

基于谱聚类的多目标进化社区发现算法研究

董明刚 弓佳明 敬超

桂林理工大学信息科学与工程学院 广西 桂林 541004

嵌入式技术与智能系统重点实验室 广西 桂林 541004

(d2015mg@qq.com)

摘要 多目标优化算法在复杂网络社区发现中具有很强的竞争力,然而,在处理社区结构较为模糊、网络数据规模大的问题时难以得到满意的效果。为克服现有多目标方法的不足,提出一种基于谱聚类的多目标复杂网络社区发现算法。该算法先用谱聚类对编码后的复杂网络进行初始种群划分,利用子图聚类特性生成高质量的初始种群。采用一种网格约简的数据归减方法在进化过程中对种群进行约减,有效降低算法复杂度,以满足大规模网络社区发现需求。在仿真网络和9个真实网络上的实验结果表明,该算法在社区发现精度性能和计算复杂度方面,都要优于MRMOEA,RMOEA,MCMOEA 3种代表性的基于多目标的社区发现算法。

关键词: 多目标进化算法;社区发现;复杂网络;大规模网络

中图分类号 TP391

Multi-objective Evolutionary Algorithm Based on Community Detection Spectral Clustering

DONG Ming-gang, GONG Jia-ming and JING Chao

School of Information Science and Engineering, Guilin University of Technology, Guilin, Guangxi 541004, China

Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin, Guangxi 541004, China

Abstract The multi-objective optimization algorithm is competitive in the discovery of complex network communities. However, it is difficult to obtain the satisfied results while dealing with the problem of fuzzy community structure and large scale of network data. To overcome the shortcomings of existing multi-objective methods, a multi-objective complex network community discovery algorithm based on spectral clustering is proposed. The proposed algorithm uses spectral clustering to perform initial population partitioning on the encoded complex network, and exploits its subgraph clustering characteristics to obtain a better initial population. A data reduction method based on grid reduction is applied to reduce the population in the process of evolution, which effectively reduces the complexity of the algorithm. The experimental results on the simulation network and the real network show that the proposed algorithm outperforms than that of the other three representative multi-target based community discovery algorithms in terms of community discovery performance and computational complexity.

Keywords Multi-objective evolutionary, Community discovery, Complex network, Large-scale network

1 引言

真实网络中个体之间存在大量复杂信息,形成复杂网络,例如社交网络、电力网络、生物网络和神经网络^[1-4]。分析与研究复杂网络的特性及潜在规律有着重大的意义,比如,通过分析知网中研究人员的研究领域,可以挖掘出同一领域的相关研究工作有助于更好地进行科研交流;生物网中,一个社区内的蛋白质可能有相似的作用和隐藏的生物信息,能够推测出相似的网络社区内未知蛋白质的作用和隐藏的生物信息^[5]。然而近年来,随着 Internet 技术的发展,网络用户数量激增,网络节点增加,节点之间的关系越发复杂,使得网络结

构变得模糊,直接应用传统的复杂网络社区发现算法会出现精度下降、计算效率低等问题。

针对结构模糊的复杂网络问题,有学者提出了基于多目标复杂网络社区发现算法,此类算法大都基于智能算法,如具有支配关系的多目标算法和遗传算法等。文献[6]提出了 MOGA-Net 算法,这个算法有两个目标函数,可以解决内部链路密度大而外部链路密度小的复杂网络。并且在不同节点度上有着不同的网络划分能力,该算法使用遗传算法,通过不断进化保留最优解,不断迭代得出最优解集。文献[7]提出了 MOPSO 算法,该算法利用多目标粒子群算法,在两个目标函数作用下通过改进进化策略,使得算法有着高收敛性,同时减

基金项目:国家自然科学基金(61563012,61802085);广西自然科学基金(2014GXNSFAA118371,2015GXNSFBA139260);广西嵌入式技术与智能系统重点实验室基金(2018A-04);广西研究生计划项目(YCSW201962)

This work was supported by the National Natural Science Foundation of China (61563012,61802085), Natural Science Foundation of Guangxi, China (2014GXNSFAA118371,2015GXNSFBA139260), Guangxi Key Laboratory of Embedded Technology and Intelligent System Foundation (2018A-04) and Guangxi Graduate Program(YCSW201962).

通信作者:敬超(jingchao@glut.edu.cn)

少了陷入局部最优的可能。文献[8]提出了 MEAs_SCN 算法,该算法通过相似度得出目标函数,然后针对目标函数之间的关系平衡最优解集,之后又提出一种新表示方法用以处理多目标问题。Zhang 提出了一种基于混合编码的多目标复杂网络社区发现算法(MRMOEA)^[9],该算法用了一种混合编码的方式来解决社区发现问题。文献[10]提出 MCMOEA 算法,该算法通过选取最密切相关的部分社区组成一个大的社区来简化社区发现问题。最近 Zhang 提出了一种基于网格约简的多目标网络社区发现算法(RMOEA)^[11],该算法利用网格约简对种群精英个体进行筛选,从而大大减少了搜索空间,提高了算法的效率和性能。以上算法都具有较高的复杂网络社区发现水平,极大地推动了基于多目标社区发现的研究。

以上算法在面向结构模糊和规模较大的复杂网络时,存在的问题有:1)能保证社区发现的准确性,却不能解决大规模的复杂网络;2)可以解决大规模复杂网络的问题,但是社区发现的准确性较低。针对这类问题,本文提出了基于谱聚类多目标复杂网络社区发现算法(Spectral clustering-Based Multi-objective Evolutionary Algorithm for Community Detection, SMOEA),通过子图聚类将复杂网络数据集进行预处理,以更好地利用节点信息;通过网格约简将一个社区整体归减为一个新的节点,同时记录其合并基因,保证网络信息不丢失,有效降低算法复杂度。实验对比分析结果表明,基于谱聚类多目标社区发现算法有着良好的效率和性能,且在结构模糊的数据集上也能较好地解决复杂网络社区发现问题。

2 相关工作

本节先对多目标社区发现算法(Multi-Objective Community Discovery Algorithm, MOCD)作以介绍^[12],然后对多目标进化算法作以说明,最后描述了谱聚类算法。

2.1 多目标社区发现问题

社区发现是将网络中节点根据节点之间的链路密度划分为不同的群体的过程。相同群体中的节点之间的链接(称为内部链接密度)需要尽可能密集,而不同群体中的节点之间的链接(称为外部链接密度)应该足够稀疏。在复杂网络中,这样的节点群体通常称为社区。图1给出了一个具有9个节点的复杂网络的示例。可以看出,网络分为两个社区,一个由5个节点{1,2,3,4,5}组成,另一个由4个节点{6,7,8,9}组成。

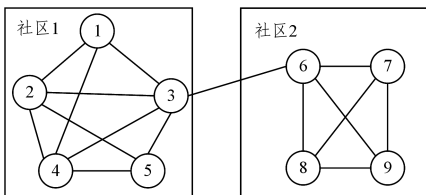


图1 社区链路图

Fig. 1 Community link diagram

为了进行社区发现,使用几个标准来测量链内和链间密度^[1],本文使用最小核 k-均值(KKM)和比率削减(RC),将社区发现定义为一个双目标优化问题。给定网络 $G=(V, E)$ ($|V|=n, |E|=m$),且 a 为 G 的邻接矩阵^[13],目标极小化问题的形式化定义为:

$$F(C) = (KKM(C), RC(C)) \quad (1)$$

$$KKM = 2(n-k) - \sum_{i=1}^m \frac{L(V_i, V_i)}{|V_i|} \quad (2)$$

$$RC = \sum_{i=1}^m \frac{L(V_i, \bar{V}_i)}{|V_i|} \quad (3)$$

其中, $C = \{V_1, V_2, \dots, V_k\}$ 是 G 的一部分,由 k 个社区 V_1, V_2, \dots, V_k 组成, $L(V_i, V_i) = \sum_{i \in V_i, j \in V_i} A_{ij}$, $L(V_i, \bar{V}_i) = \sum_{i \in V_i, j \in \bar{V}_i} A_{ij}$ 表示节点的内部和外部链路密度。如文献[14]所述,使用这两个目标是因为 KKM 是社区数量的递减函数,而 RC 相反,这也就说明 KKM 和 RC 是相互冲突的目标。KKM 和 RC 的最小化可以确保一个社区内的链接是密集的,而社区间的链接是稀疏的,这符合社区检测的要求。

2.2 多目标进化算法的简介

可以通过数学表达式表达矛盾冲突普遍存在的多目标演化问题中的多目标函数,如式(4)所示^[15]:

$$\begin{cases} \min F(x) = (f_1(x), f_2(x), \dots, f_k(x))^T \\ \text{s. t. } g_i \leq 0, i = 1, 2, \dots, p \\ h_i = 0, i = 1, 2, \dots, q \end{cases} \quad (4)$$

其中, $x = (x_1, \dots, x_m) \in \Omega$ 被称为 m 维决策向量,也是种群个体(决策向量有 m 维,决策空间的可行区域在 Ω 处)。在多目标进化中优化函数向量 $\min F(x) = (f_1(x), f_2(x), \dots, f_k(x))^T$, k 为目标函数的个数^[16]。

多目标进化算法的大概流程如图2所示。

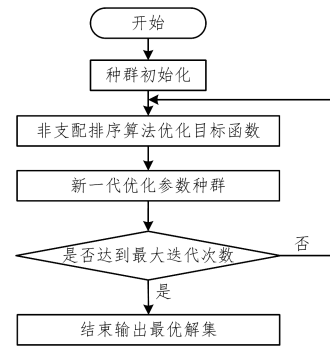


图2 多目标进化算法流程图

Fig. 2 Multi-objective evolution algorithm flow chart

2.3 谱聚类算法描述

谱聚类是一种基于图论的聚类方法^[17],可以最佳地划分任何形状的数据。高维空间中的数据被缩减为低维空间,然后通过低维空间中的其他聚类算法(例如 K-means 算法)进行聚类。因此谱聚类算法具有处理规模庞大且复杂数据的能力。

将 n 个样本点数据集中的每个点和任意两点之间的边看作无向带权图 $G=(V, E)$,其中, V 是点的集合, E 是边的集合。任意两点之间的边的权重 w 指两点之间的相似度。然后利用所有点的任意两点之间的相似度构造相似度矩阵 W 。矩阵 D 是度矩阵(对角矩阵)。相似度矩阵 W 的每一行的元素之和作为度矩阵 D 的对角元素。 L 是拉普拉斯矩阵:

$$L = D^{-1}L = D^{-1}(D - W) \quad (5)$$

计算 L 的特征值,将特征值从小到大排序,取前 k 个特征值,并计算前 k 个特征值的特征向量,将 k 列向量合为一个矩阵,并使其成为 U 的第 i 行向量,其中, k 均值算法用于聚类新的采样点。 $U = \{u_1, u_2, \dots, u_k\}$,令 $y_i \in R^k$ 是 U 的第 i 行向量,其中 $i = 1, 2, \dots, n$;使用核 k-means 聚类算法处理新样本 $Y = \{y_1, y_2, \dots, y_n\}$ 。

3 基于谱聚类的多目标社区发现算法

基于复杂网络面对的结构模糊及数据规模庞大的问题,本文给出了一种基于谱聚类的多目标社区发现算法(SMOEA),该算法包括两个主要过程:1)根据谱聚类的子图聚类特性划分初始种群,利用多目标进化算法对数据进行处理;2)在种群进化中使用网格约简对数据进行归纳缩减,从而使计算效率得到提升。最终获得最佳的社区发现结果。

3.1 基于谱聚类的初始种群划分

本文利用在任何形状的样本空间上都可以聚类的谱聚类算法来预处理数据集,以促进编码后的总体划分,从而保证在利用多目标优化算法处理大规模的复杂网络数据时,提升算法的性能。种群划分能为后续的多目标优化算法提供较好的初始种群,保证种群在进化中不会受到较多非重叠节点的干扰,保持种群的进化方向。多目标优化算法得到的结果越好,解码后得到的社区发现的精度就越高。

谱聚类算法的基本思想是谱图理论,把聚类问题视为图分割问题,其实质是利用拉普拉斯矩阵的特征向量聚类。

算法 1 解的生成

输入:通过数据样本集构建 $G=(V,E)$,用种群中的个体表示网络中的节点,个体间的相似度为网络中的边权,设种群规模为 m

输出:初始划分种群

步骤 1 计算 $\text{pop} * \text{pop}$ 的相似度矩阵 W 。

步骤 2 将矩阵 S 的每一列的元素相加,并将每一列作为元素添加到度矩阵 D 对角线上。

步骤 3 计算拉普拉斯矩阵(见式(5))。

步骤 4 计算矩阵 L 的特征值,从小到大排列特征值,计算前 k 个特征值的特征向量,并计算前 k 个特征值的特征向量,并将 k 个列向量组成矩阵 $U=\{u_1, u_2, \dots, u_k\}$ 。

步骤 5 令 $y_i \in R^k$ 为 U 的第 i 行向量,其中 $i=1, 2, \dots, n$; 对新样本点 $Y=\{y_1, y_2, \dots, y_n\}$ 使用核 K -均值聚类。

本文算法流程如图 3 所示。

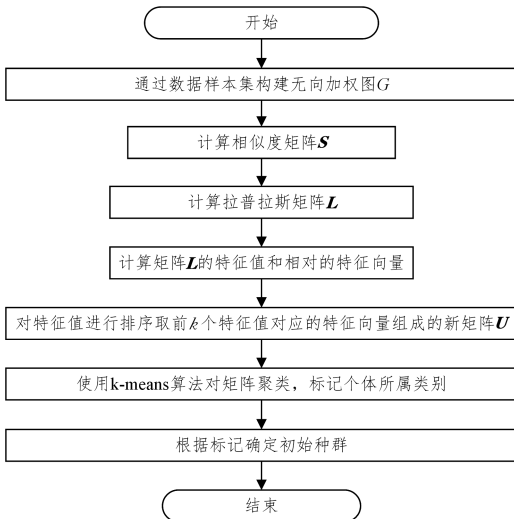


图 3 谱聚类算法划分种群的流程图

Fig. 3 Flow chart of spectrum clustering algorithm dividing population

3.2 数据归减策略

现有社区发现的多目标进化算法,在种群进化到一定程度时,个体之间会出现类似的属性,且在之后的进化中这些个

体所在的社区结构几乎没有变化,将这些社区通过 3 个个体投票的方法找到局部相同社区,同时将找到的相同社团合并为一个。以 1 个例子来说明利用网格约简对数据进行归减的过程。假设 X 包括 4 个社区 $\{1, 2, 3, 4\}, \{5\}, \{6\}, \{7\}$, Y 包括 5 个社区 $\{1, 2, 3\}, \{4\}, \{5\}, \{6\}, \{7\}$, 而社区 $\{1, 2, 3\}$ 在 X, Y 中都被表示在一个社区,因此 $\{1, 2, 3\}$ 被看作同一本地社区。对于社区 $\{4\}$, 由于社区中只包含一个节点,所以其也被认为是一个本地社区。社区 $\{5\}, \{6\}, \{7\}$ 也是如此,这样就得到 5 个本地社区 $\{1, 2, 3\}, \{4\}, \{5\}, \{6\}, \{7\}$, 从而将网络缩减为 1 个包含 5 个节点 $\{1, 4, 5, 6, 7\}$ 的网络,如图 4 所示。其中将本地社区 $\{1, 2, 3\}$ 合并为一个节点 $1'$ 减少了网络的规模。将合并基因保留,使得网络的信息不会丢失。上述的数据归减是建立在进化过程中精英个体所发现的本地社区。

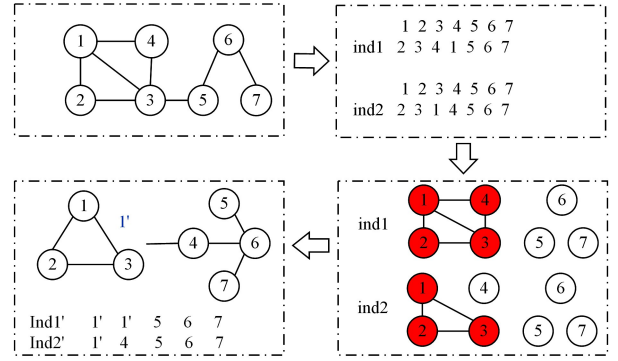


图 4 进化中归减数据的过程

Fig. 4 Evolutionary data reduction process

算法 2 进化中的归减

输入:进化中的 G 网络,随机选取个体 P'

输出:合并后的 G_R 网络

步骤 1 Num 表示 P' 中个体的数目

步骤 2 $[\text{Max_ind}, \text{Max_cum}]$

步骤 3 令 $1 \leq i \leq \text{Max_cum}$, C 为个体 Max_ind 的第 i 个社团

步骤 4 令 $1 \leq K \leq \text{Num}$

步骤 5 如果 P_k' 中有 C 的子集,合并 G 里面 C 包含的子图,在符合条件时循环上述步骤

步骤 6 G_R 为 G 合并后得到的新网络

3.3 SMOEA 算法的整体流程

先用谱聚类算法对编码后的复杂网络数据集进行初始种群划分,然后在 MOEA 算法模型中进行初始种群进化,进化过程中使用网格约简对种群归减。最后得到优化后的解集,对其进行解码得到最终结果。

算法 3 SMOEA

输入: $G, \text{gene}, \text{pop}, k, \text{gene}$ 为迭代数, pop 为种群规模, k 为进化中归减次数

输出:最终解集

步骤 1 使用谱聚类对数据进行划分,并且得到初始种群(见算法 1);

步骤 2 粒子群初始化: $P=(X_1, X_2, \dots, X_{\text{pop}})$;

步骤 3 速度初始化: $V=(V_1, V_2, \dots, V_{\text{pop}})$;

步骤 4 参考点初始化: Z^* ;

步骤 5 Pbest 初始化: $P_{\text{best}}=(X_1, X_2, \dots, X_{\text{pop}})$;

步骤 6 $N=(N_1, \dots, N_m, \dots, N_{\text{pop}})$, 通过权重向量计算每个个体的邻域个体;

步骤 7 种群进化: $G_T=G/G_T$ 为进化中变动的归减网络

for $i=1$ to gene do

步骤 8 if $i \% (\lceil \text{gene} / T \rceil) = 0$ do //迭代次数分割;
 步骤 9 $G_T = (G_T, P')$ //进化中的归减(见算法 2);
 步骤 10 for $j=1$ to pop do 随机从个体 j 的邻域 n_j 中选择一个粒子作为全局最优粒子;
 步骤 11 粒子因其速度而更新;
 步骤 12 计算目标函数值,若粒子在其邻域的粒子方向上的值低,
 步骤 13 则替代它,更新参考点 Z^* ; 如能支配 P_{best_i} , 则更新 P_{best_i} ;
 步骤 14 循环迭代选取最优,最终输出最优解集。

3.4 SMOEA 时间复杂度分析

假设在复杂网络 $G=(V, E)$ 中, n 为网络节点集 V 中的节点个数, 边的数量 E 为 m , pop 表示种群个体的数量, gene 是迭代数。从算法 1 中可以发现, 在谱聚类划分初始种群部分, 时间复杂度为 $O(n^3)$, 而在进化过程中, 因数据归减的方式, 种群个体无规律归纳约减, 时间复杂度约为 $O((m+n) * \text{gene} * \text{pop})$, 因此 SMOEA 算法的时间复杂度为 $N(O(n^3) + O((m+n) * \text{gene} * \text{pop})) \approx N(O(n^3))$ 。其比同类型 MR-MOEA 算法的时间复杂度要小, 主要原因是在进化过程中的数据归减对数据进行缩减, 使得计算效率得以提升。

4 实验结果和分析

在本文的算法之上针对节点数在 34~27000 之间的各种规模网络、不同混合参数情况下的 LFR benchmark 人工网络^[18] 和 9 个真实网络进行社区发现, 并选择 MRMOEA, RMOEA, MCMOEA 等以标准互信息和模块性作为优化目标的算法进行性能比较。

4.1 评价指标

用网络选择标准互信息 (NMI)^[19] 与扩展模块度函数 (Q_{ov})^[20] 评价算法聚类结果和重叠社区发现的精确性。

针对一个有 n 个节点的集合, NMI 和 Q_{ov} 的定义如下:

$$NMI = \frac{2 \sum_i \sum_j n_{ij} \text{lb}(\frac{n_{ij} n}{b_i d_j})}{-\sum_i b_i \text{lb}(\frac{b_i}{n}) - \sum_j d_j \text{lb}(\frac{d_j}{n})} \quad (6)$$

其中, V 代表集合, n 代表集合节点个数, $C = \{C_1, C_2, \dots, C_k\}$ 表示一个网络的划分结果, $P = \{P_1, P_2, \dots, P_{k'}\}$ 集合 C_i 和 P_j ($1 \leq i \leq k, 1 \leq j \leq k'$), $n_{ij} = |C_i \cap P_j|$, $b_i = \sum_{j=1}^{k'} n_{ij}$, $d_j = \sum_{i=1}^k n_{ij}$ 。

$$Q_{\text{ov}} = \frac{1}{2m} \sum_i \sum_{v \in C_i, w \in C_i} \frac{1}{O_v O_w} [A_{vw} - \frac{k_v k_w}{2m}] \quad (7)$$

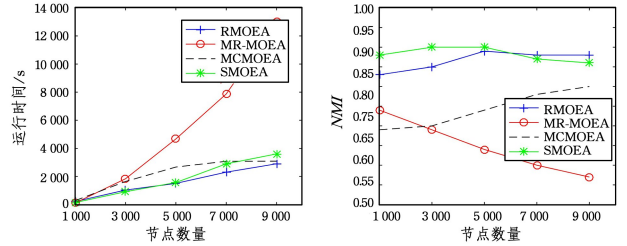
其中, A_{vw} 是邻接矩阵中的元素, C_i 表示网络划分中 i 社区的点集合, m 为网络的边数, O_v 为节点 V 所属的社区数, k_v 为节点 V 的度数。

4.2 人工网络上的实验分析与对比

人工网络实验采用在社区发现算法性能检测中普遍使用的 LFR 基准实验。LFR 基准网络的定义如下: $LFR(N, d, d_{\text{max}}, \mu, t_1, t_2, s_{\text{max}}, s_{\text{min}}, c_n, c_m)$, 其中 N 为节点数目, j 为节点度数的均值, j_{max} 为节点的最大度数, t_1, t_2 是节点度数指数和社区大小出现频率的指数, $s_{\text{max}}, s_{\text{min}}$ 为网络中最大和最小的社区规模, c_n, c_m 为节点的重叠数目与重叠节点所属的社区数, μ 为网络混合参数。 μ 的值越大, 网络的社团结构就越模糊, 反之社团结构越明显^[21]。

该测试先将各种算法的计算时间与社区发现的准确性进行比较。其他参数设置已固定 ($j=20, j_{\text{max}}=50, \mu=0.6, t_1=$

$2, t_2=1, s_{\text{max}}=100, s_{\text{min}}=20, c_n=0, c_m=1$)。图 5(a) 横坐标表示节点数为 1000 的人工网络, 节点数从 2000 加到 9000; 纵坐标是 4 组算法的计算时间。图 5(b) 表示起始节点 1000 个, 每次增加 1000 个, 直到 9000 个节点, 纵坐标是 4 种算法的 NMI 值。从图 5 显示的实验结果来看: SMOEA 算法在节点变多后, 依旧拥有较短的计算时间和较好 NMI 值, 在节点数增加时, SMOEA 算法展现出了更大的优势; 在 $N < 4000$ 时, SMOEA 算法在计算效率和 NMI 值上均优于其他 3 组算法。

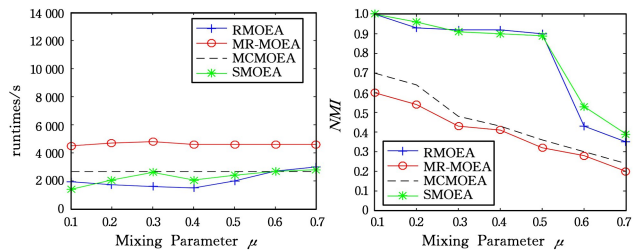


(a) RMOEA, MR-MOEA, MCMOEA, SMOEA 的运行时间随数据集大小的变化情况

(b) RMOEA, MR-MOEA, MCMOEA, SMOEA 的 NMI 精度随数据集大小的变化情况

图 5 4 种算法在不同节点数目的网络中运行时间和 NMI 值的对比
 Fig. 5 Comparison of running time and NMI accuracy of four algorithms in networks with different number of nodes

其次, 测试社区结构逐渐模糊时对每种算法的计算时间与社区发现准确性进行比较, 并设置固定参数 ($N_j=20, j_{\text{max}}=50, t_1=2, t_2=1, s_{\text{max}}=100, s_{\text{min}}=20, c_n=0, c_m=1$)。图 6(a) 横坐标为混参 μ 从 0.1 每次加 0.1 直到 0.7, 纵坐标是 4 组算法的计算时间; 图 6(b) 横坐标为混参 μ 从 0.1 增加到 0.7, 纵坐标是 4 组算法的 NMI 值。从实验对比结果来看: SMOEA 算法的运行时间平均最短; SMOEA 算法的 NMI 值和 RMOEA 算法持平, 比其他两个算法有着较大的优势。同时在 $\mu < 0.6$ 时 SMOEA 算法的时间与 RMOEA 相似, 比其他两个算法效率高, SMOEA 算法在 NMI 值上表现得更好。



(a) RMOEA, MR-MOEA, MCMOEA, SMOEA 的运行时间随混参 μ 大小的变化情况

(b) RMOEA, MR-MOEA, MCMOEA, SMOEA 的运行时间随混参 μ 大小的变化情况

图 6 4 种算法在不同混参时运行时间和 NMI 值的对比

Fig. 6 Comparison of running time and NMI precision of four algorithms under different mixing parameters

4.3 真实网络上的实验分析与对比

为了进一步验证 SMOEA 算法的性能, 本节使用 9 个常用的真实世界网络, 分别是 soc-epinions, soc-gplus, soc-anybeat^[22], football^[23], polbooks^[18], dolphin^[24], karate^[25], jazz^[26], netscience^[27]。这些网络的信息如表 1 所列。

表1 真实网络数据集的描述

Table 1 Description of real network datasets

data set	Number of nodes	Number of edges	Number of real communities
soc-epinions	27 000	100 000	16
soc-gplus	24 000	392 000	
soc-anybeat	13 000	8 000	
football	115	613	12
polbooks	105	441	3
dolphin	62	159	2
karate	34	78	2
jazz	198	2 742	
netscience	1 589	2 742	

同时选取 3 种多目标社区发现算法 MCMOE^[10], RMOEA^[11], MRMOEA^[9], 分别适用于大、中、小规模的网络。对于上述对比算法,为了算法的公平性,所有算法种群大小设置为 $pop=100$,迭代次数设置为 $Gene=100$,领域大小设置为 $ns=40$ 。其余参数设置采用各自文献中最好参数。而 SMOEA 算法中 $t_1, t_2, s_{max}, s_{min}, c_n, c_m$ 参数设置参考 4.2 节。

先验证真实网络上的计算效率和社区发现准确性。我们使用 Qov 来评估社区发现算法。4 种算法在 9 个真实网络上运行 10 次,为了验证的充分性,取 10 次社区发现精度的平均值,结果如表 2 所列。可以看出,SMOEA 算法在大多数网络上都能取得很好的 Qov 值,并且在时间上也表现出了较好的计算效率。MR-MOEA 算法在 soc-epinions, soc-gplus, soc-anybeat 3 个大型网络中花费时间过长无法得到实验结果。NMI 值越大说明算法的性能越好,反之则说明性能越差。从表 1 可以看出只有 soc-epinions, football, polbooks, dolphin, karate 5 个数据集具有真实划分,因此可以在 NMI 精度测试中使用,而不具有真实网络划分的数据集无法使用 NMI 精度指标进行测试。如表 3 所列,SMOEA 算法在大多数数据集上比其他算法效果更好。

表2 真实世界网络中 Qov 实验结果Table 2 Qov experiment results of real-world networks

Data Set	RMOEA	MR-MOEA	MCMOE	SMOEA
soc-epinions	0.623	—	0.642	0.655
soc-gplus	0.620	—	0.723	0.705
soc-anybeat	0.525	—	0.511	0.521
football	0.303	0.304	0.301	0.381
polbooks	0.260	0.273	0.233	0.266
dolphin	0.260	0.260	0.235	0.281
karate	0.220	0.230	0.201	0.233
jazz	0.2180	0.203	0.329	0.334
netscience	0.460	0.312	0.492	0.495

表3 具有真实划分的真实世界网络中 NMI 实验结果

Table 3 Real-world network NMI experimental results with real division

Data Set	RMOEA	MR-MOEA	MCMOE	SMOEA
soc-epinions	0.623	—	0.642	0.655
soc-gplus	0.620	—	0.723	0.705
soc-anybeat	0.525	—	0.511	0.521
football	0.303	0.304	0.301	0.381
polbooks	0.260	0.273	0.233	0.266
dolphin	0.260	0.260	0.235	0.281
karate	0.220	0.230	0.201	0.233
jazz	0.218	0.203	0.329	0.334
netscience	0.460	0.312	0.492	0.495

通过人工网络及真实网络的实验可知,在复杂网络结构

越模糊的情况下,SMOEA 算法的表现比其他 3 个算法更好,并且计算效率也不输于处理大规模复杂网络的算法,这也说明了 SMOEA 算法在保证处理大规模数据集的同时,又具有更好的稳定性。这主要是因为 SMOEA 采用谱聚类算法进行数据处理,通过子图聚类的方式能得到更好的初始种群。结果表明 SMOEA 算法在算法效率和算法性能上都有着较好的表现。

结束语 本文提出一种融合谱聚类的多目标社区发现算法,使用谱聚类算法对数据集进行预处理,并在编码后利用多目标粒子群算法对数据进行归减,对精英粒子进行筛选,从而得到更优的解集,通过解码得到社区检测结果。

本文算法与 RMOEA, MR-MOEA, MCMOE 等多目标粒子群社区发现算法在不同参数下 LFR benchmark 人工网络数据集及真实网络上的实验对比分析结果表明:SMOEA 算法因为使用子图聚类的方式对数据集进行预处理并在计算中采用数据归减,所以在结构模糊的复杂网络中有着较好的表现,同时在较大规模复杂网络中也有着不错的计算效率。

参考文献

- [1] PALLA G. CFinder: locating cliques and overlapping modules in biological networks[J]. *Bioinformatics*, 2006, 22(8): 1021-1023.
- [2] 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.
- [3] PASTOR-SATORRAS R, VESPIGNANI A. Evolution and Structure of the Internet: A Statistical Physics Approach[M]// *Evolution and Structure of the Internet*. 2004.
- [4] WASSERMAN S. Social network analysis: Methods and applications[M]// Cambridge University Press, 1994: 181-190.
- [5] YANG B, LIU D Y, JIN D, et al. Complex network clustering method [J]. *Journal of Software*, 2009, 20(1): 54-66.
- [6] PIZZUTI C. A Multi-objective Genetic Algorithm for Community Detection in Networks[C]// *IEEE International Conference on Tools with Artificial Intelligence*. 2009.
- [7] RAHIMI S, ABDOLLAHPOURI A, MORADI P. A multi-objective particle swarm optimization algorithm for community detection in complex networks[J]. *Swarm and Evolutionary Computation*, 2018, 4(39): 297-309.
- [8] 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*, 2017, 44(12): 2274-2287.
- [9] ZHANG L, PAN H, SU Y, et al. A Mixed Representation-Based Multiobjective Evolutionary Algorithm for Overlapping Community Detection[J]. *IEEE Trans Cybern*, 2017, PP(99): 1-14.
- [10] WEN X, CHEN W N, LINY, et al. A Maximal Clique Based Multiobjective Evolutionary Algorithm for Overlapping Community Detection[J]. *IEEE Transactions on Evolutionary Computation*, 2016, PP(99): 1-1.
- [11] ZHANG X, ZHOU K, PAN H, et al. A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community Detection in Large-Scale Complex Networks[J]. *IEEE Transactions on Cybernetics*, 2018, PP(99): 1-14.
- [12] CORNE D W, JERRAM N R, KNOWLES J D, et al. PESA-II:

- region-based selection in evolutionary multiobjective optimization[C]// Conference on Genetic & Evolutionary Computation, 2001.
- [13] ZHANG S, LIU W Q, ZHAO N. Research of Consensus in Multi-agent Systems on Complex Network [J]. Journal of Frontiers of Computer Science, 2019, 46(4): 101-105.
- [14] ANGELINI L, BOCCALETTI S, MARINAZZO D, et al. Identification of network modules by optimization of ratio association [J]. Chaos, 2007, 17(2): 175.
- [15] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.
- [16] ZHOU A, QU B Y, LI H, et al. Multiobjective evolutionary algorithms: A survey of the state of the art [J]. Swarm & Evolutionary Computation, 2011, 1(1): 32-49.
- [17] LI J Y, ZHOU J G, GUAN J H, et al. Research progress of spectral clustering algorithm [J]. Journal of Intelligent Systems, 2011, 6(5): 405-414.
- [18] NEWMAN M E J. Modularity and community structure in networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(23): 8577-8582.
- [19] LANCICHINETTI A, FORTUNATO S, KERTÉSZ J. Detecting the overlapping and hierarchical community structure in complex networks [J]. New Journal of Physics, 2009, 11(3): 19-44.
- [20] 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.
- [21] ZHANG X, ZHOU K, PAN H, et al. A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community Detection in Large-Scale Complex Networks [J]. IEEE Transactions on Cybernetics, 2018, 50(2): 703-716.
- [22] LANCICHINETTI A, FORTUNATO S, RADICCHIF. Benchmark graphs for testing community detection algorithms [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2008, 78(4 Pt 2): 046110.
- [23] DANON L, DÍAZGUILERA A, DUCH J, et al. Comparing community structure identification [M]// Research and Innovation. Policies and Strategies in the Age of Globalization, 2005.
- [24] LUSSEAU D. The emergent properties of a dolphin social network [J]. Proceedings Biological Sciences, 2003, 270(2): S186.
- [25] ZACHARY W W. An Information Flow Model for Conflict and Fission in Small Groups [J]. Journal of Anthropological Research, 1976, 33(4): 452-473.
- [26] GLEISER P M, DANON L. Community Structure in Jazz [J]. Advances in Complex Systems, 2003, 6(4): 565-573.
- [27] GONG M, CAI Q, CHEN X, et al. Complex Network Clustering by Multiobjective Discrete Particle Swarm Optimization Based on Decomposition [J]. IEEE Transactions on Evolutionary Computation, 2014, 18(1): 82-97.



DONG Ming-gang, born in 1977, Ph.D. His main research interests include intelligent computing, multi-objective optimization and machine learning.



JING Chao, born in 1983, Ph. D. His main research interests include intelligent computing, optimization and deep reinforcement learning.

(上接第 456 页)

- [6] MATATOV N, ROKACH L, MAIMON O. Privacy-preserving data mining: A feature set partitioning approach [J]. Information Sciences, 2010, 180(14): 2696-2720.
- [7] GAERTNER W. A primer in social choice theory: Revised edition [M]. Oxford University Press, 2009.
- [8] BRANDT F, CONITZER V, ENDRISS U, et al. Handbook of Computational Social Choice [R]. Cambridge University Press, 2016.
- [9] ARROW K J. Social choice and individual values [M]. Yale University Press, 1963.
- [10] KORNHAUSER L A, SAGER L G. Unpacking the court [J]. The Yale Law Journal, 1986, 96(1): 82-117.
- [11] LIST C, POLAK B. Introduction to judgment aggregation [J]. Journal of economic theory, 2010, 145(2): 441-466.
- [12] LANG J, PIGOZZI G, SLAVKOVIK M, et al. Judgment aggregation rules based on minimization [C]// Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge. ACM, 2011: 238-246.
- [13] LANG J, SLAVKOVIK M. How hard is it to compute majority-preserving judgment aggregation rules [C]// Proceedings of the Twenty-first European Conference on Artificial Intelligence. IOS Press, 2014: 501-506.
- [14] LANG J, PIGOZZI G, SLAVKOVIK M, et al. A partial taxonomy of judgment aggregation rules and their properties [J]. Social Choice and Welfare, 2017, 48(2): 327.
- [15] LANG J, MENGIN J, XIA L. Voting on multi-issue domains with conditionally lexicographic preferences [J]. Artificial Intelligence, 2018, 265: 18-44.
- [16] LI L, TANG X. A solution to the cold-start problem in recommender systems based on social choice theory [M]// Intelligent and Evolutionary Systems. Springer, Cham, 2016: 267-279.
- [17] LI L I, NIU B, TANG S X. Online marketing research based on social choice theory and cloud computing [C]// 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC). IEEE, 2015: 391-396.



LI Li, born in 1982, Ph.D, lecturer. Her research interests include modern logic and artificial intelligence.