

# 属性网络中相似性增强随机游走的多社区搜索

李青青<sup>1</sup>, 马慧芳<sup>1,2</sup>, 李 举<sup>1</sup>, 李志欣<sup>3</sup>

(1. 西北师范大学计算机科学与工程学院, 甘肃兰州 730070; 2. 桂林电子科技大学广西可信软件重点实验室, 广西桂林 541004;  
3. 广西师范大学计算机科学与信息工程学院, 广西桂林 541004)

**摘要:** 社区搜索旨在寻找与给定查询节点高度相关的个性化社区. 现有社区搜索方法多面向简单网络且处理单个查询节点或假定多个查询节点来自同一社区, 这种严格的假设使得算法灵活性受限. 据此, 提出一种在属性网络中利用查询节点随机游走路径的相似性增强的多社区搜索方法, 可以有效地定位查询节点所属的多个局部目标社区. 具体地, 有效融合网络中高阶结构与属性信息, 利用重启随机游走计算各查询节点的重要性分数向量; 计算查询节点随机游走路径的相似性并设计一种相似性增强策略, 使得在无监督学习中相似路径游走者彼此增强关联从而定位不同查询节点所属的多个社区结构; 基于结合结构和属性的并行电导值精准查询社区. 真实数据集和人工数据集的实验验证了本文方法的有效性和效率.

**关键词:** 属性网络; 高阶结构; 相似性增强随机游走; 并行电导; 社区搜索

**中图分类号:** O157.5; TP301.6 **文献标识码:** A **文章编号:** 0372-2112(2021)11-2096-05

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20200422

## Multi-Community Search Using Similarity Enhanced Random Walk in Attributed Networks

LI Qing-qing<sup>1</sup>, MA Hui-fang<sup>1,2</sup>, LI Ju<sup>1</sup>, LI Zhi-xin<sup>3</sup>

(1. College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China;  
2. Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Science and Technology, Guilin, Guangxi 541004, China;  
3. College of Computer Science and Information Engineering, Guangxi Normal University, Guilin, Guangxi 541004, China)

**Abstract:** Community search aims to find personalized communities highly related to the given query nodes. Existing community search methods are applicable to simple networks, and always assume either a single query node is given or multiple query nodes are from the same community, which limits the flexibility of the algorithm. An attributed network oriented multi-community search method, which is designed via random walk path similarity enhancement of query nodes, is proposed to effectively locate multiple local communities that query node belongs. Attribute and high-order structure information in the network are effectively fused, and the importance score vector of each query node is calculated based on random walk with restart. The similarity between random walk paths of query nodes is calculated and the similarity enhancement strategy is designed to enhance the association of similar path walkers so as to locate multiple community structures of different query nodes. Based on the combination of structure and attribute, the parallel conductance is used to accurately find the community. The experiments on both real-world datasets and synthetic datasets verify the effectiveness and efficiency of the proposed method.

**Key words:** attributed networks; high-order structure; similarity-enhanced random walk; parallel conductance; community search

### 1 引言

作为一种通用并且灵活的数据结构, 图(网络)可

以有效地编码数据对象(节点)之间的复杂关系(边). 这些关联包含着对象之间的潜在结构和交互规律等. 此外, 网络中节点往往与大量属性相关联, 提供了丰富

收稿日期: 2020-05-06; 修回日期: 2021-04-07; 责任编辑: 李勇锋

基金项目: 国家自然科学基金(No.61762078, No.61363058, No.6196604); 广西多源信息挖掘与安全重点实验室开放基金(No.MIMS18-08); 西北师范大学 2019 年度青年教师科研能力提升计划重大项目(No.NWNU-LKQN2019-2); 甘肃省自然科学基金(No.21JR7RA114); 甘肃省教育厅优秀研究生“创新之星”项目(No.2021CX2X-270)

且与节点高度相关的辅助内容信息. 社区搜索本质是局部社区发现,旨在基于用户给定的查询节点搜索与其相关性较高的个性化社区<sup>[1]</sup>. 与关注全局图划分的社区发现<sup>[2]</sup>相比,社区搜索更关注于局部网络和用户的个性化需求,只对满足指定条件的社区感兴趣,在现实世界中可以广泛应用到电子商务、社交媒体等许多任务中<sup>[3,4]</sup>.

基于随机游走的社区搜索方法被证明有效且被广泛应用. 面向无属性网络(下文称简单网络)的基于随机游走的社区搜索核心思想是通过扫描排名节点,计算电导率最小的节点集作为局部社区返回. 如经典的 PRN (PageRank-Nibble)<sup>[5]</sup>,最近提出的 MRW (Memory-based Random Walk method)<sup>[6]</sup>. 然而以上方法均适于简单网络,且仅考虑了节点间直接的交互信息,忽略了高阶邻居对节点重要性的影响. 仅依赖图拓扑结构信息的方法一旦数据收集发生偏差,图拓扑结构信息可能会包含错误信息或缺失. 另一方面,属性网络中包含丰富的信息,这些信息通常会补充网络,充分利用属性信息可促使随机游走结果更为可靠. 例如 Hsu 等人<sup>[7]</sup>设计的 AttriRank 框架等. 尽管在属性网络中的工作中考虑了网络中的结构和属性信息,但仍依赖于查询节点来自同一社区且查找单个社区的严格假设,并均会因图拓扑结构的数据收集偏差影响结果的准确性. 此外,算法设计中没有考虑到节点高阶邻居对于随机游走重要性分数的影响,无法解决社区搜索方法中的查询偏向问题.

针对以上问题,本文从查询节点可来自多个社区的假设出发,设计了面向属性网络的相似性增强随机游走的多社区搜索方法 (Similarity-Enhanced Random Walk in Attribute networks, SERWA). 具体地,基于高阶邻居与属性信息重构概率转移矩阵;利用游走者游走路径相似性设计一种相似性增强策略来更新节点的重要性分数;采取结合结构和属性的并行电导精准查询社区. 该方法在简单网络中仍适用,因为其考虑了简单网络中节点的高阶邻居信息以作为节点的内部属性,使得游走者自动地探索网络时捕获到的社区更为精确. 在各种现实世界网络和人工网络上的大量实验结果表明,本文方法在很大程度上优于最新的社区搜索方法.

## 2 准备知识

### 2.1 问题定义

给定无向属性图  $G=(V, E, H)$ , 其中  $V=\{v_i\}_{i=1, \dots, n}$  表示节点集合,  $|V|=n$ ;  $E \subseteq V \times V$  表示边集且  $|E|=m$ .  $G$  的拓扑结构记作邻接矩阵  $A_{n \times n}$ , 若  $(v_i, v_j) \in E$ , 则  $A_{ij}=1$ ; 否则  $A_{ij}=0$ .  $h_i$  是网络中节点  $v_i$  所携带外部属性向量,  $f_i$  是节点  $v_i$  的综

合属性为  $y$  维的行向量, 图  $G$  节点的外部属性矩阵与综合属性矩阵分别表示为  $H_{n \times (y-3)}=[h_1; \dots; h_n]$ ,  $F_{n \times y}=[f_1; \dots; f_n]$ . 设  $P_{n \times n}$  是结构矩阵, 其中  $P_{ij}$  是节点  $v_i$  到  $v_j$  的转移概率, 即  $P_{ij}=A_{ij}/\sum_{k=1}^n A_{kj}$ . 设  $Z=\{z_1, z_2, \dots, z_k\}$  是用户给定的查询节点集合,  $Z \subseteq V$ . 算法的目标是找到查询节点  $Z$  所在的社区集合  $C=\{C_1, C_2, \dots, C_l\}$ ,  $l \leq k$ , 其中  $C_i$  表示特定社区. 社区被定义为网络中关联紧密的节点集合, 具有社区内部节点间边数相对较多而社区内部节点与社区外部节点间边的数量相对较少的特质. 此外, 社区内部的节点往往具有类似属性或在网络中扮演着相似的角色.

### 2.2 融合结构和属性的转移矩阵构建

随机游走受限于一阶马尔科夫假设, 即节点的排名仅仅取决于相邻节点得分, 并不考虑非相邻节点. 但是, 图中存在与节点重要性高度相关的其他信息, 如一个节点的 2 度邻居的总数, 或经过该节点的最短路径总数等, 均有助于目标节点的表示. 将这样复杂的信息建模到节点属性中可以产生更可信的排名. 本文将此属性定义为内部属性, 包括: 同配性(度/邻居节点度的平均值); 距离为 2 的节点数; 距离为 3 的节点数. 其中同配性衡量节点的重要性, 距离为 2 的邻居节点数表示节点的传递性, 距离为 3 的节点数表征节点在网络中的向外拓展能力. 根据以上要求提取内部属性, 再将节点内部属性矩阵  $M$  与外部属性矩阵  $H$  结合, 可得到最终的综合属性矩阵  $F$ . 本文选择径向基核函数计算节点间的属性相似度, 即  $s_{ij} = e^{-\eta \|f_i - f_j\|_2^2}$ , 其中  $s_{ij} = s_{ji}$ ,  $0 < s_{ij} \leq 1$ , 参数  $\eta$  用于控制属性距离之间的影响. 计算列归一化属性相似度矩阵  $Q$ , 其中第  $i$  行第  $j$  列元素  $Q_{ij} = s_{ij} / \sum_{k=1}^n s_{kj}$ . 则基于属性相似度矩阵  $Q$  和结构矩阵  $P$  计算转移矩阵  $R$  如下:

$$R=(1-d)Q+dP \quad (1)$$

参数  $d$  用来调节结构和属性的相对重要性.

## 3 属性网络中相似性增强随机游走的多社区搜索方法

### 3.1 属性网络中相似性增强的随机游走

由传统基于重启的随机游走可知, 查询节点以一定的概率跳转到自身或者与其结构相似的节点. 在相似性增强随机游走方法中不仅考虑网络的拓扑结构信息, 也研究了由网络结构信息而量化得到的节点内部属性, 以及节点本身所携带的外部属性. 如式(2)所示:

$$x_i^{(t+1)} = \alpha R x_i^{(t)} + (1-\alpha)q \quad (2)$$

其中  $t$  表示迭代为固定概率分布的时刻.  $\alpha$  是调节跳转到自身或者与其相似节点的跳转参数.  $q$  表示只有查询节点的位置为 1, 其余位置为 0 的列向量. 直观地,  $q$  表

示查询节点对步行者的影响, 游走者以  $(1-\alpha)$  概率受  $q$  影响并跳转到它. 不同于传统的基于重启的随机游走公式, 式(2)考虑了节点本身所携带的外部属性和节点  $v_i$  的高阶邻居信息.

### 3.2 相似排名向量的增强

如前所述, SERWA 方法通过考虑查询节点间相似的重要性分数向量来互相增强随机游走效果. 具体地, 如果节点  $v_i$  和节点  $v_j$  在同一个社区, 则其对应的重要性分数向量  $\mathbf{x}_i^{(t)}$  和  $\mathbf{x}_j^{(t)}$  之间应该有着高相似性. 否则, 重要性分数向量之间的相似性很低. 对于高度相似的重要性分数向量, 允许其进行相互加强. 设  $\hat{\mathbf{R}}^{(t)}$  为  $t$  时查询节点重要性分数向量的相似性增强矩阵, 初始时  $\hat{\mathbf{R}}^{(0)}=0$ , 当  $t>0$  时, 对于节点  $v_i \neq v_j$ , 可得式(3):

$$\hat{R}_{ij}^{(t)} = \begin{cases} \cos(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)}), & \cos(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)}) > \theta \\ 0, & \text{其他} \end{cases} \quad (3)$$

其中  $\cos(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)})$  是两个向量之间的余弦相似度. 来自不同社区的两个查询节点的重要性分数向量具有较小相似性, 因此加强时只考虑相似性值大于阈值  $\theta$  (例如,  $0 \leq \theta \leq 0.01$ ) 的重要性分数向量的相似性值.

为了保持重要性分数向量的随机性, 将  $\hat{\mathbf{R}}$  列归一化得到  $\tilde{\mathbf{R}}$  为:

$$\tilde{R}_{ij}^{(t)} = \begin{cases} \hat{R}_{ij}^{(t)} / \sum_{i=1}^k \hat{R}_{ij}^{(t)}, & \sum_{i=1}^k \hat{R}_{ij}^{(t)} > 0 \\ 0, & \text{其他} \end{cases} \quad (4)$$

通过增加来自其他查询节点的重要性分数向量的积极影响增强重要性分数向量  $\mathbf{x}_i^{(t)}$ , 由此得强化重要性分数向量如式(5):

$$\boldsymbol{\pi}_i^{(t)} = \begin{cases} (1-\gamma)\mathbf{x}_i^{(t)} + \gamma \sum_{j=1}^k \tilde{R}_{ji}^{(t)} \mathbf{x}_j^{(t)}, & \sum_{j=1}^k \tilde{R}_{ji}^{(t)} = 1 \\ \mathbf{x}_i^{(t)}, & \sum_{j=1}^k \tilde{R}_{ji}^{(t)} = 0 \end{cases} \quad (5)$$

$\gamma(0 \leq \gamma < 1)$  是操纵相似性增强偏向的参数, 如果  $\gamma=0$ , 则不存在任何节点的影响.  $t+1$  时刻可以基于式(2)计算  $\mathbf{x}_i^{(t+1)}$  并将  $\mathbf{x}_i^{(t)} \leftarrow \boldsymbol{\pi}_i^{(t)}$ ,  $\boldsymbol{\pi}_i^{(t)}$  是将相似重要性分数向量进行加强所得的强化重要性分数向量.

### 3.3 局部社区搜索

为了搜索局部社区, 将每个向量  $\mathbf{x}_i^{(t)}$  中概率不为 0 的位置存放于集合  $S(x_i) = \{a | \mathbf{x}_i^{(t)}(a) \neq 0\}$  中, 记  $N_x = |S(x_i)|$ . 对于  $S(x_i)$  集合中每一个节点的位置, 计算对应位置节点的重要性分数值与该节点度的比值并进行降序排列, 设最后排序得  $v_1, v_2, \dots, v_{N_x}$ . 定义扫描集合为该排序生成的任意长度系列集合  $S_j^x = \{v_1, v_2, \dots, v_j\}, j \in \{1, \dots, N_x\}$ . 扫描所有的扫描集合, 返回具有最小并行电导的扫描集合作为局部社区, 并行电导<sup>[8]</sup>计算公式如下:

$$\phi(p) = \min_{j \in \{1, \dots, N_x\}} \phi(S_j^x) = \frac{\sum_{j \in S_j^x} A_{ij} + Q_{ij}}{\sum_{j \in S_j^x} A_{ij} + Q_{ij}} \quad (6)$$

$\phi(S_j^x)$  是结合属性和结构的并行电导, 节点集合  $S_j^x \subseteq V$  的容积是  $\text{vol}(S_j^x) = \sum_{i \in S_j^x} d(i)$ ,  $d(i)$  表示节点  $v_i$  的度.  $\text{vol}(V) = 2m$ , 有时用  $\text{vol}(G)$  代替  $\text{vol}(V)$ . 将计算得  $\phi(p)$  的扫描集合作为局部社区搜索的结果返回.

## 4 实验结果与分析

### 4.1 数据集

#### 4.1.1 人工数据集

基于 Lancichinetti-Fortunato-Radicchi (LFR)<sup>[9]</sup> 生成了一个允许节点隶属于多个社区的属性网络<sup>[10]</sup>, 其具有 300 个节点和 2982 条边. 人工网络的主要参数设定为  $d=5, d_{\max}=50, c_{\min}=3, c_{\max}=10$ .

#### 4.1.2 真实数据集

选取了 3 个真实世界中的属性网络验证本文方法的有效性和效率. 具体地, Amazon 以产品间的共同购买关系为边构建关系网络, 节点表示产品, 节点的属性主要有标签、销售排名、详细产品分类等. DBLP 中节点表示作者, 边表示作者之间的合著关系, 以权重来表示作者之间合作的次数. Cora 中节点表示论文, 边表示论文之间的引用关系, 通过论文的关键字将其划分到机器学习的子科目中.

### 4.2 对比方法以及参数设置

为了验证 SERWA 的有效性, 选择以下四类对比方法. 第一, 为研究本文方法和其他方法的性能, 选择了两种简单图上的随机游走方法 MWC (Multi-Walker Chain)<sup>[11]</sup> 和 MRW. 第二, 为了对比 SERWA 方法与属性网络中社区搜索方法的性能, 选择了 LOCLU (Local Clustering Unimodality)<sup>[12]</sup>. 第三, 为了分析 SERWA 在多社区搜索任务中的有效性, 选取了 S-MLC (Sparse Multiple Local Community)<sup>[13]</sup>. 第四, 设计了本文方法的变体 SERWA-noAttri (忽略外部属性) 以及 SERWA-sc (仅能挖掘单一社区).

本文选用 F-score<sup>[14]</sup> 对所有方法的有效性进行评估. 在与本文方法进行比较时, 所有方法的参数设置与其原始文章中提出的最优设置一致. 此外, MWC 采用本文处理属性的方法, 将转移矩阵替换成本文设计的转移矩阵. SERWA 方法的参数统一设定为  $\alpha=0.2, \gamma=0.3, d=0.85, \eta=1/\gamma$ .

### 4.3 性能比较

#### 4.3.1 有效性

在本节中, 评估本文方法对于单社区搜索和多社

区搜索的有效性. 实验结果如表 1 下所示.

表 1 社区搜索结果的 F-score 比较

	方法	Amazon	DBLP	Cora	Synth 1
单社区	MWC	<b>0.919</b>	0.581	0.913	<b>0.927</b>
	LOCLU	0.745	0.767	0.861	0.882
	SERWA-sc	0.893	<b>0.904</b>	<b>0.914</b>	0.926
多社区	MRW	0.895	0.531	0.884	0.910
	S-MLC	0.693	0.631	0.910	0.911
	SERWA	<b>0.911</b>	<b>0.930</b>	<b>0.961</b>	<b>0.968</b>

选择每个数据集中的 10 个社区,并相应的从中取出 6 个查询节点,通过执行各种社区搜索算法 50 次,找到查询节点所属的社区并返回平均值.从单社区搜索方法实验结果可知,LOCLU 的 F-score 低于其他社区搜索方法.究其原因在于该方法在倾斜测试时更倾向于囊括所有具有指定属性且和查询节点相似的节点,这种策略使得找到社区内的所有节点变得更容易但也更易将其他社区的节点包括在内.MWC 由于采用了本文转移矩阵的构建方式,使得方法在 Amazon 数据集上的性能表现优于本文方法.SERWA-sc 不仅考虑到了网络中的复杂关系而且采用的随机游走方法有助于捕获查询节点所在的社区,相比多数对比方法性能表现较好.

通过表 1 多社区搜索方法的实验结果观察到,显然 SERWA 实现了最佳性能.原因在于 SERWA 方法进一步考虑了网络间的复杂拓扑结构信息和节点的属性信息,并通过游走者跳转的路径相似性加强了查询节点间的关联,结合了属性和结构的并行电导搜索到的社区与基准社区较为一致.S-MLC 的性能表现与查询节点的位置紧密相关,对于较小的数据集采样较为准确,故性能表现更佳.MRW 通过记录游走者的访问历史而解决了查询偏向问题,社区搜索结果相对 S-MLC 而言较为准确,但由于缺乏属性信息的辅助,使其性能表现不如 SERWA 方法.

4.3.2 效率评估

为评估查询节点数量与 SERWA 方法运行时间的关系,一方面观察查询节点数目对于效率的影响.另一方面比较多种方法与 SERWA 方法效率.

图 1 展示了 SERWA 方法不同查询节点数目的运行时间.其从每个数据集上选取 200 个社区,在选取的社区中随机选择 2、4、6 和 8 个查询节点以搜索社区.随着查询节点数目的增多,算法的运行时间与查询节点的数目大致成线性关系.这是由于查询节点的增加会使得算法在动态更新相似性增强矩阵  $\hat{R}$  时的计算量增大.

表 2 是从每个数据集上选取 10 个社区,再从其中选择 6 个查询节点以搜索社区并取得平均运行时间返回的实验结果,时间单位为秒.从表 2 可得:(1)在面向

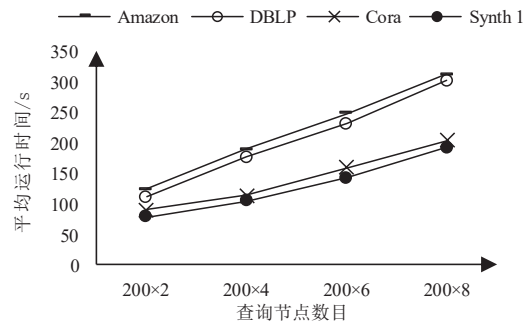


图 1 SERWA 方法的不同查询节点数量的平均运行时间

表 2 不同数据集上算法平均运行时间(s)

	方法	Amazon	DBLP	Cora	Synth 1
简单网络	MRW	92	91	72	67
	S-MLC	116	107	79	68
属性网络	LOCLU	147	135	101	92
	MWC	91	89	71	64
变体方法	SERWA-noAttri	79	70	49	47
本文方法	SERWA	91	83	69	64

简单网络的社区搜索方法中,SERWA-noAttri 相比其他方法的速度最快,这是由于该方法只考虑网络的结构信息且游走者之间的相互影响会使算法快速收敛到稳态分布.MRW 的加速策略虽可使得方法很快的收敛得到得分向量,但其需要考虑游走者的访问历史.而 S-MLC 需要通过迭代的进行稀疏非负矩阵分解来估计社区数量进而挖掘多社区,这使得相对其他简单网络中的社区搜索算法时间耗费较大.(2)在面向属性网络的社区搜索方法中,MWC 方法运行时间较短,源于多随机游走者可使得方法在较短时间内收敛.而 LOCLU 方法在指定属性下需要对网络中的所有节点进行倾斜测试以定位查询节点所在社区,这使得该方法在所有对比方法中的时间花费最大.(3)总体上,SERWA 方法的效率表现仅次于 SERWA-noAttri,这源于 SERWA-noAttri 忽略了外部属性信息的计算.此外,SERWA 方法虽与 MRW 以及 MWC 运行时间相当,但性能表现优于这两种方法.

5 总结

在面向属性网络的相似性增强随机游走多社区搜索方法中,基于查询节点可来自不同社区的假设,研究将查询节点高阶邻居信息以及外部属性信息融入到随机游走过程中,并设计了相似性增强随机游走方法,利用查询节点间游走者的交互来增强来自同一社区之间的联系.同时,采用结合属性和结构的并行电导返回多社区.通过在真实数据集和人工数据集上的多角度实验验证了本文方法优于经典以及先进的方法.

## 参考文献

- [1] 竺俊超, 王朝坤. 复杂条件下的社区搜索方法[J]. 软件学报, 2019, 30(3): 552 – 572.  
ZHU Jun-chao, WANG Chao-kun. Approaches to community search under complex conditions[J]. Journal of Software, 2019, 30(3): 552 – 572. (in Chinese)
- [2] 潘剑飞, 董一鸿, 陈华辉, 等. 基于结构紧密性的重叠社区发现算法[J]. 电子学报, 2019, 47(1): 145 – 152.  
PAN Jian-fei, DONG Yi-hong, CHEN Hua-hui, et al. The overlapping community discovery algorithm based on compact structure [J]. Acta Electronica Sinica, 2019, 47(1): 145 – 152. (in Chinese)
- [3] Wang Z Z, Yuan Y, Zhou X M, et al. Effective and efficient community search in directed graphs across heterogeneous social networks[A]. Proceedings of the 31st Australasian Database Conference[C]. Melbourne, Australia: Springer, 2020. 161 – 172.
- [4] Liu S, Xia Z. A two-stage BFS local community detection algorithm based on node transfer similarity and local clustering coefficient[J]. Physica A: Statistical Mechanics and its Applications, 2020, 537: 122717.
- [5] Andersen R, Chung F, Lang K. Local graph partitioning using pagerank vectors[A]. Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science[C]. Berkeley, CA, USA: IEEE, 2006. 475 – 486.
- [6] Bian Y, Luo D, Yan Y, et al. Memory-based random walk for multi-query local community detection[J]. Knowledge and Information Systems, 2020, 62(5): 2067 – 2101.
- [7] Hsu C C, Lai Y A, Chen W H, et al. Unsupervised ranking using graph structures and node attributes[A]. Proceedings of the 10th International Conference on Web Search and Data Mining[C]. Cambridge, UK: ACM, 2017. 771 – 779.
- [8] Freitas S, Cao N, Xia Y L, et al. Local partition in rich graphs[A]. Proceedings of the 2018 IEEE International Conference on Big Data[C]. Seattle, USA: IEEE, 2018. 1001 – 1008.
- [9] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms[J]. Physical review E, 2008, 78(4): 046110.
- [10] 刘海姣, 马慧芳, 赵琪琪, 李志欣. 融合用户兴趣偏好与影响力的目标社区发现[J]. 计算机研究与发展, 2021, 58(1): 70 – 82.  
LIU Hai-jiao, MA Hui-fang, ZHAO Qi-qi, LI Zhi-xin. Target community detection with user interest influence [J]. Journal of Computer Research and Development, 2021, 58(1): 70 – 82. (in Chinese)
- [11] Bian Y, Ni J, Cheng W, Zhang X. The multi-walker chain and its application in local community detection[J]. Knowledge and Information Systems, 2019, 60(3): 1663 – 1691.
- [12] Ye W, Mautz D, et al. Incorporating user's preference into attributed graph clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2020. DOI: 10.1109/TKDE.2020.2976063.
- [13] Kamuhanda D, Wang M, He K. Sparse nonnegative matrix factorization for multiple-local-community detection [J]. IEEE Transactions on Computational Social Systems, 2020, 7(5): 1220 – 1233.
- [14] Ding X, Zhang J, Yang J. A robust two-stage algorithm for local community detection[J]. Knowledge-Based Systems, 2018, 152: 188 – 199.

## 作者简介



李青青 女, 1995年生, 甘肃庆阳人. 现硕士研究生就读于西北师范大学计算机科学与工程学院, 主要研究方向为社区搜索.  
E-mail: 2019211784@nwnu.edu.cn



马慧芳(通信作者) 女, 1984年生, 甘肃兰州人, 现为西北师范大学计算机科学与工程学院教授, 主要研究方向为人工智能、数据挖掘与机器学习.  
E-mail: mahuifang@yeah.net