

基于双编码的重叠社团检测多目标优化方法

张 磊, 刘 庆, 杨尚尚, 杨海鹏, 程 凡, 马海平

(安徽大学计算机科学与技术学院计算智能与信号处理教育部重点实验室, 安徽合肥 230601)

摘 要: 近年来,多目标进化方法已被广泛应用于重叠社团检测问题并取得了较好的社团划分性能.如何设计合适的个体编码以及进化策略是提高基于多目标进化重叠社团检测算法性能的重要因素.为此,本文设计了一种双编码表示方法对非重叠社团结构和重叠点分别进行编码,能够有效解码得到重叠社团结构.在双编码表示的基础上,本文提出了一种基于双编码的重叠社团检测多目标优化方法(DRMOEA).在 DRMOEA 中,为了获得好的初始个体并提高算法检测性能,本文提出了一种基于社团边界点的初始化策略.除此之外,针对双编码中的重叠点编码部分,本文提出了基于精英个体边界点的交叉策略,该策略利用社团边界信息引导种群向好的方向进化,从而有效提高了算法的检测性能.最后,在 9 个真实世界网络上的实验结果表明 DRMOEA 算法优于其他 5 个代表性重叠社团检测算法.

关键词: 复杂网络; 重叠社团检测; 双编码; 多目标优化

中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112(2021)11-2101-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20201094

A Dual Representation-Based Multi-Objective Evolutionary Algorithm for Overlapping Community Detection

ZHANG Lei, LIU Qing, YANG Shang-shang, YANG Hai-peng, CHENG Fan, MA Hai-ping

(Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Computer Science and Technology, Anhui University, Hefei, Anhui 230601, China)

Abstract: In recent years, the multi-objective evolutionary methods have been widely used for solving overlapping community detection problem and have achieved good community division performance. To design appropriate individual encoding and evolution strategies is important to improve the performance of multi-objective overlapping community detection evolutionary algorithm. To this end, a dual representation method is designed to encode the non-overlapping community structures and overlapping nodes respectively, which can effectively obtain the overlapping community structures. On the basis of the dual representation, this paper proposes a dual representation-based multi-objective evolutionary algorithm for overlapping community detection (DRMOEA). In DRMOEA, an initialization strategy based on community boundary nodes is suggested to obtain good initial individuals, with the aim to improve the detection performance of the algorithm. In addition, for the overlapping part of the dual-representation, this paper proposes a crossover strategy according to the boundary nodes of elite individuals, which uses community boundary information to guide the evolution of the population towards a better direction. Finally, the experimental results on nine real-world networks show that the proposed DRMOEA is better than five representative baseline overlapping community detection algorithms.

Key words: complex network; overlapping community detection; dual representation; multi-objective optimization

1 引言

随着网络科学的快速发展,生活中很多领域的系统都可以表示为复杂网络,如蛋白质网络^[1]、社交网络^[2]、计算机网络^[3]等.其中社团结构^[4]作为复杂网络

的重要特性,对于研究与分析复杂网络有着很大帮助.在社团检测研究问题中,重叠社团检测问题受到了广泛地关注,因为现实世界中节点通常属于多个社团,比如社交网络中的每个人会有多个社交圈像班级、家庭、

收稿日期:2020-10-05;修回日期:2021-06-06;责任编辑:李勇锋

基金项目:国家自然科学基金(No.61976001, No.61876184, No.62076001);安徽省自然科学基金(No.2008085QF309);安徽高校自然科学基金项目(No.KJ2020A0036)

朋友、同事等。Palla 等人拓展了非重叠社团检测算法的研究,在 2005 年提出了重叠社团检测算法^[5]。此后,很多利用不同技术的重叠社团检测算法被提出,比如局部扩充^[6]、标签传播^[7]、模糊聚类^[8]等。近年来随着多目标进化算法^[9,10]在非重叠社团检测中取得的良好效果,越来越多的科研工作者提出基于多目标进化的重叠社团检测算法^[11-17]。在基于多目标进化优化的重叠社团检测算法中,通过同时优化多个互相牵制且充分利用网络结构信息的目标函数,最终得到一组不同层次的网络划分,决策者可以根据自己的要求来选择满足条件的社团划分。

Liu 等人在 2010 年提出 MEA_CDPs^[11] 算法, MEA_CDPs 既可以检测非重叠社团又可以检测重叠社团。具体来说,该算法提出了基于间接编码的个体表示方法,在解码过程中根据节点的适应度将节点分配到多个社团从而实现重叠社团检测。在 2014 年, Liu 等人在 MEA_CDPs 的基础上提出了针对符号网络的算法 MEAs_SCN^[12], 该算法采用间接编码使得多目标优化可以解决符号网络的重叠社团检测问题。虽然 MEA_CDPs 和 MEAs_SCN 可以解决重叠社团检测问题,但是个体解码比较耗时。2015 年, Li 等人提出了基于光谱聚类的多目标量子粒子群算法 IMOQPSO^[13] 来处理重叠社团检测问题, 该算法提出了中心边编码可以直观地解码得到重叠社团, 但是这种方法需要提前设置社团数目且进化算子难以设计。Wen 等人在 2016 年提出了 MCMOE^[14] 算法, 在 MCMOE 中提出了基于最大完全子图的编码方式, 由于完全子图之间可以共享节点, 故 MCMOE 可以处理重叠社团检测问题, 但是这种编码检测不到完全子图之外的重叠点。2017 年, Zhang 等人提出了基于混合编码的重叠社团检测多目标进化算法 MRMOEA^[15], 该算法将网络中节点分为非重叠点和候选重叠点, 并设计不同的进化策略对其优化, 最终将候选重叠点编码为 0 的节点解码为重叠点。由于 MRMOEA 只在候选重叠点范围里检测重叠点, 故会将不在候选重叠点范围里的重叠点遗失。Gema Bello-Ortiz 等人在 2018 年提出了 MOGA-OCD^[16] 算法, 这篇文章采用基于边的编码方式, 由于网络中节点可以连接多条边, 故基于边的编码方式可以自然地解码得到重叠社团, 但是网络中边的数目远大于节点的数目, 所以该算法不适用于大规模或者稠密的网络。在 2020 年, Tian 等人提出了基于模糊聚类的多目标重叠社团检测方法 EMOFM^[17], 该算法采用基于社团中心点的编码方式, 通过优化节点到社区中心的模糊阈值来解码得到重叠社团, 由于 EMOFM 需要为每个节点优化阈值来判断其是否为重叠点, 故时间复杂度较高。

以上基于多目标优化的重叠社团检测算法, 尽管

可以解决重叠社团检测问题, 但这些算法没有考虑进化过程中社团结构和重叠节点之间的关系, 从而限制了算法的检测性能。基于此, 本文提出了基于双编码的重叠社团检测多目标优化方法 (DRMOEA)。在 DRMOEA 中, 分别对非重叠社团和重叠点进行编码和进化, 并且在进化过程中考虑到社团结构与重叠点之间的关系, 充分利用网络划分得到的社团结构指导重叠点进化, 因而能够更加有效且全面地检测重叠社团。本文主要贡献包括以下 3 个方面:

(1) 本文提出了一种双编码表示方法, 具体来说, 个体编码由两部分组成: 第一部分检测非重叠社团对网络中节点采用标签编码, 第二部分检测重叠点对网络中节点采用二进制编码。解码过程中将标签编码中标签相同的节点放到同一社团内, 同时将二进制编码中编码为 1 的节点解码为重叠点, 通过这种编码能够有效解码得到重叠社团结构。

(2) 在以上双编码的基础上, 本文提出了基于双编码的重叠社团检测多目标优化方法 DRMOEA。在 DRMOEA 中, 提出了基于社团边界点的初始化策略, 以此来获得初始个体并有效提高了算法性能; 在进化过程中针对重叠点编码部分, 提出了基于精英个体边界点的交叉策略, 该策略利用社团边界信息引导种群向好的方向进化, 进一步提高了算法检测性能。

(3) 本文在 9 个真实世界网络上进行了实验, 和其他 5 个具有代表性的基于多目标优化的重叠社团检测算法进行对比, 实验结果表明 DRMOEA 算法优于其他 5 个代表性重叠社团检测算法。

2 DRMOEA 算法设计

2.1 双编码表示

基于双编码的重叠社团检测多目标优化方法的基本思想是提出了一种双编码表示方法, 个体基因由两部分组成, 第一部分是基于标签的编码, 用来检测非重叠社团结构, 第二部分采用二进制编码, 用来检测网络中的重叠点。在对第二部分进行编码前, 先解码第一部分得到非重叠社团结构, 然后获取社团的边界节点, 在本文中社团的边界节点指一条边连接的两个节点不在同一社团内, 则该边连接的两个节点为社团的边界节点。本文这样做的目的是为了在社团的边界点之间找重叠点, 因为现实表明重叠点一般位于社团的边界部分。在进化过程中将种群里精英个体第一部分标签编码解码得到非重叠社团, 然后利用社团局部信息指导第二部分编码表示的重叠点的进化, 从而提高算法性能。

图 1 展示了本文提出的双编码表示, 以及其编码及解码过程。如图 1(a) 所示网络有 8 个节点, 图 1(b) 展示了个体的两种编码方式: 第一部分表示非重叠社团结

构,采用标签编码;第二部分表示重叠点编码,采用二进制编码.在解码过程中将标签编码中标签相同的节点放到同一社团内,同时将重叠点编码中编码为1的节点解码为重叠点,得到如图1(c)所示网络划分.对于二

进制编码基因位为1的节点,表示该节点为重叠点,在解码时将其添加到邻居节点所在社团,否则为非重叠点,得到如图1(d)所示重叠社团结构,根据所提双编码最终解码为三个社团{1,2,3}、{3,4,5,6}和{6,7,8}.

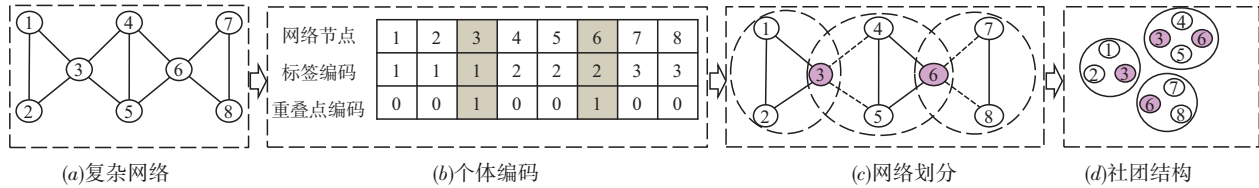


图1 个体编码及解码示例

2.2 算法描述

2.2.1 算法框架

本文提出的DRMOEA算法采用NSGA-II框架,同时优化多目标重叠社团检测算法中两个常用目标,即社团内部联系比例(KKM)^[18]和社团外部联系比例(RC)^[19].算法1为DRMOEA算法的总体框架,主要包括两个步骤:首先初始化种群,在初始化重叠点时考虑网络划分结构信息,提出基于社团边界点的初始化策略,得到一个具有良好社团结构的初始种群.接着在进化阶段,种群进行交叉变异并更新种群,算法不断迭代直到满足终止条件停止,输出最终种群.为了更好的利用第一部分标签编码得到的社团结构信息,从而引导种群进化并提高算法性能,针对第二部分重叠点编码提出了基于精英个体边界点的交叉策略.下面详细介绍上述两个策略.

算法1 DRMOEA算法流程

输入: A : 网络的邻接矩阵, popsize : 种群大小, pm : 变异概率, pc : 交叉概率, g : 控制交叉频率

输出: 最终种群 P ;

Step1: 种群初始化

1. $P \leftarrow \text{Initialization}(A, \text{popsize});$

Step2: 种群进化

2. $t = 1$; //迭代次数

3. WHILE termination condition is not satisfied DO

4. $\text{MatingPool} \leftarrow \text{TournamentSelection}(P);$

5. $\text{Elite} \leftarrow \text{Select}(\text{MatingPool});$

6. $C \leftarrow \text{CrossMutation}(\text{MatingPool}, \text{Elite});$

7. IF $t \% g == 0$ THEN

8. $\text{Elite} \leftarrow \text{Select}(C);$

9. $C \leftarrow \text{CrossMutation}(C, \text{Elite});$

10. END

11. $P \leftarrow \text{Update}(P, C);$

12. $t++;$

13. END

14. 选择最终解

2.2.2 基于社团边界点的初始化策略

本文提出的双编码表示主要是通过社团边界信息来初始化重叠点,首先初始化个体第一部分编码并解码得到社团结构,然后利用社团边界点信息对个体第二部分重叠点编码进行初始化.如图2(a)中的网络,解码图2(b)中标签编码得到两个非重叠社团结构 $C_1 \{1, 2, 3, 4\}$ 和 $C_2 \{5, 6, 7\}$, 社团的边界节点为4、5、7.通过直观的观察可以发现,重叠点一般位于社团的边界,且和其所在的每个社团的联系紧密性差距不大.得到社团边界点之后,将其解码为重叠点,计算其与除本身所在社团之外的社团的连接边数,与其在每个社团分布的平均边 $\text{avg} = d_i / (c_{\text{num}} + 1)$ (c_{num} 表示节点 i 邻居节点所在社团数目)相比较,如果该边界节点与其邻居节点所在社团连接的边数均小于平均边数,则该节点可能不是重叠点,如果存在与其他社团连接的边数大于或等于平均边数,则该节点可能为重叠点,初始化时将二进制编码基因位标为1.如图2(a)中的4、5、7节点, $\text{avg}_{i4} = 4/2 = 2$, 与其邻居节点所在社团连接边数 $L_{C_2} = 2$ 相等,故将4号节点作为重叠点保留, $\text{avg}_{i5} = 3/2 = 1.5$, 大于邻居节点所在社团连接边数 $L_{C_1} = 1$, 故将5号节点从边界节点里删除,同理将7号节点从边界节点里删除.最后只剩下4号节点,在第二部分的初始化过程中,将4号节点对应重叠点编码的基因位编码为1,表示节点4是重叠点.初始化种群流程如算法2所示.

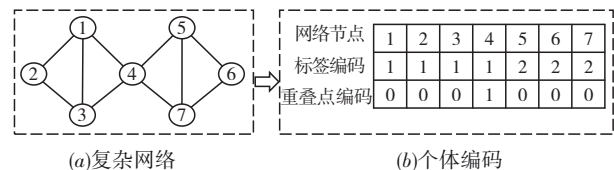


图2 个体初始化示例

2.2.3 基于精英个体边界点的交叉策略

在多目标进化算法中,遗传操作对种群的进化起着至关重要的指导作用,种群中的个体通过遗传操作向好的方向进化,遗传操作主要有交叉、变异、选择等.由于

算法2 初始化种群算法流程

输入: $A(V, E)$: 邻接矩阵, pop: 种群规模, n : 节点数目;

输出: 初始种群: $P = [P_1, P_2, \dots, P_{\text{pop}}]$;

1. FOR $i = 1$ to pop DO
2. Generate a vector-based individual P_i ;
3. Decode P_i to obtain communities;
4. Get community boundary point set B based on P_i ;
5. FOR $j = 1$ to length(B) DO
6. $\text{avg}_{ij} = d_j / (c_{\text{num}} + 1)$;
7. FOR $m = 1$ to c_{num} DO
8. IF $\text{avg}_{ij} \leq L_{C_m}$ THEN
9. $P_i[n + V_j] = 1$;
10. BREAK;
11. END
12. END
13. END
14. END

本文中采用双编码表示方法,所以对不同的编码方式采用不同的交叉策略.第一部分采用标签编码,用任意整数表示社团标签,对于标签编码本文采用文献[20]中提出的双向交叉算子,该算子选择父代中一部分节点作为交叉点,适用于整数编码且表现出较好的性能.

个体中第二部分二进制编码首先采用均匀交叉,保留父代的一些特性.然而随着个体第一部分的进化,社团结构会随之改变,重叠点的位置也会发生变化.随着迭代次数的增加,个体基因逐渐变得稳定,社团结构也随之趋于稳定.在进化过程中为了充分利用社团的边界信息指导重叠点进化,从而得到重叠社团结构,本文提出了基于精英个体边界节点的交叉操作.因为随着种群的进化,社团结构逐渐向好的方向发展,算法每隔十代从种群第一前沿面中选择精英个体,根据精英个体第一部分编码解码得到好的非重叠社团结构,然后利用这些社团结构投票得出边界点向量 B . 如果 $B(i)$ 为 1,说明节点 i 在得到的所有社团结构中都位于社团边界,然后利用投票得出的向量 B 与均匀交叉产生的子代进行交叉,基于精英个体边界点交叉策略的主要步骤如算法 3 所示.对于生成的子代 O ,本文采用随机变异,根据投票得出的边界点,对二进制编码进行随机变异,由 0 变为 1 或者由 1 变为 0.

3 实验与分析

本节中,将 DRMOEA 算法与五个具有代表性的社团检测算法在不同特性的网络上进行一系列实验,通过比较 $Qov^{[21]}$ 和 $gNMI^{[22]}$ 值来验证 DRMOEA 算法的性能. Qov 既可以用于社团结构已知的网络也适用于社团结构未知的网络, $gNMI$ 精度是对两种社区结构一致

算法3 基于精英个体边界点交叉策略算法流程

输入: 子代 O , 精英个体, $\text{Ind}\{\text{Ind}_1, \text{Ind}_2, \dots, \text{Ind}_k\}$ 交叉概率 pc, 节点数目 n ;

输出: 子代 O ;

1. Decode Ind to obtain communities;
2. Get the boundary point vector B based on Ind;
3. FOR $i = 1$ to n DO
4. IF $\text{rand} \leq \text{pc}$ THEN
5. $O(i) = B(i)$;
6. END
7. END

性的信息理论度量,用来评估算法检测到的社团和网络真实社团的一致性,故 $gNMI$ 精度只能用于社团结构已知的网络.接下来详细分析实验设置及实验结果.

3.1 实验设置

3.1.1 对比算法

本文选取了五个基于多目标进化的社团检测算法 MEAs_SCN^[12]、IMOQPSO^[13]、MCMOEA^[14]、MRMOEA^[15]、EMOFM^[17],作为对比算法和提出的 DRMOEA 进行比较.对于每个对比算法,本文都使用作者提供的代码,以及各自文章中建议的参数设置.为了公平起见,算法中种群大小均设置为 100,迭代次数为 100,所有算法的实验结果都是每个算法在每个网络上独立运行 20 次,记录最佳值,平均值.对于真实网络数据集上的实验结果,本文在 0.05 的显著性水平下采用了 Wilcoxon rank sum test 来分析算法性能的统计差异,表中的符号“+”, “≈”, “-”分别表示对比算法实验结果显著的优于、接近以及比 DRMOEA 所得结果差.所有算法的运行环境为: Intel (R) Core (TM) i5-7500 CPU 16GB RAM, Windows 10 操作系统.

3.1.2 真实世界网络数据

本文在 9 个已知且被广泛使用的真实世界网络上验证 DRMOEA 算法性能,网络的详细特征描述如表 1 所示.

表 1 9 个真实网络的信息

Networks	#Nodes	#Edges	Ave. Degree	#Communities
Karate	34	78	4.59	2
Dolphin	62	159	5.13	2
Football	115	613	10.66	12
Polbook	105	441	8.4	3
PPI_D2	1443	6993	9.69	162
Jazz	198	2742	27.70	Unknown
Netscience	1589	2742	3.45	Unknown
PPI	2445	6265	5.12	Unknown
Blogs	3984	6803	3.41	Unknown

3.2 实验结果

3.2.1 Qov 值的对比分析

表 2 展示了 DRMOEA 和 5 个对比实验在 9 个具有不同特性的真实网络上得到的 Qov 值,从表中可以看出在大多数网络上 DRMOEA 算法得到的 Qov 值比其他算法得到的 Qov 大,结果表明 DRMOEA 优于对比算法. 因为本文提出的双编码表示,在社团的边界中找重叠点,充分利用了个体第一部分编码得到的网络划分结构指导第二部分重叠点的进化,因而能够更准确全面的找重叠点,从而提升了算法检

测性能.

3.2.2 gNMI 值的对比分析

在本节中展示 DRMOEA 算法在 gNMI 值上和其他算法的对比,由于 gNMI 只能用于社团结构已知的网络,故本节只展示在 5 个已知社团结构的真实世界网络上的结果. 表 3 给出了 DRMOEA 和对比算法在 5 个已知社团结构的真实网络上得到的 gNMI 值,从表中结果可以看出 DRMOEA 在 5 个网络得到的 gNMI 值都是最大的,这些实验结果验证了 DRMOEA 算法在真实世界网络上的有效性.

表 2 DRMOEA 和 5 个对比算法在真实网络上 Qov 值的对比

Network	Metric	MEAs_SCN	IMOQPSO	MCMOEA	MRMOEA	EMOFM	DRMOEA
Karate	Qov _{max}	0.2090	0.2114	0.2096	0.2285	0.2348	0.2348
	Qov _{avg}	0.1993-	0.2091-	0.1856-	0.2242-	0.2331-	0.2338
Dolphin	Qov _{max}	0.2098	0.2540	0.2229	0.2680	0.2744	0.2746
	Qov _{avg}	0.1918-	0.2506-	0.1733-	0.2595-	0.2732≈	0.2735
Fotball	Qov _{max}	0.2582	0.2440	0.2669	0.3056	0.3067	0.3067
	Qov _{avg}	0.2314-	0.2336	0.2445-	0.3028-	0.3061-	0.3065
Polbook	Qov _{max}	0.2518	0.2440	0.2173	0.2641	0.2703	0.2702
	Qov _{avg}	0.1875-	0.2372-	0.1868-	0.2625-	0.2695≈	0.2702
Jazz	Qov _{max}	0.1482	0.0371	0.1467	0.2193	0.2259	0.2258
	Qov _{avg}	0.1448-	0.0309-	0.1335-	0.2143-	0.2252≈	0.2252
Netscience	Qov _{max}	0.4227	0.3839	0.4458	0.4698	0.4697	0.4740
	Qov _{avg}	0.4212-	0.3581-	0.4362-	0.4668-	0.4667-	0.4720
PPI	Qov _{max}	0.2147	0.1290	0.1977	0.3293	0.3437	0.3470
	Qov _{avg}	0.1938-	0.0861-	0.1903-	0.3225-	0.3414-	0.3438
Blogs	Qov _{max}	0.2450	0.3047	0.3352	0.3968	0.4106	0.4079
	Qov _{avg}	0.2431-	0.2463-	0.3233-	0.3960-	0.4093+	0.4070
PPI_D2	Qov _{max}	0.3360	0.3908	0.3391	0.4112	0.4147	0.4186
	Qov _{avg}	0.3243-	0.3885-	0.3162-	0.4049-	0.4135-	0.4169
+/-/≈		0/9/0	0/9/0	0/9/0	0/9/0	1/5/3	

表 3 DRMOEA 和 5 个对比算法在真实网络上 gNMI 值的对比

Network	Metric	MEAs_SCN	IMOQPSO	MCMOEA	MRMOEA	EMOFM	DRMOEA
Karate	gNMI _{max}	0.6071	0.7061	0.8334	1.0000	1.0000	1.0000
	gNMI _{avg}	0.5790-	0.6821-	0.7369-	0.9356-	1.0000≈	1.0000
Dolphin	gNMI _{max}	0.3313	0.8357	0.3239	1.0000	1.0000	1.0000
	gNMI _{avg}	0.3271-	0.7240-	0.2694-	0.9340-	0.9667-	1.0000
Fotball	gNMI _{max}	0.6360	0.8090	0.7340	0.8298	0.8348	0.8465
	gNMI _{avg}	0.5404-	0.7942-	0.7026-	0.8008-	0.8104-	0.8215
Polbook	gNMI _{max}	0.3210	0.4022	0.3139	0.4379	0.3514	0.4587
	gNMI _{avg}	0.2225-	0.3612-	0.2668-	0.3969-	0.3301-	0.4298
PPI_D2	gNMI _{max}	0.2483	0.2052	0.2636	0.2617	0.2177	0.2682
	gNMI _{avg}	0.2416-	0.2006-	0.2371-	0.2545-	0.2071-	0.2606
+/-/≈		0/5/0	0/5/0	0/5/0	0/5/0	0/4/1	

从这些实验结果中可以看出 DRMOEA 相较于其他 5 个对比算法可以取得更准确的社团划分效果,这是因为 DRMOEA 采用双编码表示方法,对非重叠社团和重叠点分别进行编码且采用不同的交叉算子. 算法首先根据标签编码解码得非重叠社团结构,在初始化时利用了非重叠社团的边界信息. 在进化过程中利用种群中的精英个体,解码精英个体第一部分编码得社团结构,由社团结构投票得出边界点,利用边界点指导种群中重叠点的进化,从而提高了算法的检测性能.

4 总结

本文提出了基于双编码的重叠社团检测多目标优化方法(DRMOEA)来解决重叠社团检测问题. 为了表示重叠社团结构,本文中提出了一种双编码表示方法,具体来说该编码由两部分组成,第一部分是基于标签的编码表示非重叠社团结构,第二部分是二进制编码表示重叠点,最终两部分结合在一起解码得到重叠社团结构. 在双编码的基础上提出了基于社团边界点的初始化策略来初始化二进制编码,从而得到较好的初始化个体提高算法性能. 在进化过程中对不同的编码采用不同的交叉策略,针对第二部分编码提出了基于精英个体边界点的交叉策略,利用社团结构信息指导种群的进化来提高算法的性能. 在真实网络上的实验结果表明本文的算法能够取得较好的结果,说明提出的 DRMOEA 算法能够有效的处理重叠社团检测问题. 本文提出的双编码表示方法虽然能够处理重叠社团检测问题,但是现实世界中的网络往往都是大规模的,在未来将进一步探索如何设计面向大规模的双编码表示,以此来解决大规模网络下的重叠社团检测问题.

参考文献

- [1] Pizzuti C, Rombo S E. Algorithms and tools for protein-protein interaction networks clustering, with a special focus on population-based stochastic methods[J]. *Bioinformatics*, 2014, 30(10): 1343 – 1352.
- [2] Harakawa R, Ogawa T, Haseyama M. Accurate and efficient extraction of hierarchical structure of web communities for web video retrieval[J]. *ITE Transactions on Media Technology & Applications*, 2016, 4(1): 49 – 59.
- [3] Jie Y, Zhishuai L, Qiu X. Community detection in complex networks: Algorithms and analysis[A]. *International Conference on Trustworthy Computing and Services[C]*. Berlin, Heidelberg: Springer, 2014. 238 – 244.
- [4] Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks[J]. *Nature*, 2010, 466(7307): 761 – 764.
- [5] Palla G, Derényi Imre, Farkas Illés, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. *Nature*, 2005, 435(7043): 814 – 818.
- [6] 於志勇, 陈基杰, 郭昆, 等. 基于影响力与种子扩展的重叠社区发现[J]. *电子学报*, 2019, 47(1): 153 – 160.
Yu Z Y, Chen J J, Guo K, et al. Overlapping community detection based on influence and seeds extension[J]. *Acta Electronica Sinica*, 2019, 47(1): 153 – 160. (in Chinese)
- [7] Sun H, Liu J, Huang J, et al. CenLP: A centrality-based label propagation algorithm for community detection in networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2015, 436: 767 – 780.
- [8] Sun P G, Gao L, Han S S. Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks[J]. *Information Sciences*, 2011, 181(6): 1060 – 1071.
- [9] 公茂果, 焦李成, 杨咚咚, 等. 进化多目标优化算法研究[J]. *软件学报*, 2009, 20(2): 271 – 289.
- [10] 巩敦卫, 季新芳, 孙晓燕. 基于集合的高维多目标优化问题的进化算法[J]. *电子学报*, 2014, 42(1): 77 – 83.
GONG Dun-wei, JI Xin-fang, SUN Xiao-yan. Solving many-objective optimization problems using set-based evolutionary algorithms[J]. *Acta Electronica Sinica*, 2014, 42(1): 77 – 83. (in Chinese)
- [11] Liu J, Zhong W, Abbass H A, et al. Separated and overlapping community detection in complex networks using multiobjective evolutionary algorithms[A]. *IEEE Congress on Evolutionary Computation[C]*. IEEE, 2010. 1 – 7.
- [12] Liu C, Liu J, Jiang Z. A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks[J]. *IEEE Transactions on Cybernetics*, 2014, 44(12): 2274 – 2287.
- [13] Li Y, Wang Y, Chen J, et al. Overlapping community detection through an improved multi-objective quantum-behaved particle swarm optimization[J]. *Journal of Heuristics*, 2015, 21(4): 549 – 575.
- [14] Wen X, Chen W N, Lin Y, et al. A maximal clique based multiobjective evolutionary algorithm for overlapping community detection[J]. *IEEE Transactions on Evolutionary Computation*, 2017, 21(3): 363 – 377.
- [15] Zhang L, Pan H, Su Y, et al. A mixed representation-based multiobjective evolutionary algorithm for overlapping community detection[J]. *IEEE Transactions on Cybernetics*, 2017, 47(9): 2703 – 2716.
- [16] Gema B O, Sancho S S, David C. A multi-objective genetic algorithm for overlapping community detection

- based on edge encoding[J]. Information Sciences, 2018, 462: 290 – 314.
- [17] Tian Y, Yang S, Zhang X. An evolutionary multiobjective optimization based fuzzy method for overlapping community detection[J]. IEEE Transactions on Fuzzy Systems, 2020, 28(11): 2841 – 2855.
- [18] Gong M, Ma L, Zhang Q, et al. Community detection in networks by using multiobjective evolutionary algorithm with decomposition[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(15): 4050 – 4060.
- [19] Wei Y C, Cheng C K. Ratio cut partitioning for hierarchical designs[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1991, 10(7): 911 – 921.
- [20] Gong M, Fu B, Jiao L, et al. Memetic algorithm for community detection in networks[J]. Physical Review E, 2011, 84(5): 056101.
- [21] Nicosia V, Mangioni G, Carchiolo V, et al. Extending the definition of modularity to directed graphs with overlapping communities[J]. Journal of Statistical Mechanics Theory & Experiment, 2009, 2009(03): 3166 – 3168.
- [22] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3): 033015.

作者简介



张 磊 男, 1986年5月生于安徽安庆, 现为安徽大学计算机科学与技术学院副教授、博士生导师. 主要研究方向为多目标优化及其应用、社交网络分析、约束模式挖掘、推荐系统等.
E-mail: zl@ahu.edu.cn



马海平(通讯作者) 女, 1986年5月生于安徽合肥, 现为安徽大学计算机科学与技术学院讲师. 主要研究方向为推荐系统、个性化教育、机器学习等.
E-mail: hpma@ahu.edu.cn