Journal of Computer Applications
计算机应用,2013,33(11):3129-3133
ISSN 1001-9081
CODEN JYIIDU
2013-11-01
http://www.joca.cn

文章编号:1001-9081(2013)11-3129-05
doi:10.11772/j.issn.1001-9081.2013.11.3129

基于免疫遗传算法的复杂网络社区发现

曹永春*,田双亮,邵亚斌,蔡正琦
(西北民族大学 数学与计算机科学学院,兰州 730030)
(*通信作者电子邮箱 cych33908@163.com)

摘　要:针对大部分基于智能优化算法的社区发现方法存在的种群退化、寻优能力不强、计算过程复杂、需要先验知识等问题,提出了一种基于免疫遗传算法(GA)的复杂网络社区发现方法。算法将改进的字符编码和相应的遗传算子相结合,在不需要先验知识的情况下可自动获得最优社区数和社区划分方案;将免疫原理引入遗传算法的选择操作中,保持了群体多样性,改善了遗传算法所固有的退化现象;在初始化种群及交叉和变异算子中利用网络拓扑结构的局部信息,有效缩小了搜索空间,增强了寻优能力。计算机生成网络和真实网络上的仿真实验结果表明算法可自动获取最优社区数和社区划分方案并具有较高的精度,说明算法具有可行性和有效性。

关键词:社区发现;复杂网络;免疫原理;遗传算法;单向交叉
中图分类号:TP301.6　　**文献标志码**:A

Community detection in complex networks based on immune genetic algorithm

CAO Yongchun*, TIAN Shuangliang, SHAO Yabin, CAI Zhengqi
(*School of Mathematics and Computer Science, Northwest University for Nationalities, Lanzhou Gansu 730030, China*)

Abstract: As many of the community detection methods based on intelligent optimization algorithms suffer from degeneracy, unsatisfactory optimization ability, complex computational process, requiring priori knowledge, etc., a community detection method in complex networks based on immune Genetic Algorithm (GA) was proposed. The algorithm combined the improved character encoding with the corresponding genetic operator, and automatically acquired the optimal community number and the community detection solution without the priori knowledge. Immune principle was introduced into selection operation of GA, which maintained the diversity of individuals, and therefore improved the intrinsic degeneracy of GA. By utilizing the local information of the network topology structure in initialization population, crossover operation and mutation operation, the search space was compressed and the optimization ability was improved. The simulation results on both computer-generated networks and real-world networks show that the algorithm acquires the optimal community number and the community detection solution, and has a higher accuracy. This indicates the algorithm is feasible and valid for community detection in complex networks.

Key words: community detection; complex network; immune principle; Genetic Algorithm (GA); one-way crossing over

0　引言

复杂网络的研究中,社区结构的发现目前已成为一个热点问题,近年来受到计算机、数学、生物和社会学等领域研究者的广泛关注。复杂网络中的社区是一组彼此相似并与网络中其他节点存在差异的节点构成的集合,同一社区内部节点相互连接密集,而社区间节点相互连接相对稀疏[1]。目前,在生物网、科技网和社会网等真实网络中均能发现社区结构的存在,复杂网络社区结构的发现对于复杂网络的拓扑结构分析、功能分析和行为预测具有重要的理论意义及实用价值。

由于社区结构的发现对于复杂网络的分析研究具有重要意义,许多研究者研究并提出了许多不同的社区结构发现算法。其中,Kernighan-Lin算法[2]和谱平分法[3]是较早提出的2种社区发现算法,谱平分法对社区数为2的网络社区发现效果明显,但对多社区网络的社区发现效果不是很明显;Kernighan-Lin算法则需以网络社区数作为先验知识,但该先验知识往往是很难获得的。在众多的社区划分算法中最著名的算法是Girvan和Newman提出的GN算法[4],该算法思想简单,精度较高,但是算法复杂度较高($O(n^3)$),适用于中等规模的网络。Newman对GN算法进行了改进又提出了一种Newman快速算法(Fast Newman, FN)[5],该算法与GN算法性能相当,但算法复杂度有了较明显的改进($O(n^2)$)。随着Girvan和Newman提出的定量描述社区结构的模块性Q函数被广泛使用,许多研究者提出了以Q函数作为目标函数的优化算法,由于遗传算法(Genetic Algorithm, GA)具有较强的全局寻优能力,一些基于遗传的复杂网络社区发现算法[6-10]获得了较好的划分效果,但仍存在一些问题。文献[8]采用基于聚类中心的编码方式,使算法解码困难,计算复杂,不适用于较大规模的网络,并且文献[8]需要先验知识如社区数,使算法的使用受到限制;另外一些算法[10]普遍采用随机生成初始种群、无针对地交叉变异,使算法寻优能力不强,收敛速度

收稿日期:2013-05-17;**修回日期**:2013-07-19。　**基金项目**:国家自然科学基金资助项目(11161041);2012年度国家民委科研项目基金资助项目;中央高校基本科研项目基金资助项目(31920130009,zyz2012081)。

作者简介:曹永春(1972-),男,甘肃天祝人,副教授,硕士,主要研究方向:智能算法、复杂网络;　田双亮(1965-),男,四川安岳人,教授,硕士,主要研究方向:图论及组合优化;　邵亚斌(1974-),男,甘肃天水人,副教授,博士,主要研究方向:不确定性处理的数学;　蔡正琦(1974-),男,甘肃天水人,副教授,硕士,主要研究方向:智能信息处理。

慢。

基于复杂网络社区发现算法中存在的问题，本文提出了一种基于免疫遗传算法的复杂网络社区发现算法(Community Detection in complex networks based on Immune Genetic Algorithm, CDIGA)，该算法利用改进的字符编码方式，不仅简化了算法的计算而且结合本文设计的遗传算子在不需要任何先验信息的情况下可自动确定网络社区数并获得最优社区划分方案；采用基于免疫原理的选择算子有效抑制了种群的退化现象，增强了算法的全局寻优能力；并且在种群初始化和交叉算子中采用启发式方法使算法搜索的解空间更靠近问题的最优解空间，提高了算法的收敛速度，增强了算法的性能。

1 社区结构评价标准

网络社区的发现问题就是要揭示不同类型复杂网络中存在的真实社区结构，但是如何来定量地刻画社区划分的优劣是首先要解决的问题，目前应用较多的是用 Girvan 和 Newman 提出的模块性函数来定量地描述网络中的社区，衡量社区结构的好坏。所谓模块性是指网络中连接社区结构内部节点的边所占的比例与另外一个随机网络中连接社区结构内部节点的边所占比例的期望值的差值[11]，其一种表达方式为：

$$Q = \sum_{i=1}^{k} (e_{ii} - a_i^2)$$

其中：e_{ii} 表示所连接的两个节点均在社区 i 内的边占网络中总边数的比例，a_i 表示至少有一个节点在社区 i 内的边占总边数的比例，k 为划分的社区数。对于一个特定的网络，一个社区划分方案的 Q 值越小则表示该划分的社区结构越弱，反之社区结构越强。Q 值的范围在 $-1 \sim 1$，如果 Q 值为负数则表示社区内部节点的边没有随机连接得到的边多，一个较明显的社区划分的 Q 函数值一般在 $0.3 \sim 0.7$。本文 CDIGA 采用 Q 函数作为目标函数。

2 CDIGA

2.1 编码方式

受分组遗传算法[12]启发，本文基于字符编码并加以改进作为算法编码方式。具体编码方式为，对于有 n 个节点的网络，每个个体的基因由表示社区划分信息的 n 位基因和表示社区个数的 1 位基因两部分组成，其形式化表示为：

$$r_m = [r_m^1 r_m^2 \cdots r_m^i \cdots r_m^n \mid k]$$

其中：r_m 表示种群中的第 m 个个体，r_m^i 表示网络中第 i 个节点被划分到哪个社区的信息，最后一位 k 表示该个体表示的划分方案将网络划分为 k 个社区。结合本文所设计的种群初始化方案和遗传算子的操作，通过该编码方式不仅有效解决了字符编码交叉困难的问题，而且可自动确定最优社区个数，同时获得准确率较高的社区划分方案。

2.2 种群初始化

由上述编码方案可见，对于一个给定的社区划分问题，遗传算法中不同个体所表示的社区划分方案的社区数 k 通常是未知的，复杂网络社区划分问题本质上是一个特殊的聚类问题，本文借鉴文献[13]初步确定 k 值，即如果社区数 k 未知，则 k 可为 $[k_{\min}, k_{\max}]$ 范围内的整数，其中：$k_{\min} = 2$，$k_{\max} = round(\sqrt{n})$（$n$ 为网络中节点数）。据此，种群初始化按照如下方法生成每个个体：首先随机生成一个 $[k_{\min}, k_{\max}]$ 范围内的整数 k 作为该个体所代表社区划分方案的社区个数，则每个社区可标记为 $[1,2,\cdots,k]$；然后将个体的每一位基因值随机指定为 $[1,k]$ 范围内的整数来表示该基因位所代表的节点被划分到哪个社区，这一处理过程重复 P 次，则可初步生成 P 个个体。

基于在网络中通常一个节点与其绝大多数邻居节点往往属于一个社区这一特点，为了提高初始种群的多样性，并使种群中初始个体具有一定的精度，本文利用网络的这一局部信息作为启发信息进一步优化初始种群。具体方法为：对个体 r_m，随机选择 1 个基因位(代表节点)，将其基因值(表示对应节点被划分到哪个社区)复制到该基因位所表示节点在网络中所有邻居节点对应的基因中。例如，对个体 $r_m = [33\underset{\uparrow}{2}1132\,1 \mid 3]$，若某次选择的基因位为 3(基因值是 2)，假设节点 3 在网络中的邻居节点是 2 和 5，则将个体 r_m 中第 2 和 5 两个基因位的值修改为第 3 个基因位的值(2)，即 $r_m = [3\underset{\uparrow}{2}21\,\underset{\uparrow}{2}321 \mid 3]$。对每个个体这一过程重复 αn 次(α 是模型参数，本文取经验值 0.4)则完成种群初始化。用这样的方法产生的初始种群在一定程度上使得初始解空间靠近了最优解空间，从而可提高遗传算法的收敛速度。

2.3 交叉算子

交叉算子通过互换两个个体的部分基因产生新个体的遗传操作，交叉操作产生的后代个体结合了父代个体的特征，能够产生多样化和有潜在希望成为最优解的新个体，在遗传进化进程中起着全局搜索的作用，决定遗传算法的全局搜索能力。

本文借鉴单向交叉(One-way crossing over)[14]方法并结合本文采用的编码方式加以改进作为 CDIGA 的交叉算子，具体操作过程描述如下：

第 1 步　从种群中随机选择两个个体，一个作为源个体 *src*，另一个作为目标个体 *dest*。

第 2 步　从源个体 *src* 中随机选择一个基因位，获得该基因的值，即对应网络节点被划分到哪个社区的信息 k_s，在源个体 *src* 中找出所有值为 k_s 的基因位。

第 3 步　若 $k_s \leqslant k_d$（k_d 表示目标个体 *dest* 表示的社区划分方案的社区个数)，则将与源个体 *src* 基因值为 k_s 的基因位对应的目标个体 *dest* 的基因位的值修改为 k_s；若 $k_s > k_d$，则将与源个体对应的目标个体 *dest* 基因位的值修改为目标个体基因值取值范围 $[1,k_d]$ 内的随机整数。

第 4 步　重新调整交叉后个体的各基因值，使该个体实际社区数与各基因值所表现的社区号一致。

在迭代过程中，对一个个体上述交叉操作的次数控制在 ηn 次，η 为模型参数，本文取经验值 $\eta = 0.2$，n 为网络节点数。上述交叉操作表示将在源个体中划分到同一社区中的节点的信息交叉到目标个体中，而不是像传统交叉那样盲目地进行信息交换，所以不仅能够产生多样化的新个体扩展搜索空间，而且有效促进了算法的收敛。一个具体的交叉操作实例如表 1 所示。

表 1 说明了对两个待交叉个体在 $k_s \leqslant k_d$(左)和 $k_s > k_d$(右)两种情况下单向交叉操作的过程。其中：v 表示待划分网络中的 8 个顶点，*src* 为源个体，*dest*(before) 为交叉前的目

标个体,*dest*(after) 为交叉后的目标个体。若源个体 *src* 中被选中交叉的基因位为4(表中左部),则源个体 *src* 中对应基因值③表示4号顶点被划分到第3个社区,此时 $k_s(=3) \leqslant k_d(=3)$,与该基因值相同的基因位(表示都被划分到第3个社区的顶点)还有2、8号位,则将目标个体 *dest*(before) 中对应3个基因值(4、2、8位)都修改为3,如表中 *dest*(after) 所示;若源个体 *src* 中被选中交叉的基因位为5(表中右部),此时 $k_s(=4) > k_d(=3)$,则将目标个体 *dest*(before) 中与该基因值相同的基因位(1、5、7位)都修改为 $[1,k_d]$ 范围内的随机整数(如2),如表中 *dest*(after) 所示。

表1　两种情况下交叉操作过程

$k_s \leqslant k_d$				$k_s > k_d$			
v	*src*	*dest*		*v*	*src*	*dest*	
		before	after			before	after
1	4	3	3	1	④→	3→	②
2	③→	2→	③	2	3	2	2
3	2	3	3	3	2	3	3
4→	③→	1→	③	4	3	1	1
5	4	1	1	5→	④→	1→	②
6	2	1	1	6	2	1	1
7	4	2	2	7	④→	2→	②
8	③→	2→	③	8	3	2	2

2.4　变异算子

交叉操作产生的子代个体除了继承父代个体的信息外,还会按较小概率变异,这体现了生物遗传的多样性,其作用是开辟新的搜索空间,避免算法陷入局部最优。变异算子的操作就是在给定变异率的前提下,对选定个体的每个基因值进行改变。本文利用网络的局部拓扑结构信息采用启发式的变异操作,将要变异基因的基因值变异为其大多数邻居节点所在的社区,若变异节点邻居数最多的社区不止一个,则从中随机选择一个作为变异节点所属新社区。这样有针对性的变异操作有效地缩小了问题搜索空间,减少了无意义的搜索,促进了遗传算法的收敛。

2.5　选择算子

在遗传进化进程中,选择操作起到启发进化方向的作用。本文基于免疫原理,在依据适应度选择的机制上,增加基于浓度的调节因子,从而保持群体多样性,改善一般遗传算法所固有的未成熟收敛现象。

对于规模为 m 的特定种群,定义个体 $i(i=1,2,\cdots,m)$ 的选择概率 p_i 为:

$$p_i = \frac{f_i}{\sum_{j=1}^{m} f_j} \cdot e^{-\beta \cdot c_i}$$

其中:f_i 为个体 i 的适应度,该式要求 f_i 非负,本文对目标函数 Q 函数变换获得,定义为 $f=(1+Q)/2$;β 为 $[0,1]$ 区间的反映个体适应度和浓度相对重要性的可调参数,为简单起见 β 可取1;c_i 为个体 i 的浓度。

个体浓度 c_i 表示群体中相同或相似个体在群体中所占的比例,其计算方式定义如下:

$$c_i = \frac{\sum_{j=1}^{m} S_{i,j}}{m}$$

其中:

$$S_{i,j} = \begin{cases} 1, & q_{i,j} \leqslant \varepsilon \\ 0, & \text{其他} \end{cases}$$

$q_{i,j}$ 是反映个体 i 和 j 相似性的指标;ε 是个体的相似度阈值,本文算法中取0.001。

相似性指标 $q_{i,j}$ 按式:$q_{i,j} = |f_i - f_j| / f_{\max}$ 计算,其中 f_i 和 f_j 分别为个体 i 和 j 的适应度函数值,$f_{\max}$ 为 f_i 和 f_j 中较大的值,则 $q_{i,j}$ 的值以不同个体适应度的接近程度反映了个体间的相似程度,以此计算个体的浓度。

上述个体选择概率 p_i 的计算公式表明个体适应度越大,则该个体被选择的概率越大,加速了算法的收敛;个体浓度越大,则该个体被选择的概率越小(抑制)。这样既保留了高适应度的个体,又减少了相似个体的比例,有效保持了种群中个体的多样性。

另外,对于进化过程中每一代中的最优个体采用精英保留策略直接传递到下一代继续参与进化,确保了算法在进化过程中不丢失可能最优解,本文算法精英个体保留数为3。

2.6　算法描述

综上所述,CDIGA 过程伪代码描述如下:

```
function [opt,qval] = CDIGA (graphadj,psize,gens,pc,pm)
  % 算法结果 opt 为最优个体,qval 为最优个体 opt 的 Q 函数值
  % 参数 graphadj 为网络的邻接矩阵,psize 为种群规模,gens 为
    进化代数,pc 为交叉概率,pm 为变异概率
  P = initpopulation(graphadj,psize)                    % P 表示种群
  ctimes = η * nodenum
  % nodenum 为节点数,η 为模型参数,ctimes 控制个体交叉次数
  for g = 1:gens
    for i = 1:psize/2                                   % 交叉操作
      pcr = rand()
      if pcr < pc
        random select two individuals as src and dest from P
        if the times of dest crossover < ctimes
          dest = crossover(src,dest)
        end if
      end if
    end for
    for i = 1:psize                                     % 变异操作
      P(i) = mutation(P(i),pm)        % P(i) 表示种群中第 i 个个体
    end for
    P = select(P)                                       % 选择操作
  end for
  opt,qval = findopt(P)                                 % 获得算法结果
end function
函数说明:
initpopulation(graphadj,psize)                          % 初始化种群
mutation(P(i),pm)                                       % 变异算子
crossover(src,dest)                                     % 单向交叉操作
select(P)        % 基于免疫原理的选择操作,精英保留个体数为3
findopt(P)                  % 找出种群 P 中最优个体及其 Q 函数值
evaluate(s)                                             % 计算个体 s 的 Q 函数值,
                                % 在 findopt(P) 和 select(P) 中被调用
```

CDIGA 中时间复杂度最高的操作为种群初始化和交叉操作,种群初始化中对初始种群进一步优化操作的时间复杂度为 $O(\alpha n^2)$,交叉操作时间复杂度为 $O(\eta n^2)$,由于通常认为遗传算法中最大迭代次数和种群规模都是常数,故 CDIGA 时间复杂度为 $O(n^2)$。

3　仿真实验及结果分析

为了验证本文提出的基于免疫遗传的社团划分算法的有

效性,下面以1组计算机生成的人工网络和2个真实的社会网络进行仿真实验,实验中CDIGA的各参数及取值如表2所示。

表2 CDIGA的各参数及取值设置

参数名	参数含义	参数值
psize	种群规模	200
gens	进化代数	200
pc	交叉概率	0.8
pm	变异概率	0.02
ε	计算个体选择概率时判断2个个体相似度的阈值	0.001
η	控制个体交叉次数的模型参数	0.2
α	控制初始化种群中个体优化次数的参数	0.4

3.1 计算机生成的构造网络

首先采用已成为测试社区发现算法性能基准的一组计算机生成网络[5]来测试CDIGA的性能。这组计算机生成网络的每个网络中包含128个节点,形成4个社区,每个社区包含32个节点;初始时每个节点都是独立节点,按照如下方法构造网络:对每一节点,以概率P_{in}随机地与同一社区内的其他节点连边,以概率P_{out}随机地与其他社区内的节点连边,且$P_{out} < P_{in}$。通过设置合适的P_{out}和P_{in}使网络中每个节点与其所在社区内部节点的连接数为Z_{in},与其他社区节点的连接数为Z_{out},且满足$Z_{in} + Z_{out} = 16$。随着Z_{out}的增加,该网络社区结构越来越模糊,当$Z_{out} > 8$时,通常认为该网络无明显的社区结构,

用计算机生成Z_{out}分别为0 ~ 8的9个网络测试CDIGA的性能,采用正确划分率[15]度量算法性能,这种方法通过比较算法划分得到的社团结构与网络实际社团结构之间的差距(正确划分节点数占总节点数的比例)来衡量算法的优劣。图1是CDIGA和两个非常经典且目前被广泛应用的算法GN[4]和FN[5]对计算机生成网络进行社区划分结果进行的比较,图中所用数据是对具有相同Z_{out}值的30个计算机生成网络,每个网络运行算法10次所获得的正确划分率的平均值。

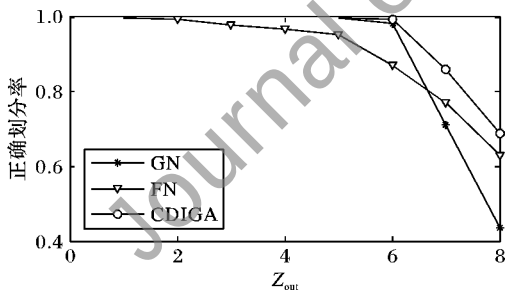

图1 CDIGA和GN及FN算法社区划分结果比较

图1中横轴为特定网络的Z_{out}值,纵轴为3种不同算法对特定Z_{out}值的网络进行社区划分的正确划分率。由图1可以看出CDIGA对Z_{out}值在0 ~ 5间的网络正确划分率为1,即均能完全正确地进行社区划分,与GN精度一样,但是对于Z_{out}值在6 ~ 8间的网络要优于GN算法,并且Z_{out}值越大效果越明显;FN算法在Z_{out}为3的网络中精度就开始有偏差,在Z_{out}值较低(3 ~ 6)的网络中精度均低于CDIGA和GN算法,在Z_{out}值较高(7、8)的网络中其正确划分率要优于GN算法,但还是明显低于CDIGA。结果表明CDIGA对计算机生成网络进行社区划分结果具有较高的精度。

3.2 Zachary's karate club网络

Zachary's karate club网络[4]是Zachary通过对一个美国大学空手道俱乐部的长期观测而构建的描述俱乐部会员之间社会关系的网络。该网络中的节点代表俱乐部成员,边代表成员之间的社会交往,它共包含34个节点和78条边。在观察过程中,该俱乐部由于意见分歧,最终分裂为以俱乐部管理者和俱乐部教练为中心的两个子集团。

CDIGA对Zachary karate club网络进行社区划分,结果如图2所示。图2显示CDIGA将网络中34个节点划分为4个社区(分别用4种不同形状的节点来表示),比真实的网络社区结构多两个社区,但通过观察分析发现图2中划分的两个社区(圆形节点和菱形节点表示)正好与一个真实社区对应(虚线左边),另外两个社区(星形节点与矩形节点表示)与另一个真实社区(虚线右边)对应。并且这样划分的Q函数值为0.4295,比真实社区结构的Q函数值0.3797还要大,说明CDIGA不仅正确发现了真实的网络社区结构,而且对该网络进行了更精细的划分。

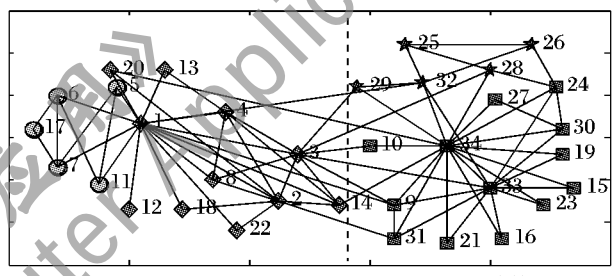

图2 CDIGA对Zachary's karate club网络的划分结果

3.3 American college football网络

American college football网络[4]是根据2000年秋季常规赛季的比赛计划构建的网络。网络中的节点代表球队,边代表两个球队之间常规赛季的比赛,该网络包含115个节点和616条边。对该网络本文除用正确划分率衡量算法精度,还用更具判别力的共同信息比较法——归一化互信息(Normalized Mutual Information, NMI)[15]来比较本文CDIGA与GN和FN算法的精度。NMI是通过比较算法得到的社区结构与真实社区结构之间的相似度来衡量算法精度,NMI的值在[0,1]区间,该值越大表示划分出的网络越接近真实网络结构。

CDIGA对American college football网络进行社区划分,运行20次的结果出现了三种划分方法,社区数分别为9,10和11,与网络的真实社区结构[8]均有一些误差。将CDIGA与GN和FN算法社区划分的结果进行比较,如表3所示。

表3 对American college football网络3种算法社区划分结果比较

算法	Q值	正确划分率	NMI值
GN算法	0.5985	0.8147	0.8682
FN算法	0.5532	0.6783	0.6912
CDIGA	0.6065	0.8783	0.8812

由表3看出,CDIGA得到的Q函数值、正确划分率值和NMI值均高于另外两种算法,优于同类社区划分算法文献[7](Q值为0.5618)的划分结果,与文献[5](Q值为0.6054,NMI值为0.8787)的划分结果非常接近,说明CDIGA对于较大规模网络社区的划分结果也有较高的精度。另外从多次运行CDIGA产生不同的社区划分结果这一点来看,说明

算法对于较大规模的网络进行社区划分时稳定性有待提高,这也是下一步要做的工作。

4　结语

针对目前大部分基于智能进化算法的社区发现方法存在的问题,本文提出了一种基于免疫遗传的复杂网络社区划分算法,利用改进的字符编码方式结合种群初始化方法和遗传算子在不需要任何先验信息的情况下自动确定网络社区数并获得最优社区划分方案;在种群初始化和交叉算子中采用启发式方法使算法搜索的解空间更靠近问题的最优解空间,提高了算法的收敛速度,增强了算法的性能;基于免疫原理,在依据适应度选择的机制上,增加基于浓度的调节因子,保持了群体多样性,有效抑制了种群的退化现象。以上 4 个方面的工作使 CDIGA 能够更有效地发现复杂网络的社区结构。通过对 CDIGA 仿真实验的结果分析,表明该算法可自动获得最优社区数和社区划分方案,并且有效提高了社区划分的质量。本文进一步要做的工作一是通过对更多真实网络上的运行结果进行参数分析,更加合理地设定算法中采用的各参数,从而提高算法的稳定性;二是增强算法局部搜索能力,进一步提高算法的精度。

参考文献:

[1] 罗锦坤,元昌安,杨文,等.基于基因表达式编程算法的复杂网络社区结构划分[J].计算机应用,2012,32(2):317 - 321.

[2] KERNIGHAM B W, LIN S. An efficient heuristic procedure for partitioning graphs [J]. Bell System Technical Journal, 1970, 49(2): 292 - 307.

[3] POTHEN A, SIMON H, LIU P, *et al.* Partitioning sparse matrices with eigenvectors of graphs [J]. SIAM Journal on Matrix Analysis and Applications, 1990, 11(3): 430 - 452.

[4] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821 - 7826.

[5] NEWMAN M E J. Fast algorithm for detecting community structure in networks [J]. Physical Review E, 2004, 69(6): 066133.

[6] 何东晓,周栩,王佐,等.复杂网络社区挖掘—基于聚类融合的遗传算法[J].自动化学报,2010,36(8):1160 - 1170.

[7] 朱大勇,侯晓荣,张新丽.遗传聚类的社团发现[J].智能系统学报,2009,4(1):81 - 84.

[8] SHANG R H, BAI J, JIAO L C, *et al.* Community detection based on modularity and an improved genetic algorithm [J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(5): 1215 - 1231.

[9] SHI C, YAN Z Y, WANG Y, *et al.* A genetic algorithm for detecting communities in large-scale complex networks [J]. Advances in Complex Systems, 2010, 13(1): 3 - 17.

[10] LIU X, LI D Y, WANG S L, *et al.* Effective algorithm for detecting community structure in complex networks based on GA and clustering [C]// Proceedings of the 7th International Conference on Computational Science. Berlin: Springer, 2007: 657 - 664.

[11] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(22): 1 - 15.

[12] AGUSTIN-BLAS L E, SALCEDO-SANZ S, JIMENEZ-FERNANDEZ S, *et al.* A new grouping genetic algorithm for clustering problems [J]. Expert Systems with Applications, 2012, 39(10): 9695 - 9703.

[13] 周世兵,徐振源,唐旭清.新的 K-均值算法最佳聚类数确定方法[J].计算机工程与应用,2010,46(16):27 - 31.

[14] TASGIN M, HERDAGDELEN A, BINGOL H. Community detection in complex networks using genetic algorithms [EB/OL]. [2013-04-16]. http://arxiv.org/abs/0711.0491.

[15] 郭世泽,陆哲明.复杂网络基础理论[M].北京:科学出版社,2012:270 - 271.

(上接第 3122 页)

[2] GIDÓFALVI G, BORGELT C, KAUL M, *et al.* Frequent route based continuous moving object location and density prediction on road networks [C]// Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2011: 381 - 384.

[3] YING J J C, LEE W C, WENG T C, *et al.* Semantic trajectory mining for location prediction [C]// Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2011: 34 - 43.

[4] ABRAHAM S, SOJAN LAL P. Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations [J]. Transportation Research Part C: Emerging Technologies, 2012, 23: 109 - 123.

[5] 张伟,柳先辉,丁毅,等.基于支持向量回归的多时间序列自回归方法[J].计算机应用,2012,32(9):2508 - 2511.

[6] 余雪岗,刘衍珩,魏达,等.用于移动路径预测的混合 Markov 模型[J].通信学报,2006,27(12):61 - 69.

[7] 彭曲,丁治明,郭黎敏.基于马尔可夫链的轨迹预测[J].计算机科学,2010,37(8):189 - 193

[8] 吕明琪,陈岭,陈根才.基于自适应多阶 Markov 模型的位置预测[J].计算机研究与发展,2010,47(10):1764 - 1770.

[9] KRUMM J. A Markov model for driver turn prediction [EB/OL]. [2013-04-22]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.153.2524&rep=rep1&type=pdf.

[10] GIDÓFALVI G, DONG F. When and where next: individual mobility prediction [C]// Proceedings of the First ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems. New York: ACM Press, 2012: 57 - 64.

[11] RON D, SINGER Y, TISHBY N. The power of amnesia: Learning probabilistic automata with variable memory length [J]. Machine Learning, 1996, 25(2/3): 117 - 149.

[12] BEJERANO G, YONA G. Variations on probabilistic suffix trees: statistical modeling and prediction of protein families [J]. Bioinformatics, 2001, 17(1): 23 - 43.

[13] LEONARDI F G. A generalization of the PST algorithm: modeling the sparse nature of protein sequences [J]. Bioinformatics, 2006, 22(11): 1302 - 1307.

[14] LIN J, JIANG Y, ADJEROH D. The virtual suffix tree [J]. International Journal of Foundations of Computer Science, 2009, 20(6): 1109 - 1133.

[15] LIN J, ADJEROH D, JIANG B H. Probabilistic suffix array: efficient modeling and prediction of protein families [J]. Bioinformatics, 2012, 28(10): 1314 - 1323.