

## 第十二章：信息计量学与网络计量学 (Informetrics & Webometrics)

赵党志, University of Alberta, Canada

**摘要:** 本章阐述信息计量学 (Informetrics) 的理论和方法在网络环境中开发增值信息服务和提高信息组织和检索效率等方面的成功应用, 以及新兴的网络化学术传播系统给信息计量学的进一步发展和丰富而带来的机会和挑战, 其中包括对网络计量学 (Webometrics) 的主要研究分支和目前研究中所面临的一些问题的介绍和探讨。

**Abstract:** This chapter discusses various applications of Informetrics theories and techniques in the development of value-added information services and in the improvement of the efficiency and effectiveness of information organization and retrieval in the networked environment. This is followed by an exploration of new opportunities that an emerging networked scholarly communication system has brought to the field of Informetrics. This chapter ends with a discussion on major webometric research areas as well as related issues and concerns.

### 12.1 引论 (Introduction)

信息计量学是研究产生于各种社会阶层的各类信息数量特征的一门学科, 其中文献计量学和科学计量学最广为人知。信息计量学具有广泛的应用价值, 其中最重要的应用之一是研究学术传播 (scholarly communication) 的模式, 例如学术成果的定量评价、学术创新的历史传播、学术传播的结构分析 等等。

互联网作为一个强大的传播媒介正在引起信息生产和交换方式的变革, 随着各种信息在互联网上的大量涌现、学术传播系统从以学术期刊为中心的模式向网络化新模式的发展, 信息计量学的理论和方法一方面正在被成功地应用于开发增值信息服务和提高信息组织和检索的效率, 另一方面也面临着新的数据和工具给其进一步发展和丰富而带来的机会和挑战。本章旨在探讨这些问题, 首先概述信息计量学在网络环境中的信息组织与检索方面的应用, 然后探讨网络环境中的学术传播系统给信息计量学的进一步发展所带来的机会, 最后介绍网络计量学 (Webometrics) 的主要研究分支和目前研究中所面临的一些问题。

### 12.2 信息计量学与网络化的信息组织与检索 (Informetrics in networked information organization and retrieval)

由于互联网上的信息一直在飞速而紊乱地增长, 探索适当的组织和检索互联网上信息的方式方法一直是、而且还将继续是非常重要的课题。这方面的努力包括采用引文网络和引文分析等信息计量学的方法和原理进行信息组织与检索。

作者通过引用相关文献以给予这些文献认可已经成为科技创作中的一个传统, 引用代表着作者对其撰写的论文与其引用的文献之间关系的一种判断。沿着引文链寻找相关文献已被证明是一种非常有效的文献检索方式, 这是因为, 这种检索方式是基于引用作者而不是标引人员对文献之间相关性的判断, 而作者比标引人员对学科及文献内

容一般要理解得深刻许多。另外一个原因是，追寻引用链寻找相关文献可以克服由不同语言及不同的分类和标引系统所造成的障碍，这对于交叉科学领域的研究极有价值。此外，引文分析可以用来评价学者、出版物、学术研究机构等，还可以用于识别科技研究的前沿领域和趋势，因此，引文分析研究的结果可以辅助信息检索和科学与文献管理。引文索引通过对引文链接的有效记录和组织，可以支持大规模的引文分析研究和沿着引文链进行信息检索的方式。

下面，我们介绍一些应用引文网络和引文分析的原理来提供增值信息服务的例子。

### 12.2.1 搜索引擎

在网络时代，许多人每天都要使用搜索引擎在互联网上寻找自己所需要的信息。Google之所以从众多的搜索引擎中脱颖而出，原因之一就是它的检索机制成功地采用了引文分析原理，把被链接 (backlinking) 次数作为衡量网页和网站质量的重要指标之一。具体地讲，Google 根据一个网页的“被引次数”（即被链接次数）来决定这个网页的序位，在其他因素相似的情况下，网页的被链接次数愈多，其在结果显示时的排序位置就愈靠前。使用这样的检索机制，搜索引擎可以聚焦于高质量及与用户需求密切相关的信息。

另外，美国的“开放引文项目” (Open Citation Project: <http://opcit.eprints.org>) 也创建了一个可以根据被引次数对搜索结果进行排序的搜索引擎 (<http://www.citebase.org/>)，这个项目的目标是将美国 Los Alamos 国家实验室的预印本文献库中的每一篇论文与其所引用的该文献库中所有其他文献链接起来，而这个具有按被引次数进行排序功能的搜索引擎是这个项目的一部分。

### 12.2.2 引文链接

电子出版远不止于仅提供印刷出版物的电子复印件。新兴的信息技术支持许多印刷出版物所不可能具备的新特性，这些特性可以增加所出版内容和所提供服务的价值。添加价值的有效方法之一就是在参考文献与其所代表的文献的全文信息之间建立可以点击的超链接，这样，读者就可以在阅读网上文章时自然而有效地追寻这些链接，在相互关联的文章之间随意浏览选择。引文互联项目 (CrossRef - <http://www.crossref.org>)、开放 URL (Open URL - [http://www.exlibrisgroup.com/sfx\\_openurl.htm](http://www.exlibrisgroup.com/sfx_openurl.htm)) 以及前面提到的“开放引文项目”都支持引文链接。

引文互联项目是一项出版社之间的合作引文链接服务，它通过出版社之间的合作，实现在单枪匹马的情况下不可能实现的一些功能，进而把用户与其感兴趣的信息内容有效地连接起来。有了这项服务，用户可以在自己单位已订阅的出版物的范围内，通过简单地点击引文超链接直接从一个出版社的信息内容过渡到被其所引用的另一个出版社的信息内容。至今为止，已经有上百万份学术和专业出版物被收入了引文互联项目。

与引文互联项目不同，开放引文项目不要求参加合作的出版商为其出版物提供源数据、并为其出版物中的参考文献建立链接，而是以专门的计算机程序自动从各种文献

形式的全文文献中为引用及被引用文献抽取源数据、并为参考文献建立可以点击的链接。这样的服务的确已向实现“把每一个曾经发表的学术著作与其所引用的所有文献或所有那些引用它的文献链接起来”的梦想前进了一步 (Hitchcock et al., 1997; Cameron, 1998)。

在这些先驱者项目的促进下，网络环境中的学术文献现已在很大程度上相互链接起来。例如，当用户通过图书馆的检索系统寻找科研论文，可通过简单地点击每一条目来获取相应的论文全文，前提是该图书馆订购了发表这些论文的期刊；当用户进入一个学术期刊的网站，寻找到一篇所需要的论文，可以简单地点击文后的参考文献来获取相应的论文全文，而不用考虑这些论文是否发表在同一期刊，前提是该期刊属于引文互联项目。也就是说，不仅科研论文中的参考文献与其相应的全文论文已被链接起来，二次文献（如图书馆目录、图书馆购买的文献数据库）中的条目也已与相应的全文论文链接起来，这给科研教学人员带来了极大的方便。

### 12.2.3 引文标引

引文链接使得在互联网上沿着引文链从目前的文章追寻到其引用的早期的文章的信息检索成为可能，若要沿着引文链既能从目前的文章追寻到其引用的早期的文章、又能从早期的文章追寻到后期引用这些文章的文献，引文索引是必不可少的工具。美国科学信息研究所 (Institute for Scientific Information – ISI, 现名为 Thomson Scientific) 出版发行的引文索引曾一度是唯一的引文索引。随着互联网在学术传播系统的应用及学术信息电子化的发展，新兴的引文索引不断涌现，CiteSeer, CiteBase, Scopus 及 Google Scholar 是几个实例代表。

CiteSeer (<http://citeseer.ist.psu.edu>) 是由美国 NEC 集团公司的研究所创建，现由该所和美国 Pennsylvania 州立大学的信息科学与技术学院 (Pennsylvania State University's School of Information Sciences and Technology) 共同维护。该索引基于互联网，结构上类似于 ISI 的引文索引，它自动在网上收集各种类型的对公众开放的计算机科学领域的学术文献，对收集到的文献进行引文标引。该索引虽然与 ISI 的引文索引相似，却又有很大的不同之处：

- (1) ISI 的引文索引主要收录期刊文献，而 CiteSeer 所收录的文献使用互联网作为传播媒体，这个媒体比期刊的传播速度快、信息流通范围广；
- (2) CiteSeer 覆盖各种各样的文献类型，例如学位论文、科技报告、会议论文、预印本、期刊文献等等，这些文献类型有可能代表着学术传播过程的不同阶段；
- (3) CiteSeer 比 ISI 引文索引提供更多的有关被引用文献的信息，比如 ISI 的索引只标引第一著者、只提供发表期刊的缩写名称及发表日期和卷页，CiteSeer 则提供全部著者的名称、被引文章的全文标题、发表期刊的全称、以及发表日期和卷页。此外，有些文章的全文及作者的网页也可以获得；
- (4) CiteSeer 的制作和维护高度自动化，而 ISI 的引文索引则建立于人工处理的基础之上；
- (5) CiteSeer 提供引用文献的全文信息，而 ISI 的引文索引只提供引用文献的文摘；
- (6) CiteSeer 对论文后的参考文献按其自身被引用次数进行排序，但不总是提供一篇论文后的全部参考文献。相比之下，ISI 的引文索引则提供全部参考文献，但不进行任何排序；

(7) 除了提供 ISI 引文索引所包括的种种信息外, CiteSeer 还提供许多 ISI 的引文索引所没有或根本不可能提供的很有价值的信息, 例如, 引文环境 (citation context)、作者的网页、文章的被引用次数按年代的分布图、基于各种标准 (如被共引、被共同浏览、来自相同网页等) 的相关论文等等。

由此可见, CiteSeer 不仅是一个有价值而独特的互联网上的学术文献检索工具, 而且还是一个引文分析的有用数据源, 与 ISI 的引文索引相比, CiteSeer 支持更丰富的研究课题, 对此我们将在下一节 (§12.3.3) 作进一步的阐述。

CiteBase (<http://www.citebase.org/>) 是前面提到的美国开放引文项目所开发和建立的引文索引, 这个项目自动从美国 Los Alamos 国家实验室的预印本文献库的全文文献中抽取引用和被引用数据, 通过引文把文献链接起来, 并在此基础上建立了引文索引, 支持按文献引用和被引用关系检索文献和排列结果的信息组织与检索方式。

Scopus 是由闻名世界的出版商 Elsevier 最近开发和生产的引文索引, 它象 ISI 的引文索引一样只收录期刊文献, 主要侧重于自然科学和生物医药科学领域中 1996 年之后出版的论文, 但对这些领域中近年出版的论文的覆盖率接近甚至超过 ISI 的引文索引。它的特别之处在于提供全文引用文献及比较完全的有关被引文献的信息, 包括论文题名及前七位和最后一位作者的姓名。这些信息对引文分析具有非常重要的价值, 例如, 使用这个索引来收集引文数据, 引文率和共引率的计算可以不象使用 ISI 的引文索引那样局限于只考虑第一作者, 这样的引文分析结果也因此会有可能更公正、明朗一些 (Zhao, 2006a; 2006b)。可惜的是, 由于使用 Scopus 收集引文数据的困难性 (例如, Scopus 不支持大批量下载检索结果), 至今还没有基于这个引文索引的大规模的引文分析研究。

Google Scholar (<http://scholar.google.com>) 是 Google 的最新发展之一, 专用于搜索学术论文。除了具有 Google 作为搜索引擎的优势之外, Google Scholar 还提供命中文献的被引次数及引用了每一篇命中文献的所有文献等等, 因此成为一个利用引文链辅助学术文献检索的有效工具, 并为引文分析研究提供了有价值的工具。研究显示, 根据 Google Scholar 提供的被引次数对科研论文进行排序的结果与根据 ISI 引文索引提供的相应数据的结果非常接近, 如果只考虑近期的出版物, 连被引次数的绝对值都非常接近 (Charbonneau, 2006; Pauly & Stergiou, 2005)。

ISI 引文索引虽然奉行其创始人 Eugene Garfield 的“文献集中法则” (Garfield's law of concentration) 不可能轻易改变其覆盖面及收录期刊的原则, 但它们在新的网络环境中却也发生了一些变化。比如, 这些索引现在的万维网版本, 即“科学网” (Web of Science), 若被引文献恰巧是这些索引的源文献, 这些被引条目可以通过超链接与源文献记录相连, 这样可获得这些被引文献的全面信息, 如所有作者而不只是第一作者的信息。

#### 12.2.4 基于文献计量学的信息检索系统

引文索引在学术传播中具有两个重要的作用: (1) 独特的信息检索系统, (2) 应用引文分析的方法研究科学技术传播规律的数据源。如前所述, 引文分析的结果不仅有助于理解学术传播的结构和过程, 而且可以辅助信息检索。例如, 评价性引文分析 (evaluative citation analysis) 的结果有助于检索出高质量的文献及核心著者、机构或国

家所发表的文献，而关联性引文分析 (relational citation analysis) 所产生的论文、作者或研究领域的聚类网络图示结果可以帮助扩展检索提问。ISI 的引文索引及 CiteSeer 等其他引文索引已经通过提供诸如每篇文章的被引次数、共引文献及期刊质量的各种指标等数据，展示出将引文分析结果应用于信息检索系统所具有的独特价值。

所谓“基于文献计量学的信息检索系统”，即寻求充分利用文献计量学的技术和研究结果，以帮助解决目前信息检索系统所面临的一些问题 (Ding et al., 2000)，这些系统常常使用可视化 (visualization) 技术动态地显示由某种词汇分析（如共词分析）所产生的概念网络或由共引分析所产生的论文或作者网络。这些网络可以帮助用户获得某一学科领域的概貌和理解概念、论文或作者之间的相互关系，因此有助于检索提问的扩展和检索的精细化。然而，结合评价性和关联性信息计量学分析结果辅助检索的“充分基于文献计量学” (fully bibliometrics-aware) 的信息检索系统至今还不存在。

虽然信息计量学特别是引文分析结果在信息检索中的应用价值并不难理解，但对基于文献计量学的信息检索系统的研究多数集中于词汇分析和概念网络上。信息计量学中最广为人知的引文分析技术在信息检索系统中的应用还没有被充分地探讨，这在很大程度上是因为绝大多数信息检索系统收录的信息很久以来根本不含有关引用和被引用之间关系的信息。随着科研论文全文信息（包括参考文献）越来越多地涌现，以及全自动引文标引工具的日益增长完善，这种由数据源所带来的障碍即将消失，对基于文献计量学的信息检索系统的研究也会大量增长，甚至有可能成为信息检索系统研究的一个新的焦点。林夏、White 及 Buzydlowski 等人已经在这方面作出了有益尝试 (Buzydlowski, White, & Lin, 2001; Lin, White, & Buzydlowski, 2003; White, Lin, & Buzydlowski, 2001)。他们使用 ISI 人文科学引文索引 (Arts & Humanities Citation Index) 1988-1997 的数据，探讨了著者共引分析结果对文献检索的价值。使用他们的实验系统，输入著者名称进行检索可以获得一幅与该著者共引次数最高的 25 位著者的共引关系图，用户可以用鼠标将其中任何著者的名称拖入检索词输入区以获得该著者的所有文献。

### 12.3 网络环境中的学术传播系统给信息计量学研究带来的机会与挑战 (Opportunities and challenges brought to Informetrics research by scholarly communication in the networked environment)

传播 (communication) 在科学研究中的重要作用 (Cole & Cole, 1973; Garvey, 1979; Meadows, 1974; Merton, 1979) 曾经吸引了从科学社会学家、传播研究人员、科学史学家到信息学家等各类研究人员的广泛注意。他们从不同的角度对学术传播进行了研究，其中很多研究使用了引文分析这个信息计量学中众所周知的技术。

信息技术的加速发展，尤其是互联网的快速成长，正在改变学术传播的结构和过程，人们对学术传播的研究兴趣也再次萌发和增长，希望能够了解正在形成的传播形式以及与人们原先所熟悉的以学术期刊为中心的学术传播模式的相似和不同之处。引文分析及其他信息计量学的理论和方法再次被成功地应用于研究这些学术传播中的新现象。

#### 12.3.1 网络环境中的学术传播

以学术期刊为中心的学术传播系统已有几百年的历史，学术期刊不仅是学术传播的有效媒介，还是评价学术贡献、确立科学发现的有用资料。随着互联网的广泛应用，大部分现有的学术期刊已经开始出版电子版，新兴的纯电子版学术期刊也如雨后春笋般地不断涌现。

与此同时，一些全新的学术传播模式也在不断出现，开放存取 (Open Access) 学术期刊和机构数字典藏 (institutional repository) 就是其中的两个例子。开放存取学术期刊与普通学术期刊本质相似，唯一不同之处在于这些期刊是免费供所有用户使用。机构数字典藏用于储存由某个或某些机构所拥有的信息资源，包括教职员工的科研论文和教学资料以及学生的学位论文等等。这些资源无需经过漫长的同行评价过程或者昂贵的出版发行渠道，有助于最新科研成果快速与广大的用户见面，进而极大地促进学术传播的效率。更多的相关介绍讨论请见本书的§2.4，数字典藏。

此外，越来越多的科研和教学人员往往把最新的资料和成果公开发表在自己或工作机构的网站上，并把参与网上学术交流活动（如学术邮件表 Listserv）作为生活和工作不可缺少的一部分。越来越多的资源数字化 (digitization) 项目正在系统化地把许多科技和历史资料全文加以组织整理，有些项目还供用户免费使用。

可以说，互联网上的学术信息资源已经而且还在继续迅猛地增长，在网上寻找科研信息已经相当普遍。在有些领域（如高能物理和计算机科学）的科研人员甚至往往把互联网作为寻找最新研究成果信息的首选资源，以保持与其他科研单位的同行或竞争对手齐头并进。

由此可见，学术传播系统正在从以学术期刊为中心的模式向更适用于网络环境的新模式发展，这给信息计量学带来了极大的机会和挑战。

### 12.3.2 研究学术传播系统从以学术期刊为中心的模式向网络化新模式的发展

如前所述，以学术期刊为中心的学术传播系统已有百余年的历史，从这一模式向更适用于网络环境的新模式的发展并非易事。例如，机构数字典藏的优势众所周知，但由于科研人员一般不主动地把自己的论文和数据存放到这些资源库中，现有的数字典藏库存量相当有限，以至于无法展现其潜在的巨大优势。因此，研究怎样促使机构数字典藏的成功成为一个非常重要的课题。

应用信息计量学研究学术传播系统从以学术期刊为中心的模式向网络化新模式的发展可以包括这样一些课题：比较新兴传播模式与传统模式的联系与区别、研究基于互联网的学术传播与以期刊为中心的学术传播之间的信息流通、探索如何将信息计量学为基础的科研评估从期刊扩展到机构数字典藏等等 (The Open Citation Project, 2001; Goodrum et al., 2001; Zhao & Logan, 2002; Zhao, 2005a; Zhao, 2005b)。

### 12.3.3 发展信息计量学的理论和方法

新兴的学术传播模式为信息计量学提供了新的数据基础。由于电子化全文科研信息的大量涌现，新的检索工具接连出现，以帮助人们搜索这些新型信息，已有的信息检索系统也由此得到了很大的提高。这些新的数据和工具给信息计量学的研究带来了新

的机会和挑战，这从前面对新型引文索引（比如 CiteSeer, CiteBase, Scopus, Google Scholar）的讨论已经可以看出，这里将进一步阐述如下。

如前所述，引文分析是一个众所周知的强大的信息计量学技术。迄今为止，绝大多数引文分析的研究使用 ISI 的引文索引作为数据源，这些引文索引曾经一直是唯一可用的引文索引，也极大地促进了引文分析的理论和方法的发展及广泛使用。然而，由于在印刷时代，生产引文索引是一项极为困难而昂贵的工作，ISI 的引文索引具有很多弊端（例如，只标引第一著者，只覆盖选择性极高的期刊论文，等等）；同样因为在印刷时代，若没有引文索引的帮助，收集信息计量学所要求的大量的引文数据近于不可能，因此，信息计量学家虽然非常了解这些弊端，却几十年来继续使用这些引文索引进行信息计量学的研究。这些引文索引的弊端已经大大地局限了引文分析理论和方法的进一步发展和应用 (Borgman, 1990; Edge, 1979; Garfield, 1979; MacRoberts & MacRoberts, 1989; Osareh, 1996; Smith, 1981)。

随着学术传播系统的最新发展，ISI 引文索引的局限性日渐严重。例如，信息技术的飞速发展正在引起信息生产和交换方式的革命性变化，学术传播系统正在向一个新模式发展，在这个新模式中，预印本、会议论文以及学术论文的网上公布成为重点，互联网成为科研人员文献检索的首选资源，这就意味着，研究这些类型文献中所呈现出的学术传播方式和规律变得越来越重要，使用只覆盖期刊的 ISI 引文索引来进行引文分析的研究具有越来越大的局限性。

对这些问题的系统研究应该对引文分析、进而对信息计量学的理论和方法的发展产生很大的影响，而新兴的互联网上的引文数据和工具为这类研究提供了有效的数据源。这些数据和工具虽然具有它们自身的问题，例如，文献收录缺乏选择性和系统性、数据自动分解 (parsing) 难以做到绝对准确，等等，但这些问题却不同于前面提到的 ISI 引文索引所具有的弊端，因此，使用这些数据和工具进行信息计量学研究，会给一些有关引文分析的理论和方法的长期争论提供有力的实验数据，这些研究所展现出的新兴数据和工具在信息计量学中的应用潜力又会引发更为深入的研究，从而推动信息计量学理论和方法的发展。

如前所述，与 ISI 的引文索引相比，互联网上的新兴数据和工具有很多不同之处，而这些不同之处也正是新的机会和挑战所在。例如，这些数据和工具使用互联网作为传播媒体，这个媒体比期刊的传播速度快、信息流通范围广，这对学术传播系统的涵义值得研究和探讨；这些数据和工具覆盖学位论文、科技报告、会议论文、预印本、期刊文献等各种各样的文献类型，它们有可能代表着学术传播过程的不同阶段，因而支持更丰富的研究课题；这些数据和工具比 ISI 的引文索引提供更多的有关被引用文献的信息，比如 ISI 的索引只标引第一著者、只提供发表期刊的缩写名称及发表日期和卷页，而这些新工具往往提供全部著者的名称、被引文章的全文标题、发表期刊的全称以及发表日期和卷页，些文章的全文及作者的网页也可以获得；这些数据和工具还提供引文环境 (citation context) 等传统引文索引难以提供的数据。

由此可见，这些新的数据和工具给许多新的研究课题打开了大门，例如，引文环境分析早已由共引分析的创始人之一 Henry Small 提出 (Small, 1982)，但由于 ISI 的引文索引不提供全文信息，使用它们进行这类研究意味着依靠大量、繁琐而费时的手工操作来获得引文环境的数据。因而，虽然 Small 的工作展示了引文环境分析的巨大潜

力，但这类研究却始终未能广泛地开展起来。随着 CiteSeer 等新型引文索引的出现，获取引文环境的数据变得比较容易，这类研究也有可能随之发展起来。又如，CiteSeer 等新工具所提供的被引用文献的全部著者的信息可以支持引文分析的研究从只考虑被引文献的第一著者发展到考虑所有著者，进而得到更公正、明了的结果 (Zhao, 2006a; 2006b)。此外，这些新工具的开放性及其引用和被引用文献之间的超链接使得采用更复杂而巧妙的算法（如 Google 的 PageRank）来评价学术贡献成为可能。也就是说，这些新的数据和工具可以给信息计量学指标的扩展、提高和创新提供有力的支持。正如 ISI 的引文索引的诞生极大地促进了引文分析理论和方法的发展，互联网所带来的新数据和工具也许会给信息计量学带来新的发展。

关于上述方面的探索刚刚开始，发展前景广大。例如，美国的“开放引文项目” (The Open Citation Project) 对 Los Alamos 国家实验室的预印本文献库进行了各式各样的分析研究，以探索这个文献库中所呈现出的学术传播方式和规律 (The Open Citation Project, 2001)。Goodrum 等人 (2001) 探讨了学术传播系统向新模式的发展，讨论了研究这些新模式的重要性和使用全自动引文索引进行这类研究的可能性，他们以 CiteSeer 覆盖的出版物来代表网上出版物、ISI 引文索引覆盖的出版物来代表期刊出版物，就高频率被引文献对文献类型和出版日期分布对网上出版物和期刊出版物进行了比较研究。他们发现，会议论文在高频被引的网上出版物中的含量远远高出在高频被引的期刊出版物中的含量，网上文献对近期文献的引用普遍多于期刊文献对近期文献的引用，另外，从两种数据源中提取的高频被引文献有很大的重合。Zhao 和 Logan (2002) 探索了采用网上科技出版物作为数据源是否为研究网络环境中的学术传播规律的有价值方法。她们对 XML 研究领域进行了著者引用和共引分析，发现了一些基于互联网的学术传播特点，及其与基于期刊的学术传播的联系与区别。正如 Goodrum 等人的研究 (2001)，Zhao 和 Logan (2002) 也发现了网上出版物对会议论文的引用比期刊对会议论文的引用要多，在网上和期刊中的高频被引文献有很大重合，然而，网上的引用文献与期刊中的引用文献之间的重合却非常小。Zhao (2004; 2005a) 进一步比较了基于互联网和基于期刊的学术传播模式，发现网上出版物与期刊文献的确代表学术传播过程中的不同阶段，其中期刊文献比网上出版物要滞后几年。这说明，采用网上出版物进行引文分析具有揭示科研最前沿状况的优势，网络环境中的信息计量学研究需要有效地结合多种数据源，而不是象传统的引文分析研究那样只使用期刊文献为数据。Zhao (2006a; 2006b) 展示了互联网上的新数据和工具给探索信息计量学新指标带来的机会，这两项研究突破了使用 ISI 引文索引的传统引文分析研究中只考虑被引文献的第一著者的局限，利用 CiteSeer 的支持，比较了基于第一著者和所有著者的引文分析结果，展示了基于所有著者的引文分析研究的优势。

这些研究展示了应用网上科技出版物及其有关工具进行信息计量学研究的可行性，揭示了一些值得进一步探讨的问题。

## 12.4 网络计量学 (Webometrics)

网络计量学是应用信息计量学的理论和方法研究与万维网有关现象的一门新兴学科。虽然网络计量学实际上也是信息计量学在网络环境中的应用之一，但由于网络计量学已经成为一门涉及多门学科、快速发展的交叉学科，本节特作专门阐述。对网络计量学的研究已经超越了信息科学的范畴，对此贡献较大的其他学科包括传播学、统



计学以及计算机科学。这里，我们将概括网络计量学的研究分支及网络计量学研究中面临的问题。

#### 12.4.1 网络计量学的研究分支

网络计量学的研究包括三个分支：超链接结构分析 (link structure analysis)、网络服务器记录分析 (Web server log analysis) 以及互联网环境中信息计量学分布规律的研究，其中前两个分支的研究占绝大部分。

##### 12.4.1.1 超链接结构分析

这方面的研究包括与传统的引文分析可以类比的超链接分析(link analysis) 和应用计算机图形学原理 (graph theory) 进行的万维网拓扑结构研究 (Web topology study)。

Rousseau (1997) 创新使用了 *sitation analysis* 一词描述使用传统的引文分析的方法将万维网超链接类比为文献中的引文而进行的分析研究。正如在传统的引文分析领域，有些研究通过计算链接到网页或网站的网页数量评价网页或网站的质量或影响程度；有些研究通过计算网页或网站的共同链接关系对一批相关的网页或网站进行聚类分析，以发现其相互关系；还有些研究试图探讨人们设立超链接的动机和模式。一个网页可以是任何一个普通网页，也可以是一篇学术论文；一个网站可以是一所大学的所有网页，也可以是一个院系或其他子单位的所有网页，还可以是一个国家的所有网页；通过选择不同的分析对象探讨不同的研究课题。例如，Chu (2001) 选择了十二所图书情报学院的网站作为分析对象，通过研究这些网站的超链接结构对这些学院进行了的质量评价和相互关系的分析研究。更多的实例可见 Thelwall, Vaughan, & Bjorneborn (2005) 和 Ingwersen (2001)。

由于万维网可以被看成一个有向或无向的图形 (directed or undirected graph)，而且是迄今为止最大的提供拓扑信息的网络，许多物理学家、计算机学家及信息学家已经采用图形学理论研究互联网，并使用万维网作为验证新模型的基地 (Albert & Barabasi, 2002; Thelwall, Vaughan, & Bjorneborn, 2005)。实际上，自从六十年代图形学理论就已经在信息计量学中被用于研究引文网络和其他信息网络（如 Furner, Ellis, & Willett, 1996; Garner, 1967; Hummon & Doreian, 1989; Nance, Korfhage, & Bhat, 1972），应用图形学理论来研究万维网只是这种应用的一个自然延伸。

许多这类研究建筑在搜索引擎 Google 的 PageRank 算法或 Kleinberg 的 Hypertext-Induced Topic Search (HITS) 算法的基础之上。Google 的 PageRank 算法使用前面已介绍的被链接 (backlinking) 机制，以网页获取的被链接次数作为衡量该网页质量和检得结果排序的标准之一。HITS 与 PageRank 的不同之处在于其以检索题目为中心的做法以及对“权威” (authority) 和“枢纽” (hub) 网页的区分。“权威”网页是指那些被认为是有关某个题目最好的信息源，而“枢纽”网页是一个链接到这些高质量信息源的超链接的集合。权威和枢纽网页的判断选择标准既包括传统的词频词距，也采用 PageRank 这样的新算法。对于任何一个检索题目，HITS 首先通过传统的词汇匹配机制检得一批相关网页，然后根据 HITS 机制检得权威和枢纽结果 (Chakrabarti, et al., 1999, p. 5)。

Bjorneborn (2003) 研究了英国学术网站的拓扑结构, 发现由 HITS 产生的权威和枢纽网页也常常具有较高的 *betweenness centrality*, 即在一个网络环境中一个节点 (如网页) 出现在任意两个节点之间最短路径上的概率。Thelwall (2003) 使用 PageRank 算法对学术网页进行了排序, 发现因为许多网页具有为导航而设置的指向自身的链接, 致使排在前面的网页常常是指向而不是拥有实际内容的枢纽类网页 (如大学的首页)。从理论上讲, HITS 和 PageRank 在测量网页的价值上优于简单地计算“被引次数” (即链接到一个网站的网页数量) 的做法, 但在实际中, 这个优势很难得到展现。因而, 探索 HITS 和 PageRank 在科研论文之间的引文网络 (而不是互联网上的超链接网络) 中的应用效果成为一个非常有意义的课题。

#### 12.4.1.2 网络服务器记录分析

信息学研究中的一个传统是通过观察和记录用户的检索活动来研究用户的检索行为, 以此探索怎样改善和提高信息检索系统的用户界面或用户培训, 这些观察和记录大都是在实验环境下、而不是在用户的实际工作中进行的。万维网服务器自动而客观地捕获真正用户的实际网上浏览过程, 为研究用户的信息检索行为、进而改进信息服务系统和网站设计提供了非常有价值的数据库资源 (Thelwall, Vaughan, & Bjorneborn, 2005)。

虽然网络服务器记录的分析研究有很大困难, 但对诸如 AltaVista 和 Excite 等搜索引擎的使用行为的研究却已有很多成果 (Jansen & Spink, 2006; Jansen, Spink, & Pedersen, 2005; Spink, et. al., 2006; Spink, Jansen, & Pedersen, 2004)。通过分析检索提问的结构和题目类型、各类检索特征的使用频率、命中结果的使用方式等等, 研究人员提高了对用户使用搜索引擎的行为规律的理解, 进而对搜索引擎进行了相应的改造不住。此外, 人们也通过对网络服务器记录的分析, 对个别网站 (特别是提供医药卫生信息的网站和图书馆信息资源网站) 上用户与信息资源的交互模式进行了研究。例如, Nicholas 等 (2004) 发现用户很少进入一个网站的深层网页, 但经常访问多个网站来满足某个信息需求, 用户也很少重新访问以前到过的网站。Chen 和 Cooper (2001) 使用了聚类技术探测用户对一个图书馆目录系统的使用模式。

网络服务器记录的分析也被用于研究科研论文的网上下载率与它们在科研论文中的被引用率的关系。例如, Moed (2005) 发现, 下载和引用的文献呈现出不同的老化规律; 下载与引用互相影响: 一篇论文被引用后的三个月内, 若与未被引用的情况相比, 其下载率增长了 25%; 下载与引用的相关程度在 0.11 到 0.35 之间, 其中以早期下载率与后期引用率之间的相关程度最低、引用率与忽略早期的情况而计算的下载率之间的相关程度最高。Moed 的结论是, 下载和引用分别与写作一篇期刊论文的信息收集和处理过程中的不同阶段相关联。Brody, Harnad 和 Carr (2006) 使用美国 Los Alamos 国家实验室的预印本文献库的英国镜象网站的记录为数据, 发现在物理学和数学领域, 论文的下下载率与引用率之间具有大约 0.4 的相关程度, 其结论是, 下载率可以为预测一篇论文的被引率提供一个早期的估计, 并可以提供另外一个测度论文影响率的指标, 这个指标所测度的科研信息的使用模式是引文率所无法做到的。

#### 12.4.1.3 万维网环境中的信息计量学分布规律的研究

典型的信息计量学法则（如布拉德福法、洛特卡法及齐普夫法）已被人们在万维网环境中进行了验证。例如，Ajiferuke, Wolfram 和 Xie (2004) 对两个大型网站进行了信息计量学研究，发现网站访问的数据遵循齐普夫定律，Nielsen (1997) 对网站数量对网站的年浏览量的分布研究也发现了齐普夫分布规律；Rousseau (1997) 则发现域名及网站间链接的分布符合洛特卡分布规律；Cui (1999) 在网上发现了布拉德福分布；Boudourides, Sigrist 和 Alevizos (1999) 以及 Egghe (2000) 则提出了是否这些典型的信息计量学法则在互联网上仍然有效的问题。

#### 12.4.2 网络计量学研究面临的问题

将在研究学术文献中所发展起来的信息计量学的方法和技术应用到互联网这样一个完全不同的环境之中，问题和顾虑自然不可避免，概括起来主要有以下几个方面：

##### (1) 超链接与科研论文中的引文在本质上不同

如前所述，科研论文中的参考文献或引文是著者本人对这些文献与自己的论文之间在内容或方法上是否密切相关的一种判断，虽然这个判断过程或许受其他一些因素的影响，如出于礼貌的引用 (Edge, 1979)，内容或方法上的相关性是不可缺少的一个重要因素。而在万维网上，超链接可以由不对网页内容负责的网站管理员出于各种各样的原因而建立。除了内容相关外，导航（如“下一页”）和其他许多与内容无关的东西（如“下载插件”、“版权信息”）也可成为超链接。而且，超链接可以是双向的，而引文一般是单向的。也就是说，超链接与科研论文中的引文具有本质的不同，因此应用引文分析的理论和方法进行超链接研究需要谨慎从事 (Egghe, 2000; Chu, 2001)。

##### (2) 作为网络计量学研究数据源的互联网与引文索引具有本质区别

由于互联网本身缺乏有效的质量控制，因此互联网上的信息良莠混杂、种类繁多。相比之下，ISI 引文索引只收录其专家顾问团慎重选择的科学期刊中发表的论文，也就是说，收录在这些索引中的论文已经过期刊编辑部、同行评审、引文索引编辑等多道质量检查，因而达到一定的质量标准、满足一定的格式要求。因为撰写学术论文本身意味着遵循一定的行为标准和规范，如引文分析理论所依据的 Merton 的规范化科学观 (Merton's normative view of science (Merton, 1942; Edge, 1979; Cronin, 1984))，以这些“精选”出来的论文与以“大众化”的论文作为引文分析的数据源相比，哪一个更有助于获得准确而有效的研究结果还有待于进一步研究和验证。但是，互联网上的信息发表不必遵循任何这样的标准和规范，原则上任何人任何时候都可以在网上发表任何信息，因此，很难说引文分析的理论和方法是否适用于这个环境。而且，互联网上的信息非常不稳定，一个网页可以随时出现，也可以随时消失，这给网络计量学的研究带来了进一步的困难 (Chu, 2001)。

##### (3) 搜索引擎作为数据收集工具所具有的问题

网络计量学的研究很大程度上依赖于搜索引擎收集数据，而这些搜索引擎自身具有覆盖率、检索功能和商业性等问题。

首先, 搜索引擎的相对覆盖率很有限, 而且自 1997 年以来还在不断降低。据研究, 在所有可公开标引的网络信息中, 一个搜索引擎能够收录的内容不超过 16% 者 (Lawrence & Giles, 1999)。另外, 搜索引擎的覆盖率还常常具有国家偏向, 比如, 一些主要搜索引擎对美国网站的覆盖程度要比对中国和新加坡的网站的覆盖程度高得多 (Vaughan & Thelwall, 2004)。

其次, 搜索引擎所支持的检索功能也各不相同、不尽人意, 检索结果变幻无常、难以依赖 (Bar-Ilan, 1999; Ingwersen, 1998; Rousseau, 1999; Snyder & Rosenbaum, 1999)。例如, Google 最初不支持逻辑“与”操作, 检索词的不同组合次序 (比如检索词 A “与”检索词 B, 或检索词 B “与”检索词 A) 会使得 AltaVista 返回不同的检索结果 (Chu, 2001)。

此外, 搜索引擎的内部操作、检索方式和结果排序的算法是商业机密, 公众对此不得而知, 这给以搜索引擎作为数据收集工具的网络计量学研究造成相当的困难。并且, 这些算法可以未加通知随时变化, 网络计量学所依赖的一些高级搜索命令常常没有足够的使用说明, 并可能随时消失。部分搜索引擎所采用的商业模式 (如以交费的多少决定检索结果的排序) 也很明显会对网络计量学研究带来一定的问题 (Thelwall, Vaughan, & Bjorneborn, 2005)。

上述问题说明, 网络计量学研究人员在使用搜索引擎收集数据时, 应该尽一切努力来评定数据的准确性、并尽可能采取补救措施。如何进行评定和补救一直而且仍然是网络计量学研究中的一个重要课题 (Thelwall, Vaughan, & Bjorneborn, 2005)。

## 12.5 结论 (Conclusions)

互联网作为一个强大的传播媒介正在引起信息生产和交换方式的变革, 随着各种信息在互联网上的大量涌现、学术传播系统从以学术期刊为中心的模式向网络化新模式的发展, 信息计量学的理论和方法一方面正在被成功地应用于开发增值信息服务和提高信息组织和检索的效率, 另一方面也面临着新的数据和工具给其进一步发展和丰富而带来的机会和挑战。本章探讨了这些问题。

信息计量学在网络环境中的信息组织与检索方面的应用广泛, 本章以搜索引擎、引文链接、引文标引以及基于文献计量学的信息检索系统为例对此进行了介绍。采用引文分析原理把被链接次数作为衡量网页和网站质量的指标之一是 Google 等搜索引擎成功的重要原因之一。引文链接使得网络环境中的学术文献在很大程度上相互链接起来, 这种链接现已不局限于科研论文中的参考文献与其相应的全文论文之间, 二次文献 (如图书馆目录、图书馆购买的文献数据库) 中的条目也已与相应的全文论文链接起来, 这给科研教学人员带来了极大的方便。支持追寻引文链进行文献检索的引文索引在网络环境中得到了可喜的发展, 与传统的 ISI 引文索引相比, 新兴的引文索引 (如 CiteSeer, CiteBase, Scopus 及 Google Scholar) 在网络技术和电子化信息的支持下发展了一些有用的检索特点, 例如提供更丰富的有关被引文献的信息, 覆盖更多的文献类型, 等等。学术信息的大规模电子化也为基于文献计量学的信息检索系统的发展奠定了基础, 概念和词汇网络在信息检索中的应用已在实际检索系统中可见, 引文分析结果在信息检索中的应用也在探讨之中。

学术传播系统正在从以学术期刊为中心的模式向更适用于网络环境的新模式发展, 本章探讨了如何应用信息计量学研究这一发展过程。这方面潜在的研究课题很多, 例

如：比较新兴传播模式与传统模式的联系与区别、研究基于互联网的学术传播与以期刊为中心的学术传播之间的信息流通、探索如何将以信息计量学为基础的科研评估从期刊扩展到机构数字典藏，等等。新的学术传播系统也推出了新的信息计量学研究数据和工具，正如 ISI 引文索引的诞生极大地促进了引文分析理论和方法的发展，这些新的数据和工具也许会进一步促进信息计量学理论和方法的扩展、提高和创新，如何将基于第一被引著者的引文分析研究扩展到所有被引著者、如何使用类似于 Google 的 PageRank 的更复杂而巧妙的算法进行基于文献的科研评估、如何利用大规模的电子化知识信息研究知识网和知识流的结构和特点，等等，即为其中的几个研究方向。

本章最后介绍了网络计量学的主要研究分支，即：超链接结构分析、网络服务器记录分析以及互联网环境中信息计量学分布规律的研究，并指出了目前研究中所面临的一些问题，其中包括搜索引擎作为数据收集工具所具有的问题，以及超链接与引文及互联网与引文索引在本质上的不同给应用引文分析的理论和方法的网络计量学研究带来的问题。如何解决这些问题是今后网络计量学的研究方向之一。

## 参考文献

- Ajiferuke, I., Wolfram, D. & Xie, H. (2004). Modeling Website Visitation and Resource Usage Characteristics by IP Address Data. *Proceedings of Canadian Association for Information Science/L'association canadienne des sciences d'information (CAIS/ACSI) 2004 Annual Conference, Access to Information: Technologies, Skills, and Socio-Political Context*, June 3-5, 2004, Winnipeg, Manitoba, Canada.
- Albert, R., & Barabási, A.L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47-97.
- Bar-Ilan, J. (1999). Search engine results over time. A case study on search engine stability. *Cybermetrics*, 2/3(1). Retrieved June 20, 2006, from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>
- Buzydowski, J., White, H. D., & Lin, X. (2001). Co-cited author analysis as an interface for digital libraries. *JCDL Workshop on Visual Interfaces to Digital Libraries*, Roanoke, VA, June 24 - 28, 2001.
- Bjørneborn, L. (2003). Small-world link structures across an academic Web space: A library and information science approach. Unpublished doctoral dissertation, Royal School of Library and Information Science, Copenhagen, Denmark.
- Borgman, C.L. (ed.). (1990). *Scholarly Communication and Bibliometrics*. Newbury Park, CA: Sage Publications, Inc.
- Boudourides, M.A., Sigrist, B., & Alevizos, P. (1999). *Webometrics and the self-organization of the European information society*. Available at <http://hyperion.math.upatras.gr/webometrics/>
- Brody, T., Harnad, S., & Carr, L. (2006). Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Association for Information Science and Technology*, 57(8), 1060-1072.
- Chakrabarti, S., Dom, B., Kumar, S. R., Raghavan, P., Rajagopalan, S., Tomkins, A., Kleinberg, J. M., Gibson, D. (1999). Hypersearching the Web. *Scientific American*, 280(6), 54-60.

- Cameron, R. D. (1998). A universal citation database as a catalyst for reform in scholarly communication. *First Monday*, 2(4). Retrieved June 20, 2006, from [http://www.firstmonday.dk/issues/issue2\\_4/cameron/](http://www.firstmonday.dk/issues/issue2_4/cameron/)
- Charbonneau, L. (2006). Google Scholar service matches Thomson ISI citation index. *University Affairs*, 40.
- Chen, H.-M., & Cooper, M.D. (2001). Using clustering techniques to detect usage patterns in a Web-based information system. *Journal of the American Society for Information Science and Technology*, 52(11), 888-904.
- Chu, H. (2001). A webometric analysis of ALA accredited LIS schools' websites. In M. Davis & C. S. Wilson (eds.), *Proceedings of the 8th International Conference on Scientometrics and Informetrics* (pp. 117-126). Sydney, Australia: The University of New South Wales.
- Cole, J.R., & Cole, S. (1973). *Social Stratification in Science*. Chicago: University of Chicago Press.
- Cronin, B. (1984). *The Citation Process: The Role and Significance of Citations in Scientific Communication*. London: Taylor Graham.
- Cui, L. (1999). Rating health Web sites using the principles of citation analysis: A bibliometric approach. *Journal of Medical Internet Research*, 1(1). Retrieved June 20, 2006, from <http://www.jmir.org/1999/1/e4/index.htm>
- Ding, Y., Chowdhury, G. G., Foo, S., & Qian, W. (2000). Bibliometric information retrieval system (BIRS): A Web search interface utilizing bibliometric research results. *Journal of the American Society for Information Science*, 51, 1190-1204.
- Edge, D. (1979). Quantitative measures of communication in science: a critical review. *History of Science*, 7, 102-134
- Egghe, L. (2000). New informetric aspects of the Internet: Some reflections, many problems. *Journal of Information Science*, 26, 329-335.
- Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics*. New York: Elsevier Science.
- Furner, J., Ellis, D., & Willett, P. (1996). The representation and comparison of hypertext structures using graphs. In M. Agosti & A.F. Smeaton (Eds.), *Information retrieval and hypertext* (pp. 75-96). Boston: Kluwer.
- Garfield, E. (1979). *Citation indexing — its theory and application in science, technology, and humanities*. New York: John Wiley & Sons.
- Garner, R. (1967). A computer oriented, graph theoretic analysis of citation index structures. In B. Flood (Ed.), *Three Drexel information science research studies* (pp. 3-46). Philadelphia: Drexel Press. Retrieved 20 June, 2006, from <http://www.garfield.library.upenn.edu/rgarner.pdf>
- Garvey, W.D. (1979). *Communication: The essence of science*. New York: Pergamon.
- Goodrum, A. A., McCain, K. W., Lawrence, S., & Giles, C. L. (2001). Scholarly publishing in the Internet age: A citation analysis of computer science literature. *Information Processing and Management*, 37, 661-675.
- Griffith, B.C. (1990). Understanding science: Studies of communication and information. In C. L. Borgman (ed.), *Scholarly Communication and Bibliometrics* (pp. 31-45). Newbury Park, CA: Sage.

- Hitchcock, S., Carr, L., Harris, S., Hey, J.M.N., & Hall, W. (1997). Citation linking: Improving access to online journals. *Proceedings of the 2nd ACM International Conference on Digital Libraries*, 1997, 115-122.
- Hummon, N.P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11, 39-63. Retrieved June 20, 2006, from <http://www.garfield.library.upenn.edu/papers/hummondoreian1989.pdf>
- Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, 54, 236-243.
- Jansen, B.J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, 42 (1 SPEC. ISS), pp. 248-263.
- Jansen, B.J., Spink, A., & Pedersen, J. (2005). A temporal comparison of AltaVista Web searching. *Journal of the American Society for Information Science and Technology*, 56 (6), pp. 559-570.
- Lawrence, S., & Giles, C.L. (1999). Accessibility and distribution of information on the Web. *Nature*, 400, 107-110.
- Lin, X., White, H.D., & Buzydlowski, J. (2003). Real-time author co-citation mapping for online searching. *Information Processing and Management*, 39(5), 689 – 706.
- Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science*, 10, 145-162.
- Long, J. S., McGinnis, R. and Allison, P. D. (1980). The problem of junior-authored papers in constructing citation counts. *Social Studies of Science*, 10, 127-143.
- MacRoberts, M. H. and MacRoberts, B. R. (1989). Problems of citation analysis: a critical review. *Journal of the American Society for Information Science*, 40, 342-349.
- McCain, K. W. (1988). Evaluating cocited author search performance in a collaborative specialty. *Journal of the American Society for Information Science*, 39, 428-431.
- Meadows, A.J. (1974). *Communication in science*. London: Butterworths.
- Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1, 115-126.
- Merton, R. K. (1979). *The sociology of science: An episodic memoir*. Carbondale: Southern Illinois University Press.
- Moed, H. F. (2005). Statistical Relationships Between Downloads and Citations at the Level of Individual Documents Within a Single Journal. *Journal of the American Society for Information Science and Technology*, 56(10), 1088-1097.
- Nance, R.E., Korfhage, R.R., & Bhat, U.N. (1972). Information networks: Definitions and message transfer models. *Journal of the American Society of Information Science*, 23(4), 237-247.
- Nicholas, D., Huntington, P., Williams, P., & Dobrowolski, T. (2004). Re-appraising information seeking behaviour in a digital environment: Bouncers, checkers, returnees and the like. *Journal of Documentation* 60(1), 24-43.
- Nielsen, J. (1997). *Do websites have increasing returns?* Available at <http://ww.useit.com/alertbox/9704b.html>

- The Open Citation Project. (2001). Mining the social life of an eprint archive. Retrieved June 20, 2006, from <http://opcit.eprints.org/tdb198/opcit/>
- Osareh, F. (1996). Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature II. *Libri*, 46, 217-225.
- Pauly, D., & Stergiou, K. (2005). Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar Service. *Ethics in Science and Environmental Politics*, 2005: 33-35, Retrieved June 21, 2006, from <http://www.int-res.com/articles/esep/2005/E65.pdf>
- Peritz, B. C. (1992). On the objectives of citation analysis: problems of theory and method. *Journal of the American Society for Information Science*, 43, 448-451.
- Rousseau, R. (1997). Sitations: an exploratory study. *Cybermetrics*. 1(1). Retrieved June 20, 2006, from <http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>
- Rousseau, R. (1999). Daily time series of common single word searches in AltaVista and NorthernLight. *Cybermetrics* 2/3(1). Retrieved June 20, 2006, from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>
- Small, H. (1982). Citation context analysis. *Progress in Communication Science*, 3, 287-310.
- Smith, L. C. (1981). Citation analysis. *Library Trends*, 30, 83-106.
- Snyder, H., & Rosenbaum, H. (1999). Can search engines be used as tools for Web-link analysis? A critical view. *Journal of Documentation*, 55(4), 375-384.
- Spink, A., Jansen, B.J., Blakely, C., & Koshman, S. (2006). A study of results overlap and uniqueness among major Web search engines. *Information Processing and Management* 42 (5), pp. 1379-1391.
- Spink, A., Jansen, B.J., & Pedersen, J. (2004). Searching for people on Web search engines. *Journal of Documentation*, 60 (3), pp. 266-278.
- Stokes, T. D. and Hartley, J. A. (1989). Coauthorship, social structure and influence within specialties. *Social Studies of Science*, 19, 101-125.
- Thelwall, M. (2003). Can Google's PageRank be used to find the most important academic Web pages? *Journal of Documentation*, 59(2), 11-20.
- Thelwall, M., Vaughan, L., & Bjerneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology*, 39, 81-133.
- Vaughan, L. and Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 40(4), 693-707.
- White, H.D., Lin, X., & Buzydlowski, J. (2001). Co-cited author maps as real-time interfaces for Web-based document retrieval in the humanities. Conference Abstracts, ACH/ALLC 2001. 2001 Joint International Conference of the Association for computers and the Humanities and the Association for Literary and linguistic Computing. June 13-17, New York City, NY (pp. 127-129).
- Zhao, D. and Logan, E. (2002). Citation analysis of scientific publications on the Web: A case study in XML research area. *Scientometrics*, 54, 449-472.
- Zhao, D. (2004). Web-based and print journal-based scholarly communication in the XML research field: a look at the intellectual structure. Proceedings of the American Society for Information Science and Technology 2004 Annual Meeting (pp 72-83), November 13-18, 2004, Providence, Rhode Island, USA



- Zhao, D. (2005a). Challenges of scholarly publications on the web to the evaluation of science – a comparison of author visibility on the web and in print journals. *Information Processing & Management*, 41(6), 1403-1418
- Zhao, D. (2005b). Scholarly Communication in Transition: Evidence for the Rise of a Two-Tier System. From Author to Reader: Challenges for the Digital Content Chain - Proceedings of the 9th ICCC International Conference on Electronic Publishing (pp. 77-82), June 8-10, 2005, Leuven-Heverlee, Belgium
- Zhao, D. (2006a). Towards all-author co-citation analysis. *Information Processing & Management*, 42, 1578-1591.
- Zhao, D. (2006b). Dispelling the myths behind straight citation counts. To appear in *Information Realities: Shaping the Digital Future for All - Proceedings of the American Society for Information Science and Technology 2006 Annual Meeting*, November 3 - 8, 2006, Austin, Texas, USA.