

追根溯源：优秀科学计量学家引用的重要文献识别及引用内容特征研究¹⁾

赵 勇^{1, 2} 武夷山³

(1. 中国科学技术信息研究所, 北京 100038; 2. 中国农业大学图书馆, 北京 100193;
3. 中国科学技术发展战略研究院, 北京 100038)

摘要 具体分析优秀科学家在学术论文中所引用文献的内容特征, 可以客观反映其掌握科学知识的深度和广度, 也可以揭示其研究的学术背景、科学依据与演进脉络。本文利用引文出版年光谱和引用内容分析方法, 识别了莱兹多夫教授在 269 篇学术论文中引用的不同时期重要文献。研究结果将有助于了解他的学术思想根源及对科学计量学发展的贡献。最后讨论了莱兹多夫教授的引用行为所带来的启示。

关键词 引文出版年光谱; 引用内容分析; 重要参考文献; Loet Leydesdorff; 学术思想

Tracing Origins: A study on the Characteristics of Important Literature Cited by a Famous Scientometrician

Zhao Yong^{1, 2}, Wu Yishan³

(1. Institute of Scientific and Technical Information of China, Beijing 100037;

2. Library of China Agricultural University, Beijing 100193;

3. Chinese Academy of Science and technology for Development, Beijing 100038)

Abstract An analysis on the characteristics of references in excellent scientific articles can objectively reflect the depth and breadth of knowledge of the authors, and reveal their academic background, scientific evidence, and evolution course in their research. The purpose of this paper is to identify the important studies cited by 269 research articles authored by Loet Leydesdorff based on Referenced Publication Year Spectroscopy (RPYS) and Citation Content Analysis (CCA). The findings can be used to better understand Loet Leydesdorff's academic ideology and contribution in the area of Scientometrics. Finally, we discuss the inspiration of Loet Leydesdorff's citing behavior.

Keywords Referenced Publication Year Spectroscopy; Citation Content Analysis; Important References; Loet Leydesdorff; Academic Ideology

基金项目: 本文系中国科协“高端科技创新智库青年项目”(编号:DXB-ZKQN-2016-030)的研究成果之一。

作者简介: 赵勇, 男, 1980 年生, 中国农业大学图书馆情报研究中心, 副研究馆员, 硕士生导师; 中国科学技术信息研究所, 博士后。主要研究方向: 情报学理论与方法研究、科学计量与科技政策研究。E-mail: zhaoyong@cau.edu.cn。武夷山, 男, 1958 年生, 中国科学技术发展战略研究院, 研究员, 博士生导师。主要研究方向: 科学计量学研究、科技政策研究。

1 引言

科学发展离不开大胆的假想与创新,更离不开后人对前人研究成果的继承、质疑和完善。卡尔·波普尔(Karl R. Popper)认为科学知识是通过科学家的猜想与反驳,不断清除错误而增长^[1]。虽然对于科学发展的原因还存在其他不同的理论观点,如弗朗西斯·培根(Francis Bacon)的累积式发展观^[2]、托马斯·库恩(Thomas S. Kuhn)的科学革命模式^[3],但科学家的研究与其所在科学共同体早期成果之间的关系却是这些科学哲学家们共同思考的问题。

与科学哲学家一般利用历史事例来探究科学发展问题的路径不同,20世纪60年代起,科学计量学的先驱者们开始重点关注科技文献之间的引用,强调了科学知识增长与文献集合之间的关系^[4]。尤其是在1963年尤金·加菲尔德(Eugene Garfield)建立科学引文索引(Science Citation Index, SCI)后,引文分析方法在探索科学发展进程、发现学科结构特征、评价论文学术影响等方面发挥了重要的作用。同时,引文分析也为科学史研究打开了一个新的趣向——科学史定量研究^[5]。2003年, Garfield在“为什么我们需要计量史学”^[6]一文中提出,可利用引文图谱分析工具(HistCite)描绘某一领域不同文献之间的引用关系,从而展示该领域的科学发展图景,发现里程碑式的重要文献及核心学者。2013年, Marx等学者^[7]提出了引文出版年光谱(Referenced Publication Year Spectroscopy, RPYS)分析方法,主要通过对科技论文的引文出版年份及被引频次分析,挖掘学科领域、研究主题和个体学者层面引用的重要文献,进而为刻画知识领域的动态演进历程或科学家的学术发展轨迹提供依据。RPYS弥补了HistCite分析中容易遗漏外部学科或其他研究领域文献的缺陷,同时将探测重要参考文献的时间范围从研究领域产生之后延伸至研究领域形成之前或萌芽期^[8]。目前,国内外学者已将RPYS方法应用于物理学中希格斯玻色子^[9]、生物学中“达尔文雀”传奇^[10]、引文分析研究^[11]以及信息计量学^[12]等研究主题或学科领域的历史根源探测,但鲜见该方法在个体学者层面的应用研究。

科学计量学是应用数理统计和计算技术等定量方法研究科学活动规律性的一门科学分支学科^①,其理论与方法论基础源自科学哲学、科学社会学、信息论、概率论等学科领域^[13],形成了共词分析、合著关系分析、引文分析等本学科特色的研究方法。同时,科学计量学也在不断地从相关学科汲取营养,进一步增强本学科的思想体系和知识体系。普赖斯奖是科学计量学界公认的最高学术荣誉,对其获奖者的学术特征进行梳理总结也是剖析科学计量学学科基础和发展脉络的途径之一^[14]。所以,本文以2003年普赖斯奖获得者路特·莱兹多夫(Loet Leydesdorff)为研究对象,综合利用引文出版年光谱(RPYS)和引用内容分析方法(CCA),识别并挖掘他在学术论文中频繁引用的不同时期重要文献,旨在为深入了解他学术思想的根源及其对科学计量学发展的贡献提供清晰的线索和有益的启示。

2 相关研究

引用内容分析(Citation Content Analysis, CCA)是研究学者引用动机和引用行为规律的方法路径之一。早在上世纪60年代, Garfield^[15]即根据引文的位置、内容、形式等,较为系统地归纳了15种典型引用动机, Lipetz^[16]则依据被引文献的内容和功能将引用动机细分为4组29项特征。当前,国外学界对引用内容

^① 国内多数学者认为科学计量学(Scientometrics)是科学学的一部分,但国际学界则较普遍地认为它是图书情报学的一部分,例如《科学计量学》杂志(Scientometrics)即被确定是国际图书情报学(Library and Information Science, LIS)的核心期刊或代表性期刊。

的研究主要涉及以下几个方面：第一，引用内容的分布位置。Voos 和 Dagaev^[17]发现参考文献的重要性与引用内容在施引文献中的分布位置有关。Halevi 和 Moed^[18]将引用位置分为引言、综述、方法、发现、结论、讨论六个部分，并对 *Journal of Informetrics* 杂志 2007 年发表的学术论文进行了分析，结果发现，在方法部分，作者引用学科内文献多于学科外文献，而在引言部分，引用的学科外文献多于学科内文献，其他四个部分的差异性并不明显。第二，引用内容的重要程度。Moravcsik 和 Murugesan^[19]将引用按照重要程度分为有组织引用和敷衍式引用两类。Chubin 和 Moitra^[20]将其扩展为基础的必要引用、辅助的必要引用、额外的补充引用、敷衍的补充引用、部分的负面引用、全面的负面引用六类。Ding 等^[21]则研究了一篇论文在施引文献中被多次引用的现象，提出了考虑多次引用的文献被引次数统计方法（CountX），从定量角度反映了参考文献在引用强度上的差异。第三，引用内容的语境语义。引用语境分析主要涉及话语分析的相关理论，Frosig^[22]即将语类理论（Genre theory）与引文分析理论相结合构建了引用语境研究模型。部分学者^[23-24]则借助自然语言处理技术对引用内容的描述性或评论性文字进行了语义分析研究。Zhang 等^[25]较为系统地提出了引用内容分析的研究框架，包含了对引用内容的语法分析和语义分析。

引用内容分析被认为是引文分析的新阶段，在多个层面上拓展和深化了传统引文分析方法。因此，该领域也引起国内学界的高度重视。一方面，国内学者关注了引用内容分析的基本框架和方法层面上的研究。刘盛博等^[26]认为引用内容分析是基于引文分析基础理论，借助文本挖掘和自然语言处理等技术，从施引文献的全文入手，聚焦于引用的片段，对引用频次、引用位置和引用文本的内容主题进行的挖掘和研究。祝清松和冷伏海^[27]对引用内容分析方法^①进行了详细综述，提出了引用内容分析主要包括数据预处理、引用内容抽取和内容深度分析三个步骤。也有学者特别关注了引用内容分析过程中的标注框架设计。陆伟等^[28]提出引用内容标注框架包括揭示引文关系抽象性质的引文分类标注体系，描述被引文献具体内容的引用对象标注体系，以及记录引文客观特征的引文属性标注体系三个方面。张梦莹等^[29]则进一步拓展，从引用对象、引文功能、引用情感、引文位置、引文重要性、标注自信度等方面设计了新的引用内容标注系统。另一方面，国内学者重点开展了基于引用内容分析的应用研究，如学术专著引用行为研究^[30]、高被引论文主题识别^[31]、论文学术影响力评价^[32]等。

从现有研究成果来看，国内外学者针对引用内容分析的整体研究框架和具体技术方法开展了大量的研究工作，也展现了引用内容分析在学术评价、信息检索、科学知识演化等领域的广泛应用价值，但目前仍存在问题有待关注和深入研究，一是对优秀科学家经常引用的重要文献的内容特征分析略显不足；二是缺少引用内容分析结果与科学家的研究主题等其他学术信息之间的关联分析。因此，本文聚焦了优秀科学计量学家的研究与科学共同体早期学术成果之间的关系，在对莱兹多夫教授引用的不同时期重要文献进行 RPYS 识别和引用内容分析的基础上，将分析结果与其研究主题相关联，综合揭示他的学术思想根源与学术贡献表现。

3 研究设计

3.1 研究对象

莱兹多夫教授是世界著名的图书情报学家、科学计量学家和社会学家，2010 年受聘为阿姆斯特丹大学传播研究学院教授，2014-2016 年连续入选“全球高被引科学家名录”（社会科学领域），是科学计量学界的

^① 祝清松和冷伏海（2013）在文章中将“Citation Content Analysis”译为“引文内容分析”，但从概念内涵上看，译为“引用内容分析”更为准确。刘盛博（2015）的研究中也明确采用了“引用内容分析”译法，本文统一使用了“引用内容分析”一词。

高产出和高影响力学者之一。莱兹多夫教授具有生物化学、哲学、社会学、图书情报学的多学科教育背景（见表1），长期多角度、多学科地思考和研究关于科学计量学的理论方法问题，并不断从事相关的实证研究以验证其理论建构。因此，以他引用的重要文献为剖析样本，对了解科学计量学学科的历史根源具有典型意义。

表1 莱兹多夫教授的基本履历信息

科学家 (ORCID)	国籍	出生日期	教育背景 (毕业时间, 毕业学校, 学科专业)
Loet Leydesdorff 0000-0002-7835-3098	荷兰	1948年8月21日	1969年, 荷兰阿姆斯特丹大学, 化学学士
			1973年, 荷兰阿姆斯特丹大学, 生物化学硕士
			1977年, 荷兰阿姆斯特丹大学, 哲学硕士
			1984年, 荷兰阿姆斯特丹大学, 社会学博士

资料来源: 履历信息来自莱兹多夫教授的个人主页 <http://www.leydesdorff.net/>。

3.2 数据来源与全文获取

相比 Web of Science 数据库, Scopus 数据库收录的期刊范围更为广泛^[33], 并且在学术论文的下载元数据中提供了参考文献的详细信息, 更能满足本研究对重要参考文献识别的需求。因此, 本研究选择 Scopus 数据库, 以莱兹多夫教授的作者身份识别号 (ORCID) 为检索点, 时间窗为截至 2016 年, 文献类型为 Article, 语种类型为 English, 累计检索到莱兹多夫教授的 269 篇学术论文。检索时间为 2017 年 6 月 1 日。此外, 在本单位购买的全文数据库、莱兹多夫教授个人网站和科研社交网站 ResearchGate 中, 查找并下载其学术论文的全文, 以满足本研究的引用内容分析需求。

3.3 分析方法与数据处理

3.3.1 引文出版年光谱分析

引文出版年光谱 (RPYS) 的主要分析原理是将特定文献集中参考文献的被引频次按照其出版年份的分布情况进行排序, 并利用可视化方法呈现引文出版年分布和 5 年中值偏差曲线, 以高峰年为中心的前后共 5 年, 均可以视为作者引用的高峰时期文献。2014 年 Leydesdorff 等学者编写了 RPYS 软件包。在此基础上, 2016 年 Thor 等学者^[34]开发了更直观的用户界面软件分析工具 CRExplorer。为了辅助用户判断重要参考文献, CRExplorer 设置了 6 类统计指标, 包括: 参考文献被引频次 (N_CR)、参考文献被引频次占同年参考文献总被引频次的比例 (PERC_YR)、参考文献被引频次占所有参考文献总被引频次的比例 (PERC_ALL)、参考文献的施引年份数量 (N_PYEAR)、参考文献的施引年份数量占比 (PERC_PYEAR)、每一施引年份中参考文献被引频次排序前 50%、75%、90% 的数量 (50N_TOP50, N_TOP75, N_TOP90)。本研究采用非参数统计方法, 将各参考文献的 N_CR、N_PYEAR 和 N_TOP90 指标数值按递减顺序排列, 三个指标排序均在前 10% 位置的文献, 就认为是作者引用的重要参考文献。

另外, CRExplorer 还提供了参考文献的数据标准化功能, 主要是基于编辑距离 (Levenshtein Distance) 算法比对两两参考文献在作者姓氏和来源文献名称上的相似度, 默认相似度阈值为 0.75, 将相同参考文献合并至同一类团组号 (ClusterID) 下, 并统计每个类团组下的文献数量 (ClusterSize)。为了进一步保证参考文献统计分析结果的准确性, CRExplorer 还设计人工清洗数据的功能选项。本研究对相似度阈值在 0.75 以上的参考文献进行了自动合并, 同时人工清洗合并了高被引参考文献的相似文献。

3.3.2 引用内容分析

首先, 本文在前期研究成果^[35]的基础上, 完善了莱兹多夫教授的研究主题分类表(见表2), 进而建立参考文献与研究主题之间的关联关系。其次, 利用 VOSviewer 软件绘制重要参考文献之间的共被引网络图谱, 可视化展现它们之间的知识关联。最后, 通过主题关联和内容分析揭示莱兹多夫教授所引用重要文献的内容特征。

表 2 莱兹多夫教授的研究主题分类与文献数量

主题类目	主题说明	文献数量
三螺旋研究	包括对知识经济背景下产学研关系、三螺旋系统的相关研究	41
科技动态研究	包括定量测度科学技术知识增长, 描述国家、机构或领域科技状况的相关研究	39
评价研究	包括国家、期刊、机构、学科或领域的科研评价研究	36
引文研究	包括引用理论、引用分析、引用指标、引用评价、引文影响等研究	34
知识图谱	包括科学地图、以及主题、期刊、作者、机构、专利等的可视化研究	26
自组织研究	包括对社会系统、自组织现象、信息交流等的相关研究	22
创新研究	包括对创新理论、创新系统、创新过程、技术创新、创新指标等的相关研究	12
科研合作	包括国家间、机构间或学者间的科研合作研究	15
跨学科研究	包括跨学科程度的测量、跨学科研究影响力的评价等研究	11
期刊分类	包括期刊分类方法、期刊分类体系的相关研究	4
其他	不适于归纳为上述类目的文章	29

4 结果分析

对 269 篇学术论文的参考文献分析结果显示, 莱兹多夫教授引用 19 世纪及以前出版的文献有 24 篇, 其中法国数学家勒内·笛卡尔于 1687 年发表的《方法论》是其研究中引用的最早文献。利用 RPYS 方法对 20 世纪以来出版的被引文献进行分析表明(见图 1), 莱兹多夫教授的引用文献出版年高峰较为明显地出现在 1948 年、1963 年、1972 年、1988 年、1994 年和 2009 年。考虑到科学计量学诞生于 20 世纪 60 年代初, 并参考 6 个引文出版年的高峰, 本研究将莱兹多夫教授的引文出版年光谱划分为三个时间段, 即 1960 年以前、1960-1990 年间、1990 年以来, 分别统计三个时期内的重要参考文献, 最终识别出莱兹多夫教授引用的 19 篇重要参考文献, 涉及莱兹多夫教授发表的 217 篇论文, 占其总论文数的 80.7%。其中, 有 167 篇论文引用了本文识别出的重要参考文献中的至少两篇。

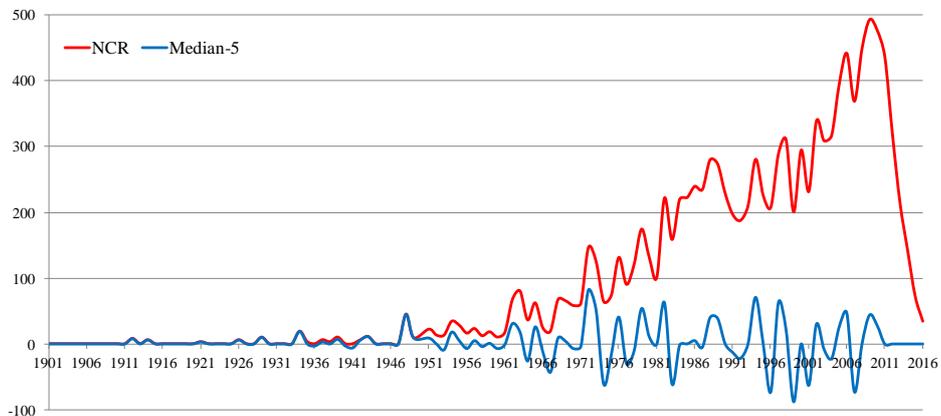


图 1 莱兹多夫教授学术论文的引文出版年光谱(1901-2016 年)

4.1 莱兹多夫教授引用的 1960 年以前发表的重要文献

表 3 显示, 莱兹多夫教授在科学计量学领域研究的知识基础可以追溯到 20 世纪 60 年代之前出版的一些经典文献。1934 年, 英国文献学家布拉福德 (Bradford) 的《专业领域的信息资源》一文提出了描述文献分散规律的经验定律, 即布拉福德定律, 是引文研究和评价研究领域的经典文献。美籍奥地利经济学家熊彼特 (Schumpeter) 是现代创新理论的奠基者, 其在 1943 年出版的《社会主义、资本主义与民主》一书中特别关注了大规模企业的技术积累性在产业技术创新中的重要作用, 形成创新的二阶段模型 (Mark I 和 Mark II), 莱兹多夫教授借用了该模型的思想, 探索从科学计量学视角来勾勒具体技术领域的非线性创新轨迹。

1948 年, 美国数学家香农 (Shannon) 发表的《通讯的数学原理》是信息论的奠基性之作。信息论也为测度知识系统所需要的非参数统计学 (熵统计学) 奠定了基础。从 20 世纪 90 年代起, 莱兹多夫教授创新性地引入香农的信息熵理论模型, 将多变量分析与时序分析相结合探究科学知识发展的动力学特征。1954 年, 美国学者麦吉尔 (McGill) 则在香农信息熵理论的基础上, 提出用“构型信息” (configurational information) 表达三维或更多维度中的互信息转接量, 成为莱兹多夫教授对“大学—产业—政府”相互作用的“三螺旋”组织结构研究的重要知识基础。

表 3 1960 年以前的高引用参考文献统计信息

编号	参考文献信息	N_CR	N_PYEARS	N_TOP90
A1	Bradford, S.C., Sources of information on specific subjects (1934) Engineering, 137, pp. 85-86	13	8	3
A2	Schumpeter, J., (1943) Socialism, Capitalism and Democracy, London: Allen & Unwin	11	11	3
A3	Shannon, C.E., A mathematical theory of communication (1948) Bell System Technical Journal, 27, pp. 379-423	45	23	4
A4	McGill, W.J., Multivariate information transmission (1954) Psychometrika, 19 (2), pp. 97-116	27	10	9

4.2 莱兹多夫教授引用的 1960-1990 年间发表的重要文献

从 1960-1990 年间出版的文献来看 (见表 4), 莱兹多夫教授继承了科学计量学先驱者普赖斯 (Price) 和加菲尔德 (Garfield) 的学术思想, 同时也大量引介了科学社会学、数学和计算机科学领域经典文献中的理论和方法, 不断探索和夯实其在科学计量学领域的理论建构。

1965 年, 普赖斯 (Price) 借助于刊行不久的《科学引文索引》在 *Science* 期刊上发表的《科学论文的网络》一文研究了科学论文之间的引证和被引证关系, 以及由此形成的引证网络, 其网络分析的思想为科学计量学迈向科学知识图谱及知识可视化奠定了基础。他在 1970 年发表的《硬科学、软科学、技术与非科学的引文计量》一文则从引文视角对硬科学和软科学的差异性进行了量化研究, 阐释了不同学科的引用差异, 也是科学学理论及科学计量学研究方面的重要文献之一。

1972 年, 加菲尔德 (Garfield) 在 *Science* 期刊上发表《引文分析作为期刊评价的工具》, 提出了影响因子 (Impact Factor) 这一计量期刊影响力的指标, 是科研评价领域的基础性文献。他于 1979 年出版的《引文索引: 在科学、技术和文学中的理论与应用》 (中文版译名为《引文索引法的理论及应用》) 一书则详细阐述了引文索引法的概念、应用原理和方法, 以及其作为检索工具的重要意义和应用方法, 是引文研究的

经典著作。

1972年,荷兰计量经济学家泰勒(Theil)出版的《统计学的分解分析》是莱兹多夫教授很多研究中采用的数学方法之基础,尤其是莱兹多夫通过概率熵分解提出了三维互信息的数学表达式,成为“三螺旋”模型的核心算法。

美国管理学家、诺贝尔经济学奖获得者西蒙(Simon,又译名司马贺)是复杂系统理论的重要奠基者,其在1973年发表的《复杂系统的组织》一文中论述了系统层次结构的近可分解性(nearly decomposable),将层级结构的高频动态过程与低频动态过程区分开来,这一系统科学的研究思想被广泛地应用于物理学、化学、生物学、管理学等学科领域。莱兹多夫教授认为引文网络具有复杂系统的特征,节点的组团划分可以降低网络的复杂性。

德国学者索尔顿(Salton)是现代信息检索的奠基人,1983年他与麦吉尔(McGill)合著的《现代信息检索导论》提出了基于向量空间模型的相似度计算方法(Salton Cosine Similarity),被广泛地应用于相似度计算领域。莱兹多夫教授则在文献计量学领域共现分析的数据标准化过程中常借用该算法。

1984年,英国学者惠特利(Whitley)出版的《科学的智力组织和社会组织》一书从组织社会学角度对比分析了不同国家、不同科学领域的知识生产组织及其智力结构的特点,其尝试从科学领域的组织形态及其演变的侧面,去理解科学的多样性、科学结构及其演变、以及科学与社会的关系等问题,该书是科学社会学领域的经典著作。同年问世的《社会诸系统》一书是德国当代社会学家卢曼社会系统理论的奠基之作,其详细论述了将社会作为交流系统的模型,为科学知识交流的相关研究提供了重要的理论基础。

1989年,日本学者卡马达(Kamada)等学者提出了一种无向图的节点布局算法(又称为K-K算法)。图布局是利用图结构解决现实世界中信息可视化问题的一种重要技术,K-K算法作为经典布局算法被广泛应用于信息可视化领域。莱兹多夫教授在知识图谱的相关研究中较多地使用了该图布局算法。

表4 1960-1990年间高引用参考文献统计信息

编号	参考文献信息	N_CR	N_PYEARS	N_TOP90
B1	Price, D.J.D.S., Networks of scientific papers (1965) Science, 149, pp. 510-515	32	19	15
B2	Price, D.J.D.S., Citation measures of hard science, soft science, technology, and nonscience (1970) Communication among Scientists and Engineers, pp. 3-22., In C.E. Nelson (Eds.), Lexington, MA: Heath	27	15	15
B3	Theil, H., (1972) Statistical Decomposition Analysis, Amsterdam/London: North-Holland	57	25	20
B4	Garfield, E., Citation analysis as a tool in journal evaluation (1972) Science, 178 (4060), pp. 471-479	39	14	11
B5	Simon, H.A., The organization of complex systems (1973) Hierarchy Theory: The Challenge of Complex Systems, pp. 1-27., H.H. Pattee (Ed.), New York: George Braziller	29	17	14
B6	Garfield, E., (1979) Citation Indexing: Its Theory and Application in Science, Technology, and Humanities, New York: Wiley	27	17	15
B7	Salton, G., McGill, M.J., (1983) Introduction to Modern Information Retrieval, Auckland: McGraw-Hill	38	11	11
B8	Whitley, R.D., (1984) The Intellectual and Social Organization of the Sciences, , Oxford University Press, Oxford	44	21	21
B9	Luhmann, N., (1984) Soziale Systeme: Grundriss Einer Allgemeinen Theorie, , Suhrkamp, Frankfurt a. M., 1995, Social Systems (Stanford	39	17	15

	University Press, Stanford, CA)			
B10	Kamada, T., Kawai, S., An algorithm for drawing general undirected graphs (1989) Information Processing Letters,31(1),pp. 7-15	54	13	13

4.3 莱兹多夫教授引用的 1990 年以来发表的重要文献

从 1990 年以来出版的文献来看（见表 5），莱兹多夫教授引用较多的重要文献中包括了他本人的一些论著，这也体现了他对科学计量学发展的主要学术贡献，即对科学计量学理论方法的研究以及三螺旋理论和模型的建构。此外，科学社会学的知识生产经典模型、文献计量学的典型计算方法也在莱兹多夫教授的研究中经常引用。

1994 年，英国学者吉本斯（Gibbons）等人出版的《知识生产的新模式：当代社会科学研究的动力学》一书对比分析了科学技术研究的历史过程与社会方式，总结了知识生产的两种模式，即在学科的、规范的术语体系和科学研究方法体系下，以单学科研究为主的传统知识生产模式 1，以及在应用环境中，置身跨学科领域，利用交叉学科研究方法的新知识生产模式 2。该著作揭示了现代科学研究活动区别于学院式科学研究活动的特征，得到了学术界的广泛认同^[36]。莱兹多夫教授的三螺旋研究正是对知识生产模式 2 的拓展和深化。

1995 年，莱兹多夫教授出版的《科学计量学的挑战：科学交流的发展、测度和自组织》一书综合运用科学哲学、科学社会学和卢曼的交流理论、信息论、概率论的方法，研究了知识的自组织化过程，介绍了共词分析、合著关系分析以及引用分析等科学计量学方法，探索了如何通过数理模型模拟科学知识的增长，该著作集中汇聚了莱兹多夫教授的前期思想成果。另外，莱兹多夫教授也是三螺旋理论和模型的创始人，2000 年他与美国社会学家埃茨科威兹教授（Etzkowitz）发表了《创新动力学：从国家系统和模式 2 到大学-产业-政府的三螺旋关系》一文，系统阐释了知识经济时代大学、产业和政府之间的关系，揭示了创新的动态性以及知识生产与传播机构之间的复杂网络。三螺旋理论与方法弥补了已有创新理论的不足，被广泛应用于科技管理学、公共管理学、管理社会学等领域^[37]。

2003 年，瑞典学者阿尔格伦（Ahlgren）和比利时信息计量学家鲁索（Rousseau）等学者针对文献计量学领域引文共现数据的非正态分布问题，提出 Cosine 系数相对于 Pearson 系数更适合作为引文共现数据标准化的方法，该文章也是莱兹多夫教授在共现数据处理中常引用的重要文献之一。2009 年，英国学者拉佛尔斯（Rafols）和莱兹多夫教授研究了基于内容和算法的两种不同期刊分类方法对学科分析所带来的影响。莱兹多夫教授在文献计量学的评价研究中较多地引用了该篇文章。

表 5 1990 年以来高引用参考文献统计信息

编号	参考文献信息	N_CR	N_PYEARS	N_TOP90
C1	Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M., (1994) The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies, London: Sage	51	19	16
C2	Leydesdorff, L., (1995) The Challenge of Scientometrics: The Development, Measurement, and Self-organization of Scientific Communications, Leiden: DSWO Press, Leiden University	49	19	18
C3	Etzkowitz, H., Leydesdorff, L., The dynamics of innovation: From national systems and "mode 2" to a triple helix of university-industry-government relations (2000) Research Policy, 29 (2), pp. 109-123	41	16	16
C4	Ahlgren, P., Jarneving, B., Rousseau, R., Requirement for a cocitation similarity	52	13	13

	measure, with special reference to Pearson's correlation coefficient (2003) Journal of the American Society for Information Science and Technology, 54 (6), pp. 550-560			
C5	Rafols, I., Leydesdorff, L., Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects (2009) Journal of the American Society for Information Science and Technology, 60 (9), pp. 1823-1835	35	8	10

4.4 莱兹多夫教授引用重要文献的知识关联分析

对 19 篇重要文献的共被引分析（见图 2）发现，莱兹多夫教授的学术研究主要集中在两大领域：一是在知识经济背景下对现代科学研究活动的动力机制研究，核心主题包括：三螺旋研究、自组织研究和科技动态研究。其中，三螺旋研究的知识基础主要来源于香农和麦吉尔的信息熵理论（A3、A4）、泰勒的统计学分解分析方法（B3）、以及卢曼的社会系统理论（B9）。在自组织研究方面，莱兹多夫教授充分吸收了卢曼的社会系统理论思想（B9），并在其《科学计量学的挑战》（C2）一书中进行了充分阐述。在科技动态研究方面，吉本斯的知识生产模型 2（C1）和惠特利对科学组织形态及其演变研究（B8）都是莱兹多夫教授在该领域研究的重要学术根源。

二是对科学计量学理论与方法的研究，核心主题包括：引文研究、评价研究和知识图谱研究。其中，在引文研究方面，莱兹多夫教授很好地传承了普赖斯和加菲尔德的核心学术思想（B1、B2、B4、B6）。在评价研究方面，莱兹多夫教授不仅常回顾经典的布拉福德定律（A1），还对评价研究的准确性问题（C5）进行了思考。在知识图谱研究方面，索尔顿的相似度计算方法（B7）、卡马达的 K-K 图布局算法（B10）以及阿尔格伦的共现数据标准化方法（C4）都是莱兹多夫教授构建知识图谱的重要方法基础。

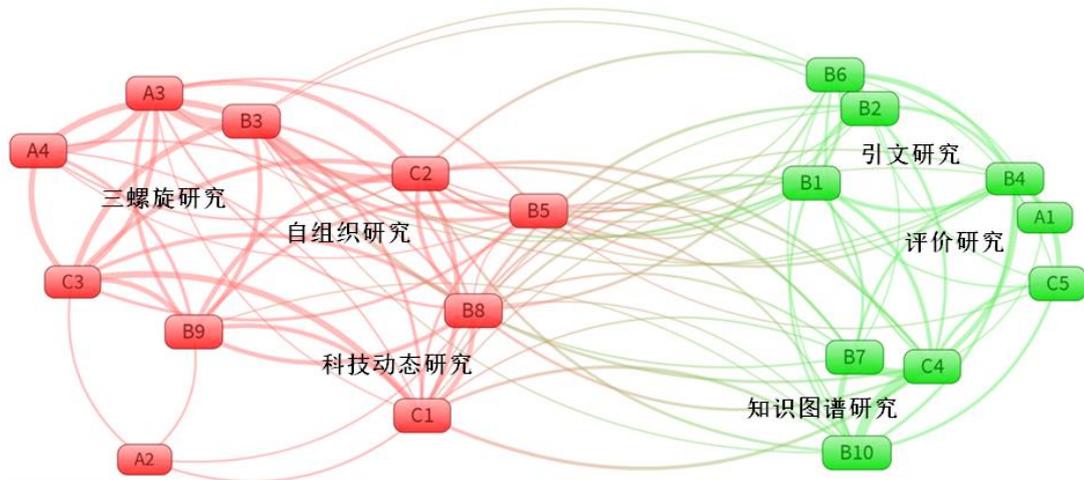


图 2 莱兹多夫教授引用的 19 篇重要文献的共现关系

参数设置：标准化方法为 Linlog/modularity，共现频次大于 10。

5 结论与启示

科学家是科学研究活动的主体，是科学理论和方法的创造者和实践者。因此，具体分析高水平科学家在学术论文中所频繁引用的重要文献的内容特征，可以客观反映其掌握科学知识的深度和广度，也可以揭示其研究的学术背景、科学依据与演进脉络。这样的工作对高水平科学家的理论和方法剖析较深，对其他

学者的相关研究工作应更具有启示性作用。莱兹多夫教授是世界著名的图书情报学家、科学计量学家和社会学家，选择他引用的重要文献为剖析样本，对于研究科学计量学学科的历史根源亦有帮助。

本文综合利用引文出版年光谱和引用内容分析方法，识别并挖掘莱兹多夫教授在学术论文中频繁引用的不同时期重要文献，研究结果表明：（1）莱兹多夫教授具有生物化学、哲学、社会学、图书情报学的多学科教育背景，其学术思想的根源可以追溯至科学哲学、科学社会学、数学、信息科学等学科领域的 19 篇重要文献，他借用的方法之多和汲取营养的学科之广，在当代图书情报学家中罕有其匹。（2）莱兹多夫教授的主要学术贡献表现在两个方面：一是长期多角度、多学科地思考和研究关于科学计量学的理论问题，并不断从事相关的实证研究以验证其理论建构，代表著作是《科学计量学的挑战》；二是对知识经济背景下现代科学交流中的自组织现象以及科学研究活动的动力机制开展研究，代表文章是《创新动力学：从国家系统和模式 2 到大学-产业-政府的三螺旋关系》。（3）莱兹多夫教授在学术研究中传承了科学计量学先驱者普赖斯和加菲尔德的学术思想，同时通过跨学科知识转移，不断创新和拓展科学计量学的思想体系和知识体系，再加上他的量化分析能力，使其可以从更广阔的学术视野审视科学和技术发展进程中的基础性问题。

莱兹多夫教授的引用行为特征也给我们带来了一些启示。一是，在学术研究中要重视经典著作或经典论文的反复细致研读。没有对原典的深入细致的研读与诠释，则很难把握本学科的精髓^[38]。图书情报学具有多学科交叉的特征，因此更需要进一步梳理和重新思考经典文献所带来的启示，避免因“嚼别人吃过的馍”而造成的理论方法的误解与误用。二是，科学计量学方法对优秀科学家的学术史研究具有较好的辅助性作用，尤其是通过引文出版年光谱分析可以清晰地掌握科学家学术发展线索，弥补了传统科学技术史主要依赖于对科学家原著的研究而未能关注其他重要施引文献的不足。因此，追溯科学家学术思想的源头，研读科学家频繁引用的经典文献，将为深入了解和全面辨析科学家的学术成就与贡献提供坚实的文献支持。

参考文献

- [1] 卡尔·波普尔. 猜想与反驳：科学知识的增长[M].傅季重，纪树立，译. 上海：上海译文出版社，2015.
- [2] 余丽嫦. 培根及其哲学[M].北京：人民出版社，1987.
- [3] 托马斯·库恩.科学革命的结构[M].金吾伦，胡新和，译.北京：北京大学出版社，2003.
- [4] Price, D.J. de Solla. Little Science, Big Science [M]. New York: Columbia University Press, 1965.
- [5] 董国豪，潜伟.普赖斯与科学史定量研究[J].科学学研究，2017，35（5）：667-680.
- [6] Garfield, E., Pudovkin, A. I., and Istomin, V. S. Why do we need algorithmic historiography? [J]. Journal of the American Society for Information Science and Technology, 2003, 54(5), 400-412.
- [7] Marx, W., Bornmann, L., Barth, A., et al. Detecting the Historical Roots of Research Fields by Reference Publication Year Spectroscopy (RPYS) [J]. Journal of the Association for Information Science and Technology, 2014, 65(4):751-764.
- [8] 李信，陆伟，李旭晖.一种新兴的学科领域历史根源探究方法：RPYS[J].图书情报工作，2016，60（20）：70-76.
- [9] Barth, A., Marx, W., Bornmann, L., et al. On the origins and the historical roots of the Higgs boson research from a bibliometric perspective [J].European physical journal plus, 2014,129（6）:1-13.
- [10] Marx, W. and Bornmann, L. Tracing the origin of a scientific legend by reference publication year spectroscopy (RPYS): the legend of the Darwin finches [J]. Scientometrics, 2014, 99(3):839- 844.
- [11] 侯剑华.基于引文出版年光谱的引文分析理论根源探测[J].情报学报，2017，36（2）：132-140.

- [12] Leydesdorff, L., Bornmann, L., Marx, W., et al. Referenced publication years spectroscopy applied to iMetrics: scientometrics, journal of informetrics, and a relevant subset of JASIST [J]. *Journal of informetrics*, 2013, 8(1):162-174.
- [13] 洛埃特·雷迭斯多夫. 科学计量学的挑战: 科学交流的发展、测度和自组织[M]. 乌云等, 译. 北京: 科学技术文献出版社, 2003.
- [14] 许海云, 方曙. 科学计量学的研究主题与发展——基于普赖斯奖得主的扩展作者共现分析[J]. *情报学报*, 2013, 32 (1): 58-67.
- [15] Garfield, E. Can citation indexing be automated? [J]. *Essays of an Information Scientist*, 1962, 1: 84-90.
- [16] Lipetz, B.A. Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators [J]. *American Documentation*, 1965, 16(2):81-90.
- [17] Voos, H. & Dagaev, K.S. Are all citations equal? Or, Did we op. cit. your idem? [J] *Journal of Academic Librarianship*, 1976, 1(6):19-21.
- [18] Halevi, G. & Moed, H.F. The thematic and conceptual flow of disciplinary research: A citation context analysis of the journal of informetrics, 2007[J]. *Journal of the American society for information science and technology*, 2013, 64 (9): 1903-1913.
- [19] Moravcsik, M.J. & Murugesan P. Some results on the function and quality of citations [J]. *Social studies of science*, 1975, 5 (1):86-92.
- [20] Chubin, D.E. & Moitra, S.D. Content analysis of references: adjunct or alternative to citation counting? [J]. *Social studies of science*, 1975, 5 (4):423-441.
- [21] Ding, Y., Liu, X., Guo, C., et al. The distribution of references across texts: some implications for citation analysis [J]. *Journal of Informetrics*, 2013, 7(3):583-592.
- [22] Frosig, R.E. Citation classification based on genre: the significance of the textual location of citations [D]. Royal School of Library and Information Science, Denmark, 2011.
- [23] Agarwal, S. & Yu, H. Automatically classifying sentences in full-text biomedical articles into Introduction, Methods, Results and Discussion [J]. *Bioinformatics*, 2009, 25(23):3174-3180.
- [24] Agarwal, S., Choubey, L. and Yu, H. Automatically classifying the role of citations in biomedical articles [J]. *AMIA Annual Symposium*, 2010(1977):11-15.
- [25] Zhang, G., Ding, Y. and Milojevic, S. Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content [J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(7):1490-1503.
- [26] 刘盛博, 丁堃, 张春博. 引文分析的新阶段: 从引文著录分析到引用内容分析[J]. *图书情报知识*, 2015 (3): 25-34.
- [27] 祝清松, 冷伏海. 引文内容分析方法研究综述[J]. *情报资料工作*, 2013, 38 (5): 97-107.
- [28] 陆伟, 孟睿, 刘兴帮. 面向引用关系的引文内容标注框架研究[J]. *中国图书馆学报*, 2014, 40 (6): 93-104.
- [29] 张梦莹, 卢超, 郑茹佳, 等. 用于引文内容分析的标准化数据集构建[J]. *图书馆论坛*, 2016 (8): 48-53.
- [30] 章成志, 王玉琢, 卢超. 学术专著引用行为研究——基于引文内容特征分析的视角[J]. *情报学报*, 2017, 36 (3): 319-330.
- [31] 祝清松, 冷伏海. 基于引文内容分析的高被引论文主题识别研究[J]. *中国图书馆学报*, 2014, 40 (1):

39-49.

- [32] 刘盛博, 王博, 唐德龙, 等. 基于引用内容的论文影响力研究——以诺贝尔奖获得者论文为例[J]. 图书情报工作, 2015, 59 (24): 109-114.
- [33] Mongeo, P., Paul-Hus, A. The journal coverage of Web of Science and Scopus: A comparative analysis [J]. *Scientometrics*, 2016, 106(1):213-228.
- [34] Thor, A., Marx, W., Leydesdorff, L., et al. Introducing Cited References Explorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization [J]. *Journal of Informetrics*, 2016, 10(2) :503-515
- [35] 赵勇, 高思嘉, 武夷山. 见微知著: 一位优秀科学计量学家采用的研究方法之分类编码分析[J]. 情报学报, 2017, 36 (5): 443-451.
- [36] Hessels, L.K. and Van Lente, H. Re-thinking new knowledge production: A literature review and a research agenda [J]. *Research Policy*, 2008, 37(4):740-760.
- [37] 叶鹰, 莱兹多夫, 武夷山. 三螺旋模型及其量化分析方法研讨[J]. 中国软科学, 2014 (11): 131-139.
- [38] 武夷山, 王海燕. 情报学的“三原法” [J]. 情报学报, 2014, 33 (5): 1.