

面向科研关系网络的发现算法研究

王卓昊 徐晨阳 江俊鹏 王东
(中国科学技术信息研究所, 北京 100038)

摘要: 采用数据挖掘等技术手段自动发现科研关系网络中蕴含的社区结构, 对于有针对性地组建科研队伍、维护项目申报和职称评审等程序的公平公正具有重要意义。现有的科研关系网络发现方法大多存在算法复杂度高、片面依赖论文合著和项目合作关系等问题, 因此本文基于经典的PageRank算法, 提出一种面向科研关系网络的发现算法。首先刻画科研关系模型, 建立科研关系网络, 然后以单个科研人员为核心通过对局部网络进行挖掘分析, 快速发现其所在的社区, 不仅降低了社区发现的复杂度, 而且提高了所发现社区的可用性, 对我国科技管理工作起到一定的推动作用。

关键词: 科研关系网络; 算法; PageRank

中图分类号: TP311 **DOI:** 10.3772/j.issn.1673-2286.2022.09.003

引文格式: 王卓昊, 徐晨阳, 江俊鹏, 等. 面向科研关系网络的发现算法研究[J]. 数字图书馆论坛, 2022(9): 21-27.

1 相关研究

根据国家统计局2019年发布的《科技发展大跨越 创新引领谱新篇——新中国成立70周年经济社会发展成就系列报告之七》, 2018年我国按折合全时工作量计算的科研人员总量已达到419万人年, 连续6年位居世界第一^[1]。如果将科研人员视为节点, 那么众多的科研人员就组成了一张巨大的图或网络, 根据论文合著、项目合作、师承等关系, 该网络可划分为若干社区。借助数据挖掘技术从海量的科研数据中挖掘出科研关系网络, 进而发现科研人员社区, 对我国的科技管理工作具有重要意义: 一方面, 借助科研人员社区可以有针对性地建立一支完备的科研队伍, 满足我国高水平科技项目的研发需求; 另一方面, 借助科研人员社区可以发现任意两个科研人员之间的联系紧密程度, 有利于践行项目申报、职称评审等程序的回避原则, 维护科研环境的公平公正。

目前, 关于科研关系网络发现的研究已经取得了一些进展: 夏欢等^[2]针对中国知网学术论文, 利用Pajek构建论文合著网络, 进而挖掘出科研社区, 并与经典的社区发现算法Girvan-Newman (GN) 算法^[3]进行了性能

对比; 罗纪双^[4]同样针对科研论文, 通过改进Louvain算法对科研合作网络进行社区划分, 并基于FR算法优化了社区内部的可视化布局; 蒲实等^[5]针对科研网络的动态特征, 提出一种基于动态科研网络表示学习的社区检测算法DANE-CD, 并与既有算法在准确率、归一化互信息和模块度3个指标上进行了对比。以上研究虽然获得了一定的成果, 但是仍然存在一些问题, 主要体现在两点: 一是现有研究大多基于论文合著或项目合作关系构建社区, 没有综合考虑到其他关系(如同事、合伙人、师承、校友、同乡等), 因此构建的科研关系网络不能全面地反映科研人员之间的联系紧密程度; 二是所使用的社区发现算法(如Louvain、GN等)在超大规模网络上的效率不够高, 社区构建和更新的效率较低。

鉴于以上问题, 本文在科研人员多种关系的基础上, 对经典的网页排名算法PageRank^[6]进行改进, 提出一种面向科研关系网络的发现算法, 能够对科研人员之间的关系进行多维刻画, 建立科研关系网络模型, 从单个科研人员出发, 快速发现该科研人员所处的局部社区。相对于现有的科研关系网络发现方法, 本文提出的算法不仅降低了社区发现的复杂度, 而且提高了所发

现社区的可用性。

2 相关技术

2.1 社区发现算法概述

在现实世界中,许多系统都可以用网络进行描述,如社交网络、万维网、公路铁路交通网等,网络中的节点表示系统中的个体,节点之间的边表示个体之间的关系。一个网络可认为是由若干个社区(community)组成的,同一社区内的节点之间的联系较为紧密,而社区与社区之间的联系较为松散。社区发现(Community Detection)算法就是用来发现网络中的社区结构,主要可分为以下3种。

(1) 传统的社区发现算法。这类算法主要分为两类:第一类是基于图分割的算法,如谱二分法^[7]、Kernighan-Lin (KL) 算法^[8]等,它们的基本思路是将图分为预定义大小的若干个簇或子图,使得子图内部的边数比子图之间的边数更密集;第二类是基于聚类的方法,包括分层聚类^[9]、划分聚类^[10]、谱聚类^[11]等,这类算法基于图的邻接矩阵表示方法,使用常规的算法对其中的子矩阵进行聚类。

(2) 基于分裂的社区发现算法。这类方法基于低相似性删除网络中的簇间的边,从而将社区彼此分离,前面提到的GN算法就属于这类算法。

(3) 基于模块度优化的社区发现算法。模块度(modularity)是用来衡量一个社区的划分是否优良的指标,可以简单将其理解为每一个社区内部的边的权重之和减去所有与社区节点相连边的权重之和。这类方法可进一步分为贪心法、模拟退火法、极值优化法、谱优化法等,前面提到的Louvain算法就属于贪心法。

上面介绍的是目前较为成熟的社区发现方法,可应用于大多数网络结构中。除此之外,新的方法也在不断被提出,例如,对于更新较为频繁的网络,学者陆续提出了多种动态社区发现算法^[12-14];再如,在一些网络中,单个节点可能同时属于多个社区,即社区之间存在明显的重叠情况,针对这种情况学者陆续提出了一些重叠社区检测方法,如Clique percolation^[15]、SVINET^[16]等。

2.2 PageRank算法

PageRank算法最早由Google创始人Larry Page和

Sergey Brin在1998年提出,是一种对网页重要性进行排名的算法,其核心思想是:被很多网页所链接的网页重要性较高,同时被重要的网页所链接的网页重要性也较高。具体来说,互联网中的网页被视为若干节点,网页之间的超链接被视为节点之间的有向边,每个节点的重要性取决于网络中其他节点与该节点的链接数量,一个节点的总链接数越多,则其重要性越高;同时,一个节点的总链接数越多,则其被指向的节点的重要性也越高。该算法通过PageRank值(PR值)来衡量节点的重要性,其计算方式如公式(1)所示。

$$PR_i = \sum_{j \in B_i} \frac{PR_j}{N_j} \quad (1)$$

其中, i 和 j 表示节点, PR_i 和 PR_j 表示它们的PR值, B_i 表示指向节点 i 的节点集合, N_j 表示节点 j 所指向的节点个数。

通过迭代,最终网络中所有的节点都会计算得到一个稳定的PR值,但在实际的网络中,可能会出现等级下沉(rank sink)的情况,它指的是节点的入度为0,即没有被任何节点所链接的情况,会导致PR值异常。为了解决该问题,可以引入阻尼系数或称衰减因子 α ($0 < \alpha < 1$)对PR值的计算方式进行修正,如公式(2)所示。

$$PR_i = \sum_{j \in B_i} \frac{PR_j}{N_j} \times \alpha + \frac{1-d}{n} \quad (2)$$

其中, α 一般取值为0.85, n 是网络中的节点总数; d 是线性组合系数,称为阻尼因子, $0 \leq d \leq 1$ 。

上面介绍了PageRank算法在网页重要性排名方面的应用,实际上该算法可以被用在和图有关的问题上,如社会影响力分析^[17]、文本聚类^[18]等。在本文中,则将PageRank算法应用到局部社区发现中。

3 面向科研关系网络的发现算法

我国拥有数量众多的科研人员,彼此之间形成了一张超大规模的关系网络,即使应用前述的Louvain等算法对该网络进行社区发现,其复杂度也难以想象。因此,本文考虑从局部网络入手,探讨如何从某个节点出发,在局部网络中找到该节点所在的社区,即所谓的局部社区发现。

本研究基于科研人员基本属性数据设计科研关系模型,定义科研人员之间存在的多种关系,提出关系紧

密度的计算方法。通过计算科研人员之间关系的紧密度, 建立由节点和边组成的科研关系网络, 形成一张有向图。提出一种面向科研关系网络的发现算法, 对科研关系网络进行关联关系挖掘和分析, 实现以单个科研人员为核心的局部社区发现。

3.1 科研关系模型

科研人员相关数据主要包括两部分: 一部分是其发表或署名的各类科技文献, 如论文、专利、申报书

等; 另一部分是其个人主页、机构官网公开或在人才数据库中存储的自我介绍, 一般包括教育背景、工作经历、获奖信息等。这些数据体现了科研人员之间可能存在的合作/合伙、行政、师承等关系, 为简单起见, 本文将其中的一对一关系作为基础, 组织形成了科研关系模型。同时, 为刻画两个科研人员之间的关系紧密程度, 参考项目申报、职称评审等程序中的回避要求, 本文对每种关系赋以权重, 用于表示关系的紧密程度。科研关系的具体描述如表1所示。

表1 科研关系描述

关系类型	具体描述	权重计算方式
科技文献合作关系	作为某篇论文或者专利的共同作者/发明人	署名顺序之差的绝对值的倒数, 例如分别为第一作者和第三作者, 则权重为1/2
项目/课题合作关系	共同完成某个项目/课题	共同完成项目权重为0.6, 共同完成课题权重为0.8
行政关系	共同隶属于某个行政单位	共同隶属于一级行政单位, 如XX大学, 权重为0.6; 共同隶属于二级行政单位, 如XX大学YY学院, 权重为0.7
师承关系	师生关系、共同导师关系以及校友关系	师生关系权重为0.9, 共同导师关系权重为0.8, 校友关系权重为0.5
合伙关系	共同创办企业、以营利性为目的的团体等	初始权重为0.9, 合伙时间每满一年, 权重加0.01, 最终权重不超过1
其他关系	如同乡等	权重统一设为0.3

需要注意的是, 在计算科技文献合作关系和项目/课题合作关系的权重时, 还需纳入时间因素, 这是由于主要考虑当前关系的亲密程度, 即距当前时间越近的合作理应取得更高的权重, 因此, 根据表1计算出每一次合作的初始权重后, 还需进行以下处理, 见公式(3)。

$$w' = w \times \left(1 - \frac{year_{cur} - year_{coo}}{year_{cur} - year_{start}} \right) \quad (3)$$

其中, w 表示每次合作的初始权重, $year_{coo}$ 和 $year_{cur}$ 分别表示合作年份和当前年份, $year_{start}$ 表示起始年份, 只有发生在 $year_{start}$ 之后的合作才会被纳入考虑。例如, 我们仅考虑近十年内的合作关系, 即 $year_{start}$ 设为2012, 那么对于两位科研人员在2019年合作的论文, 若初始权重 $w=0.5$, 则修正后的权重 $w' = 0.5 \times \left(1 - \frac{2022-2019}{2022-2012} \right) = 0.35$ 。

此外, 考虑到科研人员之间可能存在多项合作, 因此科技文献合作关系和项目/课题合作关系的权重可以进行叠加, 叠加后的结果除以该类关系的最大值进行归一化即可。

3.2 科研关系网络构建

定义好科研关系模型后, 即可进一步构建科研关系网络, 示例如图1所示。首先, 将每个科研人员视为一个个节点, 然后把科研人员之间的关系转换成无向的有权边, 即若两位科研人员存在某种关系, 则在对应的两个节点之间连上一条边, 边上的权重表示这种关系的强弱。

由于两位科研人员之间可能同时存在多种关系, 因此关系的权重或边的权重可以进行叠加, 如公式(4)所示。

$$W_{p,q} = \sum_{i=1}^n w_{p,q}^i \quad (4)$$

其中, $w_{p,q}^i$ 表示科研人员 p 和 q 所存在的第 i 种关系的权重, n 表示关系的种类总数, $W_{p,q}$ 是两者各种关系权重之和, 即对应节点之间边的权重。最终每条边的权重通过除以网络中所有边最大的权重进行归一化。

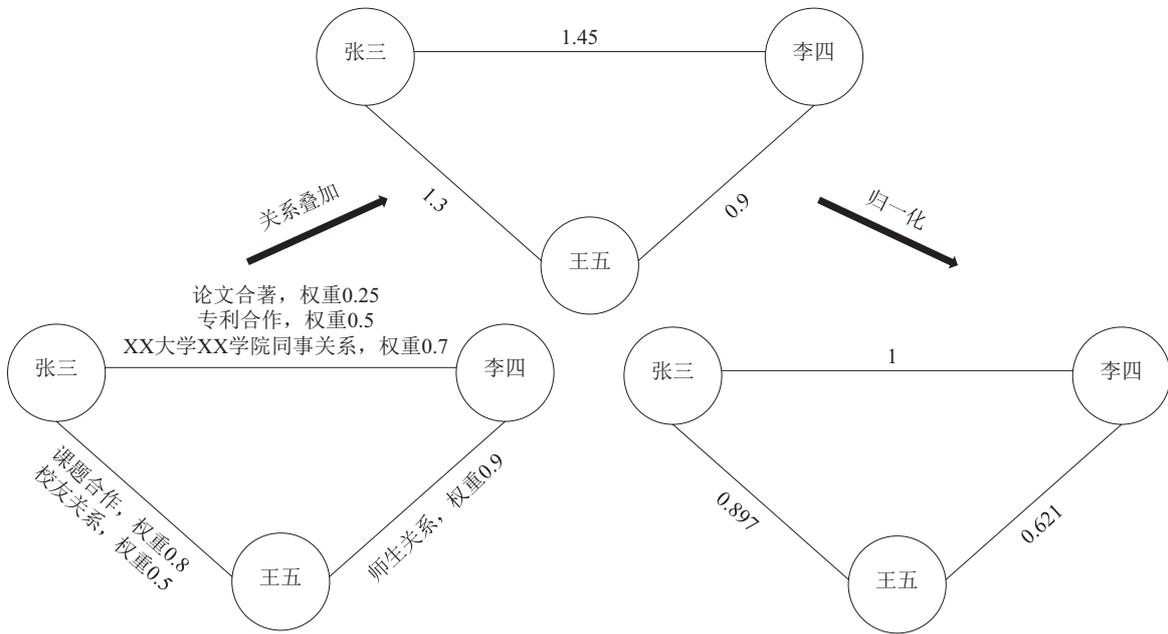


图1 科研关系网络构建过程

3.3 面向科研关系网络的发现算法

PageRank算法实际上属于随机游走模型 (Random Walk Model), 因此可以基于该模型的思想将PR值的计算方法写成公式(5)的形式。

$$\bar{\pi} = (1-\alpha) \bar{\pi}P + \alpha \bar{c} \quad (5)$$

其中, $\bar{\pi}$ 表示图中每个节点的PR值所组成的 n 维向量, n 表示图中节点的个数; P 表示图的概率转移矩阵, 如果节点 v_i 存在一条出边指向节点 v_j , 则 $P(i, j) = \frac{1}{d_{out}(v_i)}$, $d_{out}(v_i)$ 表示节点 v_i 的出度; \bar{c} 为 n 维起始向量, 表示最初各网页的PR值, 在PageRank的计算中, \bar{c} 的每一维都被设为 $\frac{1}{n}$ 。

可以看出, PageRank算法的主要目的是对所有节点按照全局重要性排序, 所以在初始化时, 算法设定所有节点的重要性都相同, 然后通过迭代对重要性不断调整。当我们进行局部社区发现时, 实际上也是对所有节点按照重要性排序, 因此从理论上来说可以基于PageRank算法实现。需注意的是, 这里的重要性并不是全局重要性, 而是相对于某个节点来说的相对重要性。因此, 在初始化时, 该节点的重要性应与其他节点有所区别, 在这个时候, PageRank算法就演化成了一种特殊形式, 即Personalized PageRank (PPR)。基于Personalized PageRank算法, 本文提出了一套面向科研关系网络的发现算法。

在本算法中, 首先选定某个节点作为源节点, 即认为该节点局部社区的中心。然后计算其他节点相对于源节点的重要性即PPR值, 如公式(6)所示。

$$\bar{\pi}_{PPR} = (1-\alpha) \bar{\pi}P + \alpha \bar{c} \quad (6)$$

其中, $\bar{\pi}_{PPR}$ 表示每个节点的PPR值所组成的 n 维向量, P 表示图的概率转移矩阵, \bar{c} 为 n 维起始向量, 对应源节点的那一项设为1, 而其他项均设为0, 也就意味着在起始状态下, 源节点相对自身的重要性最高, 其余节点相对源节点的重要性都为0。

在计算出其他节点的PPR值后, 就可以根据PPR值由大到小的顺序, 依次尝试将其他节点纳入以源节点为核心的局部社区中, 在这里我们提出一种用于衡量社区紧密程度的指标 Φ , 当某个节点加入社区 S 后 $\Phi(S)$ 变小了, 则认为该节点的加入使得社区 S 被强化了, 那么就正式纳入该节点, 否则不纳入。 $\Phi(S)$ 的计算方法如公式(7)所示。

$$\Phi(S) = \frac{\sum \text{edge weight from } S \text{ to } \bar{S}}{\sum \text{edge weight from } S \text{ to } S + \sum \text{edge weight from } S \text{ to } \bar{S}} \quad (7)$$

其中, S 表示当前考虑的局部社区, \bar{S} 表示网络中由不属于 S 社区的节点所构成的社区。

至此, 可以将面向科研关系网络的发现算法描述如下。①选定某一科研人员 s 作为源节点, 初始化目标社区 $S=\{s\}$; ②计算其他节点相对于 s 的重要程度即PPR值, 按照从大到小的顺序依次排列; ③将当前PPR值最

大的节点 t 纳入目标社区 S 中, 即 $S=S\cup\{t\}$; ④重新计算社区的紧密程度 $\Phi(S)$, 若 $\Phi(S)$ 减小了, 则将节点 t 正式纳入社区 S 中, 否则剔除节点 t ; ⑤考虑下一个节点, 执行第③④步; 若 $\Phi(S)$ 小于阈值 δ , 则认为社区 S 已经饱和, 不再接受其他节点; ⑥返回社区 S , 表示局部社区发现完成。

基于上述算法进行科研人员局部社区发现的过程

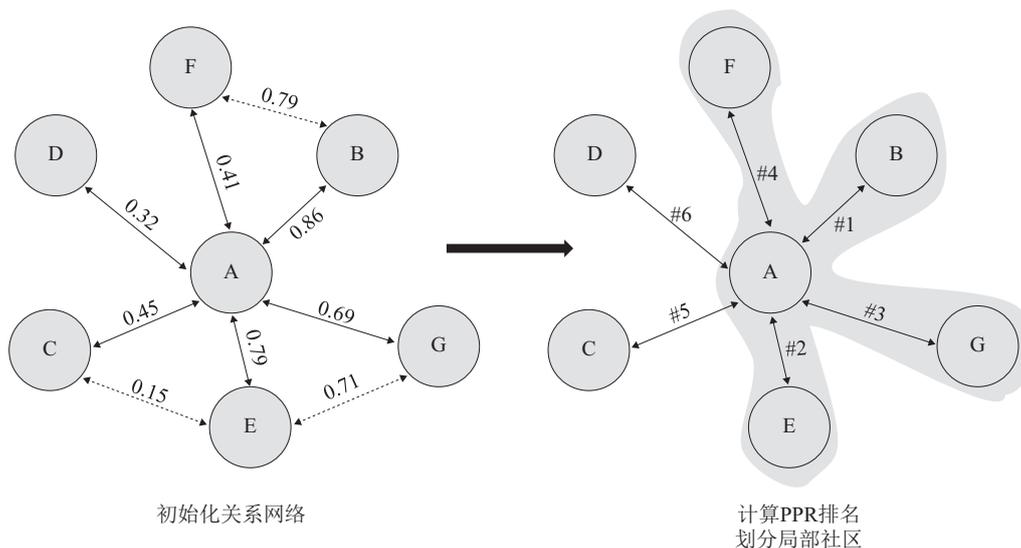


图2 科研人员局部社区发现过程

4 实验验证

为了说明本文所提出的方法的有效性, 设计实验进行验证。本实验基于科研人员关系数据, 提取255项科研人员信息, 其中包括科研人员的姓名、工作单位、论文、专利和户籍关系等。利用Python对科研人员数据进行相关关系抽取, 得到表1中定义的科技文献合作关系、项目/课题合作关系、行政关系、师承关系与合伙关系等几类主要的关系, 并根据上述关系的亲密程度度量方法计算关系的权重。采用Neo4j图数据库建立科研人员关系网络的图模型, 以该图数据为基础, 利用本文提出的面向科研关系网络的发现算法进行科研关系网络的挖掘和分析, 得到以科研人员为核心的社区。

4.1 实验环境及数据

本文实验的环境为Ubuntu 22.04.1 LTS x86_64 (Intel Xeon Silver 4208 (32) @ 3.200GHz), 32GB内存。实验工具为Python3.9.12, Visual Studio

如图2所示。其中, A到G分别表示科研人员节点, 首先根据二者之间存在的各种关系构建关系网络, 并计算出边的权重。然后, 将A视为源节点, 通过面向科研关系网络的发现算法计算出其他节点相对于A的重要程度即PPR值, 然后按照从大到小的顺序排列。最后, 依次将其他节点纳入以A为核心的目标社区中, 直至社区饱和。可以看出, 最终构建的社区包括A、B、E、G、F共5个节点。

Code 1.71.1, Neo4j Community 4.4.10图数据库, py2neo2021.2.3, numpy1.23.1, Neo4j Graph Data Science (GDS) library 2.1.11以及 MySQL5.7。

实验数据从关系数据库MySQL中导出并存储为CSV格式文件。首先, 利用Python中的Pandas库对数据进行预处理, 包含缺失值填补、异常数据清理和数据对齐处理。接着从数据中抽取科研人员的关联关系, 并利用Cypher数据库请求语句将科研人员相关关系映射至Neo4j图数据库中, 建立科研人员关系网络的图模型, 共包含694个节点以及3 369个边。

4.2 实验及结果分析

基于上面构建的科研关系网络, 采用本文提出的面向科研关系网络的发现算法, 首先计算以科研人员为中心的各节点PPR值。接着根据PPR值进行降序排序, 根据社区紧密程度指标将其他节点组织加入以源节点为核心的局部社区中。假设以科研人员41为核心, 计算PPR值见表2。以科研人员41为核心的局部社区展

示见图3。

表2 以科研人员41为核心的PPR值

科研人员ID	科研人员ID
41	1.5
294	0.187 542
292	0.187 542
295	0.187 542
296	0.187 542
293	0.187 542

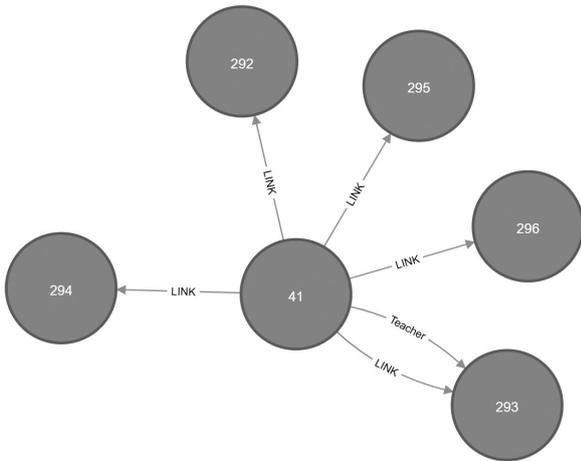


图3 以科研人员41为核心的局部社区

对上述实验结果进行分析,说明本文提出的面向科研关系网络的发现算法,能够基于科研关系网络实现局部社区的快速发现,验证了本方法的有效性。由于本方法基于局部网络进行挖掘分析,能够有效降低社区发现的复杂度,同时因为考虑的科研关系更加丰富并且通过关系紧密度对多种科研关系进行归一化处理,刻画科研人员之间的联系密切程度,使得所发现社区在可用性和实用性方面有所提升。接下来将增加实验数据、扩展实验方法、与经典方法进行对比分析,进一步验证本方法的性能。

5 结语

本文提出的面向科研关系网络的发现算法能够基于科研关系网络进行局部社区发现,其优势在于涵盖了更多的关系类型,在社区发现的复杂度和所发现社区的可用性方面具有一定优势,可进一步应用于面向重大科研需求的科研人员推荐,有针对性地组建科研队伍,同时在项目申报、职称评审等重要程序中能够更准

确地发现人员之间的关联关系,落实回避机制,对于科技管理工作具有较大意义。

接下来将进一步通过试验对上述结果进行验证。同时围绕下面3个问题进行更深入的研究。

(1) 本文所提出的算法建立在关系网络的基础上,即需要提取出科研人员之间的合作、行政、师承、合伙等关系,但是由于数据和方法上的问题,其中某些关系很难被准确、完整地提取出来,因此可能会在不同程度上影响到最终局部社区发现效果。

(2) 为了简单起见,科研人员之间的大多数关系权重系数由人为定义,不够灵活和精细,也很难保证同一套权重系数可以在不同的网络中均能取得最佳的表现,因此后续需要进一步优化权重系数的处理方式。

(3) 在本文所考虑的科研关系中,科技文献合作关系是主要的组成部分。但是许多科技文献涉及多个交叉学科,可能会导致最终的社区内部科研人员之间的研究方向的相似性有所降低,这与许多用户的直观判断产生了偏差,即一个社区内部的科研人员应基本属于同一个研究方向。

参考文献

- [1] 科技发展大跨越 创新引领谱新篇——新中国成立70周年经济社会发展成就系列报告之七 [EB/OL]. [2022-07-23]. http://www.gov.cn/xinwen/2019-07/23/content_5413524.htm.
- [2] 夏欢,刘辉,詹隽. 基于科研论文合著网络的社区发现算法研究 [J]. 重庆工商大学学报(自然科学版), 2017, 34 (5): 50-55.
- [3] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. National Academy of Sciences, 2002, 99 (12): 7821-7826.
- [4] 罗纪双. 科研合作社区亲密度分析与可视化展示 [D]. 石家庄: 石家庄铁道大学, 2019.
- [5] 蒲实,赵卫东. 一种面向动态科研网络的社区检测算法 [J]. 计算机科学, 2022, 49 (1): 89-94.
- [6] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web [EB/OL]. [2022-09-13]. <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.5427>.
- [7] BARNES E R. An algorithm for partitioning the nodes of a graph [C] //IEEE Conference on Decision & Control Including the Symposium on Adaptive Processes. IEEE, 1981.

- [8] DUTT S. New Faster Kernighan-Lin-Type Graph-Partitioning Algorithms [C] // Proceedings of 1993 International Conference on Computer Aided Design (ICCAD). IEEE, 2002.
- [9] FORTUNATO S. Community detection in graphs [J]. Physics Reports, 2009, 486 (3-5): 75-174.
- [10] HAVENS T C, BEZDEK J C, LECKIE C, et al. A soft modularity function for detecting fuzzy communities in social networks [J]. IEEE Transactions on Fuzzy Systems, 2013, 21 (6): 1170-1175.
- [11] DHUMAL A, KAMDE P. Survey on community detection in online social networks [J]. International Journal of Computer Applications, 2015, 121 (9): 35-41.
- [12] 郭昆, 彭胜波, 陈羽中, 等. 基于密度聚类的增量动态社区发现算法 [J]. 模式识别与人工智能, 2018, 31 (11): 965-978.
- [13] 张高祯, 张贤坤, 苏静, 等. 动态加权网络中的演化社区发现算法研究 [J]. 计算机应用研究, 2019, 36 (4): 992-996, 1005.
- [14] 陈羽中, 施松, 朱伟平, 等. 一种基于邻域跟随关系的增量社区发现算法 [J]. 计算机学报, 2017, 40 (3): 570-583.
- [15] RIORDAN O. Clique percolation [M]. New York: John Wiley & Sons, 2009.
- [16] GOPALAN P, BLEI D M. Efficient discovery of overlapping communities in massive networks [J/OL]. National Academy of Sciences, 2013, 110 (36) [2022-09-01]. DOI:10.1073/pnas.1221839110.
- [17] 石立新. 改进PageRank算法挖掘社交网络关键用户 [J]. 现代电子技术, 2022, 45 (12): 95-99.
- [18] 徐杨炳. 基于引用网络的文献聚类及主题变迁分析方法研究 [D]. 杭州: 浙江财经大学, 2019.

作者简介

王卓昊, 男, 1977年生, 博士, 副研究员, 研究方向: 计算机应用、信息系统建设。

徐晨阳, 男, 1992年生, 硕士, 助理研究员, 通信作者, 研究方向: 信息安全、信息系统运维, E-mail: xucy@istic.ac.cn。

江俊鹏, 男, 1992年生, 硕士, 助理研究员, 研究方向: 信息资源管理。

王东, 男, 1987年生, 博士, 助理研究员, 研究方向: 数据挖掘、知识图谱、数据可视化。

Research on Discovery Method for Scientific Research Relationship Network

WANG ZhuoHao XU ChenYang JIANG JunPeng WANG Dong
(Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)

Abstract: Using data mining and other technical means to automatically detect the community structure contained in the scientific research relationship network is of great significance for the targeted establishment of scientific research teams, the maintenance of fairness and impartiality of project declaration, title evaluation and other procedures. Most of the existing discovery methods for scientific research relationship network have problems such as high complexity, one-sided dependence on paper co-authorship and project collaboration. Therefore, this paper presents a discovery method for scientific research relationship network, which is based on the classical PageRank algorithm and considers more relationships among scientific researchers. We first describe the relationship model of scientific researchers and then establish the network among them. After that by mining and analyzing the local network with a single scientific researcher as the core, we can quickly find the community in which the scientific researcher is located. Our method can not only reduce the complexity of community discovery, but also improve the availability of the discovered community. We believe that it will play a certain role in promoting the management of science and technology in our country.

Keywords: Scientific Research Relationship Network; Method; PageRank

(收稿日期: 2022-09-13)