

利用局部离群因子算法探测核心技术的发展趋势^{*}

——以中国风能专利数据为例

李佳佳 马铁驹

(华东理工大学商学院 上海 200237)

摘要 [目的/意义]为验证局部离群因子算法在探测核心技术及核心技术的发展趋势的有效性,丰富专利分析领域的研究。[方法/过程]以中外专利数据库服务平台 CNIPR 作为数据源,分别从局部离群因子算法和社会网络分析方法两个视角对中国风能领域的专利数据进行对比分析,识别中国风能领域的核心技术以及核心技术的发展趋势。[结果/结论]结果显示,局部离群因子算法(LOF)和社会网络分析方法得出的结论基本一致:即中国在风力发电技术方面一直保持优势,未来的发展潜力集中在风能照明装置及系统,验证了局部离群因子算法在探测核心技术及核心技术的发展趋势方面的有效性。

关键词 局部离群因子算法 专利分析 共现网络 技术预测 中国风能

中图分类号 G306

文献标识码 A

文章编号 1002-1965(2017)03-0119-06

引用格式 李佳佳,马铁驹.利用局部离群因子算法探测核心技术的发展趋势[J].情报杂志,2017,36(3):119-124,195.

DOI 10.3969/j.issn.1002-1965.2017.03.021

Detecting Core Technology's Development Trend with LOF

—Based on Wind Energy Patent Data of China

Li Jiajia Ma Tiejue

(Business School, East China University of Science and Technology, Shanghai 200237)

Abstract [Purpose/Significance]The paper aims to verify the effectiveness of LOF algorithm in the detection of the core technology's development trend and thus promote the research of patent analysis. [Method/Process] Taking patent database CNIPR as patent data source, the paper conducts a contrastive analysis between LOF algorithm and the social network analysis method, identifying the core technology and the core technology's development trend for wind energy in China. [Result/Conclusion]The contrastive analysis result shows a consistency of the two different methods of LOF algorithm and the social network analysis method: China always has the advantage in wind motor and will focus on the lighting devices and equipment in the future. The effectiveness of LOF algorithm in the detection of the core technology's development tendency is verified.

Key words local outlier factor algorithm patent analysis co-occurrence network technical forecasting China's wind energy

0 引言

现代社会是信息社会,信息资源是最重要的战略资源之一。专利作为一种特殊的信息,在资源分析和学术研究中都有着特殊的地位和作用。专利携带了各种各样的信息,比如 IPC(International Patent Classifi-

cation)、主题、摘要、申请人和机构等等,研究人员对这些专利数据进行研究分析,得出有助于决策的信息^[1]。专利分析可以帮助组织及企业探测特定领域的核心技术,发现新型技术机会并且识别出未来的关键技术。近些年来,研究人员在专利分析领域提出了多种模型、方法和工具^[2]。从宏观角度看,目前专利分析主要集

收稿日期:2016-09-17

修回日期:2016-10-24

基金项目:国家杰出青年科学基金项目“技术演化与能源系统分析”(编号:71125002)。

作者简介:李佳佳(ORCID:0000-0003-3751-3814),女,1989年生,硕士研究生,研究方向:专利分析及技术预测;马铁驹(ORCID:0000-0003-1524-5484),男,1975年生,教授,博士生导师,研究方向:技术演化与能源系统分析。

中在以下四个方面:

第一,基于社会网络分析(SNA)的方法构建共现网络图,识别核心专利。Wasserman 第一次提出了社会网络这个概念,他把社区群落当做社会关系网,其中人为网络中的节点,人与人之间的关系为网络中的线^[3]。之后 Watts 和 Strogatz 提出了著名的小世界网络模型^[4]。社会网络涉及的应用领域很多,例如博客交流模式研究^[5]、心理学研究^[6]以及高校合作^[7]。在专利分析领域,张杰等以变频技术的专利数据做数据源,研究得到低压变频技术的核心技术领域为交流电机和家用空调等领域^[8]。

第二,借助假设检验和回归分析等统计学方法,研究知识的聚集和流动。Almeida 和 Kogut 发现技术的聚集现象发生在特定区域(例如美国硅谷),不同区域之间聚集程度不同,同时专利所属人的转移会影响技术和知识的流动^[9]。Almeida 通过分析美国半导体产业的专利引用,得出跨国公司对当地知识的应用远远高于相似的本土化公司^[10]。Maurseth 和 Verspagen 通过专利分析对欧洲的知识流动进行验证,发现欧洲的知识流动过程具有障碍性,专利引用受到地理、语言甚至风俗习惯等因素的影响^[11]。Yoshikane 以日本的申请专利作为主专利进行分析,得出不同类别的 IPC 分类号被后续专利引用时,后续专利分布的多样性不同^[12]。在中国,越来越多的学者也倾向于采用回归分析等统计学方法来研究专利数据,用来探测技术距离和技术创新之间的关系。周华和韩伯棠通过对专利数据进行分析,验证了企业所处的环境会影响自身的技术创新,提出企业间相互交流能够促进知识溢出效应^[13]。刘志迎和单洁含以中国大学和企业联合申请的发明专利作为测度指标,验证了技术距离对大学企业的联合创新具有显著的促进作用^[14]。

第三,技术路线图。利用专利数据制定技术路线图^[15]。例如张海波从市场需求角度出发,运用德尔菲法对我国新能源汽车技术的发展做出预见,绘制我国新能源汽车产业市场需求技术路线图,综合技术路线图提出相应的实施建议^[16]。杨璧嘉和张旭提出将专利网络分析引入技术路线图,并在此基础上研究专利网络分析在技术路线图绘制中具体的实施办法和有可能带来的好处^[17]。许崇春提出将技术路线图、专利地图等创新方法融合集成,并将其应用于产业集群技术创新的新路径^[18]。

第四,利用专利信息做技术预测。近些年,学者们大多倾向于利用专利信息进行相关领域的技术预测,研究方法逐步多样化。Jun 利用技术成熟度曲线模型对混合动力汽车领域进行技术预测^[19]。Milanez 等利用逻辑曲线对纳米技术领域的专利进行模拟,进而实

现技术预测^[20]。Altuntas 等对专利指标进行构建,依据指标的排名决定技术的投资优势^[21]。Lee 等借助文本挖掘技术对 LED(Light Emitting Diode)领域的专利数据进行分析并进行技术预测^[22]。

1 研究框架

本文的技术路线图如图 1 所示,主要包括数据收集、算法描述与实现、局部离群因子算法与传统社会网络分析方法的对比分析以及最后的结论和建议四个部分。

第一,以中外专利数据库服务平台为基础,构建风能检索式,搜集中国风能领域专利文件,得到中国风能专利数据库。第二,对局部离群因子算法进行描述,并详述其在专利分析中的实现过程。第三,以中国风能领域的专利数据为例,借助局部离群因子算法画出中国风能领域核心专利的 LOF 折线图,并与传统的网络分析方法进行对比,以此探测此算法在专利分析中的有效性。第四,验证局部离群因子算法在探测特定领域技术发展趋势的有效性,并对我国风能领域的技术发展提出相关建议。

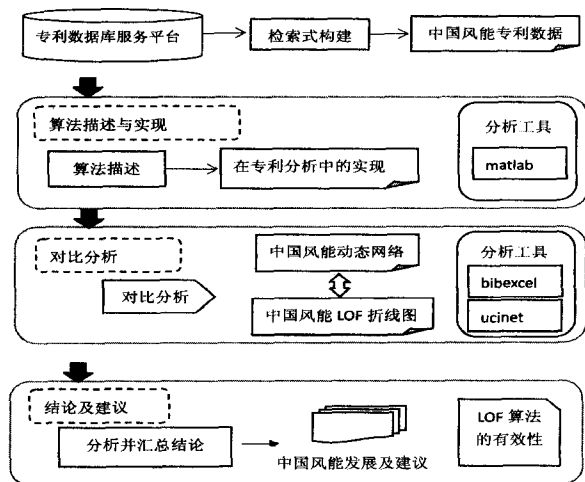


图 1 风能专利分析技术路线图

2 数据收集

利用中外专利数据库服务平台 CNIPR,对中国风能领域的专利进行检索和采集,检索时间为 2015 年 12 月,主要利用主题和摘要进行检索。搜索中国专利检索式为 TI=风能 and AB=风能,共检索出 3 125 条专利。由于检索出的专利文本中存在国外机构或者国外发明人在中国申请专利的情况,故剔除国省代码非中国的专利以及不相关专利后,剩余 2 978 条专利。

3 局部离群因子(LOF)算法在专利分析中的应用

3.1 局部离群因子算法描述 Breuing 等首先提出了局部离群因子(LOF)的概念^[23]。定义数据集中的

一个对象 p , 该算法的核心思想, 即赋予数据集中每一个对象一个表征该对象偏离程度的因子。该因子用来反映对象所处空间的一个密集程度, 离群因子的值越大, 则表明该对象所处的空间越稀疏; 离群因子的值越小, 则表明该对象所处的空间越密集^[24]。下面着重介绍本文中所涉及到的 LOF 算法的一些基本概念^[25]。

定义 1 对象 p 的第 k 距离 (k -distance of an object p)

选定一正整数 k , 对象 p 的第 k 距离可以记为 k -distance(p), 本文可以简写为 k -dis(p)。在数据集 D 中存在一个对象 o , 将对象 p 与 o 之间的距离记做 $d(p, o)$ 。满足以下条件, 则 k -distance(p) = $d(p, o)$ 。

1) 在数据集 D 中, 除对象 p 以外, 至少存在 k 个对象 o' , 满足 $d(p, o') \leq d(p, o)$ 。

2) 在数据集 D 中, 除对象 p 以外, 至多存在 $k-1$ 个对象 o' , 满足 $d(p, o') < d(p, o)$ 。

定义 2 ε 邻域 (ε -neighborhood)

对象 p 存在于数据集 D , 则对象 p 的 ε 邻域记为: $N_\varepsilon(p) = \{x \in D \mid d(p, x) \leq \varepsilon\}$

定义 3 对象 p 的第 k 距离邻域 (k -distance neighborhood of the object p)

若已知对象 p 的第 k 距离, 那么 p 对象的第 k 距离邻域就是到对象 p 的距离小于等于 p 的第 k 距离的所有对象的集合, 记为 $N_{k\text{-dis}(p)}(p)$, 本文可简写为 $N_k(p)$ 。

定义 4 对象 p 相对于对象 o 的可达距离 (reachability distance of p)

表述如下: $\text{reach-dis}_k(p, o) = \max\{k\text{-dis}(p), d(p, o)\}$

定义 5 对象 p 的局部可达密度 (local reachability density of object p)

设 k 为正整数, 则数据集中对象 p 的局部可达密度可以表述如下:

$$\text{lrd}_k(p) = \frac{k}{\sum_{o \in N_k(p)} \text{reach-dis}_k(p, o)} \quad (1)$$

该公式可以反映在特定数据集 D 中对象 p 所处区域的疏密程度。选取特定的正整数 k , 若对象 p 所处的区域比较稀疏, 则相对而言 k -distance(p) 值偏大, 进而 $\text{reach-dis}_k(p, o)$ 的取值偏大, 从而 $\text{lrd}_k(p)$ 的值偏小; 相反, 若对象 p 所处区域比较密集, 则 k -distance(p) 值偏小, 进而 $\text{reach-dis}_k(p, o)$ 的取值偏小, 从而对象 p 的局部可达密度偏大。

定义 6 对象 p 的局部离群因子 (local outlier factor of the object p)

$$\text{LOF}(p) = \frac{1}{k} \sum_{o \in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)} \quad (2)$$

对象 p 的局部离群因子 $\text{LOF}(p)$ 的值可以用来反映对象 p 所处区域在数据集 D 中的疏密程度, $\text{LOF}(p)$ 的值越大, 表示该区域越稀疏; $\text{LOF}(p)$ 的值越小, 表示该区域越密集。

3.2 利用局部离群因子算法探测核心技术发展趋势 局部离群因子算法在本文中的运用可以归纳为以下四个阶段^[23]:

(1) 对于每一个对象 p , 找出与其最邻近的 k 个对象, 算出 p 的第 k 距离。

(2) 对象 p 的局部可达距离为 $\text{reach-dis}_k(p, o) = \max\{k\text{-dis}(p), d(p, o)\}$ 。

(3) 将 $N_k(p)$ 定义为距离对象 p 最近的 k 个对象的集合, 则对象 p 的局部可达密度为 $\text{lrd}_k(p) =$

$$\frac{k}{\sum_{o \in N_k(p)} \text{reach-dis}_k(p, o)}。$$

(4) 对象 k 的局部离群因子的值为 $\text{LOF}(p) =$

$$\frac{1}{k} \sum_{o \in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)}。$$

该算法直观的计算过程如图 2 所示。

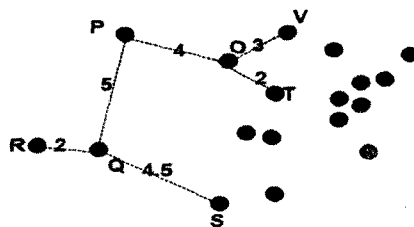


图 2 LOF 算法举例

图中每一个节点类比为数据集中的一个对象, 以对象 P 为例, 当取 k 值固定为 2 时, P 的第 k 距离为 5, k 距离邻域中包括对象 O 和 Q 。对象 P 到对象 O 和 Q 的局部可达距离均为 5, 则可以计算出 P 的局部可达

密度为 $\text{lrd}_k(P) = \frac{2}{\sum_{o \in N_k(P)} \text{reach-dis}_k(P, o)} = \frac{2}{5+5} =$

0.2。同理可以计算出另外两个对象 O 和 Q 的局部可达密度为 $\text{lrd}_k(Q) = 0.22, \text{lrd}_k(O) = 0.33$, 最终可以算

出对象 P 的 LOF 值为 $\text{LOF}(P) = \frac{1}{k} \sum_{o \in N_k(p)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(p)} = \frac{1}{2} \cdot$

$\frac{0.22 + 0.33}{0.2} = 1.375$ 。假设整个数据集的对象全部均

匀分布, 则每个对象的 LOF 值为基准值 1。若特定数据集中某个对象的 LOF 值大于 1, 则该对象区域的分布偏于稀疏; 反之, 若该对象的 LOF 值小于 1, 则该对象区域的分布偏于密集。

将局部离群因子算法运用到专利分析领域, 用来探测特定领域核心技术的发展趋势, 其基本思想为: 将数据集中的每一个对象类比为专利节点 (专利节点用 IPC 分类号表示), 专利节点的分布以及疏密程度会随

着时间发展而发生改变,因而每一个专利节点的局部离群因子 LOF 的值也会随时间的发展而发生改变。借助 matlab 计算出二维平面中专利节点随时间变化的局部离群因子 LOF 值,画出随时间变化的趋势图,以此来探测相应技术领域的技术发展趋势。

4 利用 LOF 探测中国风能领域核心技术发展趋势

4.1 局部离群因子算法在中国风能专利数据中的应用 以中国风能领域的专利数据为例,利用局部离群因子算法探测核心技术的发展趋势,其研究步骤如下:

(1) 风能专利文件数据处理:利用 Bibexcel 对中国风能专利文件进行处理,分别统计出截止到 2005 年和截止到 2015 年所有专利文件中出现的 IPC 分类号数量,每一个 IPC 分类号都代表风能领域相应的技术。

(2) 将每一个 IPC 分类号类比为为一个专利节点,而每一个专利节点则代表数据集中的一个对象。截止到 2005 年,中国风能领域的专利分类号数量为 91 个;截止到 2015 年,中国风能领域的专利分类号数量为 1 334 个。风能领域中专利 IPC 分类号出现的总频次可以反映该类技术的专利强度,与其他 IPC 分类号的总共现次数可以反映该技术的扩散潜力^[23]。因而对截止到 2005 年及 2015 年所有 IPC 分类号的总频次和总共线次数进行统计并进行数据归一化处理(归一化到 0~10 之间)。坐标轴横轴代表该 IPC 分类号出现的总频次,纵轴代表其总共现次数,可将所有的 IPC 分类号反映在二维坐标平面中,即 IPC 分布图。

(3) 将上述 IPC 分布图中的每一个 IPC 分类号类比为为一个对象,借助 matlab 程序可以计算出相应年份每一个 IPC 分类号的 LOF 值。对风能专利领域的 IPC 分类号进行总频次和总共线次数的统计之后,选出频次和共线次数都比较靠前的专利点,即代表中国风能

领域核心技术的 IPC 分类号 (F03D9/00, F03D3/06, H02J7/00, B60L8/00, F21S9/04), 其对应的技术含义如表 1、表 2 所示。

表 1 四位 IPC 分类号及对应含义

IPC 分类号	含义
F03D	风力发动机
H02J	供电或配电的电路装置或系统;电能存储系统
F21S	非便携式照明装置或其系统
F21V	照明装置或其系统的功能特征或零部件;不包含在其他类目中的照明装置和其他物品的结构组合物
H02K	电机
F21W	与照明装置或系统的用途或应用有关
B60L	电动车辆动力装置

表 2 IPC 分类号及对应含义

IPC 分类号	含义
F03D9/00	特殊用途的风力发动机;风力发动机与受它驱动的装置的组合
F03D11/00	不包含本小类其他组中或与本小类其他组无关的零件、部件或附件
F21S9/04	电源是发电机
F21S9/03	通过曝光再充电
F21W131/103	用于大街或道路
F21V23/00	照明装置内或上面电路元件的布置
F21Y101/02	微型光源,例如发光二极管

(4) 设定正整数 k 值取值为 5, 10 和 100 三种情况,分别计算出 6 个主要的 IPC 分类号从 2005 年到 2015 年的 LOF 值(截至 2005 年中国风能领域的专利 IPC 分类号共有 91 个,故当 k 取值 100 时,2005 年的数据不存在,从 2007 年开始)。三种情况下,五个主要 IPC 分类号的局部离群因子值的变化趋势相似,但是当 k 取值为 100 时,其折线图的变化更加平稳,故本文主要按照 k 取值 100 时的 LOF 折线图进行分析。图 3 左侧即为 $k=100$ 时的 LOF 折线图,对应五个主要 IPC 分类号的 LOF 值如右侧所示:

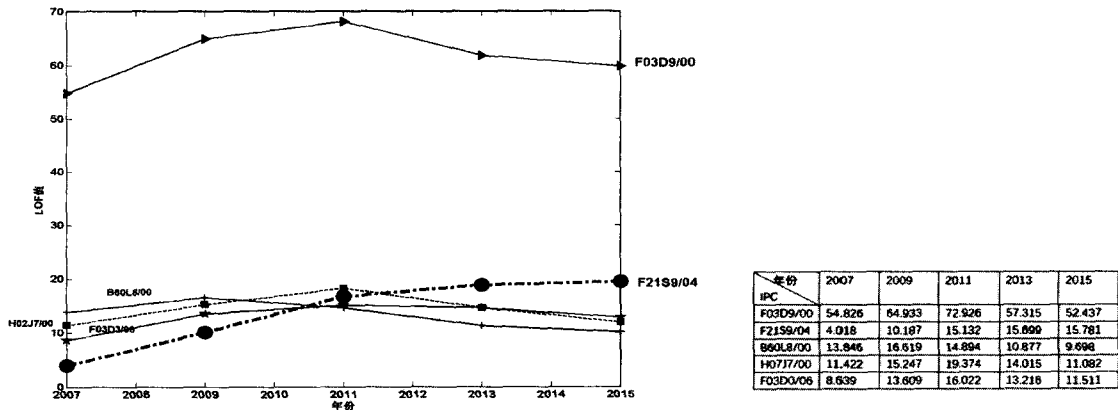


图 3 $k=100$ 的 LOF 折线图及对应 LOF 值

由图 3 左侧可知,分类号 F03D9/00 一直平稳处于最上方,即随着时间的变化,IPC 分类号 F03D9/00

的 LOF 取值一直最大,代表该专利节点所处的数据集区域的分布最为稀疏,也即处于平面数据集的右上角

区域(专利 IPC 总频次最大,总共线次数最大,代表风能领域的核心专利技术);IPC 分类号 F03D3/06, H02J7/00, B60L8/00 的变化趋势趋于平缓,表明这些 IPC 分类号代表的技术在近十年的发展中趋于平缓 and 稳定;IPC 分类号 F21S9/04 (圆形虚线)从 2007 年到 2015 年一直稳定上升,说明中国风能领域照明装置及系统方面的技术在稳步提升,专利节点随着时间推移向二维平面数据集的右上角挪动。

三种情况下,五个主要 IPC 分类号的局部离群因子值的变化趋势相似,但是当 k 取值为 100 时,其折线图的变化更加平稳,故本文主要按照 k 取值 100 时的 LOF 折线图进行分析。图 5 左分类号 F03D9/00 一直平稳处于最上方,即随着时间的变化,IPC 分类号 F03D9/00 的 LOF 取值一直最大,代表该专利节点所处的数据集区域的分布最为稀疏,也即处于平面数据集的右上角区域(专利 IPC 总频次最大,总共线次数最大,代表风能领域的核心专利技术);IPC 分类号 F03D3/06, H02J7/00, B60L8/00 的变化趋势趋于平缓,表明这些 IPC 分类号代表的技术在近十年的发展中趋于平缓 and 稳定;IPC 分类号 F21S9/04 (圆形虚线)从 2007 年到 2015 年一直稳定上升,说明中国风能领域照明装置及系统方面的技术在稳步提升,专利节点随着时间推移向二维平面数据集的右上角挪动。

为了验证局部离群因子算法(LOF)在探测技术发展趋势方面的有效性,本文与社会网络分析方法的结果进行对比,在此基础上识别中国在风能领域的核心技术并探测核心技术的发展趋势。

4.2 社会网络分析在中国风能专利数据中的应用

对收集的中国风能专利数据进行处理,借助相关软件构建共现矩阵,得到基于 IPC 分类号的共现网络。网络图的每个节点代表一个 IPC 分类号,节点越大,代表该 IPC 分类号的节点度中心度越大,即与该节点共线的节点数越多;节点之间的连线代表两个 IPC 分类号存在共现,即出现在同一条专利文本中,共现的频次越高,则连线越粗^[26-27]。

通过分析中国截止到 2005 年和 2015 年的风能专利的共现网络图可知:网络图的最大节点均为 F03D9/00,即中国的风力发电技术一直保持优势。但上述共现图只能看到整体发展态势,但并不能识别核心技术的动态发展趋势,对我国风能产业的发展没有很强的指导意义。故本文将时间序列和网络共现阈值进行结合,分阶段画出中国在风能领域的专利共现网络图。共现阈值反映两个专利节点出现在同一专利文本中的频次,频次越高,则共现阈值越大。本文通过控制节点之间的共现阈值,使每个阶段中的网络节点数目维持在 12 个(共现阈值较低的节点不会出现在网络图

中),分别绘制出中国在四个阶段的共现网络图,即截止到 1999 年、2000 年到 2005 年、2006 年到 2011 年和 2012 年到 2015 年四个阶段(分别见图 4~图 7)。

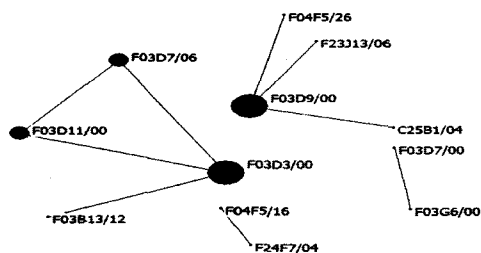


图4 中国第一阶段网络图(截止到1999)

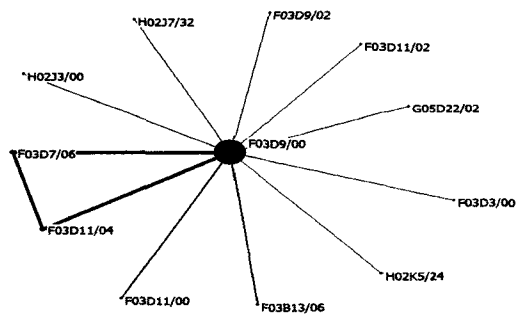


图5 中国第二阶段网络图(2000-2005)

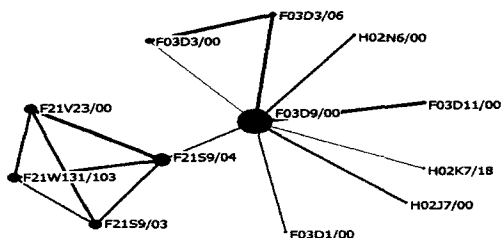


图6 中国第三阶段网络图(2006-2011)

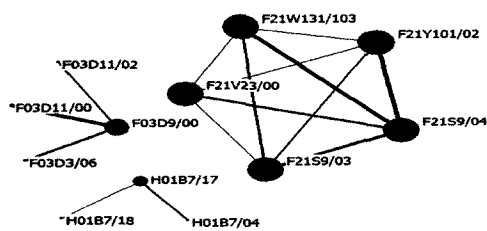


图7 中国第四阶段网络图(2012-2015)

通过分析中国四个阶段的共现网络图可得:

第一,在前三个阶段,中国的风力发电技术一直保持优势,核心节点始终维持在 F03D9/00。到第四个阶段,风力发电技术逐渐趋于弱化,在图 7 中表现为左上方较小的专利节点。从技术生命周期的角度看,中国在 F03D9/00 代表的风力发电技术方面进入了成熟期,该领域的专利申请已经趋于饱和状态或者发展空间有限^[28]。

第二,分类号为 F21 系列的节点在第三阶段开始出现并缓慢增长,在左下角形成了一个小型群落(图 6 所示)。在第四阶段,F21 系列的专利节点迅速增长并在右上角形成明显群落(F21S9/03、F21V23/00、

F21W131/103、F21S9/04、F21Y101/02),最大节点和最强连线都聚集在此群落中(图7所示)。由此可知,F21系列的专利在第四阶段发展迅速,增长空间大,可以预测,中国在照明装置及系统具有明显优势和潜力。

4.3 LOF算法与传统社会网络分析方法对比分析

将LOF算法和社会网络分析方法进行对比分析,得到以下结论:

(1) 中国的风力发电技术一直保持优势,核心专利节点始终维持在F03D9/00。在LOF算法的二维数据集平面中,该专利节点始终保持在右上角,所处区域的分布比较稀疏,从2007年到2015年期间,该专利节点的LOF值一直维持在较高水平。

(2) 社会网络分析方法中,分类号为F21系列的节点在第三阶段开始出现并缓慢增长并在第四阶段的共线网络图中形成明显群落,可以推测中国在风能领域的照明装置或系统具有明显优势和潜力;同理在LOF算法中,IPC分类号F21S9/04(圆形虚线)从2007年到2015年一直稳定上升,表示该专利节点所处的区域由密集转向稀疏,即向二维数据集平面的右上方推进(右上方即代表高频次及高共线频次的专利节点),说明中国风能领域照明装置及系统方面的技术在稳步提升,在此方面具有发展潜力。

5 结 论

本文将局部离群因子算法运用到专利分析领域,用来识别特定领域的核心技术并探测核心技术的未来发展趋势。以中国风能领域的专利数据为例,分别从局部离群因子LOF算法和社会网络分析方法两个视角对中国风能的专利数据进行研究分析,发现两个研究视角得出的结论具有一致性:a.中国在风力发电技术方面一直保持优势,核心专利节点始终维持在F03D9/00;b.中国在风能领域的照明装置及系统方面的技术在稳步提升,在此方面具有发展潜力。从而进一步验证了LOF算法在探测核心技术及其发展趋势方面的有效性。

参 考 文 献

- [1] 张 杰,崔 靖. SNA在专利分析中的应用研究[J]. 科技管理研究,2011(7):168-171.
- [2] Changyong Lee, Bokyoung Kang, Juneseuk Shin. Nolvety-focused patent mapping for technology opportunity analysis[J]. Technological Forecasting & Social Change,2015(9).
- [3] Stanley Wasserman. Social network analysis: Methods and applications[M]. Cambridge: Cambridge University Press,1994.
- [4] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature,1998.
- [5] 甘 丹. 博客世界交流模式的社会网络分析——以新浪博客为例[D]. 济南:山东大学,2007.
- [6] 马绍奇,焦 璨,张敏强. 社会网络分析在心理研究中的应用[J]. 心理科学进展,2011(19):755-764.
- [7] 钟柏昌,李 艺. 社会网络分析在教育研究领域的应用[J]. 教育研究,2013(9):25-32.
- [8] 张 杰,刘粉香,翟东升,等. 基于共现网络的核心技术领域识别研究[J]. 情报杂志,2012,31(11):35-39.
- [9] Almeida P, Kogut B. Localization of knowledge and the mobility of engineers in regional networks[J]. Management Science,1999,45:905-917.
- [10] Paul Almeida. Knowledge sourcing by foreign multinationals: Patent citation analysis in the U. S. semiconductor industry[J]. Strategic Management Journal,1996,17:155-165.
- [11] PerBotolf Maurseth, Bart Verspagen. Knowledge spillovers in Europe: A patent citations analysis[J]. Scand. J. Economics,2002,104:531-545.
- [12] Fuyuki Yoshikane. Comparative analysis of patent citations of different fields: In consideration of the data size dependency of statistical measures[J]. Social and Behavioral Sciences,2014,147:153-159.
- [13] 周 华,韩伯棠. 基于技术距离的知识溢出模型应用研究[J]. 科学学与科学技术管理,2009(7):111-116.
- [14] 刘志迎,单洁含. 技术距离、地理距离与大学-企业协同创新效应——基于联合专利数据的研究[J]. 科学学研究,2013(31):1331-1336.
- [15] 李雪凤,全允桓,谈毅. 技术路线图——一种新型技术管理工具[J]. 科学学研究,2004(21):89-94.
- [16] 张海波. 我国新能源汽车产业技术路线图研究[D]. 武汉:武汉理工大学,2012.
- [17] 杨璧嘉,张 旭. 专利网络分析在技术路线图中的应用[J]. 情报分析与研究,2008(5):61-66.
- [18] 许崇春. 基于技术路线图、专利地图和TRIZ集成的产业集群创新技术路径研究[J]. 科技进步与对策,2012(14):46-49.
- [19] Seung-Pyo Jun. A comparative study of hype cycles among actors within the socio-technical system: With a focus on the case study of hybrid cars[J]. Technological Forecasting & Social Change,2012,79:1413-1430.
- [20] Milanez D H, L I L de Faria, R M do Amaral. Patents in nanotechnology: An analysis using macro-indicators and forecasting curves[J]. Scientometrics,2013.
- [21] Serkan Altuntas, Turky Dereli, Andrew Kusiak. Forecasting technology success based on patent data[J]. Technological Forecasting & Social Change,2015,96:202-214.
- [22] Changyong Lee, Bokyoung Kang, Juneseuk Shin. Nolvety-focused patent mapping for technology opportunity analysis[J]. Technological Forecasting & Social Change,2015,90:355-365.
- [23] 辛丽玲. 基于密度差异的离群点检测研究[D]. 北京:北京交通大学,2015.
- [24] 胡彩平,秦小麟. 一种基于密度的局部离群点检测算法 DLOF[J]. 计算机研究与发展,2010,47(12):2110-2116.
- [25] Breunig M M, Kriegel H P, Ng R T. LOF: Identifying density-based local outliers[C]. Proc of ACM SIGMOD Conf. New

准计数发文数量多的研究人员其合作稳定性和合作影响力较低产出水平群体要显著得高;b.以 Hagen 加权方法计算的相对发文量划分的高、低产出水平群体在学术合作的一、二级指标中的表现有些不同,可以看出,在相对发文量(Hagen 加权)区分的高、低产出群体中,合作程度、合作稳定性和合作影响力均有显著的差异性,但在具体的合作能力指数、点度中心度和中介中心度3个二级指标上则没有显著的差异。

4 结 论

本文以图书情报领域为例,探究了作者学术合作与其科研产出的关系。本文的结论主要有四点:a.从合作程度、合作稳定性和合作影响力三个维度入手考察作者学术合作的情况,构建了作者学术合作计量指标体系,同时又选择标准计数发文量和相对发文量(Hagen 加权)作为定量考察作者的科研产出主要指标,为探索学术合作和科研产出的关系构建了定量的指标体系;b.合作能力指数是测度学术合作一个较为特殊、全面的指标,科研人员应注重与合作伙伴建立起稳定、高频的合作模式;c.如果科研人员想要提升自己的科研产出,应该着力于提升其在自身研究领域的影响力并与少数合作作者建立稳定、高频的合作关系;d.各学术合作计量指标、科研产出指标从不同角度反映作者的学术合作与产出水平,考察作者学术合作与产出之间的关系时应根据目的的不同选择不同角度的指标来分析其学术合作与科研产出水平。

综上所述,本文在探究学术合作与科研产出的关系中得出了具有理论补充和实践指导性意义的结论。但本文的研究结论是基于本文收集的图书情报领域的521位作者的数据,数据量较小;此外,本文仅用论文合作来表示科研产出水平,且忽略了科研产出与学术合作间的时间滞后性,因此存在一定的局限性。在今后的研究中,还需要进一步探究直接合作与间接合作之间的区别和联系,结合定性、定量的分析方法综合、全面考察学术合作及其影响效果。

参 考 文 献

- [1] 邱均平,温芳芳.作者合作程度与科研产出的相关性分析——基于“图书情报档案学”高产作者的计量分析[J].科技进步与对策,2011(5):1-5.
- [2] Ortega J L. Influence of co-authorship networks in the research impact: Ego network analyses from Microsoft Academic Search

[J]. Journal of Informetrics, 2014,8(3):728-737.

- [3] 王卫,潘京华,张晓梅.合作能力指数及其实证研究[J].情报杂志,2014(11):108-112.
- [4] Erjia Yan Y D. Applying centrality measures to impact analysis: A coauthorship network analysis[J]. Journal of the American Society for Information Science and Technology, 2009,60:2107-2118.
- [5] Liao C H. How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks[J]. Scientometrics, 2011,86(3):747-761.
- [6] Abbasi A, Wigand R T, Hossain L. Measuring social capital through network analysis and its influence on individual performance[J]. Library & Information Science Research, 2014,36(1):66-73.
- [7] 邱均平,曾倩.国际合作是否能提高科研影响力——以计算机科学为例[J].情报理论与实践,2013(10):1-5.
- [8] 刘俊婉.科学合作对高被引科学家论文产出的影响分析[J].情报科学,2014,32(12):77-81.
- [9] 常欢,吕瑞花,张佳静.学术谱系内合作网络研究——以刘东生为核心的第四纪学术谱系为例[J].情报理论与实践,2016,38(4):14-19.
- [10] K S. Bibliometrics Studies of Reaearch Collaboration: A review[J]. Journal of Information Science, 1983,6(1):33-38.
- [11] 朱丽娟,李丽娜.学术合作计量指标研究述评[J].情报杂志,2013(6):76-79.
- [12] Schubert A. A Hirsch-type index of co-author partnership ability[J]. Scientometrics, 2012,91(1):303-308.
- [13] Schubert A. Jazz discometrics-A network approach[J]. Journal of Informetrics, 2012,6(4):480-484.
- [14] 刘军.整体网分析讲义 UCINET 实用指南[M].上海:上海人民出版社,2014.
- [15] 林聚任.社会网络分析:理论、方法与应用[M].北京:北京师范大学出版社,2009.
- [16] 杜建,张玢,唐小利.基于作者引用与合作关系的学术影响力测度研究进展[J].图书情报工作,2013(8):135-140.
- [17] Barry Bozeman S L. The impact of research collaboration on scientific productivity[J]. Social Studies of Science, 2005,35(10):673-702.
- [18] Hagen N T. Harmonic allocation of authorship credit: source-level correction of bibliometric bias assures accurate publication and citation analysis[J]. PLOS ONE, 2008,3(e402112).
- [19] Hagen N T. Harmonic publication and citation counting: Sharing authorship credit equitably - not equally, geometrically or arithmetically[J]. Scientometrics, 2010,84(3):785-793.

(责编:王平军;校对:刘影梅)

(上接第124页)

York, ACM, 2000:427-438.

- [26] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical review E, 2004.
- [27] Newman M E J. Detecting community structure innetworks[J].

The European Physical Journal B-Condensed Matter and Complex Systems, 2004.

- [28] 李春燕.基于专利信息分析的技术生命周期判断方法[J].现代情报,2012,32(2):98-101.

(责编:贺小利;校对:王平军)