

网络引证视角的 知识交流规律研究

杨思洛 著

湘潭大学出版社

本书获湘潭大学学术著作出版资助

本书系国家社科基金项目“中外图书情报学科知识图谱比较研究”
(11CTQ015) 和湘潭大学博士科研启动项目“基于网络引证关系的知识交流
规律研究”(11QDGG04) 的研究成果

网络引证视角的知识交流规律研究

杨思洛 著

湘潭大学出版社

图书在版编目(CIP)数据

网络引证视角的知识交流规律研究 / 杨思洛著. —
湘潭: 湘潭大学出版社, 2012.6
ISBN 978-7-81128-391-4

I. ①网… II. ①杨… III. ①知识传播—研究 IV.
①G2

中国版本图书馆 CIP 数据核字 (2012) 第 070404 号

责任编辑: 魏 杰

封面设计: 刘 政

出版发行: 湘潭大学出版社

社 址: 湖南省湘潭市 湘潭大学出版大楼

电话(传真): 0731-58298966 邮编: 411105

网 址: <http://xtup.xtu.edu.cn>

印 刷: 湘潭地调彩印厂

经 销: 新华书店

开 本: 787×1092 1/16

印 张: 15.75

字 数: 384 千字

版 次: 2012 年 6 月第 1 版 2012 年 6 月第 1 次印刷

书 号: ISBN 978-7-81128-391-4

定 价: 38.00 元

(版权所有 严禁翻印)

目 录

引言	(1)
第一节 选题背景与研究意义	(1)
第二节 国内外研究现状分析	(5)
第三节 研究目标、内容和研究方法	(16)
第1章 知识交流研究的发展	(23)
第一节 知识交流及相关概念	(24)
第二节 知识交流的研究视角	(30)
第三节 知识交流的历史演进	(37)
第四节 知识交流的研究进展	(41)
本章小结	(53)
第2章 网络引证与知识交流规律分析	(55)
第一节 传统引证关系分析	(55)
第二节 网络引证关系分析	(59)
第三节 网络引证视角的知识交流特点	(67)
第四节 网络引证视角的知识交流规律的形成	(78)
第五节 网络引证视角的知识交流规律研究方法与步骤	(91)
本章小结	(99)
第3章 网络引证视角的知识交流分布规律研究	(101)
第一节 基于 P - W 网络引证关系的知识交流分布研究	(101)
第二节 基于 W - P 和 W - W 网络引证关系的知识交流分布研究	(122)
本章小结	(143)
第4章 网络引证视角的知识交流变化规律研究	(145)
第一节 基于 P - P 网络引证关系的知识交流整体变化研究	(145)

第二节	基于 P - P 网络引证关系的知识交流老化研究	(156)
第三节	基于 P - W 网络引证关系的知识交流老化研究	(170)
本章小结		(192)
第 5 章	网络引证视角的知识交流优化策略	(195)
第一节	重视网络引文的理论研究	(195)
第二节	利用网络引证关系的分布规律	(197)
第三节	改善网络知识交流的环境	(200)
第四节	加强网络引文数据库建设	(203)
第五节	构建网络引文可获得性体系	(206)
本章小结		(213)
第 6 章	结论与展望	(215)
第一节	本书的结论	(215)
第二节	研究的不足	(217)
第三节	研究的展望	(217)
参考文献		(219)
后 记		(239)

引言

知识就是力量……知识的力量不仅取决于其本身价值的大小，更取决于它是否被传播（交流），以及传播（交流）的深度和广度。

——英国哲学家、思想家、作家和科学家弗兰西斯·培根^{①②③}

第一节 选题背景与研究意义

一、选题背景

随着经济全球化的不断深入和科学技术的飞速发展，我们所处的世界正在进入一个全新的知识经济时代。知识取代了传统的资本、劳动力等要素成为重要的战略资源。“知识就是力量”，“知识就是财富”，“知识就是竞争力”，“知识就是生产力”……这已经成为知识社会人们的共识。

知识创新是国家创新体系的重要组成部分，也是获取竞争优势的关键，目前受到了国家、单位和个人的普遍重视；而知识具有累积性、共享性等特征，使知识交流成为知识创新的前提和基础，成为促进科学技术和知识经济发展的加速器。世界经合组织 OECD 在

① 袁清林．科普学概论．北京：科学技术出版社，2002：1．

② 李勇．西方公民科学素质建设的文化语境研究．理论界，2006（9）：210-211．

③ 松鹰．培根的三句名言．凤凰读书．[2011/1/31]．http://book.ifeng.com/yysl/200809/0916_3564_786869.shtml．

《国家创新系统》中系统总结了知识创新和国家创新系统的含义、意义和内容,并且认为国家创新系统的关键内容是科学技术知识在个人、企业、机构间的流动;创新和技术的发展是企业、大学、政府研究机构组成的系统,通过复杂的联系与交流的结果。^①国内,江泽民同志早在1995年的全国科学技术大会上就卓有远见地指出:“创新是一个民族进步的灵魂,是国家兴旺发达的不竭动力”。在2010年6月的中国科学院第十五次院士大会、中国工程院第十次院士大会上,胡锦涛总书记再次强调全力建设创新型国家的目标,并认为:“知识是发展永恒的重要资源,知识创新成为国家竞争力的核心要素。”^②

随着科学技术的快速发展,特别是互联网的普及应用,网络环境正改变着知识交流的手段、方式、结构和过程等。

(1) 计算机网络技术的发展与应用普及。据中国互联网络信息中心(CNNIC)最新发布的第29次中国互联网络发展状况统计报告,截至2011年12月底,中国网民数量突破5亿,达到5.13亿,全年新增网民5580万;互联网普及率较上年底提升4个百分点,达到38.3%;使用台式电脑上网的网民比例为73.4%(比2010年底降低5个百分点),手机上升至69.3%,笔记本电脑也略增至46.8%;2011年底搜索引擎用户规模达到4.07亿,在网民中的渗透率为79.4%。^③该报告还显示,我国网民的互联网应用表现出商务化程度迅速提高、娱乐化倾向有所下滑、沟通和信息工具价值加深的特点。目前,搜索引擎和各种网络互动问答平台成为人们获取知识、解决问题、寻找信息的首选之一。^④

(2) Web 2.0时代的到来。Web 2.0是以人为核心线索的网,强调用户参与和协作,实现了“草根革命”,能为用户带来真正的个性化、去中心化和信息自主权,实现网络应用从信息内容的互联到人的社会互联。它的到来是一场影响深远的草根革命,真正将互联网的主导权交还个人,从而充分发挥了个人的积极性。它使广大网民所贡献的影响和知识以及个人联系形成的社群的影响替代了原来少数人所控制和制造的影响,从而极大解放了个人的创作和贡献的潜能,使得互联网的创造力上升到了新的量级,用户贡献的数字知识资源因此海量增长。^⑤同时,Web 2.0技术也改变了人们交流和合作的手段,博客、微博、Wiki、RSS等技术应用极大地便利了知识交流的进行。

(3) 开放获取运动的开展。自从2001年12月布达佩斯会议提出了开放存取的思想并创立了《布达佩斯开放存取先导》(Budapest Open Access Initiative, BOAI)后,金色通道——(开放存取期刊)、绿色通道——(机构知识库)等风靡全球,网上的学术知识资源交流与获取得到了极大的便利。例如,2008年在Elsevier所有的25个学科期刊中,平均可开放存取比例为57.53%;所有学科可开放存取期刊的比例均超过30%,超过80%的

① OECD. National Innovation Systems. [2011/1/31]. <http://www.oecd.org/dataoecd/35/56/2101733.pdf>.

② 胡锦涛. 强调全力建设创新型国家. 人民日报(海外版). 2010, 6(8): 1

③ CNNIC. 第29次中国互联网络发展状况统计报告. [2012/1/31]. <http://tech.hexun.com/2012/cnnic29/>.

④ 杨思洛, 毕艳娜. 搜索引擎的互动问答平台及其对数字参考咨询服务的启示. 图书情报工作, 2007(2): 140-144.

⑤ 杨思洛, 毕艳娜. 基于Web 2.0的数字参考咨询服务创新. 图书馆学研究, 2007(2): 75-78.

有2个学科;介于70%—80%的有8个学科。^①

(4) 网络出版的进行。目前,知识的传递与交流纷纷依靠先进的数字技术与网络平台,电子期刊、数字图书等资源迅速发展。据统计,国内2006年数字出版产业达213亿元,2007年达362.42亿元,2008年达530.64亿元,2009年的产值达799.4亿元,是2006年产值的3.75倍,年均增长率超过55%,产业增长率继续保持高速增长速度。其中数字期刊收入6亿元,电子书收入达14亿元,数字报(网络版)收入达3.1亿元,手机出版则达到314亿元。^②现在几乎所有的国内外传统期刊都在网络数据库中有对应的数字化版本。因为数字技术的优势,许多纯网络出版的知识资源也大量出现,它们具有图文声像等多种形式,因检索准确、使用容易等特点而深受人们、特别是青少年的喜爱。

互联网使知识交流发生了质的变化,主要表现为网络信息基础环境逐步取代传统信息基础环境,而占据主导地位。知识交流系统面临着新一轮的重组与整合。互联网的发展一方面为知识交流的发展创造了机遇。它突破了时空的界限与交流的障碍,知识的创造、传递和获取都得到了极大的便利,使得全世界成为一个知识交流与共享的大平台。同时,互联网也给知识交流与共享带来了一系列问题。在网络知识海洋中,人们很难找到真正需要的、正确的、具有针对性的有效知识。“渴死在知识海洋中”形象地描述了这一严峻的现状。知识超载、知识污染、知识安全、知识产权、知识焦虑、知识鸿沟、知识老化等社会问题给知识交流带来的负面影响日益突出。^③例如,维克格伦(Wikgren)^④通过对网络上不同信息源文本(科学的、专业的、教育的、商业的文本,以及糖尿病讨论组信息文本)的研究分析,发现网络的非正式和非专业性给医药知识交流带来了新的问题。传统环境下,信息计量学的信息分布、增长与老化规律极大地优化了知识交流和知识资源管理的开展;与之相对应,要解决网络知识交流这些日益突出的问题,关键是研究网络知识交流本身,揭示知识交流新的特征、数量关系和内在规律,以指导网络知识交流的实践活动。

文献之间的引证关系较客观真实地表征了知识的累积性、连续性和继承性,可以描述、解释、预测和评价知识交流行为与过程;互联网的普及应用,引证行为、动机及引证的表现都发生了系统变化,网络信息资源将越来越多地成为人们科研参考的来源,网络引证关系应运而生。^⑤因此,从网络引证视角研究知识交流现象,系统研究网络知识交流的内容,深度揭示其规律,通过知识交流和知识创新来推动创新型国家建设是政府管理部门、学术界都十分关注的重要课题,也是摆在我们面前的一项紧迫而艰巨的研究任务。

① 黄如花,张静. Elsevier 收录期刊可开放存取情况的调查与分析. 中国图书馆学报, 2009 (3): 35-42.

② 郝振省. 2010 中国数字出版产业年度报告. [2011/1/31]. http://news.xinhuanet.com/newmedia/2010-07/22/c_12360171.htm.

③ 张永晖. 知识交流机制研究. 哈尔滨: 黑龙江大学硕士学位论文, 2005: 1-2.

④ Wikgren M. Health discussions on the Internet: a study of knowledge communication through citations. Library & Information Science Research, 2001, 23 (4): 305-317.

⑤ 杨思洛. 国外网络引文研究的现状及展望. 中国图书馆学报, 2010 (4): 72-82.

二、研究意义

在网络发展的背景下,新一代科研环境产生,基于网络引证关系来研究知识交流规律有着美好的前景和重要意义,同时也期待新的发展、新的突破、新的繁荣。本研究的意义具体体现在:

(一) 推进国家创新体系建设

我国“十五”规划《纲要》提出:“建设国家创新体系”;“建立国家知识创新体系,促进知识创新工程”;实施“跨越式发展”的宏伟战略。知识创新是其他创新活动的基础。在国家创新体系中,知识创新具有重要地位,发挥关键作用。“创新源于交流”,知识创新与其他活动不同的本质特征就是其具连续性与继承性。通过网络引证视角的知识交流规律的研究与应用,有利于营造良好的知识交流环境,构建有序的知识交流模式,促进知识创新活动。

(二) 加强网络信息资源管理

互联网作为现代社会重要的信息载体,为人类共享信息提供了极为便捷的方式和发展空间,也为广大科技研究人员阐述学术观点、交流创新思想和共享知识提供了重要的平台。但是,由于网络的自由与开放性,网上科技信息具有数量巨大、类型多样、变化频繁、结构复杂、质量混杂等特点。在人们享受着“E-Research”所带来的种种便利的同时,也深受网上科技信息诸多问题的困扰。要有效解决新环境所带来的各种问题和矛盾,保证网上科技信息资源的高效利用与可持续发展,就必须对其进行科学有效的管理。从网络引证关系的视角,系统研究网络知识交流的规律,有利于网上信息资源的开发利用,为改善科技管理和制定科技政策服务,适应网络时代科技发展的需要;利于网上知识资源的优化配置,引导科研资源流向,适应网络时代经济发展的需要。

(三) 促进学科繁荣发展

互联网正在平静而快速地改变着世界,同时也在改变着社会的生活方式和工作方式;在这场互联网引起的“变革”中,不但作为一门学科的图书馆学情报学面临巨大的挑战,而且我们所从事的图书情报实践也面临巨大挑战;不仅在中国是这样,在国外,甚至在西方发达国家都是这样,至多是程度的不同、表现形式的不同。^① 由于在科学评价、信息检

^① 黄长著等著. 网络环境下图书情报学科与实践的发展趋势. 北京: 社会科学文献出版社, 2010: 1.

索和知识发现等方面的重要应用,引证研究成为图书情报学科十分关键的领域,成为其影响其他学科的重要方面。网络引证关系是传统引文在网络环境下的延续与发展,是当前研究的重要方向和趋势。引文分析能否在新环境下重塑辉煌、为图书情报学带来新契机,有待网络引证关系研究的完善与发展。

第二节 国内外研究现状分析

知识交流有着悠久的历史,人类知识的无限性、丰富性和个人认识的局限性,以及个人之间的知识差异性,决定了人们之间进行知识交流的必要性。^①可以说,从古代的结绳记事开始就有了知识交流的萌芽;到文字的发明,改变了原始面对面的口口交流为主要方式的知识交流;纸张和印刷术的出现更是对知识交流造成了变革性的影响;到现代计算机网络技术的兴起、知识社会的到来,知识交流又到了一个新的发展时期。人们从不同角度对知识交流进行了较深入的研究,如从知识管理学角度、从科学社会学的角度、从图书馆学的角度、从情报学的角度等。

根据我国颁布的《中华人民共和国国家标准文献著录总则》(GB 3792.1-83)定义,文献是记录有知识的一切载体。^②在国内,“文献”一词最早见于《论语》。从古代开始,文献就包括了“客观知识”与“主观知识”的意思,是知识交流与传承的重要载体。^③无论是代与代之间的纵向交流还是同时段的横向知识交流,文献一直是知识交流的重要组成部分,文献之间的引证关系其实就是一种重要的知识交流形式与表征。网络环境下,文献发生了巨大变化,主要形成了传统文献(print, P)和网络文献(web, W)两大类型。一方面,传统文献仍是人们进行知识交流的主阵地;另一方面,网络资源和网络交流手段以其无可比拟的优点受到人们的青睐,它们之间组成的网络引证关系如图0.1所示。传统引证是传统环境下纸质文献间的参考行为的结果。网络引证在传统引证的基础上发展而来,但与传统引证有显著不同:在施引过程中,网络引证的作者主体、文献对象及引证行为都存在于网络环境中;另外,网络引证分析通过各种网络数据库或搜索引擎进行。网络引证与一般网络链接也有区别:首先,网络引证表征了文献之间的引证关系,突出特点是其知识性、学术性;网络链接通过超链接实现,而网络引证虽可以超链形式出现,但大多为引文格式并通过引文分析方法研究。

① 宓浩. 知识、知识材料和知识交流——图书馆情报学引论(纲要)之一. 图书馆学研究, 1983 (6): 28-35.

② 刘炜. 数字图书馆的语义描述和服务升级. 北京: 国家图书馆出版社, 2010: 17.

③ 李明杰. 中国古代图书馆学的知识论取向——从文献学路径获得的认知. 中国图书馆学报, 2010 (1): 27-34.

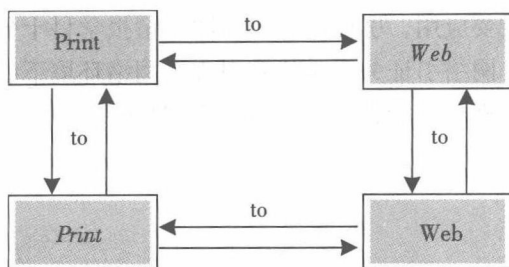


图 0.1 网络环境下文献间的引证关系模型

网络环境下，文献发生了巨大变化，网络引证关系相应地从传统文献与网络文献两个角度可分为四大类：Print - to - Print (P - P) 网络引证（是指网络环境下的传统论文间的引证关系，主要集中在各类网络传统论文数据库中引证关系的研究）；Print - to - Web (P - W) 网络引证 [更多的是探讨传统论文参考文献中的网络成分，突出表现是参考文献中有网址 (URL)]；Web - to - Print (W - P) 网络引证 [是指从网络文献引证纸质文献（包括期刊论文、会议文献、手册指南、专著等）]；^① Web - to - Web (W - W) 引证（是指源文献和引证文献都属于网络文献，在早期被认为属于链接分析的概念范畴，但它们与一般的链接有显著不同，虽然有些 W - W 网络引文以超链形式出现，但它以引文分析为基础，是一种“情报学视角的链接分析”）。^② 需要说明的是，网络引证的复杂关系远非这四种类型的简单区分关系，在一定程度上它们有交叉和重合：一方面，有许多文献既以传统形式存在也以网络文献形式存在，显著特点就是在被引时传统格式和网络引用格式两者都有著录；另一方面，许多网络文献仍以传统形式进行引用，例如网络开放获取文献的著录。

从信息计量学角度，知识交流规律又分为两大类：分布规律与变化规律。因此，网络引证视角的知识交流规律研究，可以网络 and 传统两大文献为基点，从四种网络引文形式、两大规律入手进行展开，相应的分布如图 0.2 所示。

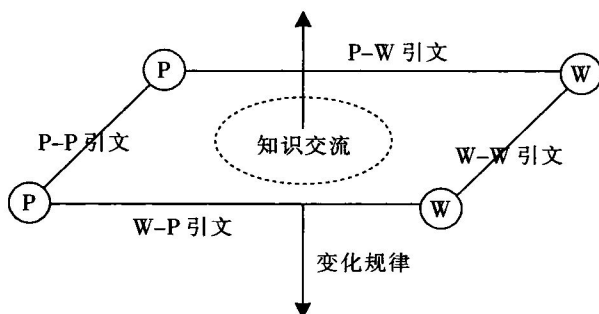


图 0.2 网络引证视角的知识交流规律分布

① Vaughan L., Shaw D. Web citation data for impact assessment: a comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, 2005, 56 (10): 1075 - 1087.

② 孙建军, 李江. 网络信息计量理论、工具与应用. 北京: 科学出版社, 2009: 29 - 30.

一、国外研究现状分析

利用 Web of Knowledge 数据库平台、ProQuest 博士论文全文数据库和 Google Scholar 等网络数据库,从多个角度检索“web citation”、“web reference”、“scholarly communication”、“scientific communication”、“knowledge communication”等关键词,浏览筛选、阅读相关论文,并对重点论文进行扩展性查找。通过对检索结果进行整理和分析及平时对相关研究内容的积累,笔者发现国外学者对网络引证视角的知识交流研究主要从以下几个方面展开:^①

(一) 网络引证视角的知识交流分布规律研究

早在 1965 年,著名学者普赖斯(Price)对传统引证网络分布规律的研究,发现了引证过程中的累积效应,通过大量数据证明引证频次符合幂律分布。^② 1983 年,“科学引文索引(Science Citation Index, 简称 SCI)之父”尤金·加菲尔德(Garfield E.)通过 SCI 引证数据,绘制了引文分布曲线,证明引文的累积总量是被引期刊数的函数,认为引文分布具有明显的聚集现象,其集中趋势比布拉德福(Bradford)定律描述的论文分布更加明显:《科学引文索引》数据库中所有参考文献的 75% 来自不到 1 000 种被引期刊;500 种期刊发表的被引文献占 SCI 收录参考文献的 70%。^③ 1998 年,雷德纳(Redner)对《物理学评论 D》(*Physical Review D*)20 年的论文分析表明,论文被引量与其排名之间的齐普夫曲线与幂律分布一致,其指数大约等于 3。^④ 2005 年,雷德纳又对 110 年内发表在《物理学评论》(*Physical Review*)上所有论文的引证情况加以分析,发现引文同样符合幂律分布。^⑤ 2010 年,彼德森(Peterson)等人提出了引文分布的双机制(即直接引证和间接引证两种)模型,高被引论文大多是通过间接引证机制形成的;由于两种机制的存在,使得引文幂律分布的指数不一致,高被引论文往往有更小的幂指数。^⑥

在基于 P-W 网络引证关系的知识交流分布方面,许多研究通过一定量的引文数据进行了初步的分析。德拉瓦尔(Dellavalle)^⑦ 选用《科学》(*Science*)等三种影响因子排名

① 杨思洛. 国外网络引文研究的现状及展望. 中国图书馆学报, 2010 (4): 72-82.

② Price D. J. D. Networks of scientific papers. Science, 1965, 149 (3683): 510-515.

③ Garfield E. Citation analysis as a tool in journal evaluation. Science, 1972 (178): 471-479.

④ Render S. How popular is your paper? An empirical study of the citation distribution. European Physical Journal B, 1998 (4): 131-134.

⑤ Redner S. Citation statistics from 110 years of Physical Review. Phys Today, 2005 (58): 49-54.

⑥ Peterson G. J., Pressé S., Dill K. A. Nonuniversal power law scaling in the probability distribution of scientific citations. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107 (37): 16-23.

⑦ Dellavalle R. P. et al. Information science. Going, going, gone: lost Internet references. Science, 2003, 302: 787-788.

前1%的高质量美国期刊,统计论文中P-W引文的频率和类型,结果显示30%论文包括至少1条P-W引文,占有所有引文的2.6%,其比例最高的论文类型为综述。奥曼(Oermann)^①对护理期刊的分析表明,平均每篇论文有3.1条P-W引文,大多P-W引文与临床实践相关。维纳(Veena)^②则选择9种电子期刊2000—2006年的论文研究,结果表明,所有25730条引文中的43.52%为P-W引文,81.49%论文有网络参考文献。除了简单介绍P-W引文使用率外,布朗(Brown)^③以8种化学期刊为样本,发现虽然P-W引文使用比例较少,但在稳步增长;P-W引文在论文结构的参考列表中出现数量最多(43%),其次是论文实验部分;P-W引文本身的类型中,纯文本的比例最大(85%),其次是文本和图像混合型;论文使用P-W引文没有影响到其总体引证率。

W-P引文主要是指被网络文献所引证的传统纸质文献出版物。目前在此角度研究知识交流分布集中在W-P引文与P-P引文的比较、W-P引文的类型及其引文数据来源。沃恩(Vaughan)^④对比了46种图书情报学期刊论文的W-P引文与P-P引文,她使用Google提取W-P引文,对W-P引文与P-P引文的异同展开研究,发现:57%的期刊的网络被引与SSCI和JCR中的被引数显著相关;许多网络引证是出于学术目的(30%的为传统论文张贴在网上,12%的为课程阅读书目);对于同一篇论文,通常W-P引文数要比P-P引文数多;对于有自己网站的期刊,其W-P引文往往更多,从1992年到1997年,期刊论文W-P引文数逐年增加。尽管W-P引文有许多优势,但她认为在近期不可能替代P-P引文分析。在后续研究中,沃恩^⑤又选择代表美国图书情报研究者学术成果的1483种出版物,用WoS(Web of Science)、Google和Google Scholar检索,结果显示:在WoS中,各类型论著(除图书外)被引的中位数为0;在Google Scholar中,期刊论文和图书的被引中位数分别为1和3;而在Google中,中位数最大的为图书,达41,最小的为会议论文,为9。通过引文分类分析,发现Google Scholar中有92%的引文是学术性引证(主要来源于期刊论文);在Google中书目服务(非学术性引证)是最大的引文源;开放获取期刊通常有更多的W-P引文,但传统期刊的W-P引文往往代表着学术性引证。卢伊特(Luyt)和唐(Tan)对Wikipedia中条目的参考文献分布进行了初步分析,发现图书占的比例较大(34.9%);而学术期刊论文很少,仅为0.8%。^⑥

① Oermann M. H. Web citations in the nursing literature: how accurate are they. *Journal of Professional Nursing*, 2008, 24 (6): 347-351.

② Veena R. Web citation behavior in scholarly electronic journals in the field of library and information science. *Webology*, 2008, 5 (2). [2011/1/31]. <http://www.webology.ir/2008/v5n2/a57.html>.

③ Brown C. The role of web-based information in the scholarly communication of chemists: citation and content analyses of American Chemical Society Journals. *Journal of the American Society for Information Science and Technology*. 2007, 58 (13): 2025-2065.

④ Vaughan L, Shaw D. Bibliographic and web citations: what is the difference?. *Journal of the American Society for Information Science and Technology*. 2003, 54 (4): 1313-1324.

⑤ Vaughan L. A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*. 2008, 74 (2): 317-330.

⑥ Luyt B., Tan D. Improving Wikipedia's credibility: references and citations in a sample of history articles. *Journal of the American Society for Information Science and Technology*. 2010, 61 (4): 715-722.

在基于链接关系的 W - W 引证视角研究知识交流分布方面, 鲁索 (Rousseau)^① 是较早对网络学术信息的引证情况深入研究的学者之一。他积极推荐采用 Sitation 术语, 系统比较了网络链接关系与传统引证关系。他认为, 从用户动机角度分析, 两者有相似之处, 但又不尽相同, 网络引证的情况要复杂得多。他对 343 个网址的分布模式和进入网页的链接情况作分析, 发现网址中的最高层域名服从洛特卡 (Lotka) 定律, 甚至对这些网址的引证也符合洛特卡定律。崔 (Cui) 以美国 19 个顶尖医学院校网站为样本, 证明网站的被引频次符合布拉德福定律。^② 史密斯 (Smith)^③ 首创将引文分析方法应用于网络期刊。他选择 22 个澳大利亚科学家经常参阅的网络期刊, 利用 AltaVista 统计从其他网站指向这些网络期刊的超链接。他认为指向网络期刊的超链接与引文有本质的区别, 因为超链接的目标是整个期刊, 而引文却是针对个别文章。在其后的研究中, 他又调查了哪些链接与传统引文相似, 通过对链出页和链入页及链接的动机的特点进行分类, 发现总体上 20% 的链接可看作学术链, 类似于传统引文的功能。^④ 基于网络论文的 W - W 引证视角, 蔻莎 (Kousha)^⑤ 提出了 URL 引文这一网络引文形式, 他定义其为网络文献的 URL 在另一网页文本中被提到 (不论是否为超链接)。通过分类, 他对图书情报学开放获取期刊论文的网络引证动机进行了检验, 提取了是正式学术交流的 URL 引文的五个特征。结果表明, 通过 Google 搜索, 2000 年 15 种期刊 182 篇论文共收到 3 045 个 URL 引文, 这些引文的 43% 有着与传统引文相同的正式学术动机, 18% 为非正式学术动机。在这些 URL 引文中, 82% 是英文, 88% 可获取全文, 58% 是非 HTML 档; 正式学术交流目的的 URL 引文主要源于非 HTML 文档中的文本类型。在后续研究中^⑥, 他选择 2001 年的 108 种开放获取期刊中的 1 080 篇论文, 对 URL/Web 引文 (定义为在一个网页中提到论文的网址或题名或两者都有) 的统计使用 Google 和 Google Scholar, 并与 WoS 引文对比, 除心理学外, 这三种引文数据显著相关。与 Google 得出的引文相比, Google Scholar 引文与 WoS 引文的相关度更高; 在社会科学领域, Google Scholar 比 WoS 中有更多的引文数; Google Scholar 与 WoS 引文的比较, 存在显著的学科差别; 虽然 W - W 引文有很好的发展前景, 但目前用它来替代 WoS 还存在许多问题。在 2007 年, 蔻莎^⑦ 又对 64 种自科领域开放获取期刊的 1 577 个 W - W 引文进行深入的分析, 结果显示: 仅有 25% 的引文为学术目的, 其中网络

① Rousseau R. Sitations: an exploratory study. *Cybermetrics: International Journal of Sociometrics, Informatics and Bibliometrics*, 1997, 1 (1): 4 - 7.

② Cui L. Rating health web sites using the principles of citation analysis; a bibliometric approach. *Journal of Medical Internet Research*, 1999, 1 (4). [2011/1/31]. <http://www.jmir.org/1999/1/e4/index.htm>.

③ Smith A. G. A table of two web spaces; comparing sites using web impact factors. *Journal of Documentation*, 1999, 55 (9): 577 - 592.

④ Smith A. G. Web links as analogues of citations. *Information Research*, 2004, 9 (4): 188.

⑤ Kousha K., Thelwall M. Motivations for URL citations to open access library and information science articles. *Scientometrics*, 2006, 68 (3): 501 - 517.

⑥ Kousha K., Thelwall M. Google Scholar citations and Google web/URL citations: a multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 2007, 58 (7): 1055 - 1065.

⑦ Kousha K., Thelwall M. How is Science cited on the web? A classification of Google unique web citations. *Journal of the American Society for Information Science and Technology*, 2007, 58 (11): 1631 - 1644.

文档引证占 23%，其他非正式学术引证占 2%；许多引证仅是一般的导航（占 45%）和自我宣传（22%）；另外，W-W 引文类型在学科间也存在差异。

（二）网络引证视角的知识交流变化规律研究

互联网对知识交流和知识利用行为产生了重大影响，一些研究者从引证角度对知识交流的总体变化规律进行研究。但由于样本和研究方法的差异，得出的结论有很大不同。例如，芝加哥大学社会学家埃文斯（Evans）在《科学》杂志发表的研究^①，调查了 1945 年至 2005 年间被引证的来自各个学科领域的 3 400 万篇论文，结果发现引证行为在向更小的论文“池”集中。需要说明的是，这种现象并不意味着期刊的总引证次数变少，只是大部分引证向少数文章集中。埃文斯强调这一趋势的严重后果是“每个人都在向已发表论文的中央集合移动”。而这种效应会更容易导致学术趋同性，并减少学术界的争鸣声音。拉里维埃（Larivière）^②以自然和工程科学、医学、人文和社会科学的论文为例，通过 3 项指标分析了 1900 年至 2007 年 WoS 中论文引文的集中度变化，结果与埃文斯的结论刚好相反：引文的分散度明显增加。在另一项研究中，拉里维埃^③也得出与埃文斯相左的结论：当前的学术论文引证越来越多的是老文献。埃格希（Egghe）^④用数学模型证实了拉里维埃的这一结论，表明参考文献年龄的均值和中位数都在增加，而普赖斯指数在下降。

从基于 P-W 网络引证关系研究知识交流变化，目前集中在 P-W 网络引文的可获得性。哈特（Harter）^⑤是此领域最早的研究者之一，他对 74 种电子期刊 1993—1995 年的 83 篇 P-W 引文加以检验，发现在 1995 年仅有 43 篇可获取，同时得出电子期刊的 P-W 引文半衰期为 1.5 年。此后有众多学者从不同年段，对不同学科的期刊论文 P-W 引文进行验证分析，例如：分析在引文库中计算机论文 P-W 引文的无效率；^⑥在两种计算机期刊中 P-W 引文的半衰期；^⑦在高影响力肿瘤期刊中的 P-W 引文可获取性；^⑧在 MED-

① Evans J. A. Electronic publication and the narrowing of science and scholarship. *Science*, 2008, 321 (5887): 395 - 399.

② Larivière V., Gingras Y., Archambault É. The decline in the concentration of citations, 1900 - 2007. *Journal of the American society for information science and technology*, 2009, 60 (4): 858 - 862.

③ Larivière V., Archambault É., Gingras Y. Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900 - 2004). *Journal of the American Society for Information Science and Technology*, 2008, 59 (11): 288 - 296.

④ Egghe L. A model showing the increase in time of the average and median reference age and the decrease in time of the price index. *Scientometrics*, 2009, 82 (2): 243 - 248.

⑤ Harter S. P. Electronic journals and scholarly communication: a citation and reference study. *Information Research*, 1996, 2 (1). [2011/1/31]. <http://InformationR.net/ir/2-1/paper9a.html>.

⑥ Lawrence S. et al. Persistence of web references in scientific research. *IEEE Computer*, 2001, 34 (2): 26 - 31.

⑦ Spinellis D. The decay and failures of web references. *Communications of the ACM*, 2003, 46 (1): 71 - 77.

⑧ Hester E. J. et al. Internet citations in oncology journals: a vanishing resource? *Journal of the National Cancer Institute*, 2004, 96 (12): 969 - 971.

LINE 数据库中 P - W 引文的稳定性与持久性;^① D - Lib 杂志论文中 P - W 引文的不可获取性;^② 在 6 种传播类期刊中 1 600 条 P - W 引文的稳定性;^③ 6 种新西兰期刊在 2002—2005 年内 P - W 引文的不可获取性;生物医学期刊的 P - W 引文的不可获得性。^④ 借鉴传统引文的半衰期概念,一些研究也研究了 P - W 网络引文的不可追溯半衰期。2005 年,斯皮内利斯 (Spinellis) 通过对 4 224 条 P - W 引文的查证,得出从出版之日起,引文的 URL 半衰期大约为 4 年。2005 年,比格贾 (Bugeja) 和季米特洛娃 (Dimitrova) 对全球顶尖的 3 种传播学期刊 4 年间的引文分析表明, P - W 引文半衰期平均为 3.02 年。^⑤ 在 2007 年,季米特洛娃和比格贾对 2000—2003 年的 1 126 条 P - W 引文进行分析,得出 P - W 引文的可追溯半衰期为 3.15 年。^⑥ 吴 (Goh)^⑦ 在 2007 年以 3 种情报学领域顶级期刊 7 年 (1997—2003) 的 P - W 引文为研究对象,对 P - W 引文的链接衰减现象进行研究,发现该学科的链接衰减半衰期为 5 年,且各类链接中 edu 域名资源的链接衰退比例最高,达到了 36%。塞利托 (Sellitto)^⑧ 查证了 1995—2003 年的 1 041 条 P - W 引文,发现平均可追溯半衰期为 4.8 年。

W - W 网络引证研究指原始文献和引证文献都属于网络文献的情况,它正逐渐从一般链接分析的研究中独立出来。目前已有的研究主要分析一般链接关系的变化规律。凯勒 (Koehler) 对网络链接的可追溯性进行较长期的系统研究。他把链接可追溯性分为一致性和可持久性两种情况,并将网页的访问状态分为三类:一直可访问的网页、间断可访问的网页和不可再访问的网页。同时他还对经典半衰期计算公式进行改进,发现网络信息的半衰期为 1.6 年。^⑨ 在后续研究中,他又发现网络信息半衰期增加到 2 年,并且随着时间的

① Wren J. D. 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics*, 2004, 20 (5): 668 - 672.

② McCown F. et al. the availability and persistence of web references in D - Lib Magazine. [2011/1/31]. <http://arxiv.org/abs/cs.OH/0511077>.

③ Dimitrova, Daniela V. Consider the source: predictors of online citation permanence in communication journals. *portal: Libraries and the Academy*, 2006 (3): 269 - 283.

④ Aronsky D., Madani S., Carnevale R. J., Duda S., Feyder M. T. The prevalence and inaccessibility of Internet references in the biomedical literature at the time of publication. *Journal of the American Medical Informatics Association*, 2007, 14 (2): 232 - 234.

⑤ Bugeja M., Dimitrova D. V. The half - life phenomenon: eroding citations in journals. *The Serials Librarian*, 2005, 49 (3): 115 - 123.

⑥ Dimitrova D. V., Bugeja M. The half - life of internet references cited in communication journals. *New Media Society*, 2007 (9): 811 - 826.

⑦ Goh D. H. L., Ng P. K. Link decay in leading information science journals. *Journal of the American Society for Information Science & Technology*, 2007, 58 (1): 15 - 24.

⑧ Sellitto C. A study of missing web - cites in scholarly articles: towards an evaluation framework. *Journal of Information Science*, 2004, 30 (6): 484 - 495.

⑨ Koehler W. An analysis of webpage and website constancy and permanence. *Journal of the American Society for Information Science*, 1999, 50 (2): 162 - 80.

变化,网络信息变得更加稳定,而不同学科、不同网域和不同时间段,半衰期的长短不同。^{①②} 马克韦尔 (Markwell) 和布鲁克斯 (Brooks) 历时 14 个月,每隔一个月查证选定的教育科学网络资源的 URL 可访问情况,结果发现 16.5% 的 URL 失效或者内容发生了变化,其中失效最快的是 .edu、.com 和 .org 域名。^③ 在后续研究中,他们又发现经过 14 个月,整个 515 个链接中的 20% 已不可打开,其中 .com 网站的不可追溯率最高,其可追溯半衰期为 2 年。^④

二、国内研究现状分析

利用中国知网 [中国期刊全文数据库 (China Journal Fulltext Database, 简称 CJFD)、中国优秀硕士学位论文全文数据库、中国博士学位论文全文数据库等]、《中文科技期刊全文数据库》和《万方数据资源系统》3 个中文数据库,对网络引证视角的知识交流研究相关论文进行检索,然后再通过《超星数字图书馆》、《书生数字图书馆》、《读秀图书搜索》查找相关的图书文献。另外,通过国家自科基金委、全国哲学社会科学规划办公室和教育部人文社科网等网站了解相关课题立项情况。^⑤

(一) 网络引证视角的知识交流分布规律研究

在 P-P 网络引证视角的知识交流分布规律方面,类似加菲尔德引文集中定律,国内对中文科学引文的分析也得出了同样的结果,发现我国的引文分布情况不仅符合该定律,而且,被引期刊集中化的趋势更加明显:《中文科学引文索引》(CSCD) 的参考文献的一半来自被引期刊总数的不到 3%;而被引期刊总数的 25% 就承担了 90% 的引文量;被引文总数的 75% 只涉及 72 种被引期刊。^⑥ 范全青通过对 CSSCI 2000—2003 年我国图书情报学引文数据的分析,揭示了引文的频数分布、时间分布、类型分布、作者分布、期刊分布和语种分布特征,说明了引文量的集中和离散趋势及程度;CSSCI 图书情报学引文的 50% 集中在 0.32% 的 15 种期刊上,被引期刊的 23% 承担了 90% 的引文量,被引文种数的 70% 只涉及 62 种期刊。^⑦ 利用 5 种科学学国际期刊,通过对科学学引文网络结构的研究,

① Koehler W. Webpage change and persistence - a four - year longitudinal study. *Journal of the American Society for Information Science and Technology*, 2002, 53 (2): 62 - 71.

② Koehler W. A longitudinal study of webpages continued: a report after six years. *Information Research*, 2004, 9 (2): 174.

③ Markwell J., Brooks D. W. Broken links: the ephemeral nature of educational WWW hyperlinks. *Journal of Science Education and Technology*, 2002, 11 (2): 105 - 108.

④ Markwell J., Brooks D. W. Link rot limits the usefulness of web - based educational material in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*, 2003 (31): 69 - 72.

⑤ 杨思洛,仇壮丽.网络引文研究现状及展望.《图书情报工作》,2009,53(10):41-44.

⑥ 邱均平.《信息计量学》.武汉:武汉大学出版社,2007.

⑦ 范全青.我国图书馆学情报学论文的引文分布规律研究.《情报杂志》,2005(6):92-94.