计算机科学
COMPUTER SCIENCE

# 基于多目标粒子群优化的属性网络局部社区检测算法

周志强, 朱焱

引用本文

周志强, 朱焱. 基于多目标粒子群优化的属性网络局部社区检测算法[J]. 计算机科学, 2023, 50(6A): 220200015-6.

ZHOU Zhiqiang, ZHU Yan. Local Community Detection Algorithm for Attribute Networks Based on Multi-objective Particle Swarm Optimization [J]. Computer Science, 2023, 50(6A): 220200015-6.

---

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220200015

# 基于多目标粒子群优化的属性网络局部社区检测算法

周志强　朱　焱
西南交通大学计算机与人工智能学院　成都 611756
(zqzhou1997@163.com)

**摘　要**　社区结构是复杂网络中的重要特征,局部社区检测的目标是查询出包含一组种子节点的社区子图。传统的局部社区检测算法通常利用网络的拓扑结构进行社区查询,而忽略了网络中丰富的节点属性信息。针对现实中广泛存在的属性网络,提出了一种基于多目标粒子群优化的属性网络局部社区检测算法。首先根据节点与其多阶邻居之间的属性相似度构造属性关系边,并根据模体结构获取网络中的高阶信息得到拓扑关系边,然后基于种子节点使用随机游走算法对两种关系边采样得到备选节点集。在此基础上,通过多目标粒子群优化算法对备选节点集进行迭代筛选,得到拓扑结构紧密和节点属性同质的社区结构。在真实数据集上的实验结果表明,所提方法有效提升了局部社区检测的质量。

**关键词**:局部社区检测;属性网络;模体;多目标粒子群优化;信息熵
**中图法分类号**　TP391

# Local Community Detection Algorithm for Attribute Networks Based on Multi-objective Particle Swarm Optimization

ZHOU Zhiqiang and ZHU Yan
School of Computing and Artificial Intelligence,Southwest Jiaotong University,Chengdu 611756,China

**Abstract**　Community structure is an important feature in complex networks,and the goal of local community detection is to query a community subgraph containing a set of seed nodes. Traditional local community detection algorithms usually use the topology of the network for community query,ignoring the rich node attribute information in the network. A local community detection algorithm based on multi-objective particle swarm optimization is proposed for realistic and widespread attribute networks. Firstly,attribute relationship edges are constructed based on the attribute similarity between nodes and their multi-order neighbours,and topological relationship edges are obtained by weighting the network structure based on the motif information,followed by sampling the two relationship edges around the core nodes using a random walk algorithm to obtain alternative node sets. Based on this,the alternative node sets are iteratively filtered by a multi-objective particle swarm optimization algorithm to obtain a topologically tight and attribute-homogeneous community structure. Experimental results on real datasets show that the proposed method improves the performance of local community detection.

**Keywords**　Local community detection,Attribute networks,Motif,Multi-objective particle swarm optimization,Information entropy

## 1　引言

随着大数据时代的不断发展,为了更全面地表达网络中的信息,越来越多的现实世界关系被建模成属性网络[1]。属性网络将网络中参与者之间的联系抽象成拓扑关系,将参与者的描述信息表示为节点属性,例如,学术合作者网络中包含作者之间的合作关系,以及作者的研究领域、社交网络中包含用户之间的好友关系和用户的兴趣等。

近年来,局部社区检测算法开始被人们所关注。对于给定的一组种子节点,局部社区检测的目的是找出包含这些节点的某个社区子图[2]。目前已经有不少局部社区检测算法被提出,如随机游走、节点扩散等。这些方法都属于结构感知型方法,其往往借助于网络的拓扑关系进行检测,得到结构合理的社区,即内部连接紧密、外部连接稀疏。然而,这类方法忽略了社区形成的另一个重要因素,即节点的属性相似性。在大多数现实网络中,属性相似的节点更容易形成社区[3],如具有相同研究领域的作者、兴趣相似的用户更可能处于同一社区。由此诞生了属性感知型社区检测方法,这类方法通过节点的属性向量对节点进行聚类,典型的是基于节点属性相似度的 $k$-means 算法。可见拓扑结构和属性信息对于复杂网络社区结构的检测同样重要,属性网络的拓扑关系和节点属性往往能为社区检测提供互补信息,如属性信息可以弥补网络结构稀疏的不足,而结构信息可以解决节点属性的噪声问题。如何有效地融合两者进行

基金项目:四川省科技计划项目(2019YFSY0032)
This work was supported by Sichuan Provincial Science and Technology Plan Project(2019YFSY0032).
通信作者:朱焱(yzhu@swjtu.edu.cn)

社区检测也越来越受到研究者的重视。

基于此，为了获得拓扑结构紧密、节点属性同质的社区结构，本文提出了一种基于多目标粒子群优化的属性网络局部社区检测算法（Local Community Detection algorithm for attribute networks based on multi-objective Particle Swarm Optimization，LCDPSO）。LCDPSO 分别捕获网络中的属性和拓扑信息，首先根据节点之间的属性相似信息构造属性关系边，同时基于高阶拓扑信息重构拓扑关系边；并基于两种边关系采样得到备选节点集；最后对备选节点集采用所提出的多目标粒子群优化算法进行拓扑结构和属性关系的优化，得到社区检测结果。在真实数据集上的实验结果表明，LCDPSO 具有良好的局部社区检测性能。

## 2 相关工作

### 2.1 局部社区检测

传统的全局社区检测通常需要借助网络的全局信息进行图划分，对于大规模网络而言计算代价是高昂的。而局部社区检测只依赖于网络的局部信息[4]进行单个社区查询任务，具有更广阔的应用场景。随机游走是一种经典的局部社区检测算法，其思想是基于种子节点不断对周围节点进行游走访问，并根据 PageRank 算法计算每个节点与种子节点的相关性指标，最后利用最小化传导度的方法截取最相关的节点集作为社区。基于此提出了各种变种方案，例如，RWR 算法[5]提出了重启策略，将游走者的部分访问概率转移到种子节点，保证了游走不会偏离社区内部，其社区检测性能受到广泛认可；MRW 算法[6]通过游走者互相作用，使游走保留在社区内部，在多种子情况下取得了一定效果。上述算法虽然能有效地进行局部社区检测，但仅考虑了网络的原始拓扑结构，忽略了网络的高阶拓扑信息以及丰富的节点属性信息。

### 2.2 属性网络社区检测

模体结构在许多研究领域中被认为是重要的高阶网络拓扑特征[7]，挖掘模体结构能够有效地发现复杂网络中潜在的结构信息。Hao 等[8]提出了一种基于模体结构的随机游走算法，并设计了模体传导度作为社区质量指标进行局部社区检测。EdMot 算法[9]根据模体结构对网络进行重构，去除不隶属于模体的边，将网络划分为多个连通子图，再利用 Louvain 算法进行社区划分。Zhao 等[10]提出了一种融合多种模体结构的异构网络元路径增强方法，通过统计不同模体结构形成的邻接矩阵并进行线性加权得到最终的边关系。

节点属性作为网络的重要组成部分，同样与社区结构有着密切关系。PWMN 算法[11]将网络拓扑结构和节点属性相似度进行线性融合计算，保存权值大于阈值的边关系，从而将属性网络转化为非属性加权网络，使其适用于传统社区检测算法。Xu 等[12]提出了基于矩阵分解的属性网络嵌入方法，分别计算节点之间的拓扑相似度矩阵和属性接近度矩阵，通过矩阵分解得到节点嵌入，用于社区检测。这类早期融合方法通常需要计算节点之间的相似度矩阵，图重构的时间复杂度高，不适用于大规模网络。

### 2.3 多目标优化算法

相较于单目标优化的局部社区检测算法，多目标优化算法针对多个社区指标进行优化，以获得更合理的社区检测结果。近年来提出了很多基于进化理论的多目标社区检测算法，例如：Zhang 等[13]选取反比率关联和割边比作为目标函数，基于五行环模型对元素进行更新得到社区结构。

粒子群算法是一种高效的多目标优化算法，其思想是在个体最优解和群体最优解的共同作用下，通过粒子的移动找出最优解，具有精度高、收敛快的特点。MOPSO-Net 算法[14]在粒子群优化的基础上，重新定义了粒子移动策略，通过对核 K-均值和割边比两个目标函数进行优化，得到社区检测结果。这些算法虽然都通过多目标优化方法解决了单一度量值下社区检测不稳定的问题，但往往都是针对社区拓扑结构进行优化，很少有研究者将其应用到属性网络中，对社区的节点属性同质性和拓扑结构紧密性同时进行优化。

## 3 基于多目标粒子群优化的属性网络局部社区检测算法

### 3.1 融合节点属性与拓扑关系的节点采样

为了充分利用复杂网络中的拓扑结构和节点属性信息，本文分别提出了基于多阶邻居的属性关系边构造方法以及基于模体结构的拓扑关系边加权方法来重构原始网络图。

#### 3.1.1 基于多阶邻居的属性边构造

传统的属性关系构造方法需要计算节点之间的相似度矩阵，计算开销大且结果与拓扑关系难以匹配。通常认为，属性相似的节点往往在网络结构中存在直接或间接联系。LCDPSO 借助于网络的局部拓扑信息，根据节点与其多阶邻居之间的属性向量相似度构造基于属性的边（简称属性边）。具体步骤为：对于某个节点 $i$，初始时将其一阶邻居节点加入邻居队列中，接着不断从队列中取出队首元素，记作节点 $j$，并计算与节点 $i$ 的属性相似度 $SA_{ij}$。若 $SA_{ij}$ 大于属性相似度阈值，则对节点 $i$ 和 $j$ 构造权值为 $SA_{ij}$ 的属性边，随后将 $j$ 的邻居节点加入邻居队列中；否则不考虑节点 $j$ 及其邻居节点。此外可以为算法设置最大遍历步长保证执行效率。节点 $i$ 和 $j$ 的属性相似度计算采用余弦相似度，可表示为：

$$SA_{ij}=\frac{\langle A_i,A_j\rangle}{\|A_i\|\times\|A_j\|} \tag{1}$$

其中，$A_i$ 表示节点 $i$ 的属性向量，$\|A_i\|$ 表示 $A_i$ 的模，$\langle A_i,A_j\rangle$ 表示 $A_i$ 和 $A_j$ 向量内积。

与传统的相似度矩阵构造相比，基于多阶邻居的属性边构造与拓扑关系的匹配度更高，更加符合拓扑紧密、属性同质的社区检测要求。在如图 1 所示的网络图中，以节点 $A$ 为例，节点 $B$ 和 $D$ 作为与节点 $A$ 属性相似的一阶邻居，首先与节点 $A$ 建立属性边，由此找出属性相似的二阶邻居节点 $C$ 并建立属性边；而邻居节点 $E$ 由于与节点 $A$ 属性相似度较低，不会建立属性边，且不需要进一步遍历节点 $F$ 和 $G$。此外，以节点 $C$ 为例，虽然其与节点 $G$ 的属性相似度较高，但由于拓

扑关联性低，不会构造属性边。

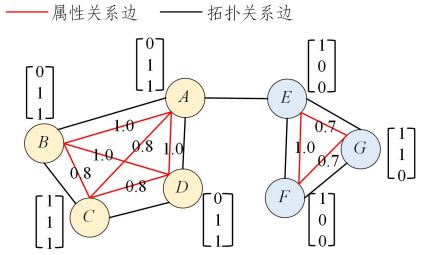



图 1　属性关系边构造

Fig. 1　Attribute relationship edge construction

### 3.1.2　基于模体结构的拓扑边加权

模体结构信息可以有效地表示网络中的高阶拓扑信息，被广泛应用于子图挖掘领域。针对传统社区检测算法缺乏高阶信息的问题，LCDPSO 挖掘了网络中的三角形模体，并融入了拓扑关系边。

三角形模体是以 3 个节点组成的子图，对表达复杂网络信息有很大帮助。如图 2 所示，3 种三角形模体可以分别表示 3 种网络结构的潜在特征，例如，$M_1$ 可以表达社交网络中的共同好友关系，$M_2$ 可以表达互联网网页之间的循环链接关系，$M_3$ 可以表达引文网络之间的相互引用关系。以好友关系为例，如果两个用户的共同好友越多，那么他们之间的联系越紧密，这种模体信息对于社区的形成是有帮助的。为了表达潜在的高阶信息，LCDPSO 计算每条拓扑边所隶属的三角形模体数，采用线性加权的方式将其融入拓扑边中，拓扑边权值计算可表示为：

$$ST_{ij}=\alpha\cdot W_{ij}+(1-\alpha)\cdot N_{ij}\cdot W_{ij} \tag{2}$$

其中，$W_{ij}$ 和 $N_{ij}$ 分别代表节点 $i$ 和 $j$ 的原始边权值和隶属的三角形模体数。$\alpha$ 作为平衡原始权值和模体权值的系数，$\alpha=0$ 表示只使用模体权值，$\alpha=1$ 表示只使用原始权值。在真实网络的社区检测中，$\alpha$ 取值较小时存在不隶属于模体的边不能被充分游走的问题，取值过大则会导致模体信息不能被有效利用，因此一般取 $\alpha=0.5$ 对两者进行平衡。融合模体权值的构造方法使得相关性高的节点之间拥有高权值的边，易于被划分在同一个社区。

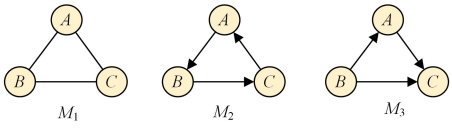



图 2　三角形模体示例

Fig. 2　Example of triangular motifs

## 3.2　基于多关系的随机游走采样

LCDPSO 利用随机游走策略分别对 3.1 节构成的属性边和拓扑边游走，并融合为备选节点集。传统的随机游走算法存在采样效果受种子节点选取影响的问题。为了解决这一问题，LCDPSO 提出了一种围绕核心节点的随机游走策略，其思想是：在每轮游走迭代完成后，将本轮采样到的访问概率处于高水平的节点加入核心节点集中；下一轮迭代中，游走会将一部分访问概率转移给核心节点。访问概率更新可表示为：

$$r^{(t+1)}=\beta\cdot P\cdot r^{(t)}+(1-\beta)\cdot c^{(t)}$$
$$P_{ij}=\frac{W_{ij}}{\sum\limits_{z\in N(i)}W_{iz}} \tag{3}$$

其中，$\beta$ 代表游走访问邻居节点的比例，$1-\beta$ 代表向核心节点转移的比例；$r^{(t)}$ 和 $c^{(t)}$ 分别代表第 $t$ 轮游走时所有节点和核心节点的访问概率；$N(i)$ 表示节点 $i$ 的一阶邻居节点；$\boldsymbol{P}$ 表示图 $G$ 的转移概率矩阵，元素 $P_{ij}$ 表示节点 $i$ 跳转到节点 $j$ 的概率值。这种游走策略能够保证即使种子节点处于社区边界，游走依然能围绕社区中心进行。此外，为了保证备选节点集的质量，LCDPSO 在随机游走中加入剪枝，避免了对无关节点的游走。基于随机游走的节点采样如算法 1 所示。

**算法 1**　基于多关系的随机游走节点采样

输入：重构图 $G_{new}$，种子集 Seeds

输出：目标社区的备选节点集 Opt

1. 将 Seeds 分别加入属性游走节点集 AttrWalkers 和拓扑游走节点集 TopoWalkers，并初始化访问概率 $r_A$，$r_T$；
2. 对于每个访问概率大于访问阈值的 walker，基于对应边关系进行游走并根据式(3)更新 $r_A$ 和 $r_T$，游走得到的新节点将加入对应 walkers 中；
3. 完成一轮迭代后，将本轮 $r_A$ 和 $r_T$ 综合排序最佳的节点加入核心节点集。重复步骤 2 和步骤 3，直到达到最大迭代次数；
4. 合并 AttrWalkers 和 TopoWalkers，过滤访问概率小于访问阈值的节点，得到备选节点集。

## 3.3　基于多目标粒子群优化的局部社区检测

### 3.3.1　适应度函数

在备选节点集的基础上，为了筛选出拓扑紧密、属性同质的社区节点，LCDPSO 将网络子图 $C$ 的属性信息熵 $H(C)$ 和拓扑传导度 $\phi(C)$ 作为多目标优化的适应度函数。属性信息熵可表示为：

$$H(C)=-\sum_{d=1}^{D}\frac{\phi(p_d)}{D\cdot\ln 2}$$
$$\phi(x)=x\ln x+(1-x)\ln(1-x) \tag{4}$$

其中，$D$ 代表节点属性向量的维度，$p_d$ 代表对于第 $d$ 维属性，具有该属性的节点在子图 $C$ 中的占比。$H(C)$ 取值介于 0～1 之间，子图 $C$ 中节点属性越相似，$H(C)$ 越小。描述子图 $C$ 拓扑结构的传导度 $\phi(C)$ 可表示为：

$$\phi(C)=\frac{cut(C)}{vol(C)} \tag{5}$$

其中，$vol(C)$ 代表子图 $C$ 的总边数，$cut(C)$ 代表子图 $C$ 中节点向外部连接的边数。$\phi(C)$ 取值介于 0～1 之间，子图 $C$ 结构越紧密，$\phi(C)$ 越小。

### 3.3.2　粒子设计

在粒子群优化算法中，粒子设计的优劣往往影响着算法的优化效果和收敛难度。粒子具有位置和速度两个基本属性，粒子的位置通常表示优化问题的解，即社区检测结果；粒子的速度用于指导粒子向最优解所在位置移动。LCDPSO 设计了一种离散化粒子表示法实现对备选节点集编码以表示社区检测结果。粒子 $i$ 的位置定义为：

$$X_i=\{x_1,x_2,x_3,\cdots,x_n\} \tag{6}$$

其中，$n$ 代表备选节点个数；$x_j$ 对应备选集中第 $j$ 个节点，$x_j$ 取值为 0 或 1，取值为 1 则代表将对应的备选节点纳入社区中，反之则不纳入社区。图 3 给出了一个简单的粒子示例，对于备选节点集 $\{A,B,C,D,E,F\}$，粒子向量 $[1,1,1,1,0,0]$ 可以表示包含节点 $\{A,B,C,D\}$ 的社区划分结果。

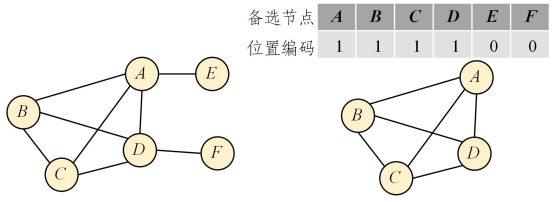



图 3 粒子位置编码示例图

Fig. 3 Illustration of particle position encoding

粒子位置的每一维取值决定了对应节点的取舍。作为控制粒子位移的速度属性，需要与粒子位置的维度对应，粒子 $i$ 的速度设计表示为：

$$V_i=\{v_1,v_2,v_3,\cdots,v_n\} \tag{7}$$

其中，$v_j$ 取值为 0 或 1，代表粒子位置的对应 $x_j$ 在下一次更新时是否发生改变。

LCDPSO 的计算瓶颈在于，在大规模、多属性的社区中，通过遍历节点集计算适应度函数造成的开销较大。为此，采用一种基于粒子迭代变化的增量式计算方法改进 LCDPSO，即基于上一轮迭代时的粒子位置、适应度函数值进行计算。这种方法需要为粒子缓存上一次迭代的状态，包括粒子坐标以及粒子代表的子图 $C$ 信息：属性前缀和 $p_d'$、拓扑信息 $vol'(C)$ 和 $cut'(C)$。在上一轮粒子的状态的基础上，对比粒子坐标编码得到粒子的增量变化，解码得到对应的增量节点并计算本轮的属性信息熵和拓扑传导度函数值。随着迭代的收敛，粒子位置只发生微小变化，增量式计算方法可以实现快速迭代。粒子的结构如图 4 所示。

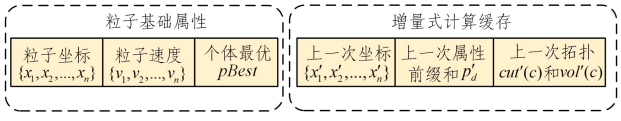



图 4 粒子结构设计

Fig. 4 Design of particle structures

### 3.3.3 粒子更新

LCDPSO 初始会产生 $k$ 个随机编码的粒子。每轮迭代中，计算种群中所有粒子的适应度函数值，适应度值越高，代表该位置越接近最优解，由此更新每个粒子的个体最优位置 $pBest$ 和整个种群的最佳位置 $gBest$。粒子根据 $pBest$ 和 $gBest$ 对自身速度进行更新，实现种群向最优解靠近，粒子 $i$ 的速度 $V_i$ 更新可表示为：

$$V_i=\omega V_i+c_1r_1(pBest_i\oplus X_i)+c_2r_2(gBest\oplus X_i)$$

$$V_{ij}\begin{cases}=1, & \text{if } \tanh(V_{ij})>\text{random}(0,1)\\ 0, & \text{if } \tanh(V_{ij})\leqslant\text{random}(0,1)\end{cases} \tag{8}$$

其中，$\omega$ 为惯性权重，$c_1$ 和 $c_2$ 是学习因子，$r_1$ 和 $r_2$ 是介于 0～1 之间的随机数，$\oplus$ 表示两个向量的异或操作，通过 tanh 函数进行取二值操作。这种方法使粒子具有向目标解靠近的趋势，且综合自身认知项和群体认知项的更新策略在保证收敛的同时也能防止陷入局部最优解。

粒子的位置更新取决于当前速度，$v_j$ 取值为 1 则 $x_j$ 进行取反操作；$v_j$ 为 0 则 $x_j$ 不变。以图 3 为例，若此时粒子速度为[0,0,0,1,1,0]，则粒子位置编码下一次更新为[1,1,1,0,1,0]。

LCDPSO 的粒子群优化方法如算法 2 所示。

**算法 2** 基于粒子群优化的局部社区检测

输入：原始图 $G_{old}$，备选节点集 Opt

输出：目标社区检测结果 Res

1. 初始化粒子群与参数设置；
2. for t←1 to MaxCycle do
3. 计算粒子的信息熵 H(C)和传导度 $\phi$(C)；
4. 更新粒子的个体最优值 pBest；
5. 更新种群的全局最优值 gBest 并存档；
6. 更新粒子的速度矢量 V 和位置矢量 X；
7. 将存档中相关性系数最高的解作为 Res

## 4 实验结果与分析

### 4.1 数据集

实验选取了 4 个真实属性网络数据集[3]，其中不仅包含节点属性和拓扑结构，还具备每个节点的真实社区标签。数据集的信息如表 1 所列。

表 1 数据集统计信息

Table 1 Datasets statistics

| datasets | Vertex | Edge | Cluster | Attr_Dim |
|---|---|---|---|---|
| Cora | 2708 | 5429 | 7 | 1433 |
| Citeseer | 3312 | 4732 | 6 | 3703 |
| WebKB | 877 | 1608 | 5 | 1703 |
| Sinanet | 3490 | 30282 | 10 | 10 |

### 4.2 属性边构造复杂度分析

为了验证 3.1 节讨论的基于多阶邻居的属性边构造法的有效性，基于 Cora 数据集设计了对比实验。首先，从数据集中随机选取指定数量的节点，分别对比了使用全局构造和多阶邻居(阶数=1,2,3)的属性边构造方案，结果如图 5 所示。

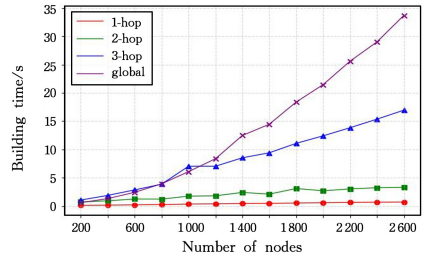



图 5 属性边构造的时间消耗

Fig. 5 Time consumption for attribute edge construction

在图 5 中，3 种基于多阶邻居的构造方案耗时与网络中的节点数基本呈线性关系，且斜率随着选取的邻居阶数增高而变大。对比基于邻接矩阵的全局构造和基于三阶邻居构造方法，当节点数为 1000 个时两者的时间消耗均在 7s 左右；当节点数达到 2600 个时，前者耗时为后者耗时的两倍。

### 4.3 局部社区检测性能分析

为了验证 LCDPSO 的局部社区检测性能，本文设计了与相关工作中介绍的 RWR[5]，MAPPR[8]，MEMP[10]，PWMA[11]这 4 种方法的对比实验。对于数据集中的每个社区，按照$\lceil|C|/10\rceil$的比例从社区中随机选取种子节点，分别使用上述算法在相同实验环境下完成社区检测任务，并采用常规的 F1-Score 作为评价指标。设 $T$ 表示目标社区 $C$ 中的真实节点集，$S$ 表示检测结果 $C'$ 的节点集，则 F1-Score 的计算式为：

$$F1\text{-}score=\frac{2*Precision*Recall}{Precision+Recall}$$
$$Precision=\frac{|S\bigcap T|}{|S|},Recall=\frac{|S\bigcap T|}{|T|} \tag{9}$$

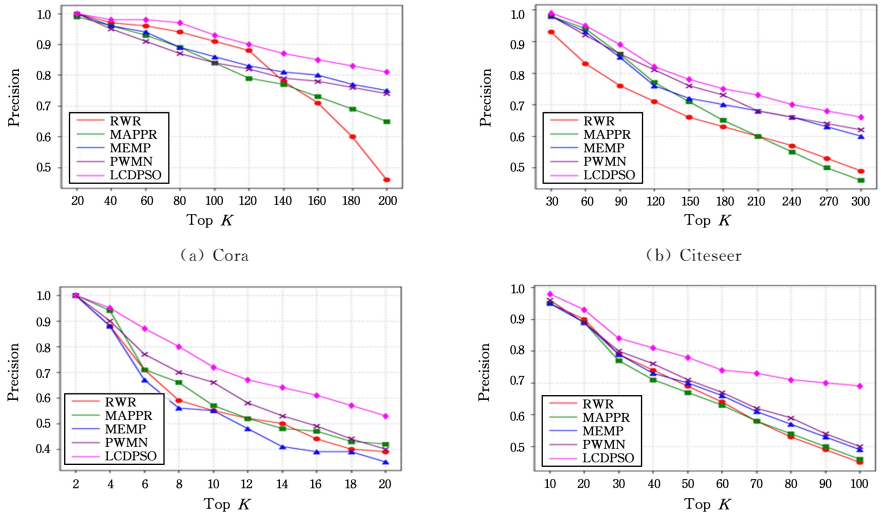
*F*1-*Score* 取值在 0～1 之间，数值越大则社区检测结果与真实社区越接近。最终评价指标为数据集中所有任务的平均 F1-Score，结果如表 2 所列。

表 2　不同算法的社区检测指标对比

Table 2　Comparison of community detection performance of different algorithms

| datasets | RWR | MAPPR | MEMP | PWMN | LCDPSO |
|---|---|---|---|---|---|
| Cora | 0.57 | 0.51 | 0.59 | 0.61 | **0.63** |
| Citeseer | 0.44 | 0.34 | 0.47 | 0.47 | **0.50** |
| WebKB | 0.30 | 0.27 | 0.32 | 0.33 | **0.41** |
| Sinanet | 0.20 | 0.22 | 0.26 | 0.34 | **0.52** |

由表 2 可知，本文的 LCDPSO 在所有数据集上的局部社区检测性能均优于对比方法，说明其在多种网络中具有较好的社区检测能力。Sinanet 数据集中的节点有密集的连边，导致基于拓扑结构的方法结果并不理想，而 LCDPSO 因多目标优化技术而取得了显著提升。为了分析不同算法结果差异的原因，对检测结果进行进一步实验分析。具体来说，将社区检测结果 $C'$ 按照节点的访问概率值进行排序，取 Top $K$ 节点集 $C_K'$，计算 $C_K'$ 中的节点属于目标社区 $C$ 的比例，即精确率。通过 $K$ 的取值变化，分析算法对于不同相关性节点的检测能力。在 4 个数据集上分别对不同算法进行实验对比，结果如图 6 所示。可见，$K$ 值较小时，所有算法的性能都处于高水平，说明所有算法对相关性高的节点具备有效的检测能力；而随着 $K$ 值的增大，所有算法的精确率都处于下降趋势，但 LCDPSO 的下降幅度较平缓，反映了 LCDPSO 仍能检测到许多弱相关性节点，这得益于围绕核心节点的游走策略和多目标优化技术。




(a) Cora　(b) Citeseer　(c) WebKB　(d) Sinanet

图 6　不同算法 Top $K$ 节点的精确率

Fig. 6　Precision of Top $K$ nodes in different algorithms

**结束语**　本文方法的两个技术贡献为：1）利用一种融合属性和拓扑关系随机游走采样方案，得到与目标社区相关的备选节点集；2）利用多目标粒子群优化算法对社区节点进行识别，得到属性同质且拓扑紧密的社区检测结果。将来的工作重点考虑依托分布式计算平台实现大规模图的并行社区检测，以及多关系网络中的社区检测任务。

## 参 考 文 献

[1] GASPARETTI F, SANSONETTI G. Community detection in social recommender systems: a survey[J]. Applied Intelligence, 2021, 51(6): 3975-3995.

[2] DILMAGHANI S E, MATTHIAS R. Community Detection in Complex Networks: A Survey on Local Approaches[J]. ACIIDS, 2021, 13: 757-767.

[3] CHUNAEV P. Community detection in node-attributed social networks: A survey[J]. Computer Science Review, 2020, 37: 100286.

[4] LUO W J, LU N, NI L, et al. Local community detection by the nearest nodes with greater centrality[J]. Inormation Sciences, 2020, 517: 377-392.

[5] TONG H, FALOUTSOS C, PAN J Y. Fast random walk with restart and its applications[C]// Proc. 6th Int. Conf. Data Mining, Hong Kong, China: IEEE, 2006: 613-622.

[6] BIAN Y C, NI J C, CHENG W, et al. Many Heads are Better than One: Local Community Detection by the Multi-walker Chain[C]// Proc. Int. Conf. Data Mining, New Orleans, USA: IEEE, 2017: 21-30.

[7] YU S, FENG Y F, ZHANG D, et al. Motif discovery in networks: A survey[J]. Computer Science Review, 2020, 37: 100267.

[8] YIN H, BENSON A R, LESKOVEC J, et al. Local higher-order graphclustering[C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 555-564.

[9] LI P Z, HUANG L, WANG C D, et al. Edmot: An edge enhancement approach for motif-aware community detection[C]// Proceedings ofthe 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 479-487.

[10] ZHAO H, ZHOU Y, SONG Y, et al. Motif enhanced recommendation over heterogeneous information network[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 2189-2192.

[11] ALINEZHAD E, TEIMOURPOUR B, SEPEHRI M M, et al. Community detection in attributed networks considering both structural and attribute similarities: two mathematical programming approaches[J]. Neural Computing & Applications, 2020, 32(8): 3203-3220.

[12] XU X, XIAO Y, LONG H, et al. Attributed Network Embedding Based on Matrix Factorization and Community Detection [J]. Computer Science, 2021, 48(12): 204-211.

[13] ZHANG Q, LIU M. Multi-objective Five-elements Cycle Optimization Algorithm for Complex Network Community Discovery [J]. Computer Science, 2020, 47(8): 284-290.

[14] 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, 39: 297-309.

**ZHOU Zhiqiang**, born in 1997, master candidate. His main research interests includes social network data mining and community detection.

**ZHU Yan**, born in 1965, Ph. D, professor, is a member of China Computer Federation. Her main research interests includes data mining and social network analysis and mining.