中国科技资源导刊 ISSN 1674-1544
2008年5月 第40卷第3期 32-37
CHINA SCIENCE & TECHNOLOGY RESOURCES REVIEW
ISSN1674-1544 Vol.40 No.3 32-37,May,2008

1998-2007年链接分析研究论文的计量分析

郑 曦 邓中华
(南京大学信息管理系,江苏南京 210093)

摘 要:本文将可视化方法引入到传统的文献计量分析中。综合使用文献计量学的基本方法、引文分析法以及共现分析法,对Web of Science数据库中1998年至2007年链接分析领域的研究论文,从年代分布、期刊分布、作者分布和主题关键词特征分布等角度进行了分析,构建出链接分析领域的关键词知识地图,并以此为依据,划分了链接分析的主要研究内容。

关键词:链接分析;共现分析;文献计量研究;定量分析;知识地图;可视化
中图分类号: G353.1 **文献标识码**: A DOI: 10.3772/j.issn.1674-1544.2008.03.006

Bibliometrical Analysis on Link Analysis from 1998 to 2007

Zheng Xi, Deng Zhonghua
(Department of Information Management, Nanjing University, Nanjing 210093)

Abstract: In this paper, method of visualization was imported to the traditional bibliometrical study. Based on the recent literatures on Link analysis collected in the database of Web of Science from 1998 to 2007, applying bibliometrical statistical methods, co-occurrence analysis and citation analysis, this paper makes a comprehensive analysis on the distribution of time, journals, authors, keywords and the characteristics of thesis, and establishes the science map of keywords.

Keywords: link analysis, co-occurrence analysis, bibliometrical study, quantitative analysis, map of science, visualization

链接分析成为目前国内外诸多学者研究的重点,也取得了令人瞩目的研究成果,本文检索了Web of Science数据库中1998-2007年链接分析领域的研究成果,并从文献的数量、期刊、著者、关键词以及主题词等方面进行了文献计量分析,以期为相关学者的研究提供参考。

1 数据来源与研究方法

1.1 数据来源

本文选择Web of Science数据库作为文献来源。Web of Science收录了世界上最有影响的经

第一作者简介:郑曦(1983-),女,江苏南京人,研究生,研究方向是信息资源管理。

收稿日期:2008年4月12日。

过同行专家评审的高质量期刊。该数据库每周更新保证原始数据的全面性和准确性。笔者使用 Web of Science 数据库[1]，设定关键词为“Topic = ("link analys*") OR ("hyperlink analys*"))”，共检索出 123 篇有效文献作为本文的分析数据（其中去除了“link analysis”作为关联分析解释的文章）。笔者发现最早的链接分析文章出现在 1998 年，因此，选取 1998～2007 年作为研究时间段。下面将对这些文献进行数量、期刊、著者、关键词以及主题方面的分析。

1.2 主要研究方法

本文综合使用文献计量方法、共现分析法、可视化方法对链接分析的研究成果进行文献计量分析。将链接分析领域的主要研究内容用可视化的图像直观地表示出来，形成链接分析领域的知识地图。知识地图的绘制主要使用 Pajek 软件。Pajek 是基于图论、网络分析以及可视化软件等发展而来的一种基于 Windows 的大型网络可视化绘制软件[2]。

1.3 数据处理

在检索到的 123 篇文献中，出现了同一作者不同署名以及同一关键词不同书写形式的现象，例如 HITS 算法的创始人 Kleinberg 有两种署名形式：“Kleinberg JM”以及“Kleinberg J”；关键词万维网出现了 3 种书写形式：“Worldwide Web”、“World Wide Web”、“Web”。针对上述情况，笔者对记录中的作者以及关键词项都做了书写形式上的统一化处理。

2 文献量与期刊分析

2.1 文献量分析

文献量是指某一领域的研究者在某一段时间内所发表论文数量的多少。一个研究领域的成长过程与该领域文献的数量和内容的构成有着密切的关系，研究论文的数量在一定程度上可以反映出该领域的研究水平和发展状况（表 1）。

由表 1 可知，在文献量上，1998－1999 年为论文的平缓增长期，2000－2005 年为快速增长期，2006－2007 年略有回落。用多项式五阶曲线对 10年内的文献量进行拟合（见图 1 中的文献量趋势线），得到方程 $y=0.0047x^5-0.1524x^4+1.6101x^3-6.5163x^2+11.171x-3.8667$，$R^2=0.98$，与原分布曲线较吻合。尽管文献量近两年内有所回落，但是链接分析领域的文献积累量在这 10 年里呈逐年上升趋势。文献累积量在时间序列上的曲线可以用多项式二阶曲线 $y=1.5379x^2-2.7833x+2.9$ 加以拟合，R^2 高达 0.9936，表明与原曲线非常吻合，可以作为文献积累量的预测曲线。积累量的二阶多项式拟合曲线说明链接分析领域的文献以较快的速度逐年递增，表明该领域的研究正处于蓬勃发展阶段。

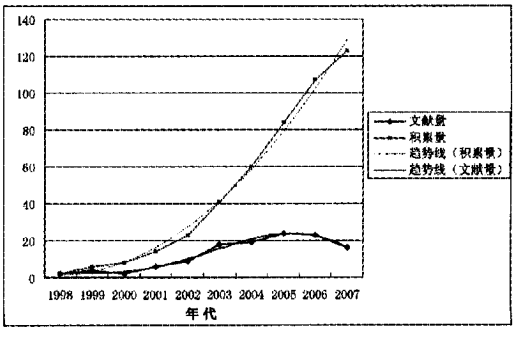

图 1 1998—2007 年 SCI 收录的链接分析文献变化趋势图

结合图 1 中文献量的分布曲线，可以将链接分析的研究分为 3 个阶段：第一阶段 1998－1999 年，为起步阶段；第二阶段 2000－2005 年，为高速发展阶段；第三阶段 2006－2007 年，发展速度

表 1 各年度文献量统计

年代	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
文献量	2	4	2	6	9	18	19	24	23	16
积累量	2	6	8	14	23	41	60	84	107	123

趋于平缓。随着科学技术的发展，网络信息开始进入学者们的研究范畴，在1998－2000年间的起步阶段，一些著名的网页排序算法被提出，如1998年提出的Pagerank算法，1999年提出的HITS算法。进入2000年后，网络技术飞速发展，网络信息海量增长，如何在网络环境下管理和搜索信息，网络环境中的链接和传统引文有何关联等问题成为人们日益关注的焦点。原有的HITS以及Pagerank算法已不能适应复杂的网络环境。众多学者们纷纷提出自己对于HITS、Pagerank算法的改进和优化思想。与此同时，1997年由Almind，T. 和Ingwersen，P. 提出的网络信息计量学也得到了众多学者的关注。基于此，2000－2005年的文献量进入了快速增长期。到了2006－2007年，文献增长速度趋于平缓。笔者认为，这一阶段文献增长率变小，并不意味着链接分析发展停滞，而是链接分析在取得了一定研究成果后进入了一个相对成熟的阶段。因此，我们可以认为链接分析领域的研究正面临新的突破，将产生出更新的发展方向。

从文献类型的角度看，在检索到的123篇文章中，研究论文所占比例最高有111篇，占总量的90.2%，会议文献9篇占7.3%，评述性文献3篇占2.4%。与其他学科相比，评述性文献所占比例要少得多。由此可知，链接分析是一项新兴学科研究领域，发展的时间不长，但是发展的潜力很大。

2.2 期刊分析

对研究论文的期刊分布情况进行分析，可以确定该领域的核心期刊。统计表明，近10年中，链接分析研究的123篇论文分别发表在78种学术期刊上，平均每种期刊发表1.58篇论文。

在本次分析中，收录3篇以上论文的期刊共有10种，具体情况见表2。从表2可以看出，虽然这10种期刊只占所有被收录期刊总数的12.8%，但是，它们收录了49篇文章，占总文献量的39.8%。这10种期刊基本构成了本学科领域核心期刊的锥形，是链接分析研究的重要信息源。

3 著者分析

3.1 发文3篇以上的著者分析

作者发文量统计如表3所示。从表3中可以看出，发文量在3篇以上的作者有14位，占作者总人数258人的5.42%。他们的发文量为38篇，占总文献数量123篇的30.9%。由此可见，这些作者构成了本领域的核心作者群。了解这些作者的科研课题和研究方向，有助于我们掌握该领域的研究重点和发展方向。

3.2 合著者分析

著者合作度是指在一定的论文集合中合著

表2 收录3篇以上论文的期刊统计

期刊刊名	载文量	国家
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	11篇	美国
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	5篇	美国
JOURNAL OF INFORMATION SCIENCE	5篇	英国
INFORMATION RETRIEVAL	5篇	荷兰
SCIENTOMETRICS	5篇	荷兰
ACM TRANSACTIONS ON INFORMATION SYSTEMS	4篇	美国
INFORMATION PROCESSING & MANAGEMENT	4篇	英国
IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS	4篇	日本
DECISION SUPPORT SYSTEMS	3篇	荷兰
INTELLIGENCE AND SECURITY INFORMATICS, PROCEEDINGS	3篇	德国

表3 作者发文量统计

发文量	第一作者	国别	著者所在机构
9	Thelwall M	英国	Wolverhampton Univ, Sch Comp & Informat Sci
5	Lempel R	以色列	Technion Israel Inst Technol, Dept Comp Sci/IBM Res Lab, Haifa, Israel
4	Ma WY	中国	Microsoft Res Asia, Beijing
4	Chen HC	美国	Univ Arizona, Dept Management Informat Syst
4	Menczer F	美国	Indiana Univ
4	Bar - Ilan J	以色列	Hebrew Univ Jerusalem, Sch Lib Archive & Informat Studies
3	Kleinberg JM	美国	Cornell Univ, Dept Comp Sci
3	Almpanidis G	希腊	Aristotle Univ Thessaloniki, Dept Informat
3	Kotropoulos C	希腊	Aristotle Univ Thessaloniki, Dept Informat
3	Vaughan, L	加拿大	Univ Western Ontario, Fac Informat & Media Studies
3	Gao YJ	加拿大	Univ Western Ontario, Fac Informat & Media Studies
3	Li XM	英国	Wolverhampton Univ, Sch Comp & Informat Technol
3	Vazirgiannis M	希腊	Athens Univ Econ & Business, Athens
3	Kitsuregawa M	日本	Univ Tokyo, Inst Ind Sci

表4 历年链接分析文献著者合作情况

年份	文章数	作者数	合作文章数	著者合作度	每篇文章平均著者数
1998	2	10	2	100%	5.00
1999	4	8	2	50%	2.00
2000	2	3	1	50.0%	1.50
2001	6	11	3	50.0%	1.83
2002	9	16	7	77.8%	1.78
2003	18	42	15	83.3%	2.33
2004	19	56	15	79.0%	2.95
2005	24	54	19	79.2%	2.25
2006	23	69	19	82.6%	3.00
2007	16	47	14	87.5%	2.94
总计	123	258	97	78.9%	2.10

文章与单作者论文的比例，这一概念可以反映一个领域内文献的合作写作的情况，进而可以反映出一个学科领域内研究的深入情况。

经过对1998－2007年链接分析文献的作者进行统计分析之后得出表4中的数据。从该表中我们可以看到，在链接分析研究的最初两年里，由于文献量较少，少数几篇文章的著者情况就可以影响到著者合作度的增减，因此，著者合作度不稳定。从2000年起，随着链接分析研究领域的不断扩大，技术的不断成熟，该项技术的研究人员有所增加，著者合作度基本上逐年递增。

从总体上看，检索到的123篇文章的作者共有258人，平均每篇文章的作者数为2.1。这与国际期刊的平均合著度相比还有一定的差距。10年里123篇文章的平均著者合作度为78.9%。由此可见，链接分析的大部分研究由多个研究人员合作完成，这些研究人员涉及了不同的学科和行业。这说明链接分析是一个跨学科的研究领域，需要综合不同学科的知识。同时，也说明了链接分析已经受到越来越多的学者的关注与研究，其应用研究领域也将会得到进一步的扩展与深入。

4 主题分析

4.1 关键词知识地图

笔者构建关键词知识地图，以空间的形式展

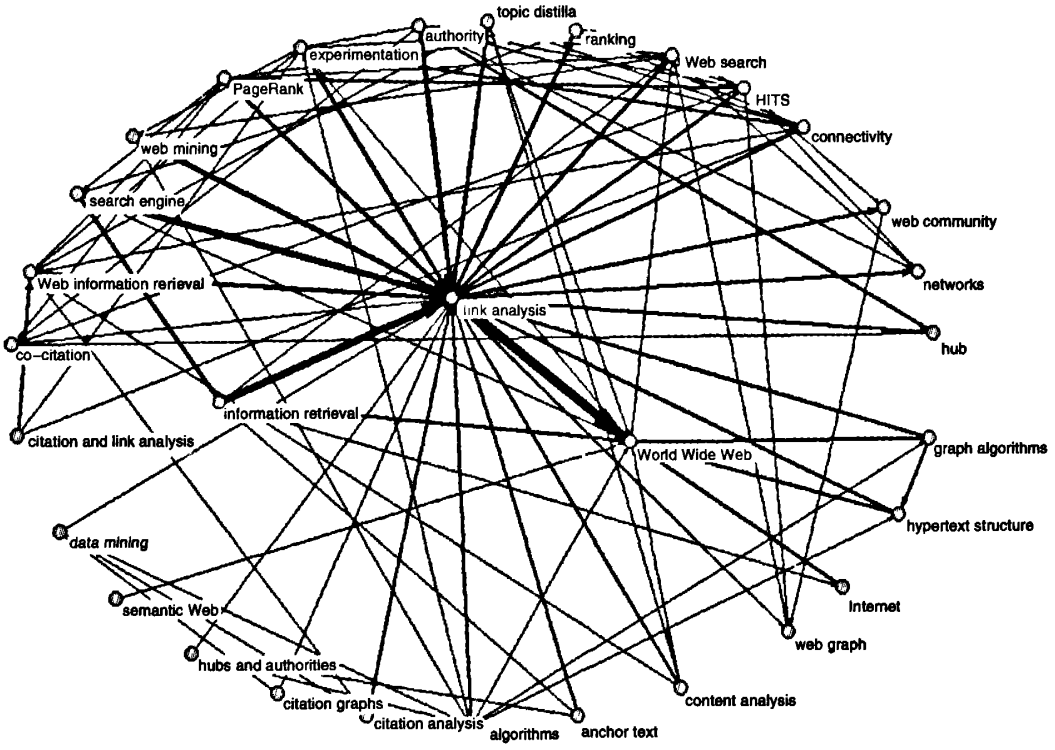

图2 链接分析关键词知识地图(1998－2007)

示关键词共同出现的频率，揭示链接分析领域的主要研究方向。首先筛选出现2次以上的关键词，其次自编程序从中萃取出共词对，然后使用Pajek绘制知识地图[3]。由于关键词构成的网络是连通网络，所以，笔者使用Fruchterman－Reingold spring embedder算法进行布局。该算法由Fruchterman和Reingold提出，将一个无向连通图看作是一个力学系统，其最终目的是要使力学系统达到平衡[4]。在构建出的关键词知识地图中，线条的粗细表示每一对关键词共同出现的次数，线条越粗，表示两组词共同出现的次数越多；箭头表示关键词在论文中的标识顺序，并不代表方向。

从图2中我们可以清楚地看出，共词对出现频率较多的是link analysis与information retrieval、search engine、web mining、pagerank、authority、connectivity、hub、HITS、ranking以及web search等。

4.2 主题分类分析

由于目前尚无这一领域分类的蓝本，本文根据上述的关键词知识地图，记录中的摘要项，同时参考一些著名学者的著作和相关学者的研究成果[5-7]，将链接分析的研究主题分为Web结构挖掘、网络信息检索、网络信息计量学三大类。

4.2.1 Web结构挖掘

链接分析在Web结构挖掘中的研究主要集中在网页排序算法的优化和改进。在网络信息的检索利用过程中，对搜索引擎结果进行相关度排序至关重要。基于Web文档内容的排序具有一定的局限性。大量的链接信息则提供了丰富的关于Web内容相关性、质量和结构方面的信息。因此，网络链接是Web结构挖掘的重要资源，Web页面的权威性可由Web页面的链接来反映。1998年Page提出了Pagerank算法，1999年Kleinberg提

出了基于 hub/authority 的 HITS 算法。此后，诸多学者使用实验法（experimentation）对这两大算法进行了对比分析以及改进和优化，排序算法由原先的静态排序逐渐向基于用户需求的动态排序算法演变。许多新算法被提出，如 EigenRumor 算法、localrank 算法、MFCRank 算法等。至 2007 年，共有 54 篇文献涉及此方面，占总文献量的 43.9%。

4.2.2 网络信息检索

在网络信息检索方面，链接分析主要被用于网页检索结果的聚类，判别网页社区（pages community），进行多媒体信息检索等，如日本学者 Wang, Y. T. 和 Kitsuregawa, M. 提出使用链接分析对搜索引擎检索结果进行聚类，以色列学者 Lempel, R. 和 Soffer, A. 提出的基于链接分析的网络图像检索系统 PicASHOW。这类文献共 26 篇，占总文献量的 21.1%。

4.2.3 网络信息计量学

链接分析在网络信息计量学方面的研究主要体现在 3 个方向：网络信息资源的评价，链接分析和引文分析的异同，网络环境下的引文分析。例如英国学者 Thelwall, M. 对 WIF（网络影响因子）的研究，加拿大学者 Vaughan, L. 等人对商业网站的共链分析，英国学者 Stuart, D. 和 Thelwall, M. 等人通过链接关系揭示英国大学科研合作关系，以色列学者 Bar - Ilan, J 利用链接评价以色列大学的网站等。该类文献共 22 篇，占总文献量的 17.9%。

5 结 语

网页排序算法优化方面的研究一直是链接分析研究的一个重要领域，链接分析在网络信息计量学方面的研究也在不断地增强并逐步走向成熟。链接分析的研究领域不仅仅局限于计算机或是图情领域，它跨越了计算机、图书情报、数学、工程学、社会学等诸多学科领域，是一个跨学科的新兴研究领域。与此同时，随着网络技术的发展和人们认识的深化，链接分析研究中也逐渐渗入其他学科的研究方法。作为一个新兴研究领域，链接分析有着巨大的发展潜力，今后将会引起越来越多学者的关注与研究，并得到进一步完善，发挥出更大的作用。

参考文献

[1] Web of Science. 1998 - 2007[DB/OL]. [2008 - 01 - 15]. http://apps.isiknowledge.com.

[2] Pajek[DB/OL]. [2008 - 03 - 08]. http://vlado.fmf.uni - lj.si/pub/networks/pajek.

[3] 李运景，侯汉清. 引文分析可视化研究[J]. 情报学报，2007,26(2):301 - 308.

[4] Fruchterman T M J, Reingold E M. Graph drawing by force - directed placement. [J] Software - Practice and Experience, 1991, 21(11): 1129 - 1164.

[5] 张洋，邱均平，文庭孝. 网络链接分析研究进展[J]. 图书情报知识，2004(6):3 - 8.

[6] 黄晓斌. 网络信息挖掘[M]. 北京：电子工业出版社，2005:67 - 73.

[7] 李江. 链接分析工具研究[D]. 武汉：武汉大学，2007:3 - 12.