



研究前沿探测及其 演化分析方法与实证研究

张丽华 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

本书由国家社会科学重点基金项目“面向领域知识发现的学科信息学理论与应用研究”（编号：17ATQ008）、ISTIC-ELSEVIER期刊评价研究中心开放基金“中国学者的会议论文产出、影响力及其对科研人员排名的影响研究——基于Scopus数据库的分析”以及山西省教育科学“十三五”规划2017年度课题“山西高校新型智库联盟建设研究”联合资助。

研究前沿探测及其 演化分析方法与实证研究

张丽华 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

研究前沿探测及其演化分析方法与实证研究/张丽华著. —武汉:
武汉大学出版社, 2017. 11

ISBN 978-7-307-19100-6

I. 研… II. 张… III. 情报处理—方法研究 IV. G254.37

中国版本图书馆 CIP 数据核字(2017)第 001950 号

责任编辑:唐伟 杨铭

责任校对:汪欣怡

版式设计:马佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.whu.edu.cn)

印刷:虎彩印艺股份有限公司

开本:720×1000 1/16 印张:15.75 字数:274千字 插页:1

版次:2017年11月第1版 2017年11月第1次印刷

ISBN 978-7-307-19100-6

定价:40.00元

版权所有,不得翻印;凡购我社的图书,如有质量问题,请与当地图书销售部门联系调换。



序

当今世界，世界科学技术发展正孕育着重大创新突破，新一轮科技革命和产业革命蓄势待发，并将引致世界格局大变革大调整，科技的巨大进步和创新性应用将成为推动人类经济和社会巨大发展的重要引擎。把握世界科技发展大势，研判世界科技发展的战略方向，敏锐抓住科技创新的突破口和新的生长点，寻找突破路径、把握创新方向、前瞻科技布局，对于我国在世界科技发展竞争中，从“跟跑”向“并跑”乃至“领跑”转变具有极其重要的意义。

如何把握和确定研究前沿，是科技发展战略和规划高度关注的问题；如何快速挖掘、探测研究前沿是科技情报和科学计量等研究中的重要议题。研究前沿探测通常以学科领域为基本研究单元，发现学科领域发展过程中最新、最有潜力、最具前瞻性和引领性的研究方向。在学科领域研究进展和演变所经历的初期、早期、进展期、成熟期和衰退期的不同发展阶段中，有关研究人员多是在进展期或成熟期进行前沿探测，此时大量研究人员已经进入该领域，探测出来的更倾向于多数研究人员已经察觉的“热点”而非前沿。若能够在研究领域快速发展之前即早期阶段就能探测到研究前沿，那么就能够更早地把握具有引领性、前瞻性、探索性的研究方向，及早投入到前沿方向和研究问题的研究中，抢占科技创新先机。围绕这一假设，本书提出研究领域发展早期的研究前沿探测问题。同时，研究领域快速发展前后，科技领域内研究主题的前沿性程度必然存在较大的差异，分析这种差异的演化规律，能够指导研究前沿的探测工作。因此，有必要探讨研究领域快速发展前后，即研究领域发展的早期阶段与进展期阶段研究主题前沿程度的演化规律。因此，本书中涉及两个基本问题：第一，研究领域发展早期的研究前沿探测方法，即如何在研究领域快速增长之前即探测到研究前沿；第二，研究领域发展早期与进展期的研究前沿变化规律。

围绕这两个基本问题，本书开展了以下研究。

(1) 研究领域生命周期不同阶段的定量划分

根据普赖斯在《小科学、大科学》中的论述,科技文献呈阶梯型增长。研究领域发展过程中的若干个S形的生命周期组成了科技文献的阶梯型增长曲线。每个生命周期包括初期、早期、进展期、成熟期和衰退期5个不同阶段。在抽取研究领域发展的不同时期的数据进行前沿探测与前沿特征演化分析之前,需要首先确定研究领域的生命周期并划分生命周期的不同阶段。本书采用拐点法确定学科领域的生命周期,将曲线的波峰和波谷作为拐点,波谷到波峰的两个拐点之间即为曲线的一个生命周期。当曲线连续两年均出现负增长时,连续负增长两年中第一年的观测值为波峰,连续负增长两年中第二年的观测值为波谷。在划分研究领域生命周期的不同阶段时,本书认为,不同阶段之间的分界线很可能并不是一个精确的值,而是一个范围。根据研究领域中每年发表论文数量占研究领域达到饱和时论文数量最大值(X)的不同比例,本书确定了研究领域生命周期不同阶段之间分界线的比例范围:初期($0\% \leq X < 1\% \sim 5\%$)、早期($1\% \sim 5\% \leq X < 10\% \sim 20\%$)、进展期($10\% \sim 20\% \leq X < 50\% \sim 60\%$)、成熟期($50\% \sim 60\% \leq X < 100\%$)以及衰退期(从 $X \geq 100\%$ 开始直到下一个生命周期的起点)。

(2) 研究领域发展早期的研究前沿探测方法构建

该方法的基本思路是研究主题的前沿特征越显著,则该研究主题成为研究前沿的概率就越高,将研究主题的前沿性特征表征为4个方面,即时效性、跨学科性、突破性和继承性;将这4个方面的前沿性特征转化为相应的文献计量学指标,分别计算每个研究主题的时效性指数、跨学科性指数、突破性指数和继承性指数;对这4个指数进行综合加权计算,获得每个研究主题的综合前沿性指数;通过综合前沿性指数来判断研究主题前沿性特征的显著性,综合前沿性指数值越大,表示研究主题的前沿性特征越显著,该研究主题就越有可能是具有前瞻性、引领性的研究方向。

(3) 研究领域发展早期与进展期主题的前沿特征演化规律分析

在研究领域发展的早期与进展期阶段,研究主题的前沿特征将呈现出不同的演化规律,根据这种演化规律能够为不同时期的研究前沿探测提供有效指导。在分析研究主题的前沿演化规律时,首先需要分析研究领域发展早期与进展期的主题演化规律。本书利用研究主题所包含论文的相互引用关系建立研究主题之间的演化关系,提出“主题演化指数”的概念来测度研究主题之间的演化强度,将主题之间的演化类型划分为分化、合并、新增、消失与融合等5种,提出“主题演化率”与“主题演化强度率”的概念定量测度研究领域发展早期与进展期主题演化关系数量与主题演化强度的变化。其

次分析研究领域发展早期与进展期主题的前沿特征演化规律。前沿特征演化规律的分析围绕7个假设进行,研究结果表明:在研究领域发展的早期和进展期,研究主题的整体前沿性水平呈现先升后降的特点;研究主题的时效性指数随着研究领域的不断发展而呈现逐渐上升的趋势;当研究主题发生演化时,研究主题的部分前沿性特征指数会同步发生演化;研究主题的演化次数越多,研究主题的前沿性特征丢失的不一定多;新增研究主题不一定是研究前沿,但新增研究主题成为研究前沿的概率较高;已确认为重大进展的研究前沿主题对其他研究主题的前沿性特征显著性的影响在不同研究领域之间存在差异;当某个研究主题从其他研究主题处获得的研究前沿特征越多时,该研究主题本身的研究前沿特征不一定显著。

(4) 实证研究

为了验证研究领域发展早期的研究前沿探测方法以及研究领域发展早期和进展期研究主题的前沿特征演化规律,本书选择了“复杂网络”和“墨累-达令河流域水资源管理研究”两个研究领域进行了实证研究。结果表明:本书构建的研究领域发展早期的研究前沿探测方法能够较准确地探测两个研究领域早期阶段的研究前沿,且研究主题的研究前沿特征演化规律具有研究领域的特定性。

作为如何在研究领域快速发展之前有效探测研究前沿的有益研究尝试,本书可以为科学计量学及相关领域的师生提供相关的研究参考。

张志强

中国科学院成都文献情报中心主任

目 录

1 引言	1
1.1 研究背景	1
1.2 问题提出	3
1.3 研究内容与意义	5
1.3.1 研究内容与目标	5
1.3.2 研究意义	7
1.4 研究思路	7
1.5 研究方法	10
1.6 本书章节安排	11
2 研究前沿探测与主题演化的理论基础与相关研究进展	13
2.1 概念界定与辨析	13
2.1.1 研究前沿的概念发展	13
2.1.2 相关概念辨析	16
2.2 研究前沿探测与主题演化的理论基础	19
2.2.1 科学计量学——引文分析、聚类分析与向量空间模型	19
2.2.2 科学学——库恩的科学革命理论	22
2.3 研究前沿探测与主题演化的相关研究进展	23
2.3.1 研究前沿探测的相关研究进展	23
2.3.2 主题演化的相关研究进展	44
2.4 存在的问题和不足	52
3 文献数据集构建与数据处理	54
3.1 验证方式与领域选择	54
3.2 文献数据集构建与清洗	62
3.2.1 文献数据获取	62

3.2.2 文献数据清洗	63
3.2.3 实证领域文献数据集	64
3.3 数据处理	67
3.3.1 研究领域生命周期划分	67
3.3.2 研究领域时间窗划分	80
3.4 本章小结	81
4 研究领域发展早期前沿探测方法构建	83
4.1 研究主题获取	83
4.1.1 直接引文矩阵构建	85
4.1.2 聚类方法选择与质量评价	87
4.1.3 研究领域发展早期的研究主题	89
4.2 研究主题的前沿特征识别	93
4.3 研究主题的前沿特征指标化	95
4.3.1 研究主题的前沿特征指标	95
4.3.2 研究领域发展早期的前沿探测结果	106
4.4 方法的有效性评价	111
4.5 主题的前沿特征指标对比分析	115
4.5.1 理论分析	115
4.5.2 实验结果分析	117
4.6 本章小结	118
5 研究领域发展早期与进展期主题前沿特征演化规律研究	120
5.1 研究领域发展早期与进展期的前沿主题分析	120
5.1.1 “复杂网络”领域发展早期与进展期的前沿主题分析	121
5.1.2 “墨累-达令河流域水资源管理研究”领域发展早期与进展期的前沿主题分析	129
5.2 研究主题演化分析	133
5.2.1 主题继承与主题演化辨析	133
5.2.2 主题演化的测度指标	134
5.2.3 主题演化的实验结果	140
5.3 研究主题前沿特征的演化分析	150
5.3.1 主题前沿特征演化的主要假设	150

5.3.2 主题前沿特征演化的实验结果	155
5.4 本章小结	168
6 总结与展望	170
6.1 总结	170
6.2 创新之处	172
6.3 研究不足与下一步研究方向	175
参考文献.....	177
附录.....	190
博士在读期间发表论文和参与科研课题情况.....	236
后记.....	237

图 目 录

图 1-1	“研究前沿”领域 1900—2014 年发表论文数量的年变化曲线	4
图 1-2	研究领域生命周期的 S 曲线	4
图 1-3	本书的研究思路	9
图 1-4	本书的组织结构图	12
图 2-1	“胶原蛋白”领域 1970—1974 年主题聚类演化图	45
图 2-2	簇演变链的线性模式和非线性模式	47
图 2-3	“里德伯”领域 1976—1985 年的主题聚类演化图	48
图 2-4	“炭疽病”领域 1981—2001 年主题的时间线演化图	49
图 3-1	“复杂网络”领域 1960—2014 年论文数量的年变化曲线	65
图 3-2	“墨累-达令河流域水资源管理研究”领域 1982—2014 年论文数量的年变化曲线	67
图 3-3	科技论文数量的阶梯形增长	68
图 3-4	“氮化镓”领域 1970—2004 年论文与引文数量的年变化曲线	69
图 3-5	“复杂网络”领域 1970—2004 年论文与引文数量的年变化曲线	70
图 3-6	“中国科技政策”领域 1949—2010 年政策文件数量的年变化曲线	70
图 3-7	“乡村地理学”领域 1900—2012 年论文数量的年变化曲线	71
图 3-8	“复杂网络”领域 1960—2014 年论文数量的年增长率曲线	76
图 3-9	“复杂网络”领域 1960—2014 年论文数量的年变化拟合曲线	77
图 3-10	“墨累-达令河流域水资源管理研究”领域 1982—2014 年论文数量的年增长率曲线	78
图 3-11	“墨累-达令河流域水资源管理研究”领域 1982—2014 年论文数量的年变化拟合曲线	79

图 4-1	研究领域发展早期前沿探测思路	84
图 4-2	引文示意图	85
图 4-3	“复杂网络”领域 1991—1995 年的直接引文矩阵 (部分截选)	87
图 4-4	“复杂网络”领域 1993—1997 年以及 1995—1999 年研究主题的 4 项前沿特征指数标准化值以及综合前沿性指数值	107
图 4-5	“墨累-达令河流域水资源管理研究”领域 1993—1997 年以及 1995—1998 年研究主题的 4 项前沿特征指数标准化值 以及综合前沿性指数值	110
图 4-6	“复杂网络”领域 1991—1999 年 Citespace 分析文献共被引 图谱	113
图 4-7	“墨累-达令河流域水资源管理研究”领域 1991—1998 年 Citespace 分析文献共被引图谱	115
图 5-1	“复杂网络”领域发展早期与进展期研究前沿演化图谱	128
图 5-2	主题继承与主题演化示意图	134
图 5-3	不同时间窗主题演化关系示意图	136
图 5-4	主题分化示意图	137
图 5-5	主题合并示意图	138
图 5-6	主题融合示意图	138
图 5-7	主题新增示意图	138
图 5-8	主题消失示意图	139
图 5-9	“复杂网络”领域发展早期与进展期主题演化图谱	141
图 5-10	“复杂网络”领域发展早期的主题演化类型	142
图 5-11	“复杂网络”领域 1995—1999 年 $T_{3,4}$ 主题的演化关系图谱	143
图 5-12	“复杂网络”领域发展的进展期主题演化图谱	143
图 5-13	“复杂网络”领域发展的早期与进展期主题演化率的 变化曲线	144
图 5-14	“复杂网络”领域发展的早期与进展期主题演化强度率的 变化曲线	145
图 5-15	“墨累-达令河流域水资源管理研究”领域发展的早期与 进展期主题演化图谱	146
图 5-16	“墨累-达令河流域水资源管理研究”领域发展的早期与 进展期主题演化率的变化曲线	148

图 5-17	“墨累-达令河流域水资源管理研究”领域发展的早期与 进展期主题演化强度率的变化曲线	149
图 5-18	研究领域发展的早期与进展期主题的整体前沿水平的变化 曲线	152
图 5-19	研究领域发展过程中主题的整体前沿水平的变化曲线	152
图 5-20	“复杂网络”领域发展的早期和进展期主题的综合前沿性 指数变化曲线	156
图 5-21	“复杂网络”领域发展的早期和进展期主题的时效性指数 变化曲线	156
图 5-22	“复杂网络”领域发展的早期和进展期新增主题图谱	159
图 5-23	“墨累-达令河流域水资源管理研究”领域发展的早期和进展期 主题的综合前沿性指数变化曲线	161
图 5-24	“墨累-达令河流域水资源管理研究”领域发展的早期和进展期 主题的综合前沿性指数时间窗均值变化曲线	162
图 5-25	“墨累-达令河流域水资源管理研究”领域发展的早期和进展期 主题的时效性指数变化曲线	163

表 目 录

表 2-1	“研究前沿”的代表性定义	15
表 2-2	研究前沿探测常用的分析工具	24
表 2-3	研究前沿探测方法	25
表 2-4	研究领域生命周期不同阶段的前沿信号强度	44
表 2-5	主题演化的代表性研究	51
表 3-1	“复杂网络”领域的重大标志性进展	58
表 3-2	“墨累-达令河流域水资源管理研究”领域的重要事件	61
表 3-3	生命周期划分依据	74
表 4-1	“复杂网络”领域发展早期各个时间窗的研究主题	91
表 4-2	“墨累-达令河流域水资源管理研究”领域发展早期各个 时间窗的研究主题	92
表 4-3	评委专家对 4 个前沿特征指标重要程度的评定结果	105
表 4-4	“复杂网络”领域 1993—1997 年以及 1995—1999 年研究 主题的各项前沿特征指标值	107
表 4-5	“墨累-达令河流域水资源管理研究”领域 1993—1997 年以及 1995—1998 年研究主题的各项前沿特征指标值	109
表 5-1	“复杂网络”领域发展早期各个时间窗的研究主题及主题的 综合前沿性指数	121
表 5-2	“复杂网络”领域发展的进展期各个时间窗的研究主题及 主题的综合前沿性指数	122
表 5-3	“墨累-达令河流域水资源管理研究”领域发展早期各个 时间窗的研究主题及主题的综合前沿性指数	129
表 5-4	“墨累-达令河流域水资源管理研究”领域发展的进展期各个 时间窗的研究主题及主题的综合前沿性指数	130
表 5-5	“复杂网络”领域主题的前沿特征演化指数与前沿特征指数的 相关性分析结果	157

表 5-6	“复杂网络”领域主题的前沿特征指数与主题演化指数的相关性分析	159
表 5-7	“复杂网络”领域主题的前沿特征演化指数与前沿特征指数的相关性分析	160
表 5-8	“墨累-达令河流域水资源管理研究”领域主题的前沿特征演化指数与前沿特征指数的相关性分析结果	163
表 5-9	“墨累-达令河流域水资源管理研究”领域主题的前沿特征指数与主题演化指数的相关性分析	165
表 5-10	“墨累-达令河流域水资源管理研究”领域主题的前沿特征演化指数与前沿特征指数的相关性分析	165
表 5-11	“复杂网络”领域和“墨累-达令河流域水资源管理研究”领域主题的前沿特征演化规律的对比分析	167
表 6-1	本书提出的各项指数及说明	174

1 引言

1.1 研究背景

1965 年, Price^① 最早提出“研究前沿 (research front)”的概念。研究前沿能够描述研究领域发展的动态本质, 揭示领域发展过程中的重大突破、重大发现与重要研究进展, 是具有前瞻性、引领性和先进性的研究方向。在科学计量学中, “研究前沿”基于大规模的论文数据集, 通过对论文的数量、内容及引用关系的定量分析研究, 描绘科学结构, 追踪研究领域的产生、发展、分化、相互渗透的现状 & 动态, 从而判断、挖掘前沿研究方向。

当今世界科技发展呈现多点突破、交叉汇聚的景象^②。随着科技革命的迅猛发展, 许多基本科学问题面临突破, 学科间交叉、融合与汇聚, 不断催生新的学科生长点。重大科学问题的解决及科技领域内的重大进展、重大突破、重要发现能够极大推动科学与社会发展的进程, 甚至可能会深刻影响人类的生产和生活方式, 进而影响国家的国际竞争优势和地位。在这种情况下, 世界各国纷纷将希望寄托在科技的革命性突破上, 把握科技革命可能发生的领域和方向, 谋求新一轮发展的主动权。越来越多的国家和政府认识到, 在制定统筹安排、整体推进的科技战略规划中, 遴选“优先领域”或“重点领域”是加速科技进步、推动社会经济协调发展的关键所在。

“研究前沿”成为政府科研资助部门、科研机构与科技创新企业等重点关注的话题。我国 2006 年发布的《国家中长期科学和技术发展规划纲要 (2006—2020 年)》中提出要发展前沿技术及战略产业, 将科技前沿分为重点

① Price D J D. Networks of Scientific Papers [J]. Science, 1965, 149 (3683): 510-515.

② 中国科学院. 科技发展新态势与面向 2020 年的战略选择 [M]. 北京: 科学出版社, 2013.

领域与优先主题、前沿技术、科学前沿问题、面向国家重大战略需求的基础研究^①。美国国家科学基金会(NSF)于2014年3月发布的《投资科学、工程和教育,开创国家未来——国家科学基金会2014—2018年战略规划》(*Investing in Science, Engineering, and Education for the Nation's Future—NSF Strategic Plan for 2014—2018*)中提出的其中一个战略目标是变革科学和工程前沿。拓展并探索人类知识的前沿,提高国家应对未知挑战的能力,创建科学、技术和经济的新范式与新能力^②。世界顶级科学期刊 *Science* 于2005年公布了125个最具挑战性的科学前沿问题^③。*Science* 认为,在未来25年中,科学家应该能够很好地解决这些问题,或者至少应该知道如何去解决。全球领先的专业信息提供商汤森路透于2013年发布了名为《2013年研究前沿:自然科学与社会科学的100个学科领域》的报告,确认多个学科领域的100个重要研究前沿^④。汤森路透和中国科学院文献情报中心于2014年联合发布了《2014研究前沿》报告,遴选了2014年排名最前的100个热点研究前沿和44个新兴研究前沿^⑤。

研究前沿的探测主要通过两种途径实现,一种是以专家为核心,通过会议、通信等多种方式,汇集不同领域专家的知识与智慧,形成最终的前沿预见。^⑥如 *Science* 通过咨询本刊高级编辑委员会、审稿编委会以及编辑和作者的建议,提出了125个未来25年待解决的关键科学问题。^⑦另一种是以文献为核心,利用科学计量学方法,对各类学者发表的期刊论文、论著进行挖掘,探测领域内在的规律和特点,进而揭示学科内部结构和下一步发展方向,完成基于文献的研究前沿探测。^⑧如汤森路透在遴选100个研究前沿时,以 Essential Science Indicators(ESI)数据库中的近8000个研究前沿为起点,首先选出那些核心文献被引用次数最多的研究前沿,再对上一步选出来的研究前沿,按核心文献的平均出版年排序,找出核心文献最新的研究

① http://www.gov.cn/jrzq/2006-02/09/content_183787.htm.

② <http://www.nsf.gov/pubs/2014/nsf14043/nsf14043.pdf>.

③ <http://www.sciencemag.org/content/309/5731/75>.

④ <http://thomsonreuters.cn/news-ideas/pressreleases/press20130628>.

⑤ <http://thomsonreuters.cn/news-ideas/pressreleases/researchfronts2014>.

⑥ 张英杰. 科技领域前沿计量探测方法研究[D]. 北京: 中国科学院大学, 2011.

⑦ <http://www.sciencemag.org/content/309/5731/75>.

⑧ 张英杰. 科技领域前沿计量探测方法研究[D]. 北京: 中国科学院大学, 2011.

前沿。^①

自 20 世纪 90 年代以来,对于“研究前沿”的研究和关注一直是学术界的热门话题。作者在 Web of Science 数据库中使用主题词途径对“研究前沿”的相关研究进行检索,检索式为:

TS=(“research front*” or “frontier” or “scientific front*” or “emerging field*” or “emergent field*” or “emerging area*” or “emergent area*” or “emerging research area*” or “emerging scientific field*” or “emerging research field*” or “emerging scientific area*” or “emergent research area*” or “emergent scientific field*” or “emergent research field*” or “emergent scientific area*”)

时间跨度为 1900—2014 年,检索时间为 2015 年 4 月 30 日,获得“研究前沿”领域年发表论文数量的变化曲线(见图 1-1)。从图 1-1 中可以看出,自 20 世纪 90 年代起,有关“研究前沿”领域的研究呈急剧增长态势。科学计量学、科学社会学、科学学、科技规划与科技战略学以及各学科领域学科规划均将前沿领域的研究作为本学科领域的重要选题。“前沿”的发展不仅能够在学科领域内部形成科学研究的“富矿区”,而且能够促进相关学科及学科群甚至整体科学的突破,它在科学发展中的地位举足轻重。

1.2 问题提出

研究领域的发生、发展、成熟和衰退遵循生命周期理论。研究领域发展的生命周期演化可以用文献量的增长规律近似刻画。根据每年发表的论文数量,研究领域的生命周期可以表示为图 1-2 中的 S 曲线^②。在领域发展的初期,论文数量可能在较长时间内始终处于较低水平,相关研究零星出现;在领域发展的早期,论文数量还是处于一个较低水平,但已经出现增长的势头,增长速度缓慢;在领域发展的进展期,大量研究人员进入该领域,相关研究快速增长,推动了领域的蓬勃发展;随着越来越多的研究出现,领域逐渐进入成熟期,形成了相对稳定的知识结构和体系,对该领域的研究可能更

^① <http://ip-science.thomsonreuters.com.cn/media/wok20130628.pdf>.

^② Shibata N, Kajikawa Y, Matsushima K. Topological analysis of citation networks to discover the future core articles[J]. Journal of the American Society for Information Science and Technology, 2007, 58(6): 872-882.