

Methodological Exploration and
Empirical Research on Patent Analysis

专利分析的 方法探索与实证研究

中国科学技术信息研究所◎编著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

专利分析的方法探索与实证研究

中国科学技术信息研究所 编著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北京 ·

图书在版编目 (CIP) 数据

专利分析的方法探索与实证研究 / 中国科学技术信息研究所编著. —北京: 科学技术文献出版社, 2016. 8

ISBN 978-7-5189-1748-8

I. ①专… II. ①中… III. ①专利—分析 IV. ①G306

中国版本图书馆 CIP 数据核字 (2016) 第 173601 号

专利分析的方法探索与实证研究

策划编辑: 周国臻 责任编辑: 赵 斌 责任校对: 赵 媛 责任出版: 张志平

出 版 者 科学技术文献出版社
地 址 北京市复兴路15号 邮编 100038
编 务 部 (010) 58882938, 58882087 (传真)
发 行 部 (010) 58882868, 58882874 (传真)
邮 购 部 (010) 58882873
官 方 网 址 www.stdp.com.cn
发 行 者 科学技术文献出版社发行 全国各地新华书店经销
印 刷 者 北京高迪印刷有限公司
版 次 2016年8月第1版 2016年8月第1次印刷
开 本 710×1000 1/16
字 数 235千
印 张 14.5
书 号 ISBN 978-7-5189-1748-8
定 价 68.00元



版权所有 违法必究

购买本社图书, 凡字迹不清、缺页、倒页、脱页者, 本社发行部负责调换

《专利分析的方法探索与实证研究》编写组

主 编	佟贺丰	傅俊英	
撰写人员	陈 亮	梁琴琴	徐 峰
	杨冠灿	高 楠	邵黎明
	赵 龙	雷孝平	曹 燕
	贡 强	郑 佳	孟 浩

前 言

在市场经济和国际化竞争的大环境下，知识产权对科技创新发挥创值与增值、激励、规范、评价和保护作用，对创新成果的权利归属、利益分配及维系创新可持续性的作用越来越突出。专利信息作为知识产权信息重中之重，是集技术、经济、法律信息于一体的综合性信息，也是关系国家产业发展安全的基础性、战略性资源。

开展专利信息分析与服务是实施国家“创新驱动发展”战略的重要组成部分，也是国家科技创新体系有效运转的重要支柱。在技术研发、应用与扩散全过程的创新链条所有环节，以及企业、科研院所和政府部门等在内的各类创新主体均对专利信息服务产生了大量需求。

中国科学技术信息研究所自2009年开展专利信息分析与服务以来，进行了专利信息分析方法的多种探索和尝试。从2013年开始，以本研究团队为主，形成了一个专利分析方法研究小组，并吸收了中国科学技术信息研究所的部分研究生成为小组成员。我们的研究聚焦于：技术从过去到现在的发展路径是什么样的？目前的研究前沿有哪些？未来的发展方向有哪些？我国和国外的技术差距有多大？这些判断都要基于专利分析，其中一个重要基础就是要识别出核心专利，并且能够识别出空白的技术点，来支撑研发。经过两年的相关部署及讨论，通过方法探索与实证研究，完成了以下研究内容，并形成本书。

佟贺丰、傅俊英负责整个研究的技术路线与任务分工。在成书过程中，佟贺丰撰写导论部分，陈亮、杨冠灿、雷孝平、郑佳撰写第1章，赵龙、孟浩撰写第2章，邵黎明、佟贺丰撰写第3章，高楠、傅俊英撰写第4章，傅俊英、佟贺丰、曹燕撰写第5章，梁琴琴、负强撰写第6章，徐峰撰写第7章。希望本书的出版对专利分析方法的探索能够起到抛砖引玉的作用，为提升国内专利分析和服务水平做出一点贡献。

佟贺丰

目 录

0 导论	1
0.1 专利分析的重要意义	1
0.2 中国科学技术信息研究所的专利分析与服务工作	2
0.3 本研究的主要内容	6
1 基于专利文献的技术发展主路径和发展阶段研究	8
1.1 技术发展主路径研究方法	8
1.2 实证研究	11
1.3 讨论	18
参考文献	19
2 基于专利耦合和文本挖掘的技术演化分析	20
2.1 技术演化分析方法的研究现状	20
2.2 基于专利耦合和文本挖掘的技术演化分析方法设计	25
2.3 实证研究	32
2.4 讨论	57
参考文献	59
3 基于专利文献和知识图谱的技术预测方法研究	63
3.1 技术预测方法的定义及发展	63
3.2 专利文献和知识图谱结合的技术预测方法	70
3.3 实证研究	87
3.4 讨论	97
参考文献	99

4	基于专利耦合和共被引融合方法的研究前沿识别	104
4.1	研究前沿的定义及概况	104
4.2	基于引用关系识别研究前沿的进展	106
4.3	基于专利耦合和共被引识别研究前沿	109
4.4	实证研究	119
4.5	讨论	161
	参考文献	162
5	基于专利耦合的技术前沿和技术差距研究	167
5.1	技术前沿和技术差距研究进展	167
5.2	文献耦合分析方法的研究进展	169
5.3	实证研究	170
5.4	讨论	184
	参考文献	185
6	核心专利识别方法研究	187
6.1	核心专利的概况及识别方法进展	187
6.2	本章拟采用的核心技术识别方法	189
6.3	实证研究	190
6.4	讨论	201
	参考文献	202
7	基于计量的专利技术形态分析	204
7.1	专利技术形态分析方法概述	204
7.2	基于计量的专利技术形态分析方法	209
7.3	实证研究	214
7.4	讨论	220
	参考文献	222

0 导 论

0.1 专利分析的重要意义

创新已成为国家的发展战略。在党的十八届五中全会确定的“五大发展理念”中，“创新”被排在首位，并处在“国家发展全局的核心位置”。“创新是引领发展的第一动力”，这是十八届五中全会对创新在国民经济、社会发展中地位和作用的全新概括。创新驱动的本质是指依靠自主创新，充分发挥科技对经济社会的支撑和引领作用。技术创新是创新的重要组成部分。专利与技术创新的关系十分密切，是世界上最大的技术信息源。专利既是技术创新活动中重要的制度安排，又是创新过程的重要产出。专利对技术创新可以发挥创值与增值、激励、规范、评价和保护作用，对创新成果的权利归属、利益分配及维系创新可持续性的作用越来越突出。

专利分析与服务是技术创新的重要支撑。专利分析与服务通过对专利文献中包含的技术信息、经济信息、法律信息进行科学的加工、整理与分析，形成具有较高技术与商业价值的专利情报，服务各类决策。专利分析与服务面向整个创新链的全过程，服务对象涵盖政府、企业、高校、科研院所等创新主体，从法律、经济等各个层面促进创新成果的运用和扩散，从而提高创新的绩效，维系创新的可持续发展。首先，专利分析与服务对企业创新活动中的知识产权创造、运用、保护和管理起着重要的促进作用。企业可以通过专利信息了解竞争对手的技术水平，跟踪最新技术发展动向，提高研发起点，加快产品升级和防范知识产权风险。其次，专利分析与服务为我国产业或行业标准制定和实施“自主创新战略”发挥重要作用。专利信息作为一种重要的科技资源，其有效利用可以掌握当今世界科技发展的进展、动向和趋势。最后，专利分析与服务还可以辅助政府的公共科技资源配置。

0.2 中国科学技术信息研究所的专利分析与服务工作

中国科学技术信息研究所（以下简称中信所，英文简称 ISTIC）自 2009 年起投入大量人力、物力，从事面向深度分析的专利信息资源加工、工具与平台开发、专利分析与服务等一系列工作。具体的工作包括以下几方面。

0.2.1 ISTIC-专利信息资源建设

专利信息资源是专利信息服务的基础和保障，中信所的专利信息资源建设目标是建成具有自主知识产权的深加工专利信息资源储备。中信所专利信息资源按加工内容可分为以下几种。

① ISTIC-专利分析数据库：是中信所在综合了专利题录信息数据源、专利法律状态信息、规范化企业信息的基础上，建设的具有自主知识产权的面向统计分析、学术研究、政策领域研究的分析性深加工专利数据库。在 ISTIC-专利分析数据库的建设层次上，划分为基础库、重点国家库、重点技术领域库和预研专利数据库 4 个层次，分别对应于宏观统计、中观分析、领域研究、探索性方法研究等不同层次的研究需求。

② ISTIC-专利权属机构权威控制数据库：建立专利权属机构的权威控制数据库，规范化专利权属机构名称，解决同一专利权属机构在不同专利中不同名的问题；根据机构的隶属关系和沿革，建立专利权属机构的层级关系。

③ ISTIC-产业/领域技术分类体系：围绕重点科技产业/领域，根据产业/领域分析需要，建立基于国际专利分类号和关键词的多层次技术分类体系。

④ ISTIC-产业/领域技术专家数据库：以重点科技产业/领域的专利发明人为线索，建立技术专家数据库。

⑤ ISTIC-产业/领域主题词表：以重点技术产业/领域为核心，对该产业/领域专利信息中的同义词、近义词、多义词及上下位概念词等不同类型的词，进行严格的控制和规范，并构建和规范词与词之间的关系。

0.2.2 ISTIC-专利分析方法与技术研究

专利分析是对专利文献蕴含的技术、法律和商业信息等内容，经分解、加工、标引、统计、分析、整合和转化等信息化手段处理后，进行全面统计、分析、解读的过程。中信所在研究过程中，主要关注以下方法和技术。

①专利信息资源整合与加工的关键技术与规范研究:研究专利信息资源的整合、清洗、标引及专利衍生信息的加工方法,以建立可以规范化专利信息资源整合加工全流程的关键技术与规范。

②专利质量评价方法研究:研究通过可获取的专利信息评价专利质量的方法和技术。

③专利知识抽取技术研究:研究从专利信息资源中识别、发现和提取出技术、方法、工艺、技术进步的内容及其相互关系和约束规则的方法及技术。

④基于文本挖掘的专利技术聚类方法研究:研究通过识别一批专利信息资源中的技术主题并对其进行聚类,来寻找技术热点的方法和技术。

⑤专利信息融合分析核心技术研究:整合专利文献中的申请人、法律状态、同族专利、引证文献等关键信息,研究专利信息动态关联分析关键技术。

⑥专利信息智能检索与语义导航关键技术研究:研究基于概念检索的海量专利数据智能搜索引擎、多维索引体系的构建、交互性立体式检索结果揭示、个性化推荐等关键技术。

⑦基于专利的科技创新测度指标体系:研究以专利为基础的评价科技创新能力的指标体系。

0.2.3 ISTIC-专利分析与服务工具开发

在形成一定的研究方法与技术以后,中信所又对这些方法与技术工具中进行集成,以利于实现工程化的专利信息服务。ISTIC-专利分析与服务工具开发以 ISTIC-专利分析与服务技术研究为基础,为 ISTIC-专利信息资源建设和专利分析与服务平台建设提供工具支持。

① ISTIC-专利数据加工、整合系统开发:开发能够按照既定数据规范对多种来源、种类的专利信息进行加工、整合的工具系统。

②专利知识抽取工具:开发通过识别一批专利信息资源中的技术主题并对其进行聚类来寻找技术热点的工具。

③专利信息自动标引工具集:开发根据专利内容特征和给定规则,对专利信息自动进行技术标引的工具。

④专利评估与预警分析工具:开发基于专利特征信息和语义处理技术的专利评估与专利预警相关工具。

⑤专利分析结果可视化工具:开发能够以图像的方式对专利分析结果进行展示,帮助用户理解专利分析结果内在含义的可视化工具。

0.2.4 ISTIC-专利分析与服务平台建设

ISTIC-专利分析与服务平台是 ISTIC-专利信息资源建设、技术研究及工具开发积累的相关成果向外提供服务的窗口。

① ISTIC-专利信息检索分析平台：以 ISTIC-重点国家专利数据库为基础面向统计分析、创新决策及专利分析方法研究的专利信息检索分析平台。该平台以中信所自有知识产权的 ISTIC-专利分析数据库为基础，向用户提供专利信息检索、数据标引、专利分析等服务，并向研究人员提供管理数据、协同工作的环境。

② ISTIC-重点技术领域专利分析与服务平台：以 ISTIC-重点领域专利数据库为基础，向用户提供经过专家标引的领域专利信息资源，并提供领域技术分类导航、领域研发机构导航及领域专利信息检索分析功能的创新信息服务平台。

0.2.5 ISTIC-专利分析与服务方向

ISTIC-专利分析与服务能力建设按照信息服务的方式，可以分为专利信息资源服务、技术服务、咨询服务 3 类。

(1) ISTIC-专利信息资源服务

ISTIC-专利信息资源服务包括数据服务和平台服务。

数据服务是指以中信所积累的自有知识产权的专利信息资源为基础，向用户直接提供裸数据和定制专利数据库的服务。

平台服务是指以 ISTIC-专利信息检索分析平台、ISTIC-重点技术领域专利分析与服务平台为基础，以账号管理的方式向用户提供专利信息资源的检索、分析及协同工作环境等信息服务。

领域平台服务，集成数据、检索、分析及其他信息为一体的细分领域平台，包括电动汽车、印刷电子、新能源、生物医药等领域。

(2) ISTIC-专利信息技术服务

ISTIC-专利信息技术服务是指以中信所积累的专利信息技术和开发的专利分析与服务工具集为基础，向用户提供专利信息资源加工工具、专利分析工具、专利信息自动标引工具等系列工具的使用权，或者根据用户需要定制相关工具的服务。

(3) ISTIC-专利信息咨询服务

在相关资源、技术、平台的基础上，以 ISTIC-专利分析团队为核心为科技管理部门和创新主体提供相关的分析报告，包括：

①各国专利数据的统计分析：结合论文统计，对专利数据在国家层面进行统计发布、监测分析。与论文统计结果一同发布，获得社会影响力。

②基于区域的专利统计与分析报告：以区域为统计维度，以新兴产业技术为统计对象的专利深度分析研究，明确定位于中观的空间分析视角，区别于宏观层面上的国家统计和微观层面上的专利权人统计。

③重点技术领域专利分析报告：以专利等事实数据为基础，采用定性与定量相结合的分析方法，系统研究特定、重点技术领域的发展趋势，并结合我国科技发展现状，提出相关领域技术研发及产业化发展策略。

ISTIC-专利分析与服务体系由资源、技术、工具、平台等几部分共同构成，如图0-1所示。其中，专利信息资源建设为专利分析与服务提供了信息基础，技术研究为资源建设、工具/平台开发及信息服务提供了理论和方法的支撑，专利分析与服务工具为平台与服务提供了技术的保障，而信息服务平台则为信息服务的输出提供了桥梁与出口。

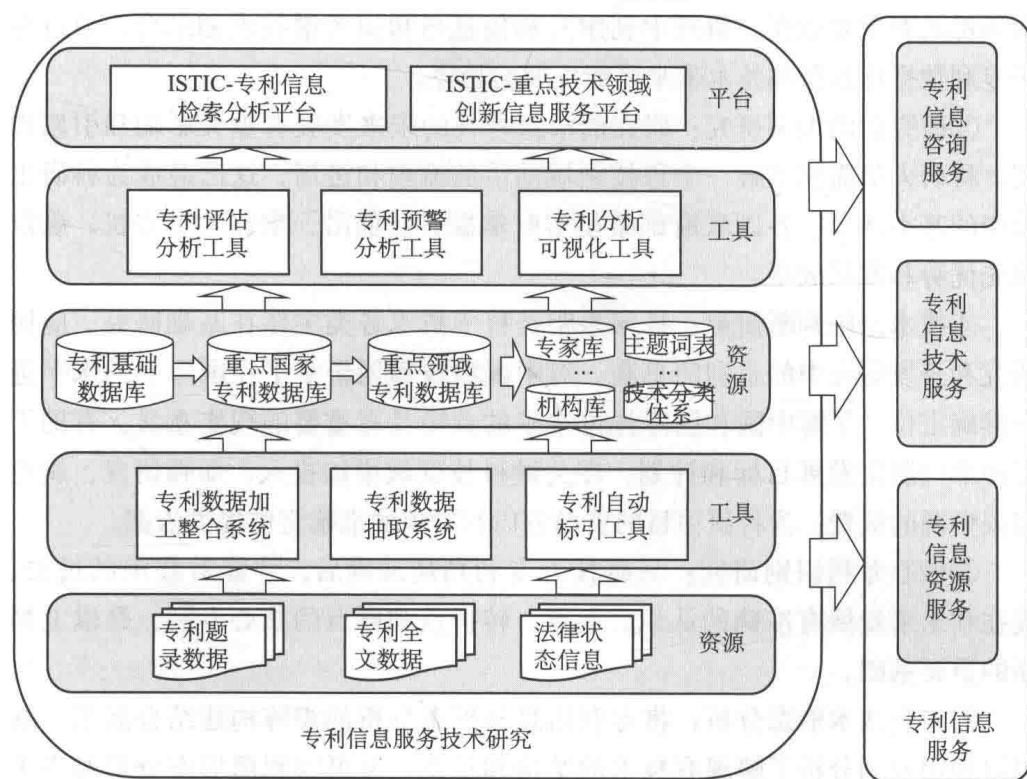


图0-1 ISTIC-专利分析与服务体系框架

0.3 本研究的主要内容

中信所早期的专利分析方法以统计分析为主,经过几年的研究准备,逐步向技术分析方法前进。本书主要从以下几个角度进行了研究。

①技术发展路径分析:全球新一轮科技革命和产业变革加速演进,以智能、绿色、泛在为特征的群体性技术革命将引发国际产业分工重大调整,颠覆性技术不断涌现。理清技术发展的过程,把握技术改进的规律,明确发展历程中的技术变迁路径,对企业和政府决策者都有很强的现实意义。研究中探索了两种方法,基于专利文献的技术发展主路径和基于专利耦合和文本挖掘的技术演化分析方法。

②技术预测研究:关注的是未来技术发展的趋势。技术预测对于国家确定技术开发重点和技术结构,制定技术政策和技术规划,对企业确定研发方向等都具有重要意义。研究中利用专利信息与知识图谱技术相结合,通过分析专利数据标签预测技术未来可能发展的趋势。

③研究前沿识别研究:研究前沿对科技的未来发展有重大影响和引领作用。科研人员需要了解一个科技领域前沿的概貌和进展,这已经成为科研工作基本环节。各国政府都希望实时掌握科技前沿,掌握发展先机,确定领先优势和发展战略。

④技术差距判断研究:技术差距是科学技术各类主体在基础研究、应用研究和开发研究中的差别的总和。对中国科技和创新水平在国际上的水平进行准确定位,了解中国和国际先进水平的差距具有重要的现实意义。有助于管理部门制定发展目标和计划,在关键科技领域增加投入、加强研发,避免有限资源的浪费,为科研项目的资助方向和对象的准确定位提供依据。

⑤核心专利识别研究:只有核心专利判断准确后,才能对技术的过去、现在和未来发展有准确的认识。起点、转折点和终点的核心专利,是以上判断的重要基础。

⑥专利技术形态分析:将专利信息与形态分析的矩阵构建结合起来,既可以利用专利分析了解现有技术的结构和状态,又可以利用形态分析对未来技术可能的形态进行分析和优选,提供有关未来技术形态组合的优选方案,为技术专家和技术管理专家在制定技术开发战略或专利战略决策时提供参考。

这些研究具有一定的探索性质,有些方法可能不一定具有在全部领域推

广应用的普遍意义,但对专利分析的前沿是有益的尝试。随着时代的发展,以检索、查新、简单统计为主的专利分析已不能满足创新主体和科技管理部门日益深化的科技创新信息需求,中信所这样的科技信息公共服务机构,有必要在专利分析过程中引入新的信息处理和分析技术,在方法和工具上有所突破,探索和研究新的专利分析方法,提升专利信息服务水平。

1 基于专利文献的技术发展主路径和发展阶段研究

特定学科（领域）的引文网络从时间上看是承载历史演化的时序图，从空间上看则是枝蔓丛生、错落有致的主题群落，它展示了学科（领域）知识在时空上的演化与分化特征。时间序列上的这些主题群落不是简单的随机堆砌，而是由该学科（领域）活跃研究人员间相互引用形成的精细结构。

引文网络刻画并展示了该学科的发展历史和演化路径。早期基于引文网络的研究大多以高被引论文为对象，通过论文间的相似性（如共词、同被引、共引等）聚类来探索学科的演化结构。在引文网络中，判断一个节点是否为关键节点，不能仅关注该节点的单一属性（如被引数量或参考文献数量），还应考察它在整个网络中的“整体连通性”，由此得到更具可靠性和说服力的结果。主路径分析正是基于“整体连通性”来离析特定时刻引文网络关键路径的方法，是不同于以高被引论文为基础的引文网络结构分析新视角。

1.1 技术发展主路径研究方法

1.1.1 主路径方法的提出

主路径方法最早起源于 Garfield 对于科学史的思考，Garfield 认为科学史是一系列基于现有技术的改进，而涌现的新发明的有序排列^[1]。1989 年，Hummon^[2]首次提出了主路径分析方法。他从一个由 40 个节点组成的、代表 DNA 学科的引文网络中提取出一些起中心作用的节点，并由这些节点之间的连接代表了 DNA 学科发展的主要思想变迁。这些研究成果都是应用引文网络分析方法提取特定学科领域的高被引论文，通过共词、共引、引文耦合等节点相似性计算方法来构建关系结构，并结合可视化方法来定义与揭示相应的科学结构，期望能预测出科学结构的未来演化发展趋势。但高被引论文遴选过程中的阈值确定一直是困扰人们的难题之一，因此，另辟蹊径地开发一种更具客观性的关键文献识别方法就摆在了研究者的面前。科学史研究及科学

结构揭示的关键文献识别需求是引文网络主路径分析产生的现实土壤。而基于节点相似性的引文网络分析方法在解决这些现实问题时,遭遇了高被引论文阈值确定的主观性瓶颈。因而,如何根据这种实践需求开发满足需求的方法成为科学计量学领域必须解决的重要问题。

1.1.2 主路径方法的研究内容

主路径分析不同于传统的诸如文献耦合、共被引分析等方法,它更关注网络中节点之间的连接而非节点本身。主路径分析方法最早由 Hummon 和 Doreian 提出。他们提出了3种度量网络中节点之间连线权重的指标:节点对投影数(Node Pair Projection Count, NPPC)、搜索路径链接数(Search Path Link Count, SPLC)、搜索路径节点对(Search Path Node Pair, SPNP)^[3],并在此基础上将深度优先搜索算法和穷举法相结合,来完成主路径的查找。之后,研究者们还陆续开发了诸如多起点路径(network of main path)^[4]、主路径成分(main path component)及多路径(multiple main path)^[5]等方法。

早期的主路径分析方法是先通过指标计算获取网络中所有连接的权重,之后从所有起始点出发,找出下一个连接中权重最大的连接,并以该连接的终点作为起始点,重复此过程直到最后终点。从最初起始点到最后终点之间的路径就是主路径。需要补充的是,如果有若干个连接同时具有最大权重,则同时入选。此方法之所以被称为“局部主路径”,是因为这一算法的每一步只关注当前连接的权重,它强调的是知识传播过程中的局部重要性。

相对于局部主路径还有一种“全局主路径”法,此方法试图在网络整体中找到一条连接权重之和最大的路径。相比较于局部主路径,全局主路径更关注知识传播的整体重要性,而非局部重要性。一般来说,在同一网络中,全局主路径的长度比局部主路径要大一些;另外,由于全局主路径的路径搜索策略并非每一步都是选取以当前节点作为出发点的最大权重连线,因此,其出现路径分叉的机会要小于局部主路径。

(1) 多起点路径

Hummon N P 首次提出主路径方法时发现,在由40个节点组成的引文网络中,从所有有出度的点(即起点和中间点)出发得出的(局部)主路径,最后都收敛于所得到的干流(main stream),再一次验证了主路径方法的代表性。受此思想的启发, Verspagen 从所有的起点出发,得出了所有起点的全局主路径,这些主路径形成了一个“主路径网络”。在这个网络中有多个大小不

等的成分。对此, Fontana 等^[5]解释认为, Hummon 的引文网络过于简单(只含有 40 个节点), 而他所用的网络较大(3371 个节点, 15 506 个连接), 因此并不能保证所有的主路径都收敛于所得到的干流。所有有代表性的路径基本上都收敛于主路径, 因此, 所得到的全局主路径是非常具有代表性的。

(2) 主路径成分^[3]

根据主路径方法, 较易获得一个主路径成分。选取一个 0~1 的阈值, 将所有 SPC 值小于此阈值的连接去除, 并去掉孤立点后得到的网络即为主路径成分。之所以称其为主路径成分, 是因为此阈值往往参考主路径上所有连接的最小 SPC 值(通常选取稍小于最小 SPC 值作为阈值), 以保证主路径被包含在此主路径成分中。当然, 这一步也可能得到多个成分, 可以仅考虑大的成分。由于主路径成分的提取取决于主路径上所有连接的最小 SPC 值, 因此, 提取出的主路径成分往往不能取得满意效果。这是因为主路径上所有的连接并非都是重要的连接, 最小 SPC 值与最大 SPC 值往往相差甚大, 最后得到的主路径成分通常包含过多的点, 不具有代表性。

(3) 多路径

该方法首先在 Liu 和 Lu^[6]的文献中出现。为了能更清晰地表现知识演化历史及最近的研究活动, 需要逐渐增加全局主路径的条数, 直到得到满意效果为止。多路径及主路径成分出发点一致, 都是在得到主路径以后, 为了更清晰地表现学科发展规律, 以主路径为中心进行扩展的, 只是所用的方法略有差异。

近年来, 主路径方法研究还产生了一系列的进展, 实际上主要是围绕如下 5 个问题展开的: ①主路径的选择是基于全局还是基于局部; ②搜索路径是基于关键路径还是从起点到终点的遍历方式; ③是仅考虑单一的主路径还是多种主路径方法的组合; ④主路径的呈现方式; ⑤主路径上连线权重的确定^[6,7]。

1.1.3 主路径方法的不足

(1) 单纯的主路径分析难以实现技术演化的精细结构揭示

主路径分析以网络结构的连通性作为根本出发点, 通过遴选出网络结构中从源点到汇点的最大连通通路来揭示网络构成中的关键路径, 其表现形态通常为时间序列上文献间的线性连接关系。因此, 主路径分析的最大优势在于通过对大型网络的主路径提取来实现海量数据的降维处理, 从而得到领域