanlin<sup>1,2</sup>, MA Sugang<sup>1,2</sup>, WANG Zhongmin<sup>1,2</sup>

(1. School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;

2. Shanxi Key Laboratory of Network Data Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

**Abstract:** Community structure analysis of complex networks can be modeled as an optimizing problem, and then be solved by Evolution Algorithm (EA). One fundamental issue of EA is how to encode a candidate solution into an evolution individual. In this paper, the Indexed Locus-based Adjacency Representation (ILAR) of evolution individual encoding for the community detection problem is proposed. Therefore, a community detection problem can be converted to a discrete integer optimization problem. Based on the ILAR, a community detection algorithm that uses Differential Evolution (DE) as the search engine is developed. A number of experiments are conducted on synthesized and real-world networks to verify the performance of the proposed algorithm, and the results are compared against those of two typical community detection algorithms based on Genetic Algorithm (GA). The experiment results show that the community structure discovered by the proposed DE-based algorithm generally has better quality than those of the two compared algorithms as the community structure of the analyzed network is sound.

**Keywords:** community detection; community structure analysis; differential evolution; complex network

### 1 引言(Introduction)

社团是复杂网络中一组节点的集合, 组内节点具有共同的属性或在网络中具有相似的功能, 拓扑结构上表现为组内节点间具有更紧密(更多)的连接边, 而组内成员与网络其余节点的连接边相对稀疏。社团结构从中观(middle-scope level)层次上揭示了网络的结构特性, 对理解网络性质具有重要意义。

社团结构检测可抽象为一个优化问题, 利用进化算法(Evolution Algorithm, EA)进行求解, 即定义一种度量社团结构质量的目标函数, 再利用进化算法求解函数的极大值或

极小值。用进化算法进行社团检测的主要优点是不需要事先知道社团结构的属性(例如社团的数量)。常用于社团检测的进化算法有遗传算法<sup>[1-5]</sup>(Genetic Algorithm, GA)和差分进化<sup>[6-8]</sup>(Differential Evolution, DE)算法。

进化算法的一个基本问题是如何将一种可能的候选解决方案编码到进化个体中。本文提出了索引局部邻接进化个体表示法, 将社团检测转化为整数优化问题; 在此基础上, 提出一种以DE为搜索引擎的社团检测算法。DE被认为是求解实数优化问题的最优进化算法之一<sup>[9]</sup>。

### 2 进化个体表示(Evolutionary individual representation)

在用进化算法求解社团检测问题时，有两种典型的个体编码方案，即“线性组表示法”(String-of-Group Representation,SGR)和“局部邻接表示法”(Locus-based Adjacency Representation,LAR)<sup>[10]</sup>。这两种表示方法中，进化个体都是一个 $N$ 维向量， $N$ 是网络中的节点数，向量的每一维代表了网络中的一个节点。在SGR中，每个维度的值是该维代表节点所属社团的社团标识符；而在LAR中，则是该维代表节点的某个邻居的节点标识符(即该邻居节点在向量中的维序号)，通过邻接关系连接在一起的一组节点构成一个社团。

文献[6]—文献[8]中基于DE的社团检测算法都采用了SGR表示法。差分进化算法中，变异操作中要进行个体求差运算，显然对社团标识符进行求差没有实际意义。同样，若采用LAR表示法，对邻居的节点标识符求差同样没有意义。本文对局部邻接表示法进行改进，提出“索引局部邻接表示法”(Indexed Locus-based Adjacency Representation, ILAR)，从而将社团检测问题转化为整数优化问题，这样就可以在差分变异操作中直接使用传统的求差运算。

索引局部邻接表示法中，个体同样是一个 $N$ 维向量，每一维也代表网络中的一个节点。与LAR不同的是，向量每一维的值是该维代表节点的某个邻居节点的邻居索引标识符，而不是邻居的节点标识符。一个节点可能有若干邻居，把所有邻居按一定顺序排列，例如按邻居的节点标识符升序或降序排列，则该节点某个邻居的邻居索引标识符就是此邻居节点在邻居列表中的序数。一个节点可以是多个节点的邻居，对不同邻居节点，它的邻居索引标识符是不同的。因此，采用索引局部邻接表示法后，网络中的每个节点会同时有多个标识符——一个节点标识符和若干邻居索引标识符，邻居索引标识符的数量与节点的邻居的数量相同。索引局部邻接表示法可以看作是一种归一化的局部邻接矩阵表示。

解析索引局部邻接表示个体中编码的社团结构分为两步：首先把向量每一维的邻居索引标识符替换为相应的邻居节点标识符，然后再找出通过邻居关系关联在一起的节点组，每个组即是一个社团。第二步操作与解析局部邻接表示个体中社团结构的过程相同。

图1给出了一个简单的用索引局部邻接表示法表示进化个体的例子。图1(a)是一个简单网络的拓扑结构图，该网络包含9个节点，它们分为两个社团，分别是{1,2,3,4,5}和{6,7,8,9}；图1(b)列出了每一个节点的邻居节点列表，按邻居的节点标识符升序排列，并给每个邻居分配了邻居索引标识符；图1(c)给出了一个采用索引局部邻接表示法的可能的进化个体，其中节点1连接到其第1个邻居(节点2)，节点2连接到

其第3个邻居(节点4)，节点3连接到其第1个邻居(节点1)，其余维类似；图1(d)则是用节点标识符替换相应维邻居索引标识符后得到的进化个体，即采用局部邻接表示法的个体；图1(e)是在该个体中编码的社团结构的图形化表示。

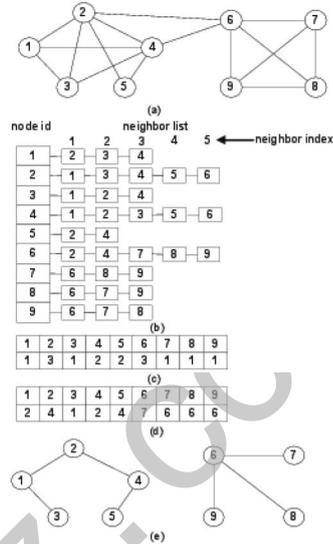


图1 采用索引局部邻接表示法表示进化个体的例子  
Fig.1 An example of an evolution individual by Indexed Locus-based Adjacency Representation

### 3 算法描述(Algorithm description)

差分进化算法包括四个步骤，即初始化、变异、交叉和选择，其中初始化只在算法开始时执行一次，而后面三步操作则迭代多次，直到算法结束<sup>[11]</sup>。在索引局部邻接个体表示法的基础上，本文提出基于差分进化的社团检测算法DECDILAR(Differential Evolution Community Detection Algorithm based on ILAR)。

#### 3.1 初始化

在初始化操作中，差分进化算法随机生成一个有 $NP$ 个个体的种群，其中每个个体都是一个 $N$ 维实数向量，代表了一种维优化问题的候选解。由于每一维都关联到优化问题的一个物理参数，其取值范围通常必须限制在一个范围内，且每一维的取值范围可能不同。初始化种群中个体的每一维都要尽可能均匀、随机地分布在它的取值范围内。

采用索引局部邻接表示法时，社团检测可以转化成一个个体的第 $j$ 维被随机初始化为1到第 $j$ 个节点的最大邻居数(记为 $MV_j$ )之间的某个数值。记第 $i$ 个个体的第 $j$ 维为 $x_{j,i}$ ，则初始化可以表示为：

$$x_{j,i,0} = 1 + R(\text{rand}_{r,i}[0,1] \cdot (MV_j - 1)) \tag{1}$$

其中 $\text{rand}_{r,i}[0,1]$ 是一个均匀分布在区间上的随机数，该随机数对每个个体的每一维都重新独立生成一次；函数 $R(\cdot)$ 将一个实数转换成最接近的整数； $x_{j,i,0}$ 的第三维下表“0”表示当前进化代数是初始化。

### 3.2 差分变异

差分进化算法中，从当前种群中选择一个父辈个体称为“目标向量”(target vector)；通过差分变异(mutation)操作生成的变异个体称为“捐赠向量”(donor vector)；而通过捐赠向量和目标向量的交叉(crossover)操作生成的后代个体称为“试验向量”(trail vector)。

变异指个体的某些维突然发生改变。与其他多数进化算法不同，差分进化采用向量(个体)间的差异来探索搜索空间。实际上，不同差分进化算法的区别就在于变异操作中使用个体差异的模式不同。DECDILAR在生成捐赠向量的变异运算中采用了一种最简单的变异模式“DE/best/1”，其中“DE”代表差分进化，“best”表示选择当前种群中的最优个体作为变异的基向量(base vector)，“1”表示在扰动基向量时只考虑了一个差分向量(即个体间的差异)。为了防止种群早熟，我们对“DE/best/1”进行了改进，通过引入一个新参数“变异率”(mutation ratio)，设计了一种受控的变异运算。记第 $i$ 个个体 $X_{i,G}$ 的捐赠向量为 $V_{i,G}$ ( $G$ 表示当前进化代数)，则这种受控差分变异操作可表示为：

$$V_{i,G} = \begin{cases} X_{\text{best}} + R \cdot (X_{r_1',G} - X_{r_2',G}) & \text{if } \text{rand}[0,1] \leq Mr \\ X_{r_1',G} & \text{otherwise} \end{cases} \quad (2)$$

其中 $X_{\text{best}}$ 是从当前种群中选择的最好个体，作为变异的基向量； $R$ 是参数“变异速率”(mutation rate)； $X_{r_1',G}$ 和 $X_{r_2',G}$ 是两个从当前种群中随机选取的个体； $\text{rand}[0,1]$ 是均匀分布在 $[0,1]$ 区间上的一个随机数，该随机数在生成每个个体的捐赠向量时重新生成一次。

需要注意的是，通过差分变异运算获得的捐赠向量中，某些维的取值可能超出了其限制区间 $[L, NN_i]$ 。如果第 $j$ 维的取值超出了其限制区间，则将其随机地设置为一个有效值。

### 3.3 交叉

交叉操作的作用在于增强种群的多样性。通过交叉操作，捐赠向量 $V_{i,G}$ 与相应的目标向量 $X_{i,G}$ 交换部分维的值，从而生成一个新个体——试验向量 $U_{i,G}$ 。差分进化常用的一种典型交叉运算是“二项式交叉”(binomial crossover)，即对试验向量的第 $j$ 维，当一个在 $[0,1]$ 区间上产生的随机数小于或等于给定的“交叉速率” $Cr$ (crossover rate)时，其值来自于捐赠向量的第 $j$ 维，否则来自于目标向量的第 $j$ 维。二项式交叉可表示为：

$$u_{j,i,G} = \begin{cases} v_{j,i,G} & \text{if } \text{rand}_{j,i}[0,1] \leq Cr \text{ or } j = j_{\text{rand}} \\ x_{j,i,G} & \text{otherwise} \end{cases} \quad (3)$$

其中 $\text{rand}_{j,i}[0,1]$ 是针对第 $j$ 维生成的一个均匀分布在区间 $[0,1]$ 上的随机数；而 $j_{\text{rand}}$ 则是一个随机选择的维数标识，条件 $j = j_{\text{rand}}$ 保证了试验向量中至少有一维来自于捐赠向量； $j_{\text{rand}}$ 在每代进化中针对一个个体(即目标向量)生成一次。

通过多次试验，我们发现“均匀交叉”(uniform crossover)运算在社团检测中很有效。均匀交叉等可能地在搜索空间中探测各个方向。差分进化中，等效的均匀交叉可通过设置二项式交叉的交叉速率为0.5来实现。由于此时试验向量的每一维等可能地来自于捐赠向量或目标向量，而又无法事先假设哪种取值更好，因此我们为每个目标向量同时生成并保留两个试验向量，其中一个向量的一维来自于捐赠向量，而另一向量的相应维则来自于目标向量。

### 3.4 选择及优化目标函数

差分进化算法中选择(selection)操作用于确定是目标向量还是其对应的试验向量在下一代种群中存活。对最大化优化问题，选择操作可表示为：

$$X_{i,G+1} = \begin{cases} U_{i,G} & \text{if } f(U_{i,G}) \geq f(X_{i,G}) \\ X_{i,G} & \text{otherwise} \end{cases} \quad (4)$$

其中 $f(\cdot)$ 是优化目标函数。注意当试验向量的目标函数值与目标向量的目标函数值相等时，差分进化算法也用试验向量取代目标向量。

DECDILAR算法采用Newman和Girvan提出的模块度(modularity)<sup>[12]</sup>作为优化目标函数。模块度值越大，表明社团结构的质量越好。注意在交叉操作中为每个目标向量保留了两个试验向量，因此选择操作是从目标向量和两个试验向量中选择最好的个体在下一代中存活。

### 3.5 算法框架

基于差分进化的社团检测算法DECDILAR的框架如下：

**算法:**DECDILAR

**输入:**网络;最大进化代数 $G_{\text{max}}$ ;种群规模 $NP$ ;变异率 $Mr$ ;变异速率 $F$ ;交叉速率 $Cr$

**输出:**网络的社团结构

**过程:**

- 01)给网络中的每个节点分配一个节点标识符;
- 02)按邻居节点的节点标识符升序或降序列出每个节点的邻居，并给每个邻居分配一个邻居索引标识符，即其在邻居列表中的序号;
- 03)基于索引局部邻接表示法 ILAR 生成规模为 $NP$ 的初始化种群;
- 04)设置当前进化代数 $G=0$ ;
- 05)WHILE ( $G \leq G_{\text{max}}$  && 未达到结束条件) {
- 06)FOR (当前种群中的每一个个体 $X_{i,G}$ ) {
- 07)通过变异操作生成其捐赠向量 $V_{i,G}$ ;
- 08)通过交叉操作生成两个试验向量 $U_{i,G}^1$ 和 $U_{i,G}^2$ ;
- 09)通过选择操作从 $X_{i,G}$ 、 $U_{i,G}^1$ 、 $U_{i,G}^2$ 中选择最好的个体存活到下一代种群中;
- 10) }
- 11)设置当前进化代数 $G=G+1$ ;
- 12) }
- 13)选择当前种群中最好的个体，解析并输出其中编码的社团结构;

## 4 实验结果及分析(Experiment results and analysis)

在一组合成网络和真实网络上对DECDILAR算法的性能进行测试。为了对比不同搜索引擎(GA和DE)，以及进化个体表示方法对社团结构检测的影响，实验中也实现了两种基于GA的算法——GACDSGR和GACDLAR，其中GACDSGR中个体表示采用了线性组表示法，而GACDLAR中则采用了局部邻接表示法。关于这两种算法更多的细节请分别参考文献

[2]和文献[3]。社团结构质量的度量则采用了模块度和归一化互信息(Normalized Mutual Information, NMI)<sup>[13]</sup>。如果网络的真实社团结构已知, 可用NMI指示算法检测到的社团结构与真实社团结构的相似程度, 值越大, 表明两种结果越相似。

#### 4.1 实验设置

参考相关文献及通过多次试验, 在两种基于GA的算法中, 设置交叉率为0.8, 变异率为0.2; 在DECILAR算法中设置变异率( $M_r$ )为0.2, 变异速率( $F$ )为0.1, 交叉率( $C_r$ )为0.5。三种算法的种群规模( $NP$ )均设置为300, 最大进化代数( $G_{max}$ )也设置为300。

对每一个测试的网络, 所有算法都执行10次; 对每次运行所发现的社团结构计算模块度或归一化互信息, 最后计算10次运行结果的平均值及标准差。

#### 4.2 合成网络实验结果及分析

实验中用LFR模型<sup>[14]</sup>生成五个无向、无权重的复杂网络, 其混合参数( $\mu$ )分别取0.1、0.2、0.3、0.4及0.5。混合参数指明了社团外部连接数占社团成员节点总连接数的比率, 值越大, 表明网络的社团结构越模糊。LFR模型的其余参数设置为: 网络节点数( $N$ )200, 平均节点度( $k$ )20, 最大节点度( $k_{max}$ )50, 最大社团规模( $c_{max}$ )40, 最小社团规模( $c_{min}$ )20, 节点度负指数分布指数( $t_1$ )2, 社团规模负指数分布指数( $t_2$ )1。各种算法在这些网络上10次运行结果的平均模块度和平均归一化互信息如表1所示。

表1 LRF合成网络的平均模块度及平均归一化互信息

Tab.1 Average modularity and NMI of LFR synthesized networks

网络	算法	平均模块度	平均归一化互信息
$\mu = 0.1$	GACDSGR	0.6931 ± 0.0050	0.9865 ± 0.0048
	GACDLAR	0.7037 ± 0.0000	1.0000 ± 0.0000
	DECILAR	0.7037 ± 0.0000	1.0000 ± 0.0000
$\mu = 0.2$	GACDSGR	0.5683 ± 0.0053	0.9360 ± 0.0103
	GACDLAR	0.6141 ± 0.0000	1.0000 ± 0.0000
	DECILAR	0.6141 ± 0.0000	1.0000 ± 0.0000
$\mu = 0.3$	GACDSGR	0.3882 ± 0.0114	0.7964 ± 0.0275
	GACDLAR	0.4614 ± 0.0000	1.0000 ± 0.0000
	DECILAR	0.4614 ± 0.0000	1.0000 ± 0.0000
$\mu = 0.4$	GACDSGR	0.3420 ± 0.0128	0.7626 ± 0.0510
	GACDLAR	0.4244 ± 0.0002	0.9987 ± 0.0042
	DECILAR	0.4245 ± 0.0000	1.0000 ± 0.0000
$\mu = 0.5$	GACDSGR	0.2001 ± 0.0094	0.3717 ± 0.0402
	GACDLAR	0.2443 ± 0.0158	0.5061 ± 0.0977
	DECILAR	0.1903 ± 0.0224	0.3538 ± 0.0713

从表中可以看出, 当混合参数( $\mu$ )取0.1、0.2和0.3时, 算法GACDLAR和DECILAR均能发现正确的社团结构, 且它们的结果都远好于算法GACDSGR。当 $\mu$ 取0.4时, 两种基于局部邻接个体表示及其改进的算法检测到的社团结构质量仍远好于GACDSGR, 且DECILAR比GACDLAR稍好。而当 $\mu$ 增加至0.5时, 算法DECILAR发现的社团结构质量却最差, 而GACDLAR算法发现的最好。但从归一化互信息看, 此时即使是最好的社团结构, 其与网络真实社团结构的平均相似度也只有0.5061。在实际应用中, 这种质量水平也是不可接受的。

上述试验结果表明, 局部邻接个体表示法及其改进比线性组表示法更适合于社团检测问题。此外, 当网络社团结构较为清晰(实验中 $\mu \leq 0.4$ )时, 遗传算法和差分进化两种搜索引擎都能够发现高质量的社团结构, 且DE更有可能发现更好的结果。

#### 4.3 真实网络实验结果及分析

真实网络的拓扑结构特性可能比合成网络更为复杂。我们也用三个真实的网络上测试了三种算法。这些网络分别是: 宽吻海豚网络(bottlenose dolphins network)<sup>[14]</sup>, 共有62个节点、159条边; 美国学院足球比赛网络(American college football network)<sup>[15]</sup>, 共有115个节点、616条边; 新陈代谢网络(metabolic network)<sup>[16]</sup>, 共有453个节点、2025条边。其中宽吻海豚网络和美国学院足球比赛网络的真实社团结构已知(由人工建立), 分别包含两个和19个社团。对这三个网络各算法10次运行结果的平均模块度及平均归一化互信息如表2所示(新陈代谢网络的真实社团结构未知, 因而无法计算其归一化互信息)。从表中可以看出, 结果与合成网络类似: 两种基于局部邻接个体表示及其改进的算法发现的社团结构质量要好于GACDSGR算法, 而DECILAR发现的社团结构质量要好于GACDLAR。

表2中也给出了宽吻海豚网络和美国学院足球比赛网络的真实网络社团结构的模块度值。可以看到, 三种算法检测到的社团结构的模块度值均比真实模块度值大, 尤其是宽吻海豚网络。表2也给出了各算法发现的社团的数量, 可以看到, 对宽吻海豚网络, 三种算法发现的社团数量比真实社团数量(2个)更多, 平均接近5个, 表明算法发现的社团结构更为精细; 而对美国学院足球比赛网络, 尽管算法发现的社团数量小于真实社团数量(19个), 但真实社团结构中包含了8个只有1个成员节点的社团, 主要社团的数量(11个)与算法发现的社团数量相当(平均接近10个)。模块度表明算法发现的社团结构质量更好。

表2 真实网络的平均模块度、平均归一化互信息及检测到社团数量

Tab.2 Average modularity, NMI of real networks and number of detected communities

网络	算法	真实模块度	平均模块度	平均归一化互信息	平均检测到社团数
	GACDSGR	0.3735	0.5259 ± 0.0021	0.5574 ± 0.0427	4.8
Dolphins	GACDLAR		0.5283 ± 0.0005	0.5850 ± 0.0047	4.9
	DECDILAR		0.5283 ± 0.0006	0.5856 ± 0.0054	4.9
	GACDSGR	0.5744	0.5977 ± 0.0044	0.9112 ± 0.0263	9.6
Football	GACDLAR		0.6046 ± 0.0000	0.9328 ± 0.0000	10
	DECDILAR		0.6046 ± 0.0000	0.9328 ± 0.0000	10
	GACDSGR	-----	0.3627 ± 0.0084	-----	24.7
Metabolic	GACDLAR		0.4455 ± 0.0013	-----	9.9
	DECDILAR		0.4470 ± 0.0017	-----	9.8

图2是DECDILAR检测到的一种宽吻海豚网络的社团结构，其中不同的节点形状表示真实的社团，不同形状及填充颜色表示算法发现的社团。可以看到，算法共发现了5个社团，其中社团1、3、4、5节点形状相同，它们合并后就是一个真实的社团，即算法发现的社团结构更为精细。表3是DECDILAR算法检测到的一种美国学院足球网络的社团结构及其与真实社团结构的对应关系(由于该网络的节点数较多，我们通过列表方式给出它的社团结构)。可见，算法正确发现了6个真实社团(社团2、3、4、5、9及10)；发现的社团1是两个真实社团(Big West和Mountain West)的合并；3个社团(社团6、7及8)与8个只包含1个成员的社团混合在了一起。

综上所述，DECDILAR是一种有效的社团检测算法。

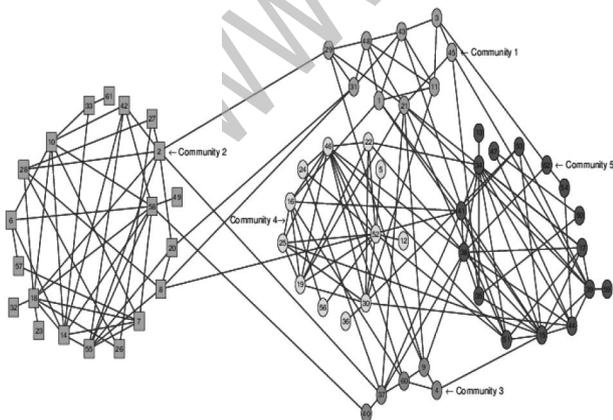


图2 一种检测到的宽吻海豚网络的社团结构

Fig.2 A detected community structure of the bottlenose dolphins network

表3 一种检测到的美国学院足球比赛网络的社团结构

Tab.3 A detected community structure of the american college football network

序号	成员数量	发现的社团	包含的真实社团	
		成员名称(成员ID)	社团名称	成员数量
1	14	BrighamYoung(1), NewMexico(5), UtahState(91), SanDiegoState(10), Wyoming(17), Idaho(51), BoiseState(29), ColoradoState(42), AirForce(94), NorthTexas(12), Utah(24), ArkansasState(25)	Big West Mountain West	6 8
3	11	Iowa(3), PennState(7), Wisconsin(16), Illinois(65), Michigan(33), Purdue(40), MichiganState(101), Northwestern(14), Minnesota(61), Indiana(107), OhioState(48)	Big Ten	11
5	10	ArizonaState(9), Washington(52), Stanford(78), UCLA(22), Oregon(69), WashingtonState(79), Arizona(23), OregonState(109), California(112), SouthernCalifornia(8)	Pacific Ten	10
7	16	Auburn(18), Vanderbilt(63), LouisianaState(97), Alabama(21), Kentucky(57), MississippiState(66), Florida(28), LouisianaTech(59), Arkansas(114), LouisianaMonroe(60), LouisianaLafayette(98), MiddleTennesseeState(64), Mississippi(88), SouthCarolina(71), Tennessee(77), Georgia(96)	Southeastern MiddleTennesseeState LouisianaTech LouisianaMonroe LouisianaLafayette	12 1 1 1 1
9	9	Louisville(58), Memphis(67), Cincinnati(93), EastCarolina(45), AlabamaBirmingham(113), Houston(49), SouthernMississippi(76), Army(92), Tulane(87)	Conference USA	9

## 5 结论(Conclusion)

本文在局部邻接表示法的基础上,提出了一种改进的索引局部邻接进化个体表示法,从而将复杂网络的社团结构分析问题转化为一个离散整数优化问题,并以差分进化算法为搜索引擎,设计了一种进化社团检测算法DECEILAR。在一系列合成网络和真实网络上验证了该算法的性能,并与以遗传算法为搜索引擎、采用线性组表示或局部邻接表示进化个体的进化社团检测算法进行了对比。实验结果表明,应用进化算法求解社团检测问题时,局部邻接表示法及其改进比线性组表示法更适用于进化个体的表示,差分进化比遗传算法的搜索性能更好(在网络社团结构较为清晰时)。

## 参考文献(References)

- [1] Han Liu,Fan Yang,Ding Liu.Genetic algorithm optimizing modularity for community detection in complex networks[C].Proceedings of the 35th Chinese Control Conference,2016:1252-1256.
- [2] Clara Pizzuti.A multiobjective genetic algorithm to find communities in complex networks[J].IEEE Transactions on Evolutionary Computation,2012,16(3):418-430.
- [3] Zhang Xiaohong,Zhang Bin,Zhang Changsheng,et al.A multi-objective hybrid genetic algorithm for detecting communities in complex networks[C].2016 12th International Conference on Natural Computation,Fuzzy Systems and Knowledge Discovery,2016:691-695.
- [4] Mursel Tasgin,Haluk Bingol.Community detection in complex networks using genetic algorithm[EB/OL].<https://arxiv.org/abs/1509.00556>,2017-11-02.
- [5] Clara Pizzuti.Boosting the detection of modular community structure with genetic algorithms and local search[C].The 27th Annual ACM Symposium on Applied Computing,2012:226-231.
- [6] G. Jia,Z.Cai,M.Musolesi,et al.Community detection in social and biological networks using differential evolution[C].The 6th International Conference on Learning and Intelligent Optimization,2012:71-85.
- [7] Wang Guoshun,Zhang Xuan,Jia Guanbo,et al.Application of Algorithm used in Community Detection of Complex Network[J].International Journal of Future Generation Communication and Networking,2013,6(4):219-229.
- [8] Leal Thiago P,Gonçaves Amanda C A,Vieira Vinícius Da F,et al.Decode—differential evolution algorithm for community detection[C].2013 IEEE International Conference on Systems,Man,and Cybernetics(SMC),2013,13-16:4635-4640.
- [9] Das Swagatam,Suganthan Ponnuthurai Nagaratnam.Differential evolution:A survey of the state-of-the-art[J].IEEE Transactions on Evolutionary Computation,2011,15(1):4-31.
- [10] Y J Park,M S Song.A genetic algorithm for clustering problems[C].Proceeding of the 3rd Annual Conference Genetic Algorithms,1989:2-9.
- [11] Newman M.E.J.,Girvan M.Finding and evaluating community structure in networks[J].Physical review E,2004,69(22):1-15.
- [12] Leon Danon,Jordi Duch,Albert Diaz-Guilera,et al.Comparing community structure identification[J].Journal of Statistical Mechanics:Theory and Experiment,2005(9):P09008.
- [13] Lancichinetti Andrea,Fortunato Santo,adicchi Filippo. Benchmark graphs for testing community detection algorithms[J].Physical Review E,2008,78(4):046110.
- [14] David Lusseau,Karsten Schneider,Oliver J.Boisseau,et al.The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations[J].Behavioral Ecology and Sociobiology,2003,54(4):396-405.
- [15] Tim S Evans.Clique graphs and overlapping communities[J].Journal of Statistical Mechanics Theory & Experiment,2010(12):257-265.
- [16] Duch Jordi,Arenas Alex.Community detection in complex networks using extremal optimization[J].Physical Review E,2005,72(2):027104.

## 作者简介:

孙韩林(1980-),男,博士,副教授.研究领域:复杂网络,大数据分析.

马素刚(1981-),男,硕士,高级工程师.研究领域:图像处理.

王忠民(1967-),男,博士,教授.研究领域:大数据分析与应用,嵌入式系统,机器人技术.