

DOI: 10.11992/tis.201706054

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.tp.20180423.1019.006.html>

WSB-EA 进化算法的符号网络弱结构平衡分析

常新功, 赵雅娟

(山西财经大学 信息管理学院, 山西 太原 030006)

摘 要: 由于大多数真实符号网络更满足弱结构平衡理论, 并且求解符号网络的弱结构平衡问题是 NP 难问题, 因此提出了基于进化算法的符号网络弱结构平衡计算方法——WSB-EA 算法。该方法将弱结构平衡定理的能量函数作为适应值函数, 首先利用启发式的方法初始化种群, 经过锦标赛选择、单路交叉、单点变异、局部搜索 4 个阶段, 迭代有限次之后得到最优解。在此算法中, 提出了大型符号网络的存储方法和增量计算方式。通过大量实验, WSB-EA 算法得出了 4 个小型符号网络和 2 个大型符号网络的弱不平衡度。并且与其他算法相比, WSB-EA 算法能更快收敛得到最优解, 具有较高鲁棒性。

关键词: 符号网络; 进化算法; NP 难问题; 结构平衡理论; 弱结构平衡理论; 单路交叉; 局部搜索; 弱不平衡度
中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1673-4785(2018)05-0783-08

中文引用格式: 常新功, 赵雅娟. WSB-EA 进化算法的符号网络弱结构平衡分析[J]. 智能系统学报, 2018, 13(5): 783-790.

英文引用格式: CHANG Xingong, ZHAO Yajuan. Weak structure balance analysis of signed network based on WSB-EA evolutionary algorithm[J]. CAAI transactions on intelligent systems, 2018, 13(5): 783-790.

Weak structure balance analysis of signed network based on WSB-EA evolutionary algorithm

CHANG Xingong, ZHAO Yajuan

(Faculty of Information Management, Shanxi University of Finance and Economics, Taiyuan 030006, China)

Abstract: The weak structural balance (WSB) theory is suitable to solve the weak structure balance problem of most signed networks, and it is an NP-hard problem. Here a WSB evolutionary algorithm (WSB-EA) is proposed, which computes the global unbalanced degree of signed network based on evolutionary algorithm. In this method, the energy function of WSB theory is used as the fitness function. First, a heuristic method is used to initialize the population. After the tournament selection, single crossing, single point variation, and local search, the optimal solution is obtained after a finite number of iterations. The algorithm involves a storage of large signed network and incremental calculation. Through several experiments, the weak unbalanced degree of four small signed networks and two large signed networks are derived from WSB-EA algorithm. Compared with other algorithms, the WSB-EA algorithm can converge to the optimal solution faster and has a higher robustness.

Keywords: signed network; evolutionary algorithm; NP-hard problem; structural balance theory; weak structural balance theory; single cross; local search; weak unbalanced degree

符号网络^[1-2]是指具有正边或负边的网络, 其中正边表示积极的、正面的意义, 负边表示消极的、负面的意义。在现实生活中, 可以将很多网

络抽象为符号网络。例如, 国际关系网^[3]中合作和敌对的关系、社交网络^[4]中朋友和敌人关系、生物网络^[5]中促进和抑制作用等。1946 年 Heider^[6]提出三角形关系中正关系与负关系的相互作用模式, 将符号网络带入大家的视线。随着符号网络的兴起, 网络的全局平衡性引起众多学者的关

收稿日期: 2017-06-14. 网络出版日期: 2018-04-24.

基金项目: 山西省哲学社会科学“十二五”规划 2015 年度课题项目; 山西省自然科学基金项目 (2013011016-4).

通信作者: 赵雅娟. E-mail: 707238065@qq.com.

注。根据计算得到的全局平衡性可以有效地进行个性化推荐、态度预测等。一般全局平衡性用不平衡度来衡量,不平衡度是指一个网络从不平衡到平衡的距离,若将这些引起不平衡的边的符号取反,网络就变为平衡网络。然而对于大多数符号网络而言,结构平衡定理太过严苛。Leskovec等^[7]也证实了弱结构平衡定理更适合真实网络。因此本文从弱结构平衡定理出发,来研究网络的弱不平衡性。孙一翔^[8]提出 Meme-SB 算法,将结构平衡定理的能量函数作为适应值函数,利用混合遗传算法求得真实符号网络的不平衡度。在 Meme-SB 算法基础上,本文提出了 WSB-EA 算法,将研究范围扩展到符号网络的弱不平衡性。即将弱结构平衡定理的能量函数当作适应值函数,利用进化算法原理,初始化种群,选择、交叉、变异,最终得到网络的弱不平衡度。4 个小型数据集的实验表明,WSB-EA 算法比其他算法能够更快收敛得到最优解。在实验最后部分计算得到两大符号网络 Epinions 和 Slashdot 的弱不平衡度。

1 背景知识

从 20 世纪 40 年代,Heider 提出三角形关系中正关系与负关系的相互作用模式,到 Cartwright 等^[9]用图论语言描述这一理论将其推广到整个网络。越来越多的研究者对符号网络的结构和演化问题感兴趣,致力于研究社会群体的派系结构和发展过程。然而,有很大一部分真实网络并不符合 Heider 提出的结构平衡理论,由此 Davis^[10]放宽结构平衡理论的约束条件,提出弱结构平衡理论。

1.1 弱结构平衡定理

符号网络用数学符号 $G(V, E)$ 表示, V 表示网络节点集合, E 表示网络边的集合。其中任意一条边 $J_{ij} \in E$, 其符号为“+”或“-”。“+”代表节点 i 与节点 j 之间的边是积极关系,“-”则代表消极关系。Heider 指出如果网络中任意一个三角形的符号乘积为正,则网络是平衡的。如图 1 所示, (a) 和 (b) 是平衡三角, (c) 和 (d) 是不平衡三角。在此基础上 Cartwright 和 Harary 等提出判断符号网络是否平衡的一个充要条件——“如果一个网络中的节点能够被分割为两个子集,每个子集内的所有边均是正边,子集间的边均为负边,这样的网络就是平衡的。”

然而对于大多数符号网络而言,结构平衡定理的要求太过严苛。之后 Davis 放宽结构平衡理论的约束条件,提出了弱结构平衡理论。这一理论将图 1(d) 三条边都是负号这样的三角结构也认

为是平衡结构。因此产生了一个新的概念, K-平衡网络。这一理论认为,一个符号网络是弱平衡的,当且仅当可以将这一网络中的节点分为 k 个子集,子集内部都是正边,子集之间都是负边。结构平衡定理是弱结构平衡定理中 k 为 2 的特殊情况。K-平衡网络的结构如图 2 所示,图中集合内部连边都是正边,集合之间连边都是负边。

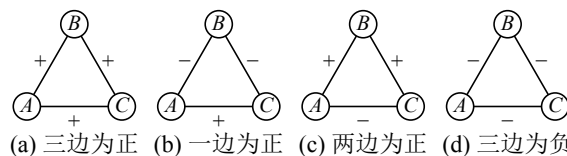


图 1 符号网络中 4 种三角形结构^[1]

Fig. 1 Four triangle signed network structures^[1]

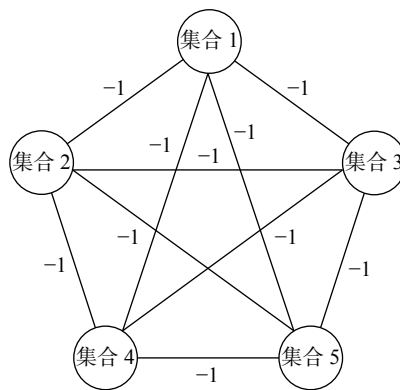


图 2 弱结构平衡网络

Fig. 2 Weak balance network structure

1.2 弱结构平衡定理能量函数

根据弱结构平衡定理,计算符号网络的弱不平衡度就是寻找一种集合划分,使得集合内部之间的负边数和集合外部之间的正边数最少。例如,节点 1 和节点 2 之间的连边为正,如果将这两个节点分到同一个集合,这两个节点就没有对弱不平衡度做出“贡献”,如果把它们分到不同的集合,就需要计算这两个节点对弱不平衡度做的“贡献”。同理,两个节点之间的边是负边时,将其分到同一集合,也要计算这一部分的“贡献”。因此计算弱不平衡度就是遍历整个符号网络,找到所有对弱不平衡度做出“贡献”的节点对。由此提出弱结构平衡定理的能量函数:

$$h(s) = \sum_{ij} (1 - J_{ij} \delta(s_i, s_j)) / 2 \quad (1)$$

在弱结构平衡定理中,一个无向符号网络被分为 k 个子集。因此,网络中每个节点都有一个所属的子集编号,在式中用 s_i 来表示这一编号。 $\delta(s_i, s_j)$ 的计算结果由 s_i 和 s_j 决定,如果 s_i 和 s_j 的值相同则计算结果为+1,否则为-1。

由此,求解符号网络的弱结构平衡问题就转变为最小化能量函数的优化问题。能量函数的最小值代表了导致符号网络弱不平衡的最少边的数目。若将这些边的符号取反,可以将此网络转变为弱结构平衡网络。如果能量函数值为0,说明此网络已是弱结构平衡网络。

1.3 进化算法

进化算法^[11-12]或称演化算法(evolutionary algorithms, EAS),是一个算法簇,它们产生的灵感都来自于大自然的生物进化,但它有很多的变化,有不同的遗传基因表达方式,不同的交叉和变异算子,特殊算子的引用,以及不同的再生和选择方法。与传统的基于微积分的方法和穷举法等优化算法相比,进化计算是一种成熟的具有高鲁棒性和广泛适用性的全局优化方法,具有自组织、自适应、自学习的特性,能够不受问题性质的限制,有效地处理传统优化算法难以解决的复杂问题。

2 WSB-EA 算法描述

2.1 WSB-EA 算法主要流程

根据问题定义以及进化算法原理,本文将弱结构平衡定理的能量函数作为目标函数,在Meme-SB算法的基础上,初始化种群,经过选择、交叉、变异得到最优解。具体框架如算法1所示。

算法1 WSB-EA 算法

输入 邻接矩阵 J , 种群规模 P , 迭代次数 M ;
输出 最优解的适应值, 即网络弱不平衡度。

```
1) put  $J$ (); //存储邻接矩阵
2) Pop ← Population( $P$ , true); //初始化种群
3)  $a=0$ ;
4) repeat
5) individual = Pop.getFittest(); //计算得到
```

种群最优个体

```
6) Pop ← evolvePopulation(Pop, individual); //种群进化, 选择、交叉、变异操作
```

```
7)  $a=a+1$ ;
```

```
8) until  $a=M$  //达到最大进化次数时停止
```

2.2 编码

本文使用的编码方式是整数编码^[13]。种群中的染色体由 $X_j = \{x_1, x_2, x_3, \dots, x_i, \dots, x_n\}$ 表示, 其中 n 为符号网络的节点个数, x_i 表示第 i 个节点所属的子集编号。假设符号网络中有 k 个子集, 则 x_i 的取值范围是 $0 \sim k-1$ 。例如第 j 个染色体可以表示成 $X_j = \{1, 2, 0, \dots, k-1, \dots, 1, 0\}$ 。

2.3 初始化

一般进化算法中的初始化方法都是随机产生

每条染色体, 但这样的初始化方式效率不高。如果初始化的染色体对应的适应值很低, 要经过很多多次迭代才能得到最优解, 收敛速度太慢。因此本文使用一种简单的启发式方法, 保证在初始化时每一个节点与它的邻居节点对能量函数没有“贡献”。即在初始化之后, 随机选择一个节点, 如果此节点与它的邻居节点对能量函数没有“贡献”, 则不改变这一节点的取值, 否则将其改为它的正邻居节点所属的子集编号。这样的操作重复 n 次, 并且保证每次选择的节点是之前没有选择过的。

2.4 遗传操作

本文的遗传操作过程如算法2所示。遗传过程分为3步: 首先将上一代种群中的优秀个体保存到新一代种群中, 这个过程称为精英保留; 接下来利用锦标赛方式选取父代染色体, 经过单路交叉; 单点变异得到新一代种群。

算法2 evolvePopulation(individual) 算法

输入 原始种群 Pop, 最优个体 individual, 锦标赛规模 tournamentSize, 交叉概率 uniformRate, 变异概率 mutationRate;

输出 新一代种群 newPop。

```
1) newPop.saveIndividual(0, individual); //精英保留
```

```
2) for( $j=0; j < \text{Pop.size}(); j++$ )
```

```
3) indiv1 = tournamentSelection(Pop); //选择父本1
```

```
4) indiv2 = tournamentSelection(Pop);
```

```
5) newIndiv = crossover(indiv1, indiv2); //交叉算法
```

```
6) newPop.saveIndividual( $j$ , newIndiv);
```

```
7) for( $j=0; j < \text{Pop.size}(); j++$ )
```

```
8) mutate(newPop.getIndividual( $j$ )); //变异算法
```

2.4.1 选择操作

在进化算法中轮盘赌是最常用也是最简单的选择算子。但对于本文而言, 轮盘赌不利于保持种群多样性。为了解决这一问题, 本文选用锦标赛方法选取优秀个体进行繁衍。首先确定锦标赛规模 t , 随机从种群中选择出 t 个个体, 计算这些个体的适应值, 将适应值最小的个体作为父代进行繁衍。

2.4.2 交叉操作

一般在进化算法中经常使用的两点交叉算子很有可能会破坏父代染色体中优秀的基因结构, 因此为了保存父代染色体中的优秀基因结构, 本文选择单路交叉方式^[14]作为交叉算子。首先通过锦标赛选择算法选择两个优秀父代 A 、 B 。随机产

生一个 $0 \sim k-1$ 的整数 m , 记录 A 中所有 m 所在的位置, 然后将 B 对应位置的基因值改为 m 。例如, 图 3 中, 每一条染色体有 15 个基因, 选择两条父代染色体, 随机产生一个数 2, 将 B 中对应位置的取值都改为 2。

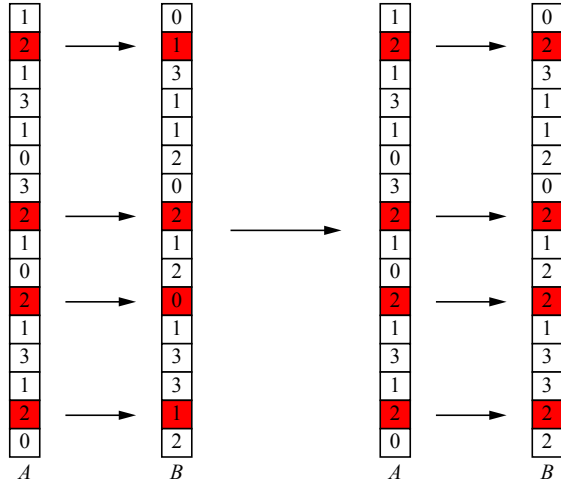


图3 单路交叉

Fig. 3 Single cross

2.4.3 变异操作

本文的变异算子是单点变异。从父代种群中随机选择一条染色体, 然后从这一染色体中随机选取一个基因, 对其重新赋值, 但要保证变异之后染色体的能量函数值没有增加。这样的操作要循环 n 次, 因此变异也可以看作一种局部搜索, 将染色体变异为它的邻居染色体, 试图找到局部最优的染色体。

2.5 符号网络存储

符号网络在计算机中一般有 3 种存储方式: 邻接矩阵、三元组和链表方式。邻接矩阵是普遍使用的一种方式。然而对于大型符号网络来说, 邻接矩阵达到 $10^5 \times 10^5$ 的数量级, 存储这样的邻接矩阵对内存要求太高。因此本文选择邻接链表^[15]来存储网络信息。首先为每一个节点都创建一条链表, 头结点是节点本身, 后面链接与这一节点有联系的节点信息, 包括节点编号以及边的符号。例如, 图 4 中, 这一链表存储节点 1 的连边信息。节点 1 与节点 2 的连边为 1, 节点 1 与节点 6 的连边为 -1。虽然每条边都存储了两次, 但相对于邻接矩阵的存储方式, 大大降低了内存占用。

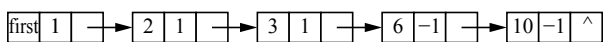


图4 二维链表示例

Fig. 4 Two-dimension link structure

2.6 增量计算

在一般的遗传算法中, 每一次的交叉和变异

之后都需重新计算一次适应值, 这样重复操作使得算法的时间复杂度太高。基于此, 本文在定理 1 的基础上提出一种增量计算的方式, 在重新计算适应值时, 只计算个体中因基因改变使得适应值增减的部分。

定理 1 假设个体 ind 在位置 h 处变异, s_h 由 old 变为 new , 则个体 ind 的新适应值 $h(\text{ind})_{\text{new}}$ 为

$$h_{\text{new}} = h_{\text{old}} + \sum_{v_j \in N(v_h)} J_{hj} (\delta(\text{new}, s_j) - \delta(\text{old}, s_j))$$

其中 $N(v_h) = \{v_k | (v_k, v_h) \in E\}$ 是节点 v_h 的邻居节点集合。

$$\text{证明} \quad h = \sum_{v_i, v_j \in E} J_{ij} \delta(s_i, s_j) =$$

$$\sum_{v_h, v_j \in E} J_{hj} \delta(s_h, s_j) + \sum_{v_i, v_j \in E \wedge i \neq h} J_{ij} \delta(s_i, s_j) = \quad (2)$$

$$\sum_{v_j \in N(v_h)} J_{hj} \delta(s_h, s_j) + \sum_{v_i, v_j \in E \wedge i \neq h} J_{ij} \delta(s_i, s_j)$$

由于式 (2) 等号右侧第 2 部分在变异前后不变, 所以

$$h_{\text{new}} - h_{\text{old}} = \sum_{v_j \in N(v_h)} J_{hj} \delta(\text{new}, s_j) - \sum_{v_j \in N(v_h)} J_{hj} \delta(\text{old}, s_j) = \sum_{v_j \in N(v_h)} J_{hj} (\delta(\text{new}, s_j) - \delta(\text{old}, s_j)) \quad (3)$$

上述推导表明, 当个体 ind 在 h 位置变异后, 新的适应值通过旧适应值加一个增量就可以得到, 而这个增量只需要遍历 v_h 的邻居节点就可以算出, 这样计算一个个体的适应值的时间复杂度从 $O(m)$ 降为 $O(d_{\text{avg}})$, 其中 d_{avg} 为平均度数。

基于定理 1, 计算新的适应值就可以简便为计算改变子集编号的节点与其邻居节点对适应值的“贡献”。具体分为两种情况: 1) 原本对适应值没有“贡献”的边, 在改变基因之后对适应值有“贡献”; 2) 原本对适应值有“贡献”的边, 在改变之后对适应值没有“贡献”。增量计算大大地降低了时间复杂度, 在交叉算法中的增量计算只需循环若干次, 而在变异算法中只需运行一次。

算法具体操作如算法 3 所示。算法第 1 行将 increaseFitness 设置为 0, 第 2 行是将 c 赋值为被改变基因所对应的节点的链表头结点, 为了找到这个节点的所有邻居, 之后遍历整个链表。算法第 4~8 行是在判断邻居为正邻居之后, 再次判断如果邻居的基因与旧基因相同, 但与新基因不同时, increaseFitness 加 2; 如果邻居的基因与旧基因不同, 但与新基因相同时, increaseFitness 减 2。算法第 9~13 行是在判断邻居为负邻居之后, 再次判断如果邻居的基因与旧基因相同, 但与新基因

不同时, increaseFitness 减 2; 如果邻居的基因与旧基因不同, 但与新基因相同时, increaseFitness 加 2。

算法 3 $\text{ncreaseFitness}()$ 算法

输入 需要计算适应值的个体 individual , 基因改变位置 id , 旧基因 old , 新基因 new ;

输出 增加的适应值 increaseFitness 。

```

1)  $\text{increaseFitness}=0$ ;
2)  $c=\text{Data.node}[\text{id}].\text{first}$ ;
3) 遍历节点  $\text{id}$  的所有邻居节点;
4) if (正邻居)
5) if( $\text{getGene}(c.\text{data})==\text{old}\&\&\text{getGene}(c.\text{data})!=\text{new}$ )
6)  $\text{increaseFitness}=\text{increaseFitness}+2$ ;
7) else if( $\text{getGene}(c.\text{data})!=\text{old}\&\&\text{getGene}(c.\text{data})==\text{new}$ )
8)  $\text{increaseFitness}=\text{increaseFitness}-2$ ;
9) else if(负邻居)
10) if( $\text{getGene}(c.\text{data})==\text{old}\&\&\text{getGene}(c.\text{data})!=\text{new}$ )
11)  $\text{increaseFitness}=\text{increaseFitness}-2$ ;
12) else if( $\text{getGene}(c.\text{data})!=\text{old}\&\&\text{getGene}(c.\text{data})==\text{new}$ )
13)  $\text{increaseFitness}=\text{increaseFitness}+2$ ;
```

2.7 复杂性分析

在整个算法中, 时间复杂度最高的是遗传操作这一部分, 因此只分析这一部分的时间复杂度。首先定义几个基础概念, 节点个数 n 、边个数 m 、种群规模 M 、锦标赛规模 t 以及迭代次数 G 。选择父本的时间复杂度是 $O(t)$ 。交叉算法所花费的时间复杂度是 $O(n)$ 。变异算子的时间复杂度是 $O(1)$ 。计算适应值的时间复杂度分两种情况, 若不用增量计算, 计算适应值时要遍历整个网络的所有边, 此时的时间复杂度是 $O(m)$ 。如果使用增量计算的方式, 在计算适应值时只需遍历网络中的若干条边。假设改变的是节点 i 的子集编号, 而节点 i 的度数为 d_i , 则需遍历 d_i 条边。在网络中每个节点的度数是不相等的, 因此用平均度 d_{avg} 来代表某一个节点的度最合理。所以此时的时间复杂度是 $O(d_{\text{avg}})$ 。总体算法的时间复杂度是 $O(MG(n+m))$ 。

3 实验结果与分析

3.1 参数设置

表 1 列出了 WSB-EA 算法中使用到的所有参

数及其取值。算法的运行环境是 Intel(R)Core(TM) i3-2330m, 运行内存 4 GB, 操作系统 Windows7 旗舰版, 使用的软件是 eclipse4.5。

表 1 参数设置
Table 1 Parameter setting

参数	含义	取值
G	迭代次数	50
M	种群规模	100
tournamentSize	锦标赛规模	5
uniformRate	交叉概率	0.9
mutationRate	变异概率	0.1

3.2 小型符号网络实验结果

本文使用的小型数据集有 4 个: 斯洛文尼亚政党网络 (SPP)、Gahuku-Gama 部落网络 (GGS)、社交圈网络 (SC) 和 yeast 网络 (yeast)。

斯洛文尼亚政党网络: 这是一个由斯洛文尼亚 10 个议会政党之间的关系组成的网络, 1994 年由一些研究政治的学者提出^[16]。10 个议会政党的英文名字缩写分别是 SKD、ZLSD、SDSS、LDS、ZS-ESS、ZS、DS、SLS、SPS-SNS 和 SNS。这一网络中有 10 个节点, 45 条连边。

Gahuku-Gama 部落网络^[17]: 这一网络中有 16 个节点, 代表 16 个部落。边数为 59, 代表部落之间的联盟和对抗。

社交圈网络: 这一网络是根据现实生活中人与人之间的关系得到的实际网络, 节点有 28 个, 边数为 42, 代表节点之间的朋友关系或敌人关系。

Yeast 网络: 这是一个酵母菌的基因调控网络^[18], 该网络包含 690 个节点和 1 080 条边。由于这一网络的节点个数较多, 在实验中将种群规模增加到 500。

为了验证 WSB-EA 算法的准确性以及健壮性, 在这 4 个数据集上与孙一翔提出的 Meme-SB 算法和没有使用增量计算的 WSB-EA 算法的实验结果作对比分析。每一个数据集都做 30 次实验, 记录求得最小适应值时的迭代次数和时间。实验结果如表 2 所示。同时, 为了验证增量计算对算法的贡献, 将 WSB-EA 算法与没有使用增量计算的 WSB-EA 算法进行时间比较, 即得到迭代相同的次数所用的时间。结果如表 3 所示。

从表 2 可知, 无论是只有 10 个节点的 SPP 网络还是有 690 个节点的 yeast 网络, 在求得相同适应值的情况下, WSB-EA 算法的迭代次数较少, 也就说明 WSB-EA 算法能更快收敛到最优解。这

也证明了,本文使用的锦标赛选择、单路交叉、单点变异和局部搜索这些遗传操作,不仅保证了算法的正确性,还加快了寻找最优解的速度。

表 2 对比实验结果 1
Table 2 Comparison results 1

网络	迭代次数		适应值	
	WSB-EA	Meme-SB	WSB-EA	Meme-SB
SPP	1	1	2	2
GGs	1	5	7	7
SC	1	4	0	0
Yeast	200	500	45	48

表 3 对比实验结果 2
Table 3 Comparison results 2

网络	WSB-EA	无增量计算的算法
SPP	0.070	0.060
GGs	0.106	0.060
SC	0.082	0.056
Yeast	1.194	0.048

从表 3 的实验结果可以看出,没有增量计算的 WSB-EA 算法的运行时间大大超过了有增量计算的 WSB-EA 算法。在规模比较小的 SPP、GGs、SC 网络上时间差距没有太大,因为这 3 个网络节点少,有无增量计算没有太大差别。但是当节点个数增加,在 yeast 网络上增量计算就体现出了优势。这也验证了本文提出的增量计算对与求解大型符号网络的弱不平衡度有很大的帮助。

接下来使用 WSB-EA 算法求得 4 个网络的弱不平衡度。在实验开始前无法获得 k 的取值,因此只能在实验之初先测试 k 的取值,即先取不同的 k 值,看哪一个 k 值在算法运行时的收敛效果好。做法是选取 k 值从 3 ~ 20,每一个 k 值都做 10 次实验,迭代次数设置为 50。比较最后适应值,适应值最小的 k 值就是最适合的 k 的取值。确定每一个网络的 k 的取值之后,对每个网络作 30 次实验,求解这个网络的弱不平衡性,实验结果如表 4 所示。由表 2 和表 4 中的适应值这一列可以看出,弱不平衡度值更小,也就说明,弱不平衡定理更适合真实符号网络,更具有现实意义。由表 4 中 k 的取值这一列可知, k 的取值都不大,所以对于一个网络而言,在处于弱结构平衡时不会被划分成太多的阵营。

表 4 4 种小型符号网络的弱不平衡度实验结果
Table 4 The experimental results of unbalanced degree of four small signed networks

网络	k 的取值	弱不平衡度
SPP	2	2
GGs	3	3
SC	2	0
Yeast	3	37

3.3 大型符号网络实验结果

本文使用的大型数据集有两个: Epinions 网络^[7]和 Slashdot 网络^[7]。根据数据集的规模来修改迭代次数和种群规模这两个参数。由于数据集中节点个数太多,以及内存的限制,种群规模只能设置为 200,所以很大程度地增大迭代次数,设置为 10 万次。

Epinions 网络: Epinions 网络是从 Epinions 获取的, Epinions 网站是一般消费者评论网站,成立于 1999 年。游客可以对各种物品进行评价,也可以通过阅读各种物品的评价来帮助他们作出购买决策。该网络有 131 828 个节点, 841 372 条边。每条连边表示节点之间的信任关系。此数据集是有向网络,因此预处理数据时需将其改为无向网络,即将节点之间关系冲突的连边删除,例如节点 1 认可节点 2 评论,但节点 2 不认可节点 1 的评论,在预处理数据时要删除这样的连边。

Slashdot 网络: Slashdot 网络是从 Slashdot 网站获取的, Slashdot 网站是一个资讯科技网站。它每天都会更新主页的新闻数次,网站使用者可以对公布在该站的新闻发表意见。该网络有 81871 个节点, 545 671 条边。预处理的方法与处理 Epinions 数据集方法一样。

在大型符号网络上,将 WSB-EA 算法与 HRT-SB 算法的实验结果进行比较。对于每一个数据集做 30 次实验,求解网络的弱不平衡度,实验结果如表 5 所示。表 5 中前 3 列由 WSB-EA 算法计算得出,最后一列数据由 HRT-SB 算法^[19]得出。由表 5 可知,对于大型符号网络而言,即使节点个数有 10 万这么多, k 的取值仍然不大。也就是说,无论这个网络有多大规模,划分的阵营个数都不会太多。由弱不平衡比可知,每一个网络中不平衡的边占很少一部分,整个网络是处于弱结构平衡的。由弱不平衡度和不平衡度这两列数据可以看出,弱不平衡度的值是远远小于不平衡度的值的,也说明了弱结构平衡定理相对于

结构平衡定理来说,更适合研究符号网络的结构性质。

表5 两种大型符号网络的弱不平衡度实验结果

Table 5 The experimental results of unbalanced degree of two large signed network

网络	k 的取值	弱不平衡度	弱不平衡比/%	不平衡度
Epinions	5	48 885	6.90	50 806
Slashdot	6	70 119	13.04	73 604

4 结束语

在符号网络中,结构平衡是一个很重要的研究领域,然而现实世界中存在的大多数网络都不是严格的结构平衡网络。因此研究网络的弱结构平衡性就显得尤为重要。在前人的基础上,本文提出了弱结构平衡定理的能量函数,利用进化算法的繁衍得到网络的弱不平衡度。通过大量的对比实验证明,本文提出的WSB-EA算法实验效果不错,且能更快得到最优解。然而,由于本文提出的算法是用来解决大型网络的,当网络节点太多的时候,想要找到最优解需要的时间太长,因此在未来的工作中可以继续改进算法的进化能力,让算法可以更快地找到最优解,及在后续工作中可以寻找WSB-EA算法的应用领域。例如,将晋商文化中的商帮当作节点,商帮之间的经济往来当作联系,将此组成晋商网络,分析网络中商帮之间的相互作用关系,以及如何达到一种稳定的发展态势,将这些信息整合,为现代企业发展提供参考。

参考文献:

- [1] 程苏琦,沈华伟,张国清,等.符号网络研究综述[J].软件学报,2014,25(1):1-15.
CHENG Suqi, SHEN Huawei, ZHANG Guoqing, et al. Survey of signed network research[J]. Journal of software, 2014, 25(1): 1-15.
- [2] 蓝梦微,李翠平,王绍卿,等.符号社会网络中正负关系预测算法研究综述[J].计算机研究与发展,2015,52(2):410-422.
LAN Mengwei, LI Cuiping, WANG Shaoqing, et al. Survey of sign prediction algorithms in signed social networks[J]. Journal of computer research and development, 2015, 52(2): 410-422.
- [3] AMARAL L A N, SCALA A, BARTHELEMY M, et al. Classes of behavior of small-world networks[J]. arXiv: cond-mat/0001458, 2000.
- [4] MOORE M. An international application of Heider's balance theory[J]. European journal of social psychology, 1978, 8(3): 401-405.
- [5] PARISIEN C, ANDERSON C H, ELIASMITH C. Solving the problem of negative synaptic weights in cortical models[J]. Neural computation, 2008, 20(6): 1473-1494.
- [6] HEIDER F. Attitudes and cognitive organization[J]. The journal of psychology, 1946, 21(1): 107-112.
- [7] LESKOVEC J, HUTTENLOCHER D, KLEINBERG J. Signed networks in social media[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Atlanta, USA, 2010: 1361-1370.
- [8] 孙一翔.基于进化算法的符号网络结构平衡分析[D].西安:西安电子科技大学,2014:25-36.
SUN Yixiang. Structural balance analysis in signed networks based on evolutionary algorithm[D]. Xi'an: Xidian University, 2014: 25-36.
- [9] CARTWRIGHT D, HARARY F. Structural balance: a generalization of Heider's theory[J]. Psychological review, 1956, 63(5): 277-293.
- [10] DAVIS J A. Clustering and structural balance in graphs[J]. Human relations, 1967, 20(2): 181-187.
- [11] FOGEL D B. An introduction to simulated evolutionary optimization[J]. IEEE transactions on neural networks, 1994, 5(1): 3-14.
- [12] FOGEL D B. Evolutionary algorithms in theory and practice[J]. Complexity, 1997, 2(4): 26-27.
- [13] 廖美英,郭荷清,张勇军.一种整数编码的改进遗传算法[J].计算机工程与应用,2003,39(1):103-105,120.
LIAO Meiying, GUO Heqing, ZHANG Yongjun. An ameliorative integer coded genetic algorithm[J]. Computer engineering and applications, 2003, 39(1): 103-105, 120.
- [14] 邓琨,张健沛,杨静.利用改进遗传算法进行复杂网络社团发现[J].哈尔滨工程大学学报,2013,34(11):1438-1444.
DENG Kun, ZHANG Jianpei, YANG Jing. Community detection in complex networks using an improved genetic algorithm[J]. Journal of Harbin engineering university, 2013, 34(11): 1438-1444.
- [15] YANG Bo, CHEUNG W, LIU Jiming. Community mining from signed social networks[J]. IEEE transactions on knowledge and data engineering, 2007, 19(10): 1333-1348.
- [16] KROPIVNIK S, MRVAR A. An analysis of the slovene parliamentary parties network[M]//FERLIGOJ A,

KRAMBERGER A. Developments in Statistics and Methodology. FDV, Ljubljana: Metodološki Zvezki, 1996.

[17] READ K E. Cultures of the central highlands, new guinea[J]. Southwestern journal of anthropology, 1954, 10(1): 1–43.

[18] MILO R, SHEN-ORR S, ITZKOVITZ S, et al. Network motifs: simple building blocks of complex networks[J]. Science, 2002, 298(5594): 824–827.

[19] FACCHETTI G, IACONO G, ALTAFINI C. Computing global structural balance in large-scale signed social networks[J]. Proceedings of the national academy of sciences of the United States of America, 2011, 108(52): 20953–20958.

作者简介:



常新功,男,1968年生,教授,博士,CCF会员,主要研究方向为社会网络分析、数据挖掘、进化算法。主持多项山西省重点课题。发表学术论文30余篇。



赵雅媚,女,1993年生,硕士研究生,主要研究方向为符号网络分析、进化算法。参与两个省级课题。发表学术论文4篇。

第四届云计算与大数据分析国际会议 (ICCCBDA 2019) 2019 the 4th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA 2019)

2019 the 4th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA 2019) which will be held during April 12–15, 2019 in Chengdu, China. ICCCBDA 2019 is one of the leading international conferences for presenting novel and fundamental advances in the fields of Cloud Computing Big Data Analytics. It also serves to foster communication among researchers and practitioners working in a wide variety of scientific areas with a common interest in improving Cloud Computing and Big Data Analytics related techniques.

ICCCBDA is organized by Sichuan Institute of Electronics, co-organized by Xihua University, technically supported by Southwest Jiaotong University.

Cloud computing and Big Data Analytics is a very hot topic in recent years, as well as a very important technology. By the advent of Big Data, cloud computing as the representative of technological innovation, it will set off a wave of reform in many areas, and will gradually create more value for human.