

网络信息计量学 理论与实证研究

张 洋 著



科学出版社

www.sciencep.com

网络信息计量学理论与实证研究

张 洋 著

教育部人文社会科学研究项目成果(批准号:07JC870002)

科学出版社

北 京

内 容 简 介

本书作为教育部人文社会科学研究项目“网络信息分布与变化规律的基础理论研究”(批准号:07JC870002)的研究成果之一,在总结国内外网络信息计量学研究的基础上,针对目前存在的主要问题,采用理论研究和实证研究相结合的方法,从理论、方法、应用三个方面对网络信息计量学进行了系统的研究。书中构建了网络信息计量学的学科体系,研究了网络信息计量学的研究方法和应用问题,探讨了网络环境下的布拉德福定律的适用性与新特点和网络指标在大学评价中的有效性问题。

本书既适合于高等院校的信息管理、科学评价、信息计量、新闻与传播、网站设计、电子商务、管理科学与工程等有关专业的师生教学和科研使用,也可供广大信息工作者、网络工作者、科技工作者和有关管理人员学习参考。

图书在版编目(CIP)数据

网络信息计量学理论与实证研究/张洋著. —北京:科学出版社,2009

ISBN 978-7-03-025516-7

I. 网… II. 张… III. 计算机网络-文献计量学-研究 IV. G257

中国版本图书馆 CIP 数据核字(2009)第 161366 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

丽源印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2009年9月第 一 版 开本:B5 (720×1000)

2009年9月第一次印刷 印张:11 3/4

印数:1—1 500 字数:223 000

定价:38.00 元

(如有印装质量问题,我社负责调换〈长虹〉)

序 言

进入 21 世纪以来,“网络化”日益成为经济、社会 and 科技发展的重要因素和显著特征,并呈现出不断加强的趋势。加强网络信息资源的开发利用和网络管理已成为当务之急,而实施定量化管理则是其主要的途径之一。“网络信息计量学(Webometrics)”正是在这种国际趋势和科学前沿背景下兴起的新兴学科,它是由网络技术、网络管理、信息资源管理与信息计量学等相互结合、交叉渗透而形成的一门交叉性边缘学科,也是信息计量学的一个新的发展方向 and 重要的研究领域。

20 世纪 60 年代正式诞生的“文献计量学和科学计量学”,为图书情报科学提供了有力的定量研究工具,推动整个学科进入了一个高速发展的“黄金时期”。在这一时期逐步发展、确立起来的一系列文献信息分布和变化规律成为整个学科基础理论的重要组成部分,为图书情报科学和信息管理科学学科地位的确立做出了突出的贡献。随着人类社会逐步向信息社会转变,科学领域中的各个学科都无法回避“网络化”带来的冲击,原有的科学体系面临着新一轮的分化组合。在这种形式下,各个学科纷纷抢占网络研究的“制高点”,希望能够在“网络时代”占有一席之地。而在这一过程中,计量科学再次成为推动相关学科发展的“主力”。网络信息计量学以传统计量科学的理论和方法为基础,以网络信息的分布和变化规律为研究内容,其诞生是顺应网络社会发展趋势的必然结果。作为信息管理科学和图书情报科学的新成员,它正在成为一座桥梁,将整个学科引入了一个更为广阔的天地——“因特网”,为学科本身的建设和发展提供了宝贵的契机,使传统的科学学科焕发出新的生命力。从这个意义上讲,张洋博士以“网络信息计量学理论与实证研究”为主题展开全面系统的研究,并撰写专著,这无疑能够促进整个信息管理科学和图书情报科学在网络时代的发展,具有重要的科学理论价值。

作为一门“应用性”特征十分明显的计量学科,“网络信息计量学”有着强大的应用潜力和广阔的应用前景。这也是本书作者重点关注之处。正如书名所反映的,本书的研究思路是“理论与实证分析相结合”。首先,在理论层面上,对当前网络信息计量学的主要研究方法和主要应用领域进行系统的总结,探讨了方法论问题 and 应用问题。其次,通过两个具体的实证研究,检验理论研究的成果,并探讨了实践中的若干关键问题,得出一系列具有实用价值的研究结论,可直接用于指导管理实践。以该书第 6 章“网络环境下布拉德福定律的实证研究”为例,作者通过这个实证研究,不仅在理论上证明了“布拉德福定律”在网络环境下同样具有适用性,而且在研究过程中实际应用了文献计量学的“等级排列技术”等具体的研究

方法,为其他研究者提供了可资借鉴的参考。此外,作者还得出“‘期刊 Web 下载总频次’等网络指标同样能够反映期刊对于该专业的贡献大小和在本学科中的地位”这一结论,可直接用于改进当前的“期刊评价”工作。从这个意义上讲,作者的研究工作还具有重要的现实指导意义。

“网络信息计量学”诞生于 1997 年,迄今已走过 11 个年头。在国内外研究者的共同努力下,网络信息计量学研究已积累了相当多的成果。可以说,该学科目前已经渡过了起步阶段,逐步进入高速发展的时期。但从总体上看,网络信息计量学仍然是一个崭新的研究领域,在理论、方法、应用等方面的研究都尚未成熟,还有许多问题有待研究和解决,尤其是缺乏针对学科本身的系统化的研究,这严重阻碍了该学科的继续发展。在这种情况下,作者将本书的研究目标定位为“解决网络信息计量学作为一门学科存在而必须解决的基本问题,构建其内容体系和方法论体系”,无疑是十分必要和及时的。围绕此目标,作者进行了全面细致的文献调查工作,基本涵盖了目前国内外网络信息计量研究领域的主要研究成果。然后,在总结国内外科学研究工作实践的基础上,对网络信息计量学进行深入的探讨。从理论研究到实证分析,从基本理论、学科体系、研究方法到应用研究,涉及网络信息计量学的多个方面,对该学科的基本问题进行了系统全面的研究,填补了我国网络信息计量学理论研究的空白,为该学科的进一步发展奠定了基础。

张洋博士在攻读博士学位之初,就确定了以“网络信息计量学”作为主要的研究领域。参加工作后,又一直专注于计量科学的教学和研究工作,先后在《中国图书馆学报》、《情报学报》等权威刊物上发表了一系列高水平的研究论文,并主持本学科领域的多项国家级、省部级科研项目。持续、连贯、系统的学习和钻研使得他能够对“网络信息计量学”这一新兴学科进行相当精深的研究,也使本书达到了较高的学术水平。全书的研究思路清晰,结构合理,层次清楚,观点新颖,内容丰富,资料翔实,既有理论深度,又有具体的实证分析,是一部科学性、系统性、创新性都十分明显的研究力作,也是网络信息计量学领域的一项开创性研究成果。当然,任何科学研究都不可能完美无缺,对于新学科而言就更是如此。本专著所探讨的问题大多是新出现的和开创性的,自然有许多需要改进和完善的地方,有待于作者继续努力探索。我作为张洋博士的导师,对他近几年来在科学研究方面取得的长足进步和突出成果,以及《网络信息计量学理论与实证研究》一书的正式出版感到由衷的高兴!在此,我也衷心地希望张洋博士能够取得更多更好的成绩,为整个学科的发展做出更大的贡献。

邱均平

2009 年 2 月 28 日于武汉大学

前 言

自 20 世纪 90 年代以来,迅速发展的因特网(Internet)对人类社会产生了越来越大的影响。人们在享受“网络化生活”所带来的种种便利的同时,对日益庞杂的网络信息管理也提出了日益紧迫的要求。在这种需求的推动下,“网络信息计量学”应运而生。这一新兴学科的根本目的是通过对网上信息的计量研究,为网上信息的有序化组织和合理分布、为网络信息资源的优化配置和有效利用、为网络管理的规范化和科学化提供必要的定量依据,从而改善网络的组织管理和信息管理,提高其管理水平,促进其经济效益和社会效益的充分发挥,推动社会信息化、网络化的健康发展。因此,针对网络信息计量学展开研究,既是网络时代社会、经济和科技发展的需要,也是情报科学、信息管理科学和信息计量学学科本身发展的需要,具有重要的科学理论价值、实际指导意义和广阔的应用前景。

网络信息计量学诞生于 20 世纪 90 年代后期。虽然时间不长,但该学科发展十分迅速。无论是国外还是国内,每年均有大量的成果发表,且呈不断增长的趋势。时至今日,网络信息计量学已成为信息科学领域令人瞩目的研究热点和重点之一。而经过十余年的发展,相关学术论文的数量已具备相当规模,在理论上和实践上均积累了一批有重要价值的研究成果。但是,从当前来看,网络信息计量学还缺乏明确的学科体系、完善的基础理论、可靠的研究方法、深入的应用研究,这些问题严重阻碍了该学科的进一步发展和建设。

本书的研究内容正是在这种国际趋势和科学前沿背景之下提出来的。本书在总结国内外网络信息计量学研究的基础上,针对目前存在的主要问题,采用理论研究和实证研究相结合的方法,从理论、方法、应用三个方面对网络信息计量学展开系统的研究。第 1 章为绪论。第 2~7 章为本书主要内容:第 2 章在对若干基本问题进行研究的基础上构建了网络信息计量学的学科体系;第 3,4 章对网络信息计量学的研究方法进行研究;第 5 章对网络信息计量学的应用问题进行研究;第 6,7 章是实证研究,分别探讨网络环境下的布拉德福定律的适用性与新特点和网络指标在大学评价中的有效性问题。本书各章的主要内容如下:

第 2 章:网络信息计量学的学科体系构建。本章首先追本溯源,从文献计量学、科学计量学和信息计量学的发展历程和相互关系出发,循着网络信息计量学产生和发展的轨迹,分析其产生背景,并探讨了该学科名称的由来和相应的中英文术语。其次,探讨了该学科的定义,深入剖析其学科性质,指出该学科既是信息计量学的分支学科,又是一门交叉性的边缘学科,还是一门应用性特征明显的学科。然

后,将该学科的研究对象分为网络特征信息、网络拓扑信息和网络文献信息三种类型进行深入探讨。最后,从发展历程、研究层次、研究内容等三个维度,构建了统一的网络信息计量学学科体系。

第3章:网络信息计量学的方法论研究。本章主要探讨网络信息计量学的研究方法问题。具体分为三个部分:首先,从科学研究方法论的基本原理出发,按照哲学方法、一般科学方法、特征方法等三个层次探讨网络信息计量学的研究方法,构建了网络信息计量学的方法论结构体系;其次,按照数据收集和数据分析两个主要研究环节,对目前应用到网络信息计量研究中的具体研究方法进行总体上的分析;最后,分节重点探讨几个重要研究方法的概念定义、理论依据、来源出处、发展历程以及与网络信息计量学的关系等关键问题,包括数据收集的重要工具“搜索引擎”,以及“网络内容分析法”和“网络数据挖掘方法”两个数据分析方法。

第4章:网络链接分析法。本章对网络信息计量学最重要的研究方法——“网络链接分析法”进行了系统全面的研究。首先,界定本章所讨论的网络链接分析研究的范围,将所探讨的网络链接分析法定义为:“一种以网络链接为研究对象,根据引文分析法的基本原理和方法,对网络链接的分布规律和网络信息单元之间的链接规律进行分析研究的网络信息计量学的特征研究方法”。其次,从网络链接分析法与引文分析法的关系出发,探讨其产生过程、理论依据、内在机理、局限性和优点等问题。再次,依据术语学的基本原理,在分析总结现有习语及其来源出处的基础上,探讨网络链接分析的相关术语集合,并构建了标准术语体系。最后,将网络链接分析的主要研究内容分为网络链接分布规律研究、同链聚类分析研究、网络影响因子研究、网络链接分析工具研究等主要组成部分,并分别对其研究进展情况进行评述。

第5章:网络信息计量学的应用研究。本章主要对网络信息计量学的应用问题进行系统研究。首先,从总体上对网络信息计量学的应用研究情况进行分析,得出结论认为目前主要存在实际应用不够、研究广度不够、研究深度不够等问题,并指出“网络信息资源管理”和“科学评价”是当前该学科最为集中和最为重要的应用领域。其次,探讨了网络信息计量学与网络信息资源管理的关系问题,并以数字图书馆为例,探讨网络信息计量学在网络信息资源管理领域的具体应用。最后,在对科学评价进行概述的基础上,从“三计学”与科学评价的关系出发,论述了网络信息计量学在科学评价中应用的必然性,并对目前有关的研究情况进行了分析。

第6章:网络环境下布拉德福定律的实证研究。本章首先对目前有关网络信息变化规律与分布规律的研究情况进行分析,指出揭示网络信息的数量特征和内在规律,寻求统一的理论基础和更准确的数学描述,乃是网络信息计量学理论研究的主要和重要内容之一。然后,通过实证研究,对网络文献的被下载次数在期刊中的分布规律进行计量分析,深入探讨文献计量学经典定律之一——布拉德福定律

在网络环境下的适用性问题。研究表明,网络信息同样具有布拉德福分布特征,但同时也具有新的特点。最后,论述了该研究在理论成果、研究方法、实际应用等方面的意义和有待进一步研究的问题,为研究网络信息的集中与离散分布规律打下基础,填补了网络信息计量学基础理论研究中的一块空白。

第7章:大学评价中网络指标有效性的实证研究。本章首先针对目前大学评价所面临的主要问题,从完善指标体系、扩展数据来源、改进分类方法等三个方面,论述了将网络信息计量学应用到大学评价中所具有的重要意义。其次,对目前有关网络指标在大学评价中的应用研究特点进行分析,并以“网络影响因子”为例,总结了目前主要的研究成果。然后,通过实证研究,对大学网站的“总入链数”、“外链数”、“网络影响因子”和大学学报的“Web 下载总频次”等网络指标在大学评价中的有效性问题进行深入探讨,并在数据收集与预处理、大学网站指标的有效性、大学学报网络指标的有效性、网络指标对“大学声誉”的影响以及不同排行榜的比较等几个方面,得出了一系列有意义的结论。既是对本书理论研究成果的实际应用,又为网络信息计量学研究实践提供了可资借鉴的参考。

在本书创作过程中,得到了许多单位和个人帮助。第一,我要深深地感谢武汉大学信息管理学院的邱均平教授,作为我的博士导师,是他把我带入了科学的殿堂,并教会了我如何学习和思考,他严谨的治学态度、渊博的学识和海人不倦的育人精神更使我深受教育和启迪。第二,我要感谢中山大学资讯管理系的各位领导和同仁,他们的无私帮助给予了我大力的支持。第三,我要感谢所有关怀和指导过我的信息管理界前辈学者,他们的教学和指导是我所有科研成果的基础。第四,我要感谢科学出版社的工作人员,他们的辛勤付出是此书能够顺利出版的保障。此外,我还要感谢所有被引用文献的作者、国内外的同行们,以及所有在本书创作过程中给予帮助和支持的友人。最后,我要特别感谢我的家人,他们的爱使我走到了今天,也是支持我继续前进的动力,谨以此书献给我最亲爱的家人。

本书是教育部人文社会科学研究项目“网络信息分布与变化规律的基础理论研究”(批准号:07JC870002)的研究成果之一。尽管本书即将付梓,但它不是终点,甚至连结局的开端也不是,它只是序幕的尾声。作为一门新兴学科,网络信息计量学所涉及的很多问题均是新出现的,而无论是因特网还是现代信息社会都在不断发生着丰富多彩的变化,因此,还有很多问题有待进一步的探索。恳请各位同行批评指正,共同参与,推动我国信息管理事业不断向前发展。

张 洋

2009年1月28日于中山大学

目 录

序言

前言

1 绪论	1
1.1 网络信息计量学的价值	1
1.2 网络信息计量学研究概述	3
1.3 本书的研究特色	9
2 网络信息计量学的学科体系构建	11
2.1 网络信息计量学的产生与发展	11
2.2 网络信息计量学的定义、性质与研究对象	20
2.3 网络信息计量学的体系结构	28
3 网络信息计量学的方法论研究	39
3.1 网络信息计量学方法论体系的构建	39
3.2 网络信息计量研究数据的收集与分析	43
3.3 数据收集工具——搜索引擎	46
3.4 网络内容分析法	59
3.5 网络数据挖掘方法	66
4 网络链接分析法	78
4.1 网络链接分析法的基本概念	78
4.2 引文分析法与网络链接分析法	79
4.3 网络链接分析的术语体系	85
4.4 网络链接分析的研究内容及进展	102
5 网络信息计量学的应用研究	111
5.1 概述	111
5.2 网络信息计量学与网络信息资源管理	113
5.3 网络信息计量学与科学评价	119
6 网络环境下布拉德福定律的实证研究	125
6.1 研究背景与目的	125
6.2 布拉德福定律概述	128
6.3 数据来源	132
6.4 研究方法	133

6.5	研究结论与讨论	142
7	大学评价中网络指标有效性的实证研究	144
7.1	研究背景与目的	144
7.2	文献回顾	147
7.3	数据来源	148
7.4	“大学网站”指标的有效性分析	150
7.5	“大学学报”指标的有效性分析	159
7.6	研究结论与讨论	162
参考文献		165

1 绪 论

自然界存在着一个无限发展的螺旋，其中越来越多的严密的理论通过更加精细的计量而日益得到越来越充分的检验。^①

——（比）Leo Egghe and Ronald Rousseau

1.1 网络信息计量学的价值

自 20 世纪 90 年代以来，迅速发展的因特网对人类社会产生了越来越大的影响。时至今日，网络化已成为经济、社会 and 科技发展的重要因素和显著特征，并呈现出不断加强的趋势。人们在享受“网络化生活”所带来的种种便利的同时，对日益庞杂的网络信息管理也提出了日益紧迫的要求。在这种需求的推动下，“网络信息计量学”（Webometrics）应运而生。这一新兴学科是由网络技术、网络管理、信息资源管理与信息计量学等相互结合、交叉渗透而形成的一门交叉性边缘学科，也是信息计量学的一个新的发展方向和重要的研究领域，其根本目的主要是通过网上信息的计量研究，为网上信息的有序化组织和合理分布、为网络信息资源的优化配置和有效利用、为网络管理的规范化和科学化提供必要的定量依据，从而改善网络的组织管理和信息管理，提高其管理水平，促进其经济效益和社会效益的充分发挥，推动社会信息化、网络化的健康发展。

本书的研究工作正是在这种国际趋势和科学前沿背景之下提出来的，其价值主要表现在以下方面：

（1）有利于加强和改善网络管理，拓展和完善国家的网络安全管理体系，适应网络时代社会发展的需要。随着网络化的快速发展，网络的影响已经渗透到人类社会活动的各个方面，“如何对网络进行管理”和“如何利用网络进行管理”已成为政府管理部门非常关注的重要课题。首先，网络信息计量学的研究成果能够为网络管理的规范化和科学化提供理论指导和科学依据，为信息产业和信息服务业的有关管理部门的决策提供定量依据。其次，随着网络信息资源的重要性日益凸现，对网络信息资源的统计成为国家统计的重要组成部分。通过网络信息计量研究，有利于建立网络统计体系，全面完成国家的统计任务，有利于完善国家

^① Egghe L, Rousseau R. 情报计量学引论. 田苍林, 葛赵青译. 北京: 科学技术文献出版社, 1992. 1.

的网络安全管理体系。最后,通过网络信息的计量分析,有助于了解社情民意,分析和发现某些热点问题,追踪某些问题的发展动态,为社会管理决策服务。

(2) 有利于网络信息资源的优化配置和开发利用,适应网络时代经济发展的需要。信息资源网络化和网上信息数量激增,是网络社会的显著特征。对网络信息的规律性进行研究,可以充分挖掘其经济价值,有力推动社会经济的发展。首先,网络信息作为网络经济的重要战略性资源,对其进行深入研究有利于网络信息的开发利用,有利于将“资源”转化为现实生产力,从而促进网络经济的发展。其次,通过对网络信息资源的计量评价和分布规律的研究,不仅能使人们全面了解和准确把握有关信息的分布和交流状况,而且还能为网络信息资源的优化配置和开发利用提供依据。最后,通过针对网络用户的信息需求和上网习惯的定量研究,可以为优化网站的结构和信息服务提供参考,推动电子商务和电子政务的发展。

(3) 有利于学术信息的开发利用和科学评价,促进科学研究和信息交流,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。目前,网络已经成为发表科研成果、交流科技信息、展示科研实力的重要途径,当前的科研评价也必然要顺应这一趋势。采用计量方法研究网上学术信息交流和学科知识结构的特点,分析并建立新的科学评价的指标体系和方法,开展有关的网上评价活动,探讨网络环境下科学交流和信息传播的规律,有利于掌握学科发展状况,预测学科发展趋势,为制定有关科技政策和管理措施提供参考依据。另一方面,通过科学评价,测定某一学科领域的核心网站、网络影响因子等,可以指导用户更快、更准地找到所需要的信息,满足科学研究或其他信息需求,有利于促进科学研究和信息交流。

(4) 情报科学、信息管理科学和信息计量学学科本身发展的需要。网络信息计量学的产生是顺应现代社会经济和科学技术发展趋势的必然结果,它作为信息计量学的分支学科,已成为信息计量学的一个新的发展方向和研究领域。网络信息计量学作为一座桥梁,将情报科学、信息管理科学引入了一个更为广阔的天地——“因特网”,为学科本身的建设和发展提供了宝贵的契机,使传统的科学学科焕发出新的生命力,从而能够在新的时代为人类社会和科学社会做出更大的贡献。

综上所述,“网络信息计量学”的研究具有重要的科学理论价值、实际指导意义和广阔的应用前景。作为教育部人文社会科学研究项目“网络信息分布与变化规律的基础理论研究”(批准号:07JC870002)的重要组成部分,本书的研究工作正是顺应这一学科发展潮流所做的一个探索,希望能对我国网络信息计量学的发展有所裨益。

1.2 网络信息计量学研究概述

网络信息计量学的研究始于20世纪90年代后期,以1997年丹麦学者T. C. Almind和Peter Ingwersen提出“Webometrics”这一概念为其诞生的标志(Almind and Ingwersen, 1997)。从那以后,这一名词使用频率越来越高,得到国内外学术界的响应和广泛使用,并引起了政府、企业、传媒、教育等社会各界的广泛关注。下面,从总体上对当前国内外研究状况进行描述和分析,总结主要的研究成果,并指出其中存在的主要问题。

1.2.1 国外研究进展

国外网络信息计量学的研究与国内相比起步相对较早,研究也更为深入全面,在该领域处于领先的地位。“Webometrics”自诞生以来,就一直受到了国外学术界的广泛关注,不仅出现了专门的学术网站和电子期刊^①,美国信息科学技术学会学报(journal of the American society for information science and technology)还在近年专门出版了网络信息计量学的专刊^②,这表明网络信息计量学已成为信息科学领域的研究热点。近年来,在这一领域涌现出了一批重要的专家学者,例如丹麦的Peter Ingwersen和Lennart Björneborn、英国的Mike Thelwall、比利时的Ronald Rousseau和Leo Egghe、加拿大的Liwen Vaughan、新西兰的Alastair G. Smith、荷兰的Leot Leydesdorff、以色列的Judit Bar-Ilan等人,他们从不同的角度、在不同方面进行了大量的研究,取得了丰硕成果,为推动网络信息计量学的发展做出了巨大贡献。

1. 理论研究

在理论研究方面,丹麦皇家图书情报学院^③(Royal School of Library and Information Science)的研究团队所作的工作最有代表性。

1997年,该学院的T. C. Almind和Peter Ingwersen首先提出了“Webometrics”的概念,被公认为网络信息计量学诞生的标志(Almind and Ingwersen, 1997)。1998年,Irene Wormell从信息计量学在网络上的应用的角度,论述了网络时代图书情报科学所面临的挑战,指出网络信息计量研究蕴含着巨大的

^① 例如, Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, 网址: <http://www.cindoc.csic.es/cybermetrics/>

^② Special topic Issue: Webometrics. In: Journal of the American society for information science and technology, 2004, 55 (14) .

^③ 网址: <http://www.db.dk/>

潜力 (Wormell, 1998)。同年, Peter Ingwersen 受文献计量学中的“期刊影响因子”(journal impact factor)的概念启发提出了网络影响因子(Web impact factor, WIF)的概念,用来分析一定时期内受关注的网站或网页平均被引情况,以此来评价网站在网上的影响力。这篇论文在网络信息计量学的发展历史上具有十分重要的意义。首先,他开创了一个全新的研究领域,此后众多学者从概念定义、测度方法、实证分析、应用实践等众多方面对网络影响因子进行了广泛深入的研究,使其成为网络信息计量学的前沿热点和重要内容。更为重要的是,他开创了一种新的研究思路,证明了文献计量学的经典概念和研究方法可以被成功地应用到网络信息研究中(Ingwersen, 1998)。

此后,研究者们纷纷效仿,引入了更多的概念和方法,既丰富了网络信息计量学的理论内容和研究手段,又明确了网络信息计量学作为情报学分支学科的性质。2001年, Lennart Björneborn 和 Peter Ingwersen 对网络信息计量学的研究热点和发展趋势做了全面的论述,介绍了知识发现(knowledge discovery)、问题追踪(issue tracking)、图论方法(graph theoretic approaches)等新的研究方向,拓展了网络信息计量学的研究领域(Björneborn and Ingwersen, 2001)。2004年, Lennart Björneborn 和 Peter Ingwersen 在信息计量学和文献计量学的框架范围内,构建了一个较为全面的网络信息计量学的基础概念体系,包括一系列专用术语和符号,标志着网络信息计量学的发展经过多年的积累,已进入了成熟的发展阶段,一个全新的学科终于要“瓜熟蒂落”了(Björneborn and Ingwersen, 2004)。

2. 方法研究

在方法研究方面,主要包括数据收集方法和数据分析方法的研究。

对于任何计量科学的研究来说,研究数据的收集整理工作无疑是其开展各项研究工作的基本前提,网络信息计量学自然也不例外。数据的查全率和查准率直接影响研究结果的可靠性和合理性,因此,对数据收集方法本身所作的研究也成为网络信息计量学研究的重要内容。从最早的 T. C. Almind 和 Peter Ingwersen (1997)所作的研究开始,一直到今天,研究者们主要依靠商业搜索引擎来收集数据。所用到的搜索引擎也是种类繁多,特点各异,包括 AltaVista、All-TheWeb、Northernlight、Google、Excite、Lycos、HotBot、Infoseek 等在内的众多搜索引擎都曾被应用于网络信息计量研究的数据搜集工作中。相应地,出现了许多以搜索引擎为研究对象的,以评价、应用和改进为目标的研究工作,例如 Clarke 和 Willett (1997)、Judit Bar-Ilan (1998)、Ronald Rousseau (1999)、Herbert Snyder 和 Howard Rosenbaum (1999)、Paul Wouters 和 Diana Gerbec (2003)等人所做的研究。虽然搜索引擎已成为事实上的主要数据搜集渠道,成

为网络信息计量研究不可或缺的工具,但其查全率和查准率一直深受质疑,其搜索结果不稳定、技术机制不透明已成为公认的缺陷。为了解决这一难题,许多学者进行了许多有益的尝试。例如, Lawrence 和 Giles (1998) 指出,为了克服单一引擎的局限性,从而获得全面的检索结果,一方面可将几个主要引擎结合起来使用,也可通过利用一些具有自动抓取功能的研究型搜索引擎来获得信息。Alastair G. Smith 和 Mike Thelwall (2001) 则采用了自己设计的爬行器和搜索引擎进行了比较研究。Mike Thelwall (2002d) 还提出了“替代文献模型 (alternative document models, ADMs)”,以目录、域名以及整个网站为单位来合计链接数量,而不以网页为单位,以避免单一链接动机所带来的数据异常现象。但从总体上看,这些方法都只能起到有限的作用,网络数据的收集仍然缺乏普遍有效的方法,可以说,网络数据的获取已成为制约网络信息计量学发展的瓶颈。

在数据分析方法上,由于网络信息计量学是网络技术、统计学、文献计量学理论三合一的产物,从某种意义上来说,网络信息计量学就是文献计量学、科学计量学在网络上应用的一门学科,因而在文献计量学、科学计量学中得到广泛应用的文献信息统计分析法、数学模型分析法、引文分析法、书目分析法、系统分析法、关键词统计分析法、关联数据分析法(包括聚类分析、共词分析、同域分析等)、计算机辅助文献信息计量分析法等定量方法在网络信息计量研究中都得到了广泛的应用。当然,由于网络环境的特殊性,研究者们应用过程中都对上述方法进行了不断的调整 and 变化,使其最终成为网络信息计量学的专用方法。在已开展的研究中,应用最广泛的是“网络链接分析法”(hyperlink analysis)。网络链接分析源自于“网络链接”与“科学文献的引文”之间天然的相似性,这使国外的一些文献计量学家找到了文献计量和网络的契合点,创造性地将 Web 网页链接(hyperlink)与传统文献中的“参考”(reference)和“引用”(citation)联系起来。1996年,美国爱荷华州立大学图书馆的理论馆员 Gerry McKiernan 根据文献计量学中引文的含义,首次提出了“Sitiation”的概念,来描述网站(site)之间相互链接的行为,他指出:Cited Sites=Sitiation,即所谓 Sitiation 就是被引用的站点(McKiernan, 1996)。此后 Isidro F. Aguillo 在 1996 年比勒菲尔德召开的 4S/ EASST 会议上引用了这一概念。1997 年 Ronald Rousseau 发表的论文首次将“Sitiation”一词正式应用到学术论文题名当中,标志着“网络链接分析法”的确立(Rousseau, 1997)。此后,“网络影响因子”、“网站双引聚类分析”(co-sitation)等一系列“移植”的新概念和新方法不断被提出来。从目前来看,在有关网络信息计量学的学术论文中,关于网络链接分析的论文占了绝大多数,这充分说明,Web 网页链接分析是目前网络信息计量学的研究热点和重点,对 Web 网页链接进行分析是网络信息计量学一项重要的研究内容,链接分析法已成为网络信息计量学研究的重要方法。有的学者更进一步认为,网络链

接研究是促进网络信息计量学产生和发展的重要动力。此外,一些学者还引入了计算机、人工智能、拓扑学、社会学、图论等其他学科领域的研究方法,例如,新闻传播领域的“内容分析法 (content analysis)”、数据库技术领域的“数据挖掘 (data mining) 方法”、理论物理学领域的“复杂网络理论 (complex network theory)”、计算机科学领域的“Web 信息检索算法 (Web information retrieval algorithms)”、社会学中的“社会网络分析方法 (social network analysis)”、数学中的“图论方法 (graph theoretic approaches)”和“拓扑学方法 (topology approaches)”等都已经应用到网络信息计量研究中。这些方法一方面充实了网络信息计量学的方法论体系,促进了网络信息计量研究,另一方面,也带来了基本概念混乱、学科界限不明等问题。

3. 应用研究

在应用与实证研究方面,是网络信息计量学研究工作最为集中的领域,参与者遍布各个国家和地区,涉及广泛的研究对象,采用了各种各样研究方法和工具,所取得的成果也最为丰富,可以说是“百花齐放”、“硕果累累”。究其原因,首先是因为社会实践需要是学科产生的根本动力,网络信息计量学就是顺应网络社会发展需要而产生的,在应用领域对其展开研究自然就是学科发展的第一步,这也符合辩证唯物主义“实践—方法—理论”的“科学的实践观”。其次,这与网络信息计量学所具有的“应用性学科”的学科性质密不可分。网络信息计量学作为一门交叉性的边缘学科,有着广泛的应用领域。它既可在图书情报领域内应用,又可以应用于科学学、社会学、人才学、历史学、未来学等许多相关学科的研究,还可以用于图书情报工作、网络管理、信息资源管理、科学评价、科技管理与预测、电子商务等具体社会生产实践活动中。

从目前情况来看,网络信息计量学的应用研究大多还停留于理论探讨阶段,与真正应用到社会生产实践中还有相当的差距,但相关的实证研究已经蓬勃地开展起来,尤其在网络信息资源管理和科学评价两个方面,已经取得了长足的进步。在国外众多的“实践者”中,最为突出的是英国伍尔弗汉普顿大学计算机与信息技术学院^① (School of Computing and Information Technology, University of Wolverhampton) 的 Mike Thelwall。他在实证研究领域的研究成果可以用“汗牛充栋”来形容,据笔者的统计,仅 2002 年一年,他就发表了学术论文近 20 篇,且大多是独著并发表在各种权威期刊上,可以说,他已成为国际网络信息计量学应用与实证研究的权威科学家之一。Mike Thewall 和其他众多的研究者们所进行的应用研究工作取得了许多重要的研究成果,不仅为网络信息计量学

^① 网址: <http://www.scit.wlv.ac.uk/>.

的实际应用奠定了基础,同时也促进了该学科的理论体系和方法体系的发展。

1.2.2 国内研究进展

国内的网络信息计量学研究始于1999年。据笔者查阅的资料,徐久龄等人(1999)在中国国防科学技术信息学会出版的《情报学进展》(第三卷)上发表的“网络计量学的研究”一文,是最早向国内引入网络信息计量学的学术论文。从那以后,国内的学者在理论、方法、应用等方面都取得了一定的成果,相关论文正在逐步增多,但从总体上看,国内的网络信息计量学研究尚处于起步阶段,主要以介绍国外研究成果和进展,以及理论探讨性的文章为主,开创性研究和探索性研究所见不多。例如,在“中国期刊网”上,“题名/关键词/摘要”中包含“网络(信息)计量学”的论文,综述类的就占到了一半以上。

在国内网络信息计量学研究领域,最为突出的研究者是武汉大学的邱均平教授。他在2000年发表于《情报理论与实践》上的“信息计量学(一)”一文中,首次对网络信息计量学的发展过程、定义、学科性质、根本目的、产生背景、研究对象、内容体系等做了深入的分析(邱均平,2000)。该文不仅是在国内最早对网络信息计量学进行系统、全面的研究,在国际上也是开创性的。该文发表以后,被国内同行广泛引用,尤其是网络信息计量学的定义,被公认为是目前最为准确、全面的定义。此后,邱均平教授率领的研究团队在这一领域开展了大量的研究工作,发表了一系列学术论文,内容涵盖网络信息计量学的理论、方法、应用与实证,其中不乏开创性的研究成果。例如,网络信息计量学及其应用研究(邱均平,陈敬全,2001)、试论网络信息计量学在数字图书馆中的应用(邹菲,2001)、WWW网页的链接分析及其意义(邱均平,黄晓斌,2002a)、网络用户使用纪录的计量分析(邱均平,黄晓斌,2002b)、Web页面链接动机分析及链接测度研究(俞培果,邱均平,2003)、网络信息的增长机制研究(侯经川,赵蓉英,2003)、中文期刊影响因子与网络影响因子和外部链接数的关系研究(邱均平,安璐,2003)、中国大学网站链接分析及网络影响因子探讨(邱均平等,2003)、中美学术型网站链接特征的比较研究(马大川等,2003)、关于网络信息老化研究的若干问题(王宏鑫,邱均平,2004)、网络链接分析研究进展(张洋等,2004)、网络信息计量学的兴起及其哲学思考(张洋,邱均平,2004)、网络信息计量学综述(邱均平,张洋,2004)、网络链接分析与网站评价研究(段宇峰,2004)等等,尤其是在2005年,邱均平等人在《情报学报》、《中国图书馆学报》等权威期刊上发表了一系列的论文,更加奠定了其在国内该研究领域的领先地位(邱均平,段宇峰,2005;邱均平等,2005;段宇峰,2005a,2005b;段宇峰,邱均平,2005a,2005b,2005c;段宇峰等,2005;赵蓉英等,2005)。

此外,国内网络信息计量学领域比较重要的研究工作还有金岩(2001)、李