

专利引用网络主路径方法研究述评与展望*

■ 张娴^{1,2} 方曙¹

¹ 中国科学院成都文献情报中心 成都 610041 ² 中国科学院大学 北京 100190

摘要: [目的/意义]对现有专利引用网络主路径方法研究内容进行总结梳理,为今后应用该方法解决技术演化进程中的关键性专利技术识别和主流线索提取提供理论支撑。[方法/过程]系统梳理相关研究成果,从算法研究、应用研究、方法优化扩展研究三个方面总结现有研究内容与特点,分析当前研究的局限性,探讨未来研究发展方向。[结果/结论]当前研究的主要局限性在于:对路径发展驱动力的多元性与系统性揭示不够、忽视不同引证关系对路径演化的影响差异性、对演化的动态性关注不足、多主路径方法本质上仍属单目标搜索。未来研究将关注以下几个方向:对算法思想进行实质性与创新性拓展、更强调动态性与未来预测性、优化计算效率以增强适用性与实用性、发挥专利引用主路径在产业化扩散研究中的独特优势。

关键词: 专利引用 引用网络 主路径 技术路径 技术演化

分类号: G306 F062.4

DOI: 10.13266/j.issn.0252-3116.2016.20.017

1 专利引用网络主路径的内涵

专利文献作为创新活动的完整记录,被视为是技术、产品、工艺创新的核心内容和基础,它能反映出一项技术自诞生之日起不同时期的开发活动状况,因此,专利信息详实地反映了技术演化的线索^[1]。专利引文是当某件专利文献公布时,在专利文件中列出的与该专利申请相关的其他文献。专利引用关系反映了某件专利的技术基础和科学基础,体现技术间的累积、继承关系。相对于经济、市场数据,专利引用信息具有基础数据容易获取、便于定性定量描述等特点;基于专利引用网络可以分析专利技术发展历程,研究技术发展的历史脉络,为开展技术评价、选择及预测活动提供重要参考^[2]。

依据技术进化思想,每一件专利个体可被看成隐藏着知识的零散片段,专利引用网络呈现了这些知识片断继承关系,网络的“关键路径”则可被视为技术进步的主干。1989年N. P. Hummon与P. Doreain^[3]提出了“主路径”思想,基于“整体连通性”来离析特定时刻引文网络关键路径,主要目标是通过识别出引文网络中具有最大连通度的系列文献来概述研究领域的发

展态势以及领域演化过程中的主要文献、主要人物与主要事件。主路径分析方法不仅可在时间维度上揭示学科(领域)的收敛与扩散态势,也为揭示特定领域演化过程的主要文献与重要理论提供了研究工具与方法^[4]。之后,主路径方法被引入专利引用网络分析当中,用以识别专利引文网络中的知识流动主路径,绘制技术演化轨迹^[5]。与此前常常采用的历史性、描述性研究方法相比,主路径方法更好地弥补了技术轨道研究的“工程性”视角,对于解决技术演化进程中的关键性专利文献识别和主流线索提取具有重要意义,已成为技术演化研究的重点之一。

2 相关研究现状

以Web of Science、CNKI为检索源,将专利引用、网络分析、技术路径3组概念的相关检索词进行组配检索,对检索结果进行判读,筛选出相关文献。通过归纳梳理,分析了当前国内外相关研究成果的主要内容。

2.1 主路径方法的提出

1989年N. P. Hummon与P. Doreain发表了关于科技文献引文网络“关键路径”的应用研究,提出了“主路径”思想。他们采用深度优先搜索算法(depth

* 本文系中国科学院科技服务网络计划(STS计划)项目“中国科学院知识产权信息服务”(项目编号:KFJ-EW-ST-032)研究成果之一。

作者简介:张娴(ORCID:0000-0002-6297-1190)副研究员,博士研究生;方曙(ORCID:0000-0002-4584-7574)研究员,博士生导师,通讯作者,E-mail:fangsh@clas.ac.cn。

收稿日期:2016-06-28 修回日期:2016-09-14 本文起止页码:140-148 本文责任编辑:易飞

first search) 与穷举搜索算法(exhaustive search algorithm) 结合的方法来寻找网络中所有可能的搜索路径,以遍历数(traversal counts) 优先来定义引证网络的主路径——对于网络中的节点,选择其输出连边中具有最高遍历数的连边作为下一路径,重复应用遍历计数最大法则,直至定义出全网络中最常用的路径,即是反映文献知识主流的主路径^[3]。

N. P. Hummon 与 P. Doreain 提出了 3 种主路径搜索的边权重指标——节点对投影计数(Node Pair Projection Count, NPPC)、搜索路径连接计数(Search Path Link Count, SPLC)、搜索路径节点对计数(Search Path Node Pair, SPNP)^[3]。V. Batagelj 进一步提出了各种连接性权重指标的有效计算方法,深化了主路径分析方法,使之得以运用于大型引证网络主路径分析; V. Batagelj 还提出了新方法——搜索路径数(Search Path Count, SPC)^[6]。

自上述研究之后,主路径方法越来越受到关注,学者们在 N. P. Hummon、V. Batagelj 等研究的基础上,在论文引用、专利引用网络中开展了越来越多的研究。总体而言,目前国内外学者对于引文网络主路径方法的研究,可大体划分为 3 类: ①主路径搜索算法的设计,迄今已形成若干基本得到普遍认可的主流算法; ②主路径算法的应用研究,通过在具体技术领域的应用来对比验证算法的合理性与优劣; ③对算法的优化与扩展研究,以满足不同的需求。

2.2 主路径分析的主流算法

2.2.1 搜索路径统计数(SPC) SPC 算法是当前主路径搜索中最常采用的经典算法。SPC 通过计算相邻两节点之间的连边被网络中所有的路径所遍历的次数,来衡量该连边在网络中的重要性。所含连边的 SPC 和达到全局最优的路径,即是基于 SPC 法的全网主路径。

如图 1 所示^[7],网络中存在源点 A、B,汇点 C、D、E 与 F。对于连边 HJ 来说,连接了 AH、BH 与 JC、JD。因此在网络所有路径中,节点 H、J 之间的连边,共有 4 条路径遍历(路径 A→C、A→D、B→C 和 B→D),则连边 HJ 的 SPC 值为 4。

V. Batagelj 定义了 SPC 的算法^[6]: 令 $N^-(v)$ 表示网络中自源点 s 至节点 v 的路径 $s \rightarrow v$, $N^+(v)$ 表示节点 v 至汇点 t 的路径 $v \rightarrow t$, 则对于任意一条自 s 到 t 的经由连边(u, v) 的路径 π 可如式(1)所示:

$$\pi = \sigma \cdot (u, v) \cdot \tau \tag{1}$$

其中 σ 表示从 s 到 u 的所有路径 $N^-(u)$, τ 表示从 v

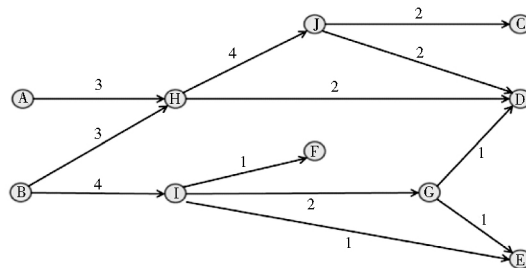


图 1 基于 SPC 算法的路径识别原理

到 t 的所有路径 $N^+(v)$ 。则,连边(u, v) 的权重 $N(u, v)$ 的计算如式(2)所示:

$$N(u, v) = N^-(u) \times N^+(v) \tag{2}$$

连边(u, v) 的标准化 SPC 权重计算方式如式(3)所示:

$$\omega(u, v) = \frac{N(u, v)}{N(s, t)} \Rightarrow 0 \leq \omega(u, v) \leq 1 \tag{3}$$

2.2.2 搜索路径节点对统计数(SPNN) SPNN 算法也是基于起始节点发出的所有搜索路径来计算相邻两节点之间的连边在所在路径上连接的所有的节点对数目。由于路径“内部”的连边与路径起点或终点附近的连边相比,连接了更多的节点对,因此根据 SPNN 方法,位于搜索路径中部的连边将比路径两端的连边收到更高的遍历数。

SPNN 识别原理如图 2 所示^[8],在网络中的所有路径中,节点 A、C 之间的连边,使 A 与 C、D、E、F、G、H、I、J 等 8 点相连,因此连边 AC 的 SPNN 值为 8; 节点 C、D 之间的连边 CD 连接了 A、B、C 三点指向 D,使 A、B、C 与 D、E、F、G、H、I、J 等 7 点相连,因而连边 CD 共连接了 21 对节点对,故连边 CD 的 SPNN 值为 21。

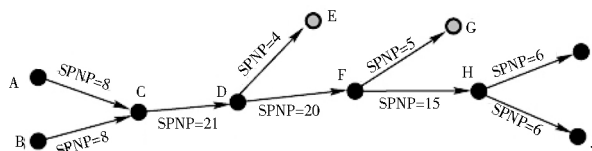


图 2 基于 SPNN 算法的路径识别原理

许琦定义了专利引用网络中 SPNN 的计算方法^[9]: 通过有向连边(u, v) 连接的上游节点和下游节点组成的节点对数目计算(u, v) 的 SPNN 值,如式(4)所示:

$$SPNN(u, v) = L^-(u) \times L^+(v) \tag{4}$$

式(4)中 $L^-(u)$ 表示所有直接或间接被专利 u 引证的节点数,包括专利 u 本身; $L^+(v)$ 表示所有直接或间接引证专利 v 的节点数,包括专利 v 本身。

2.2.3 最优主路径演化网络(NETP) NETP(network

of the evolution of top path) 算法由 B. Verspagen 于 2007 年提出,揭示网络最优路径在一定时间间隔之后发生的变化^[10]。

NETP 算法的基本思路是:定义专利引用网络 C , 采用 SPLC(或 SPNP) 方法提取主路径网络 C' ; 析出在 t 年前(含 t 年当年)的专利引文网络 C_t , 采用 SPLC(或 SPNP) 方法确定 C_t 的主路径网络 C_t' 。然后,以连边最大值确定 C_t' 中的单主路径(single main path) P_t 。最后,合并不同 t 时段的路径 P_t , 即令 $t = T_0 + a + T_1$, 其中 T_0 代表数据集中的最早专利年, T_1 代表数据集中的最晚专利年, a 为非负数。将结果标识为网络 P , 代表原始网络 C 中的主路径的时间演化情况。

NETP 方法实质上是划分不同时间间隔来计算基于 SPNP 或 SPLC 的最优主路径(top path), 也即连边的 SPNP 总值最高的专利引证路径, 从而通过对比时间段 (t, T) 、 $(t, T+1)$ 中处于最优路径上的专利的变化情况, 来分析技术随着时间进展的延续、旁落或消亡, 考察这些最优路径上的专利发明所蕴含的知识随时间演化的利用与发展情况^[11-12]。

2.2.4 前向引证节点对统计值算法(FCNP) 前向引证节点对统计值算法(Forward Citation Node Pair, FCNP) 由 C. Choi 与 Y. Park 于 2009 年提出, 是由前向引证节点对数确定连边的权值来识别主路径。C. Choi 与 Y. Park 通过计算相邻两节点之间的连边所连接的前向引证的节点对数, 得出该连边的权值^[13]。

FCNP 算法的基本原理是: 对于连边所连接的相邻两个节点, 分别统计各自所连接的前向引证的节点数(含该节点自身, 以免路径终端的节点的前向引证节点数统计为 0), 再将二者的前向引证节点数相乘即可得, 如图 3 所示:

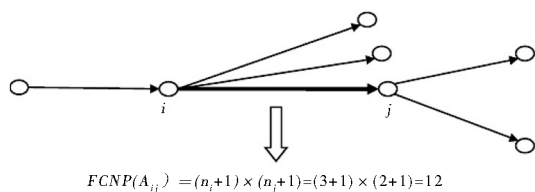


图 3 基于 FCNP 算法的路径识别原理

在图 3 所示的专利引用网络中, 专利 j 引用了专利 i 。 n_i 、 n_j 分别表示专利 i 、 j 各自的前向引用数。 $FCNP(A_{ij})$ 表示连边 $i \rightarrow j$ 的权值。

FCNP 算法设计的基础思想是考虑在专利引证网络中, 前向引证与后向引证具有不同的代表意义——后向引证(施引关系) 用于考察专利与技术之间的知

识流动或溢出关系; 前向引证(被引关系) 则是从技术或经济价值的角度来衡量发明的创造质量^[14-15]。 C. Choi 的 FCNP 算法考虑的是节点的流出路径, 即节点的被引关系, 旨在通过由此选择出的主路径反映专利的技术与经济价值。

2.2.5 节点对投影统计数(NPPC) 节点对投影统计数(NPPC) 方法由 N. P. Hummon 与 P. Doreain 提出, 建立了一个连接节点对的所有子图的邻接矩阵来计算每个连接的遍历计数值。这些矩阵“堆栈”为相对应的行、列节点。遍历计数值就是所有相连节点对在基本矩阵上的投影。最终的投影矩阵的值为网络中抽取出的子图与节点对的连接次数, 这便是计算遍历次数的节点对投影计数方法。节点对投影计数的遍历数值反映了在一个搜寻路径中, 一个连接所继承的、所衍生出的所有连接^[16]。 NPPC 方法的内涵是计算所有从网络中衍生出来的子图中每个顶点对所连接的链接数, 即每个连接所包含的全部连接节点对的数目^[17]。

2.3 主路径方法的应用研究

2.3.1 主路径方法应用效果验证 上述主路径搜索算法被学者们陆续应用于不同技术领域, 提取具体领域的主路径, 筛选路径上高价值专利节点(中间专利、终点专利、汇聚专利), 考察路径演化的连续性、稳定性、分化点等, 以此来验证算法的合理性, 对比不同优劣。

在大量的领域应用研究中, 以 SPLC/SPC、SPNP 方法最为常用。 B. Verspagen 通过燃料电池领域的技术路径识别, 证实了 SPNP、SPC 两种路径识别算法在专利引文网络中的可行性^[10]。 J. C. Guan 等采用 SPC 方法提取了纳米技术专利跨国引用网络的技术路径, 结合“小世界”模型考察了引用网络的路径可达长度和聚类特点, 验证了方法的可行性^[18]。 Z. K. Yang、Y. S. Xu 与 F. Yuan 等总结了时间间隔划分的两种经验方法——TCL 法、基于专利申请时间拐点法, 使 SPLC 方法更具有动态性特点, 通过电动汽车、航空发动机、风力发电等领域的应用研究, 比较了技术路径上突破性创新(radical innovation) 与累积式创新(incremental innovation) 的特点^[19-21]。

NETP、FCNP 方法也取得较好的应用效果验证。 R. Fontana 等利用专利引证网络研究以太网的技术变革, 采用类似 NETP 方法分析专利引证网络的结构连通性, 通过关键路径来绘制技术轨道图, 发现可能危及或减缓创新进程的瓶颈技术^[22]。 B. T. David 采用 NETP 算法识别了不同时间间隔内人工椎间盘领域的

技术演进主路径以及其中专利节点代表的技术热点,并根据市场状况检验其合理性^[11]。A. Martinelli 将 NETP 方法应用于电信交换行业路径分析^[23]。

对于 NPPC 算法,由于时间复杂度比较高,在处理大型网络时可能引起计算溢出,适用性较差,因此后续应用较少^[17 24]。

总体而言,SPC、SPNP 是受认可度最高的主路径方法,在经典的社交网络分析软件 Pejek 与 Ucinet 中都提供有基于这两种方法的相应计算功能。有研究认为二者应用结果总体上基本一致,也存在细微区别,当研究演化主路径或核心专利时,SPNP 算法更适合;当研究路径的衍生时,SPC 算法更适合^[8]。SPC 算法得到的结果路径长度略长且主路径构成组分更多,被推荐为主路径分析的优选方法^[6]。从计算的时间复杂度来看,SPC 算法最为经济^[17]。

2.3.2 主路径方法的应用技巧 在面向具体技术领域的主路径算法应用效果研究中,学者们还从主路径分析的不同环节、角度,总结出一些辅助性方法与手段来改善应用效果。例如:

在专利引文网络的构建环节,采用 C#编程方法提取和处理德温特专利数据,利用基础专利替换具有相同技术信息的同族专利,从而实现了同族专利的归一化处理,有效清洗了数据源^[25];当数据源很大时,通过增加主成分分析步骤,可以尽快缩小研究范围、提高分析效率,在应用 FCNP 方法研究碳捕获与封存技术发展路径时,结合主成分分析法从复杂的专利引用网络中识别出 5 类主要碳捕获与封存技术发展路径^[26];在分析路径演化特点时,采用时间序列和技术生命周期相结合的方法来划分网络的时间阶段,通过解读不同阶段的子网主路径,强调动态性地对比技术演化路径^[27];在选取关键路径的中间专利时,综合直接引用和间接引用信息,引入了专利引证强度指标,更全面地找出对后续技术影响作用较大的关键技术节点^[28];结合基于 TRIZ 思想的“技术难题—解决方案”启发式方法,分析各条技术进化轨迹的连续性和稳定性、关键专利中阐述的技术背景与技术难题是否发生变化,为技术范式发生了演化的定性表示提供依据^[9];将主路径分析法与引文时序分析法结合使用,能够充分发挥个体主义方法与网络分析方法的优势,更好地分析知识产生和转移、揭示学科发展历史过程^[29];利用 NPPC、SPLC、SPNP 提取主路径,以主路径为种子文献提取与主路径关联的最大核心弱组分及凝聚子群,与主路径关联的最大核心弱组分和凝聚子群能够展示丰富的领

域演化结构,是一种全新的基于连通性的领域演化结构识别方法^[30]。

这些实验总结,分别从数据预处理、网络构建、路径提取步骤、路径节点选择、路径诠释等不同的角度或环节,对主路径方法的应用效果与效率进行了一定程度的改善。

2.4 主路径方法的优化与扩展研究

除了上述主流算法及其应用研究,学者们还结合各自的实际需求,对主路径方法进行不同的优化与扩展,分别取得了不同程度的进展。

2.4.1 连通性遍历权数的加权调节 J. S. Liu 等^[31]引入了对不同引证情况的相关度考虑,利用法律数据库 WestLaw 中 KeyCite 标引项下的 4 类情况(examined, discussed, cited 和 mentioned)来表征 4 个级别的引证相关度(4-star, 3-star, 2-star 与 1-star)。当执行了传统 SPC 主路径搜索算法中第一步计算引证网络中连边的遍历数之后, J. S. Liu 等引入 4 级引证相关度对连边遍历数进行相关度加权调节,之后再利用加权遍历数来确定主路径。该研究中分别设计了 3 种不同的加权调节因子:平坦加权、线性加权、指数加权(见图 4)。

假定 $r(u, v)$ 是文献 u 与文献 v 间的引证相关度,则平坦、线性、指数三种加权调节因子分别如公式(5)、(6)、(7)所示:

$$a(u, v) = 1 \quad \text{式(5)}$$

$$a(u, v) = r(u, v) \quad \text{式(6)}$$

$$a(u, v) = e^{r(u, v)} \quad \text{式(7)}$$

假定 $a(u, v)$ 为文献 u 与文献 v 间的引证连边的调节因子,则连边 UV 的重要性测度值如式(8)所示:

$$w_{SPC}^*(u, v) = a(u, v) \times w_{SPC}(u, v) = a(u, v) \times N_{sources\ as\ origins}^-(u) \times N_{sinks\ as\ destinations}^+(v) \quad \text{式(8)}$$

J. S. Liu 等认为,指数加权可能存在调节过度(overkill)之嫌,但考虑到大型网络计算中 SPC 值可能会出现较大级差,因此仍然保留了指数型加权调节设计。该研究以商标淡化诉讼案为例开展实证,检验不同调节类型的适用度,对比主路径识别效果,认为引证相关度信息有助于主路径分析法揭示出重要性较高的法律案件。

2.4.2 主路径算法的变体及其综合运用 在传统主路径法基础上, J. S. Liu 与 L. Y. Y. Lu 提出了增强主路径分析的概念(enhances main path analysis),设计了 4 种变体^[32]: ①全网主路径(the global main path):指全网最优 SPC 路径; ②后向局部主路径(the

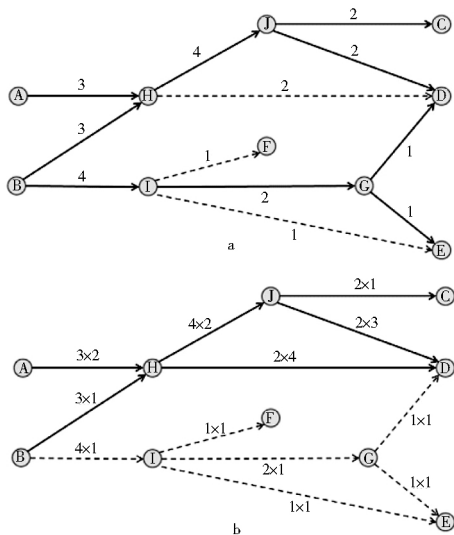


图 4 J. S. Liu 等的引证相关度加权调节主路径

注: a 为经典 SPC 路径, b 为引证相关度(线性)加权调节主路径

backward local main path): 与全网最优路径可能相同或有所差异, 当二者相一致时, 代表目标领域具有较好的技术融合度; ③多主路径(multiple main paths): 指放宽搜索限制条件, 选出满足一定阈值的次优遍历路径(该研究确定为满足前 20% 遍历权数的路径); ④关键主路径(key-route main paths): 首先找出全网中的遍历数最高连边, 从该连边终点开始, 前向搜索直到网络路径终点, 再从该连边起点开始, 后向搜索直到网络路径的起点, 将连边两端路径合并即得。该研究认为综合运用这 4 种变体, 能够揭示传统方法不能捕获到的一些路径, 其中尤以关键主路径变体特别有用。

2.4.3 基于主题关联强度的路径搜索 为了克服基于连接度算法的主路径方法对引文网络拓扑属性的依赖, 学者们采用了主题关联性指标代替连接度作为路径搜索的参考变量, 因而产生了基于主题关联度算法的主路径方法。

许琦基于孟德尔遗传定律, 采用知识遗传分解的方法, 提出了“知识适应指数”概念, 用以评价专利所承载的知识在技术进化中的适应能力, 量化早期专利对当前专利的技术贡献度^[33]。该文认为知识适应能力评价方法不同于连接度算法, 侧重于分析当前最新专利的知识起源, 重点是把握起始节点对终止节点的知识贡献度和知识传播能力。以 SAO(subject-action-object)三元组为基础作为知识基因的表现形式, 许琦还提出一种基于 SAO 三元组的知识基因提取方

法——在专利引证网络中, 应用连接度算法建立知识进化轨迹; 利用文本语法分析技术, 从专利权利要求书中提取 SAO 三元组; 计算语义相似度, 绘制知识基因图谱^[34]。

陈亮^[35]等将文本相似度引入主路径搜索过程, 计算文档间的主题相似度, 据此进行专利引文网络路径搜索, 用技术主题路径上全部节点对的语义相似度之和作为衡量该路径重要性的权重指标, 通过路径权重指标计算所得主路径的权重。

2.4.4 多主路径方法研究 由于局部主路径、全网主路径都是抽取网络中最大权重的单一路径, 学者们认为将最重要路径与其他相对重要路径相结合, 有利于发现领域发展的细节, 揭示领域的演化情况。多主路径(multiple main paths)研究由此产生。J. S. Liu 与 L. Y. Y. Lu 通过调整满足搜索限制条件的阈值范围, 在全局主路径方法的基础上, 选出最优与若干次优遍历路径, 抽取多主路径^[32]。J. S. Liu 与 L. Y. Y. Lu 还在数据包络分析领域进行了实证研究^[36]。祝青松则在此基础上提出了基于多主路径的关键路径搜索方法, 将局部路径搜索方法和全局路径搜索方法相结合, 通过降低遍历权重阈值得到由全局主路径、全局次路径、基于起点的局部主路径、基于终点的局部主路径等融合形成的多主路径^[37]。陈亮提出了基于语义的多主路径抽取方法, 在通过文档的主题相似度抽取主路径并计算出主路径权重后, 首先挑选出权重最大的技术主题路径作为目标网络, 然后按照路径权重降序依次将其余技术主题路径逐一并入目标网络, 直至得到包含路径较少但弱独立子网数目较多的目标网络, 形成技术主题骨架^[35]。

3 研究现状评述

总体而言, 当前的引文网络主路径研究, 主要关注经典路径搜索算法的应用, 比较经典算法的应用效果, 通过实证来辨析算法的合理性与可行性; 同时, 也针对不同的应用目的, 从不同角度出发对经典算法进行调节, 以期改善主路径提取效果。当前主路径研究过程的基本思路大多遵循如下步骤: 构建专利引证网络、析出最大子图、运用某一搜索算法提取主路径、通过对比解读不同时间间隔的网络主路径形态来总结技术演化主路径特点。

综观之下, 当前的引文网络主路径研究主要存在下述不足之处:

3.1 未充分反映技术演化驱动力的多元性与系统性

现有研究大多从单一角度出发来选择技术主路径的演化方向,如根据引文网络连边的连通性大小或网络节点主题相似性大小,来选取路径发展的下一节点。事实上,技术演化是技术领域内部动力、外部环境发展推力、需求拉力等多种合力综合作用下的发展结果,技术演化过程正是由于这些综合作用而呈现出了技术结构性、路径选择性、发展持续性等特点。因此,专利引证技术演化路径分析不应仅仅局限于从专利引证关系来看待技术的继承与整合,而应该将技术演化发展视为一个多目标作用下的多因素群组决策问题。当前研究中,仅仅依据单一目标属性值来提取主路径,忽视了不同影响要素对技术演化进程的综合影响效果,不能充分反映出技术演化驱动力的多元性与系统性特点。因此,有必要对影响主路径产生的诸多因素进行系统梳理,从系统性、工程化的角度解构其间的不同作用要素与作用机理,研究多因素作用下的路径识别方法,笔者所在研究团队正在开展相关工作。

3.2 未区分不同引证动机对引文主路径发展的影响差异

专利引文主要包括两种类型:一是审查对比文件,指专利审查员在审查过程中参考或引用的专利文献或非专利文献;二是发明人引用文献,指发明人在发明创造过程中参考的专利文献与非专利文献^[38]。专利法规定,专利引文是审查专利权利要求范围的参照物,每一条专利引文都对应着拒绝或限定一项权利要求的决定^[39]。因此,发明人可能会出于利益驱动,在专利申请文件撰写过程中本能地回避某些相同或相似的在先技术,或者列举其他技术存在的缺陷,以免妨碍自己的申请获得授权;审查员的施引行为多发生在专利审查过程中,引用对比文献作为说明知识来源、判断专利申请“三性”、限定授权范围的基准^[40-42]。由此可见,专利引用行为是受专利法约束、由垄断利益驱动并导致创造性毁损的复杂社会行为^[43]。不同引证主体的专利引文关系所揭示的专利技术继承性与相关性程度差异较大。已有研究陆续指出,发明人的他引信息与专利申请的相关程度远低于审查员的引文信息,审查员引文信息与技术本身相关的概率接近发明人引用信息的2倍^[44-46]。

因此,当前主路径方法均平等地对待不同施引主体的专利引用关系,而忽略了专利申请人和专利审查员截然不同的引用目的使所出具的专利引文信息的代

表意义具有的偏差和不确定性,在技术路径演化方向上,不可避免地带来结论性的误差。李睿通过区分审查员施引、发明人自引、发明人他引等3种不同类型专利引文揭示科学-技术关联的程度差异,改进了相应的关联探测模型^[47],该研究验证显示了辨别引证动机对引用关系研究效果产生的影响,其研究思路与方法对于引用主路径研究具有参考意义。

3.3 对引用网络主路径演化的动态特性关注不够

专利引文网络是专利技术发展的时序网络。实验发现,由于引证现象中不可避免的时间累积效应,无论专利还是科技论文,高被引文献往往更多地出自领域的早期产出。由于偏好依附是驱动网络演化的一个重要机制^[48],这类具有首动优势^[49]的文献在网络演化过程中更容易取得较大的节点连通性,也即在时间累积效应影响下,早期文献更容易获得高被引机会,在网络路径中具有更高的遍历频次,从而体现出连接重要度。相应地,高被引文献的网络关系显得更为凝聚,主要围绕特定节点形成关系网络,主路径组分也会展示更多的时间传递性与分块性^[50],引文网络主路径构成组分也呈现出一定的先发效应,早期文献占据了较大优势^[51]。

当前大部分专利引文网络主路径分析都关注静态网络研究,基于特定拓扑结构下展开。这些研究忽略了技术演化进程中时间因素对路径重要性的影响,缺乏一个清晰而成熟的含时框架来进行式地描述含时网络中的专利技术路径的提取和预测。已有学者关注引文网络路径的动态特征研究,如吉林大学宁景博^[24]将普赖斯的学科领域文献老化理论引入引文网络最优路径探测中,利用负指数曲线拟合引文被引用频次分布,测度引文权重随时间的变化。但总体而言,对路径随时间演化的动态特征及对链路预测研究的重要性,缺乏深入的刻画,对于有向、含时、加权的技术演化网络的路径研究较少。

3.4 多主路径方法本质上仍是对单目标搜索结果的阈值调控

当前的多主路径研究方法主要分两类情况,一类是通过调节满足搜索条件的阈值范围,抽取出基于某一目标的最优与若干次优遍历路径;另一类则是抽取出符合某一目标条件的局部主路径、全局主路径,融合形成多主路径。上述两类研究,其本质上都是基于单一目标(路径连通性或主题相关性)进行路径搜索,抽取的多主路径只是基于对单目标条件最大化的满足程

度与范围的调控。

技术创新活动具有持续性、选择性、累积性特点,在此影响下,技术路径识别的判断目标除了路径连通性、主题相似度之外,也有可能是路径总长度、路径时间跨度等,不同搜索目标各有优劣,互为相长。同时,各目标又可能处于冲突状态,例如:当路径满足连通性最大目标时,路径的相似度总和未必是最高的;当路径相似度总和最高时,未必是全网最长的演化路径。表面上,对于这些相互冲突、不可公度的目标,很难存在使所有目标同时达到最优的绝对最优解。因此,建立一套主路径问题模型与解决方案,实现从不可公度甚至可能冲突的不同目标出发、在合理的时间内找出高质量的主路径集,这是笔者所在研究团队正在着力解决的问题。

4 未来研究展望

针对当前研究中存在的局限,本文从算法思想、约束条件、方法适用性与实用性、独特性4个方面提出关于未来专利引用网络主路径研究的几点思考:

(1) 对算法思想进行实质性、创新性拓展,更全面地反映技术路径发展的系统性与多元性。主路径方法的最大优势在于通过对大型复杂网络的主路径提取、识别来实现对复杂网络的降维与简化,找出网络中最重要的主干网。但当前研究中存在搜索目标单一化、算法固定性与过度选择性等缺点,对主路径结果揭示技术领域演化主流的准确性、演化结构的精细性造成了固有约束。技术演化是多因素的“合力”作用结果,专利技术演化路径是多元驱动力下发展方向的综合性选择,因而专利技术演化路径问题应该被视作一个多目标决策问题,技术演化主路径的发展方向,体现了技术领域演化过程中各阶段影响力的多目标优化结果。今后在算法基本思想上,应该更多地借鉴多目标进化思想,尽可能兼顾技术路径演化的不同目标,尤其是不可公度甚至可能有所冲突的技术演化目标,研究同时满足多目标优化的技术演化主路径问题模型与相应的解决方案。

(2) 把握网络分析法的约束条件,提升主路径方法与结果的科学性与权威性。仅就当前单纯的节点相似性方法或节点间连接方法,是不足以完全反映出引文网络路径的本质结构特征和发展趋势的,必须把握各类约束条件对引文网络发展的动态性、客观性、未来性的影响,才能完整有效地把握引文网络演化趋势与

主路径走向。例如:在网络静态拓扑特征研究基础上,进一步关注引证行为的时间累积效应,加强专利引用网络的动态性特征、对技术主路径发展的约束和影响、未来前景与趋势推演研究,有助于进一步提高专利引文主路径方法的科学性与未来预见性;深入辨析施引主体的动机差异对技术路径方向选择的影响,有助于增强主路径提取结果的客观性与公允性;除内容无涉的分析方法外(如网络连通性指标),结合更多与内容有关的技术演化属性(譬如基于专利路径的语义关系揭示),方能增强主路径研究结果的领域专业性与权威性。

(3) 兼顾算法的效果与效率,增强主路径方法的适用性与实用性。起源于 N. P. Hummon 和 P. Dorian 的主路径方法的基本思想是基于网络理论的路径穷尽搜索,当引证网络规模较大时,算法的时间复杂度高,容易引起计算溢出,算法适用性差。当前虽有研究通过增加主成分分析预处理步骤来缩小数据源范围,或是改善初始搜索起点专利的选取方法来提高分析效率,但这些都是基于单目标穷尽搜索原理的枝节性修正,未能从方法本质上克服问题。未来研究重点之一将是对主路径方法的本质改进,兼顾搜索效果与效率最优的同时,实现在合理时间内找出质量较高的主路径或多主路径。在多目标情形下通常得到的不是一个最优解而是一个帕累托最优解集(pareto optimality set),传统的精确算法难以在合理时间内得到问题的帕累托最优解集,因此,采用启发式算法求解引文网络主路径问题的帕累托近似最优解集(pareto approximation set),具有较强的适用性与实用性,将为专利引文网络多主路径研究提供新视角,笔者所在研究团队在此方面已取得一些阶段性进展。

(4) 深挖专利信息内涵,发挥专利主路径方法在产业化扩散研究中的独到性。专利引用研究视角能够提供论文引用研究视角下不具备的独特线索:专利文献内容多反映技术、产品、法律信息,与产业和市场的关系更加紧密,专利引用流中蕴含着发明创造的技术、产业与市场扩散信息,专利引用主路径除了揭示技术领域演化过程外,还蕴含着专利技术的产业化扩散途径,譬如一项技术创新由发明创造原理到技术实现再到产品成型,从技术方法保护到产品保护再到用途保护,从技术起源地流到目标市场等。因此,专利引用网络主路径研究应该开展更多有别于论文主路径的独有探索与思考,例如通过专利引用主路径研究技术创新

的产业化途径,通过路径关键节点筛选识别关键共性技术、跨界点,通过路径演化预测研究技术的产业化前景与发展趋势,这些都是专利引用网络主路径研究的独到优势,具有广阔的研究与应用空间。

参考文献:

- [1] 刘倩楠. 基于专利引文网络的技术演进路径识别研究——以“以太网技术”为例[D]. 大连: 大连理工大学, 2010.
- [2] 刘小玲. 基于专利网络的技术演进研究[D]. 北京: 中国科学院研究生院, 2011.
- [3] HUMMON N P, DOREAIN P. Connectivity in a citation network: the development of DNA theory[J]. *Social networks*, 1989, 11(1): 39-63.
- [4] 韩毅, 金碧辉. 基于连通性的引文网络结构分析新视角: 主路径分析[J]. *科学学研究*, 2012, 30(11): 1634-1640.
- [5] 陈亮, 张志强. 技术演化研究方法进展分析[J]. *图书情报工作*, 2012, 56(17): 59-66.
- [6] BATAGELJ V. Efficient algorithms for citation network analysis [EB/OL]. [2015-10-14]. <http://www.researchgate.net/publication/1956732>.
- [7] HOM H C, LIN V H, LIU J S. Exploring knowledge diffusion among nations: a study of core technologies in fuel cells[J]. *Scientometrics*, 2014, 100(1): 149-171.
- [8] 彭爱东, 黎欢, 王洋. 基于专利引文网络的技术演进路径研究——以激光显示技术领域为例[J]. *情报理论与实践*, 2013, 36(8): 57-61.
- [9] 许琦. 基于专利引证网络的技术范式分析——以半导体制造领域为例[J]. *图书情报工作*, 2013, 57(4): 112-119.
- [10] VERSPAGEN B. Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research[J]. *Advances in complex system*, 2007, 10(1): 93-115.
- [11] DAVID B T, FERNANDO J S, ITZIAR C M. Mapping the importance of the real world: the validity of connectivity analysis of patent citations networks[J]. *Research policy*, 2011, 40(3): 473-486.
- [12] MARTINELLI A. Technological trajectories and industry evolution: the case of the telecom switching industry [EB/OL]. [2015-10-14]. <https://pure.tue.nl/ws/files/3279369/Metis222759.pdf>.
- [13] CHOI C, PARK Y. Monitoring the organic structure of technology based on the patent development paths[J]. *Technological forecasting and social change*, 2009, 76(6): 754-768.
- [14] TRAJTENBERG M. A penny for your quotes: patent citations and the value of innovations[J]. *RAND journal of economics*, 1990, 21(1): 172-187.
- [15] MOGEE M E, KOLAR R G. International patent analysis as a tool for corporate technology analysis and planning[J]. *Technology analysis & strategic management*, 1994, 6(4): 485-503.
- [16] 王燕玲. 技术轨道识别研究——以专利引文网络主路径分析为方法[D]. 武汉: 武汉大学, 2013.
- [17] 韩毅. 基于引文网络主路径的领域演化结构与知识扩散研究[D]. 北京: 中国科学院研究生院, 2011.
- [18] GUAN J C, SHI Y. Transnational citation, technological diversity and small world in global nanotechnology patenting[J]. *Scientometrics*, 2012, 93(3): 609-633.
- [19] YANG Z K, HUANG Y, LIU J, et al. Research on identification of technological trajectory based on patent citation network—taking wind motors technology as an example [C]//Proceeding of the 2012 IEEE ISMOT. Hangzhou: IEEE Press, 2012: 216-220.
- [20] XU Y S, HUA X F. Mapping technological trajectories as patent citation networks: taking the Aero-Engine industry as an example [C]//Proceedings of the 2014 ICMET. Kanazawa: IEEE, 2014: 2827-2835.
- [21] YUAN F, MIYAZAKI K. Understanding the dynamic nature of technological change using trajectory identification based on patent citation network in the electric vehicles industry [C]//Portland international conference on management of engineering & technology. Kanazawa: 2014.
- [22] FONTANA R, NUVOLARI A, VERSPAGEN B. Mapping technological trajectories as patent citation networks: an application to data communication standards[J]. *Economics of innovation and new technology*, 2009, 18(4): 311-336.
- [23] MARTINELLI A. An emerging paradigm or just another trajectory? understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry [J]. *Research policy*, 2012, 41(2): 414-429.
- [24] 宁景博. 基于最优路径的学术网络重要文献探测[D]. 长春: 吉林大学, 2015.
- [25] 范维熙, 费钟琳. 基于德温特专利引文网络的技术演进路径研究——以太阳能电池技术为例[J]. *情报杂志*, 2014, 33(11): 62-66.
- [26] 张伟. 基于专利引用的碳捕获与封存技术发展研究[D]. 北京: 北京工业大学, 2013.
- [27] 黄鲁成, 石媛嫻, 吴菲菲. 基于专利引用的技术轨道动态分析——以太阳能电池为例[J]. *科学学研究*, 2013, 31(3): 358-367.
- [28] 潘颖. 基于专利引证强度的关键技术发展路径研究[J]. *情报理论与实践*, 2014, 37(12): 71-75.
- [29] 董克, 刘德洪, 江洪, 等. 基于主路径分析的 HistCite 结果改进研究[J]. *情报理论与实践*, 2011, 34(3): 113-116.
- [30] 韩毅, 周畅, 刘佳. 以主路径为种子文献的领域演化脉络及凝聚子群识别[J]. *图书情报工作*, 2013, 57(3): 22-26, 55.
- [31] LIU J S, CHEN H H, HO M H C, et al. Citations with different levels of relevancy: tracing the main paths of legal opinions [J]. *Journal of the Association for Information Science and Technology*, 2014, 65(12): 2479-2488.

- [32] LIU J S, LU L Y Y. An integrated approach for main path analysis: development of the hirsch index as an example[J]. Journal of the American Society for Information Science and Technology, 2012, 63(3): 528-542.
- [33] 许琦. 一种面向技术进化的知识适应能力评价方法: 基于专利引证网络的知识遗传分解[J]. 情报理论与实践, 2013, 36(3): 68-76, 55.
- [34] 许琦, 顾新建. 一种基于 Subject-Action-Object 三元组的知识基因提取方法[J]. 浙江大学学报(工学版), 2013, 47(3): 385-399.
- [35] 陈亮, 杨冠灿, 张静, 等. 面向技术演化分析的多主路径方法研究[J]. 图书情报工作, 2015, 59(10): 124-130, 115.
- [36] LIU J S, LULY Y, LU W M, et al. Data envelopment analysis 1978-2010: a citation-based literature survey [J]. Omega, 2013, 41(1): 3-15.
- [37] 祝青松. 语义增强的引文分析方法与实验研究[D]. 北京: 中国科学院大学, 2015.
- [38] 杨铁军, 曾志华. 专利信息利用技能[M]. 北京: 知识产权出版社, 2011.
- [39] GARFIELD E. Patent citation indexing and the notions of novelty, similarity and relevance [J]. Journal of chemical documentation, 1966, 6(2): 63-65.
- [40] JAFFE A B, TRAJTENBERG M, HENDERSON R. Geographic localization of knowledge spillovers as evidenced by patent citations [J]. Quarterly journal of economics, 1993(8): 577-598.
- [41] HALL B. Innovation and market value[M]//Productivity, innovation and economic performance. Cambridge: Cambridge University Press, 2000: 177-198.
- [42] 李睿, 孟连生. 论专利引用行为与期刊论文引用行为在揭示知识关联方面的差异[J]. 情报学报, 2010, 29(3): 474-478.
- [43] 李睿, 孟连生. 论专利间引用关系分析中存在的问题[J]. 情报理论与实践, 2009, 32(7): 39-43.
- [44] BHATTACHARYA S, KRETSCHMER H, MEYER M. Characterizing intellectual spaces between science and technology [J]. Scientometrics, 2003(2): 369-390.
- [45] CRISCUOLOA P, BART V. Does it matter where patent citations come from? inventor vs. examiner citation in European patents [J]. Research policy, 2008, 37(10): 1892-1908.
- [46] ALCACER J, GITTELMAN M. Patent citations as a measure of knowledge flows: the influence of examiner citations [J]. The review of economics and statistics, 2006, 88(4): 774-779.
- [47] 李睿. 基于专利引文分析的科学与技术关联探测模型改进[D]. 北京: 中国科学院研究生院, 2011.
- [48] LI M H, GAO L, FAN Y, et al. Emergence of global preferential attachment from local interaction [EB/OL]. [2015-10-14]. <http://iopscience.iop.org/article/10.1088/1367-2630/12/4/043029/pdf>.
- [49] NEWMAN M E J. The first-mover advantage in scientific publication [EB/OL]. [2015-10-14]. <http://iopscience.iop.org/article/10.1209/0295-5075/86/68001/pdf>.
- [50] 樊志伟, 韩芳芳, 刘佳. 引文网络的主路径特征研究——以富勒烯领域为例[J]. 图书情报工作, 2013, 57(3): 17-21, 60.
- [51] 韩毅, 童迎, 夏慧. 领域演化结构识别的主路径方法与高被引论文方法对比研究. 图书情报工作, 2013, 57(3): 11-16.

作者贡献说明:

张娴: 研究方法设计, 研究过程实施, 论文撰写;

方曙: 研究选题, 论文修订。

Research Review on the Main Path Analysis of Patent Citation Networks

Zhang Xian^{1,2} Fang Shu¹

¹Chengdu Library and Information Center, the Chinese Academy of Sciences, Chengdu 610041

²University of Chinese Academy of Sciences, Beijing 100190

Abstract: [Purpose/significance] In this paper, some typical methods of the main path analysis (MPA) of the patent citation network are reviewed in order to provide new ideas for further research. [Method/process] Firstly, the significance of MPA of the patent citation network for technology evolution research was elaborated. Then, a comprehensive literature review of the MPA research based on the patent citation network was conducted. [Result/conclusion] The research achievements are summarized in three aspects: the algorithm research, the applied research, and the method-optimization research. The limitations in current MPA research are discussed. Finally, the possible research work in the future are described.

Keywords: patent citation citation network main path analysis (MPA) technological development path technology evolution