

· 学术论文 ·

从文献计量学到网络计量学^{*}

邱均平 杨瑞仙 陶雯 李雪璐 译

(武汉大学信息管理学院, 武汉 430072)

【摘要】从1958年以来,不知不觉中文献计量学已经发生了根本的变化:它独立成为一个学科,在图书馆和信息科学学院中被广泛教授,并成为世界各地科学评价研究组的核心研究内容。这些都因为尤金·加菲尔德的研究成果和他的科学引文索引而变得可能。本文回顾了文献计量学从1958年到现在的发展历程,对比了早期的文献计量学和它的当前研究内容,综述了一大批最新的研究方向和进展。比如专利分析、国家科研评价实践、新型可视化技术、创新应用、新在线引文索引以及数字图书馆创建相关的发展过程等。还对网络计量学,一门现代迅速发展的文献计量学的分支进行详细的评述。最后,对文献计量学和网络计量学的未来发展趋势也进行了探讨。

【关键词】文献计量学;网络计量学;学术出版

【中图分类号】G350

Bibliometrics to Webometrics

QIU Junping YANG Ruixian TAO Wen LI Xuelu

【Abstract】Bibliometrics has changed out of all recognition since 1958: becoming established as a field, being taught widely in library and information science schools, and being at the core of a number of science evaluation research groups around the world. This was all made possible by the work of Eugene Garfield and his Science Citation Index. This article reviews the distance that bibliometrics has travelled since 1958 by comparing early bibliometrics with current practice, and by giving an overview of a range of recent developments, such as patent analysis, national research evaluation exercises, new visualisation techniques, new applications, new online citation indexes, and developments related to the creation of digital libraries. Webometrics, a modern, fast-growing offshoot of bibliometrics is reviewed in detail. Finally, future prospects are discussed with regard to both bibliometrics and webometrics.

【Key words】bibliometrics; webometrics; scholarly publishing

1 前言

在过去的五十年里,伴随着无数相对较次要的发展,在学术出版界发生了两大技术性的变革,在研究的计量分析方法上发生了两次巨大的改变。出版上的两大变化一是印刷过程(论文的提交,查阅和出版)的计算机化使得费用急剧降低,更多的期刊和书籍可以被出版;二是整个出版周期向互联网转变,从而允许更快,可能还更廉价的全程通讯。从历史上来看,定量分析学术出版(即文献计量学)发展中的第一个重大发展是科学信息机构创建引文数据库(ISI,现为汤姆森科技),在1962年开始运作,同时战后社会学通过这个引文数据库对其科学成果的影响

力进行了评价。从那以后,大学里计算机可用资源一直连续上升,使得很多计量研究成为可能。文献计量学的第二个重大发展是在急速扩大范围地研究相关文献的网上出版,从论文到电子讨论记录,产生了一大类新型针对它们获取和使用的计量学。

本文重点是科学计量。两大重要的变革分别发生在我们要考虑的时期的开始后和结束前,方便了这期专题的讨论,但在两次变革间文献计量学已经发展成为一门公认的科学领域:在大学作为信息科学的一部分课程进行教授,有着一系列实质性的方法,一些理论知识,以及一组国内外的科学评价专家。本综述主要关注两点:总述文献计量学和网络计量学领域(一个源自文献计量学的新的研究领

^{*} 原文出处:Mike Thelwall. Bibliometrics to Webometrics[J]. Journal of Information Science, 2007, 34(4).

Mike Thelwall, School of Computing and IT, Univeristy of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, U K; m.thelwall@wlv.ac.uk.

域)。第一部分探讨在 1958 - 2008 前半段诞生的文献计量学,第二部分选取了更多近期的发展,第三部分主要论述网络计量学。

2 文献计量学

文献计量学包括文献属性,和文献相关过程的计量。文献计量学的方法包括字词频数统计、引文分析、共词分析和简单的文献累积计量,例如单个作者、研究组或者是国家的总发行物数量。然而,在实践中,文献计量学主要应用于科学相关文献,所以与科学计量学有明显的交叉。

尽管熟知的文献计量学方法已经应用了至少一个世纪,文献计量学作为一个科学领域是由于尤金·加菲尔德创建的科学信息机构 (ISI) 的科学引文索引发展,是他一直努力支持科技文献搜索的合理结果。科学引文索引 (SCI) 是作者标记的参考文献数据库,在早期的文献中和顶级科技期刊中发表的论文,最开始都是集中在科学总论和基因学。科学家阅读了相关文献,知道哪些文献引自哪篇并从中获益,因为这些文献可能包含了相似的主题,可以据之更新或者修正原始文献,这个基本道理到今天依然适用。科学引文索引的重要性与 Bradford 的传播原则是一致的:尽管一个科学家可以通过阅读所有可见的相关期刊,对某一研究专业保持最新动态的了解,但是少部分相关文献依然通过一些非紧密相关的期刊进行传播。因此,引文检索帮助研究人员避免了在非核心期刊中相关文献遗漏的情况。

尽管引文检索几乎是科学引文索引的一个附属品,后来的社会科学引文索引 (SSCI) 和人文艺术类引文索引 (A&HCI) 还是轻易地造就了一类新的统计学:不只对给出论文的引文数量进行统计,而且通过 SCI 其他部分的使用,对总的出版物和引文数量进行统计。总的统计数据包括一种期刊中所有论文的引文总数或者是单个作者、研究组或者国家所有论文的引文的总数。其中一些统计量通过相关理论支持,进一步发展成为指标并有了合理的被人们广泛接受的标准阐释。最有名的可能就是期刊影响因子。

自从科学引文索引发行以来,出现了两种文献计量学的应用类型:评价类和关联类。评价类文献计量学致力于评价学术成果的影响力,通常是比较两个或多个个人或组织的相关的科学成果。这些评价结果有时被用来制定研究政策和帮助确定研究基金的实施方向。相对地,关联计量学致力于揭示研

究中的一些关联,比如研究领域的认知结构,新的科研前沿领域,或国内外作者合著方式等。

2.1 评价文献计量学

大部分的评价文献计量学技术采用引文数据作为其原始数据。理论基础源于 Robert Merton 的社会科学,他假设引用量是学者了解重要前人成果的途径,在这个基础上,引文量的计算就能被用来作为评价科学价值的一个指标,因为越是重要的成果被引用就会越多。实际上,“impact”这个词现在被认为是适合以引文量作为指标来计量和表示的。后来的研究表明 Merton 的观点是把现实情况的一种简单化:引用文献有很多不同的原因,从很多可用的文献里选择哪些文献也有很大的影响因素。从另一个角度来说,De Solla Price 发现对高被引论文来说,可能存在累积优势,高被引的论文后来再被引用可能不是因为它们的内在价值,而是因为它们被引用过而已。这与 Merton 在科学上的“Matthew 效应”不谋而合,有声望的学者会获得与研究成果不合比例的荣誉。尽管存在上述复杂的情况,基于引文量的统计还是被广泛采用。

期刊影响因子,在 1960 年早期被引入,是科技信息机构 (ISI) 中建立索引期刊的文献在出版后的第 X 年到第 X-1 和第 X-2 年两年的被引总数除以期刊在第 X-1 和第 X-2 年引文总数。在 Merton 的基础上,有着较高期刊影响因子的期刊趋向于出版较高影响力的研究成果,因此也更趋向获得好评。然而,似乎大家一致认同,即使是在比较离散的学科中,基于期刊影响因子评价的期刊排名还是有问题的。此外,基于期刊影响因子的价值性,期刊编辑似乎都倾向于推荐作者多引用本期刊中的文章,以提高期刊的影响因子。

第二个广泛的应用是关于任职和升职的决策上考虑,学术成果发表的期刊影响因子或者是出版物的被引用数量。然而,这并不为很多文献计量学家所推荐,因为个人作者水平的引文量计算是不可靠的,还有那些作决策的人可能忽略了不同学科间的差异性。

第三个应用是有文献计量学专家通过出版物的引文量来比较学术院系的论文的学术水平。即使是合理的基于整个学术院系出版物的总计,并且很谨慎地构建的文献计量学指标,也仍然需要结合一些证据资源(如科研经费,来源的可信度,同行评议,陈述)来对重要决策比如那些经费相关的决策给出依据。

2.2 关联文献计量学

虽然早期有很多从 ISI 的数据中检测科学中相联系的文献计量学方法的探索尝试,但是可能由于早期计算机能源的匮乏,特别是可视化,关联分析方法的发展受到阻碍。尽管如此,通过简单的方法,早期关联分析对科学的结构产生了一些有趣的见解,比如一些主要文献的引文数的网络波动图。这个方法显然是基因学家 Gorfdon Allen 教授发明的,他把引文量图表交给 Garfield。期刊引文图表是早期的另一个发明:它可以揭示同一学科领域内期刊间的联系,发现那些交叉学科的期刊以及分辨核心和非核心期刊。

有时归功于 Garfield,共引可以被作为是测度相似性的一种重要的关联测度法。测度的基础是经常出现同一参考文献的文章之间在某一方面可能很相似。这意味着如果文献的收集依据它们共引数的计算,那么就应该产生一种反应认知科学联系的方法。作者共引分析(ACA)是一种相似性测度技术,在于从作者的文献共同被引的频数来测度两两作者间的相似性。ACA 从一个足够高的累计角度成为描绘学科结构的实用工具。

3 文献计量学现状

随着网络和网络相关技术的发展,文献计量学主流与其说经历了革命性的巨变,不如说在正常的演变。虽然基于核心引文量的影响测度仍在应用,但如今已被一类其他的技术所替代。此外,因为有一体系理论和案例分析可供利用,所以实践文献计量学可以合理地断定:可以找到很好的方法从引文量中创建更好的指标,解决任何常见的问题,同时知道怎样去解释结果。特别地是,最近关于使用引文量来测度影响力的争论在热烈进行中,同时引用动机理论的发展最近也被广泛的综合讨论。

除了核心引文量分析方法,文献计量学中最大的改变来源于新的关于学术交流的信息源的可用性,比如专利、网页、数字图书馆的使用统计。当然,文献计量学从来就没有对学术给予特别的关注,同时还使用其他数据作为质量指标,例如经费、质量指标,如同行评议意见。

在文献计量学的最近发展中,尤其是在引文分析中,可能有三大主要趋势。分别是通过高级的计量学方法和谨慎的数据整理,提高测量结果的质量;对新的任务发掘不同的计量方法;将文献计量学应用于更广泛的问题中,特别是描述文献关联性(可以

从后面章节知识领域的可视化中为后者找到例子)。

3.1 h 指数和现在的文献计量学指标

最重要最新的评价计量方法可能要数 h 指数了,对一个科学家来说,h 值的最大化是指他或她至少有 h 份出版物被引用了至少 h 次。较高的 h 值显示了一位科学家发行了很可观的被高度引用的出版物。这种方法之所以产生是因为易于计量并且理解直观。已有很多关于 h 指数的研究,评价,提出经过修正意见或者将其应用于一系列学者的评价。比如英国的 h 指数排名表和美国的图书情报科学教授排名表都引起了该领域的研究者的极大兴趣。

除了 h 指数,最重要的评价文献计量学指标似乎也在逐步地演变。例如,因为不同领域有基于其不同的广泛引文标准形式,因此在不同学科领域间的引文量的比较是不合适的。所以,当使用引文量来评价学术院系时,最好的方法是采用该学科标准化后的引文指标。即使在同一学科的一系列院系的比较,采用原始引文量累计值或者是每位学者的平均引文量都不能准确反映他们的引文影响力,因为每个院系专深领域可能有所不同,因此也有着不同的平均引文率。所以,除非指标被标准化,例如将每个院系的引文量除以那个研究领域的平均值,不然在学术研究领域有着较高平均引文累计数的院系取得的优势结果是不公平的。因此,评价引文分析的目标已经从研究影响力评价向相对于一个领域的研究影响力的评价方向转变。

3.2 国家科研评价实践

系统研究评价实践对于文献计量学来说似乎是重要但又仍然很有争议的应用领域。如今已有四个国家拥有周期性的国家科研评价实践报告用来决定很大一部分他们科研经费的方向。英国的研究评价实践(RAE,见 www.rae.ac.uk)是首次实施的,分别在 1986、1989、1992、1996 以及 2001 进行,2008 年也将实施。这主要是基于同行评议,挑选每个大学授予级别的相关专家组成一组专科专家。除了同行评议,通常还基于四大顶级出版物的每个研究者意见,这个小组将其他的因素也列入考虑,比如经费、博士完成情况,以及论述(narrative)。尽管文献计量学还没有担任正式的角色,但其可以包括提交时的论述部分,被专家作为决策过程的一部分。有趣的事实证据表明很多学科也发展了不正式的出版物索引,其中列举的目标杂志由期刊影响因子决定。2008 年实施的研究评估实践将结合文献计量学方法一起实施,然而,后来的研究评估实践将更多地融

入文献计量学因素,但基于学术出版物本身已经经过同行评议而有争议。向文献计量学方向的转变的理由是文献计量学指标,比如引文量和期刊影响因子更加明确与同行评议相比也较便宜。然而,从本身的角度来说文献计量学还是不够合适的,所以系统研究评价实践很可能还得一直纳入其他的因素,很可能要纳入同行评议作为最后的评价标准。

新西兰的基于研究基金的绩效(Performance Based Research Fund)始于2003年,在2006年又部分开始运作,预计在2012年将又开始运作。除了一直对学术界的个人进行评价而不是针对整体的提交过程外,其他都与RAE极为相似。2008年英国的国家科研评价实践正在向个人评价转变而不是以组为单位,但是在新西兰体系中是每位学者都有一个个人证据组合文件夹,而不是所有出版物的简单组合,尽管每位学者的组合文件夹中的核心组成因素是大概4篇提名的研究成果。

在澳大利亚,机构补贴计划(IGS),完全替代了早期的研究量化评价(Research Quantum),是一种有效的国家评价实践,澳大利亚的评价从来没有包括一向重要的同行评议因素,但其一直主要基于外部基金用于研究。从2002年开始,该经费主要基于“对研究生的吸引程度(经费的30%),对其他研究经费的吸引程度(60%)和出版的研究成果的质量和数量(10%)”。

世界科学系统最大的国家,美国,并没有国家科研评价实践。相反的,美国的研究经费是基于一个个项目来竞争分配的,事前的评价作为次要重要的,由相关的基金机构执行。Netherlands的研究评价过程是一种不同的尝试,包括不同基金机构和利益相关者的组合。还有不同是先前对研究发展基金会的评价系统(FRD),如今已经被在南非的国家研究基金会(NRF)所代替,在授予基金补贴之前,结合了新西兰以前第二阶段个人研究者的评价过程(研究者未来的计划)。目前的NRF系统很相似,前期通过同行评议评价个人研究者,通过评定的研究者才能够申请基金补助。最后,撇开上述陈述不说,大多数国家采用的是所谓的自主评价而不是系统实践,意大利就是这样。

那么研究评价系统有多有效和实用呢?很明显这是个很有争议的问题,并且没有一个简单的答案。然而在澳大利亚,有证据对一个简单的策略(评价方法)提出了异议:出版物的简单累计而不评价其质量的消极影响导致了质量较低的期刊论文数量的增

多。英国的RAE基于英国的绩效和世界相关的文献计量学的证据取得了成功,但是它的评价是否为大多数人所相信还有待考证。

3.3 新文献计量数据库:Google Scholar 和 Scopus

1992年,科学信息机构(ISI)被Garfield和其他合伙人卖给一个公司,后来成为Thomson Scientific公司,延续了引文索引。近些年来,ISI索引由于大规模的网上学术论文数据库的出现而面临重大的挑战,如Google Scholar和Scopus(Elsevier)也包含引文信息。此外,还有其他较小规模的专科数字图书馆和包含引文索引的档案,比如计算机科学的CiteSeer和首创CiteBase对网上免费学术出版物建立引文索引,包括(arXiv)网站(物理、数学、计算机科学和生物计量学)。

有一篇论文通过完成一间图书信息科学学院的师资评定等级任务中,以评价基于不同数据来源引文分析规模和程度为明确目标,比较了Web of Science(使用ISI数据),Google Scholar和Scopus。结果表明,大规模的引文分析采用Google Scholar是很困难的,其他两个总的来说结果相似。然而,对某些领域的覆盖面的缺陷在于一些师资人员严重的不足,因为其使用的数据库的选择。所以,将两者结合起来互促互成可以得到最公正的结果。此外,Web of Science、Scopus与Google Scholar相比,会议覆盖面低表明了它们都不能给注重在会议上发文的学者以公正的结果,比如计算机科学、计算机语言学。另一调查比较了不同数据库社会科学研究覆盖面,发现Scopus提供了较好的覆盖范围。其他一些对更多学科覆盖面的研究也表明Google Scholar的覆盖是不稳定的,对一些学科来说覆盖面很低或者特别的不可靠。

3.4 知识领域可视化

对成熟可视化的大量使用可能是关联文献计量学最重大的发展,并导致了一门新领域的创建:在信息可视化研究领域中的知识领域可视化。这包括了重要的计算资源,是更广泛的“电子研究”利用计算机达到社会科学研究目标趋势的一部分。除了Chen对个人研究领域的三维丰富信息可视化研究外,其他人通过ISI数据库中的引文量实施了描绘大型科学领域的雄伟计划。

早期的关联文献计量学的研究可能只是手工描绘作者,期刊或者论文的引文量图表,后来的学者开发了软件来自动完成这一过程。例如,Olle Persson的Bibexcel可以导入ISI的引文数据,然后得出一

系列二维表格,比如一个由研究者们对一位既定作者的最强的引用关系组成的自我网络(ego network)(这个术语是从社会网络分析中借用的)。与之相似的,Loet Leydesdorff 有一系列的软件可以将 ISI 的数据转换成一种可以生成图表的格式,特别是可以用来说明各个期刊间引用关系。

有很多系列可视化软件由于计算机学家的意义重大的介入可以免费使用,并且可以很容易地处理 ISI 数据,得出三维可视图。Katy Borner 的 Info Viz Cyberinfrastructure 是一套通用目的的开放资源软件,采用了很多算法进行处理和表示数据。一个很特别的优势是其处理大量数据的能力。例如,Boyack 通过文献计量学结合超过一百万在科学引文索引中的论文,得出了科学与技术结果的图例。Chaomei Chen 的 Citepace 软件特别注重于文献计量学研究,能够产生引文网络的漂亮的三维可视图。一些 Chen 的网络图中有趣的特点是它不仅仅包括最基本的结构而且还包括让研究者通过颜色和其它特性的使用,来表达其他层面的额外信息的功能。例如,表示一篇论文在一个论文组成的网络中的一个圆圈,可以通过图上不同的颜色层表示论文每年的引用数量。

3.5 专利

专利是通常由政府专利局颁发的,有一定时限的对一项发明的专有的一系列权利。专利这个术语也可以用做官方注册发明的描述。这些文件与学术论文在某些方面很相似,比如也包括了一系列的参考文献。基于有时学术研究者可能会直接投身于有用技术开发中的共识,专利的价值被作为科学价值的指标。

为了奖励那些有着潜在商业价值的创新研究研究者被授予专利权,从计算研究者的专利数的意义上来说,专利指标是可以很直接的。但通过专利的参考资料来识别被引用的学术成果,然后被认定为有应用价值,这样专利指标就不直接了。专利分析同样被用来评价一个国家技术发展的表现以及识别科学和技术之间的知识转变流向。例如:有关荷兰的专利研究实例表明结果并没有与现今的大学和工业的关系的理论模型相吻合,因此还有待重新考证。

3.6 数字图书馆的使用数据

从长远来看,文献计量学面临最重大的挑战可能是第一次由数字图书馆产生的大规模的学术论文使用情况的数据证明。在一些情况下编辑已经收到除出版商那里获得的影响因子之外的其他使用统计

数据,似乎很可能这两类数据又可以给出互补的有用信息。在一些情况中对日志文件进行研究可以将使用情况和用户人口统计学结合起来,得出用户和信息检索方式的一些其他见解。

有两个重要的问题是关于开放存取出版对文献可视化的影响和出版商的利益的影响。在一篇研究 ArXiv 中的数学论文阐明了上述两个问题,并指出开放存取的论文经常更容易被引用,但是原因是更多被引用的文章更趋向于存储在 ArXiv 中,而不是因为存储在那里而必然吸引了更多的引用。此外,还有一些证据表明后来发表的开放存取论文从出版商网站的下载次数相对减少(减少了 23%)。

数字图书馆使用数据能够与引文量计量相关联。例如,早期的读者数据对于一篇文献的未来预计引文量计量有一定的帮助。然而,也有很多文献的引文量和使用统计数据相差甚远。这提高了使用数据可能被用作一种新的影响力依据的可能性。例如,在一些学科中将科研用于本科生教学中可以被认为很有用,采用统计学数据进行评价比引文量更有价值。也许在将来我们同时有着“经典引文量”和“经典使用性”指标。物理学或者信息科学中,出版商通常将网上下载最多的文献列表分发到期刊编辑栏中,但是缺乏使用性统计数据的标准化阻止了通用列表的建立。而且,目前可用的使用性统计数据并不是没有问题的:例如出版商已经注意到一些个人的文献很容易获得高的使用概率是因为被一个大班的老师推荐作为必读文章。

4 网络计量学

网络计量学是一门借鉴信息计量学相关方法,定量地分析各种网络信息现象的学科,它所研究的问题通常与文献计量学属同一范畴。当网络成为海量信息,尤其是学术信息的存储库时,网络计量学便应运而生。与传统计量学不同的是,网络计量学的各项指标数据可以通过商业搜索引擎获得,这也为人们的研究活动提供了便利。事实上,几家大型搜索引擎提供将指标数据自动传送给研究者的电脑程序的服务,这使得大规模研究成为可能。人们最熟悉的一项网络计量学成果就是基于校园网站分析和在线影响力分析(Online Impact)的世界大学排名。

网络计量学的研究方法包括链接分析、网上引文分析(Web Citation Analysis)、搜索引擎评价(Search Engine Evaluation)和纯描述性网络研究(Purely Descriptive Studies of the Web)。除了最

近兴起的有关 web 2.0 的分析,下文将对上述方法的应用进行回顾。此外,本文对有关网站建设(如网站的易用性、网站内容等)的计量学方法也不予介绍。

4.1 链接分析

链接分析是研究网页间超链接关系的定量方法。在 Ingwersen 借鉴 JIF(Journal Impact Factor)的原理提出 WIF(Web Impact Factor)以及人们接受了超链接相当于计量学中的引文关系这一假设后,文献计量学引入了链接。WIF 用来计算某一网络空间(如网站、网络社区等)内部网页的平均外来链接数。早期的链接分析是基于这样的假设:指向某个学术网站的链接数可能与该网站所有者(大学、某个部门、学术组织或个体研究者)的学术生产力相对应。因为总体上,高产的研究员越多,则网站的内容越多,尽管这些网页的内容质量并不能提高网页的平均链接数。然而除非进行大量的研究,否则发现两者间的变化模式并不容易,因为它们之间不存在直接的因果关系。例如,有些作者的研究成果具有较高的链接数,而其他作者的同水平的脱机成果在网上的关注度却偏小。

后来的超链接分析引入了新的计量学方法,如改良的计数方法(如 Alternative Document Models)其应用范围也有所扩大。这时的链接分析研究主要集中在方法改进和案例分析。由于链接产生的原因多种多样,并且与引文关系不同的是链接关系在任何一门学科中都不处于中心地位,导致超链接分析很少用于评价研究。但相对于引文分析和专利分析,链接分析能更好地描述学科领域内科研组织的发展和各组织间的关系。这种方法对不同情景下的网络利用情况的研究,如不同学科机构的利用情况,也很有价值。

然而,所有链接分析方法都面临这样一个问题:网络总是处在不断变化和扩张的状态,因此网络计量的学科发现可能很快被淘汰。澳大利亚、新西兰和英国都曾对大学网站展开时间跨度研究,其成果的有效期都证明了这一点。这些大学网站经过若干的快速增长,直至 2001 年其规模才趋于稳定。然而通过各年份各网站间的链接数的比较,研究人员发现网站规模上的稳定掩盖了链接上的变化,但也同时证明有代表性的定量研究结果可以有较长的有效期。

4.2 网络引文分析

相当数量的网络计量的研究重点并不是网站,而是利用网络期刊文章被引频数来研究网络学术出版物。它的基本原理在一定程度上为 ISI 数据的用

途提供了新的视角,也为扩大网络研究的范围(包括非正式的学术交流、商业应用等)提供了可能。大量的研究证明,虽然网络引文的数量庞大,网络引文分析的结果与 ISI 引文数据产生的结果却在众多学科中表现出十分紧密的联系。然而,许多网络引文,例如引用本期刊的文章而不是外部文章,对研究都是无意义的。如果可以实现自被引和外部引用的自动区分,网络引文分析可能代替 ISI 引文指标。

4.3 搜索引擎

大量研究是利用网络计量学通过商业搜索引擎完成,研究主题主要集中在两个方面即搜索引擎的查全率和查准率,搜索引擎的算法和信息检索过程则不属于网络信息计量的研究范围。关注搜索引擎评价的研究人员有两类:一是利用搜索引擎收集数据的学者,另一类是要了解搜索结果含义的学者。

搜索引擎很早以前就成为用户通向网络的门户,因此它的覆盖面理所当然地成为人们关注的焦点。1999 年,研究人员对当时主要的搜索引擎调查后发现,所有的被调查对象的覆盖率都不大于可索引网络的 17.5%,这里的“可索引网络”是指在理想状况下搜索引擎可以找到的所有网页集合。理想状况是指搜索引擎不但可以找到某个网站内的所有网页,还可以跟踪网页上的链接,找到网站外的其他网页。此外,搜索引擎之间的重复结果也惊人得少。1999 年以后,类似的数据就没有出现,这是因为:(1)模糊超文本传输协议技术和虚拟服务器技术的产生使劳伦斯和盖尔的取样方法(the Sampling Method of Lawrence and Giles)不再有效。(2)动态网页的出现使计算“网页总数”成为不可能的事情。(3)鉴于搜索引擎的查全率不可能达到 100%,所以若非查全率有很大的变化,精确的查全率是没有意义的。但这项调查很好地证明利用元搜索引擎可以收集到更多的结果。然而,Google 在搜索引擎领域已占有绝对份额,有人认为这是因为 Google 完成了搜索引擎一项最重要的工作,即将最相关的结果排在首页,而不是分散在各页中。

搜索引擎不可能绝对客观,那么其覆盖网页的数量会不会在某些重大原因作用下发生变化呢?这个问题之所以重要是因为在新时代网络经济中,作为网民和网上信息中介者的搜索引擎能够带来巨大的商业利益。事实上,越早运用网络的国家搜索引擎返回结果的查全率就越高。这是由搜索引擎收集网页方式产生的副作用,而并非由检索策略引起。

搜索引擎的查准问题涉及到很多方面。Bar-

Ilan和 Peritz 认为搜索引擎自身不会将相同的搜索结果返回给用户。他们对 Google 进行了跨年分析,在 Google 中检索语句“Information OR Informetrics”,结果发现返回的结果只是其数据库的一部分。虽然相互重复的网页被省略,但这样终究会造成结果的丢失。一项关于 Microsoft Live Search 的相关性分析发现信息丢失的一个原因可能是 Microsoft Live Search 的政策规定每个网站最多返回两个网页。

许多网络计量研究已经开始使用由搜索引擎在结果页面上提供的估计结果数(如,数据“50 000”来自“约 50 000 结果中的 1~10 个”)而不是使用符合匹配的 URL 列表。例如,Ingwersen 在研究几组国家间的超链接数时使用的就是这类估计数据。但这类数据往往不可靠,甚至有出入,例如,扩大检索式的范围反而得到更少的结果。在网络计量学产生初期,这类估计值非常不稳定,尽管后来情况有很大的改善,但还是有人建议采用技巧来减少这种出入。

最近,一项 Live Search 估计结果数的准确性的研究结果令人惊讶。它对结果较大(>8 000)和较小(<300)的估计较稳定,而对结果在两个数字之间的估计却不稳定。这可能是由于较大的估计值是搜索引擎根据匹配出的结果总数制定的,而较小的估计值则是搜索引擎在去除匹配结果中的相同、相似网页以及同样的多余网页后得到的结果。中等大小估计值之所以不稳定是估计方法没有完全转换的结果。对于搜索结果数量偏大或偏小的网络计量研究来说,估算原理的不同是摆在它们面前的一道难题。

4.4 网络描述

基于网络的重要性,一些学者也展开了对网络的纯描述性分析。至今已经出现了基于各种调查方法的统计报告,包括:网页平均大小、平均元标记数、元标记种类以及技术的平均使用情况(如 Java、JavaScript)。此外许多商业性网络情报公司也公布了诸如使用人数、网页数、服务器数等基础统计结果。然而,公布于众的描述性分析报告只有两类,链接结构分析和时间跨度研究。

关于网络链接已有两项重大发现,分别与网络整体结构和链接的发展规律有关。AltaVista 的专家将抓取的网页复制,构建出网络链接结构的整体图景。他们发现了“蝴蝶结”模型(如图 1 所示)。位于模型中心的通过一次或多次链接就可以与彼此相连的网页集,被称为“强连接部分”(Strongly Connected Component, SCC),占全部网页的 28%。它相当于整个网络的“心脏”,在其中定位相对容易,而且

包含链接较好门户网站,如 Yahoo! Dictionary、Dmoz Open Source Directory 等。此外,从“SCC”中的网页出发,经过一次或多次链接可以到达另外 21%的网页,但无法按原链返回“SCC”,它们被称为“OUT”部分。这部分包含了许多由 Yahoo! 或其他“SCC”网页链接到的网站,但这些网站没有指向外部的链接。与“OUT”部分相对,“IN”部分则是那些直接或间接与“SCC”部分相连的网页,但却不为“SCC”所知的网站,它约占全部网页的 21%。最后,有一些组织不与任何外部网站连,约占全部网页的 8%(DISCONNECTED)。还有较大一部分网页是有很多外部链接的网页,占全部的 28%(TENDRILS)。

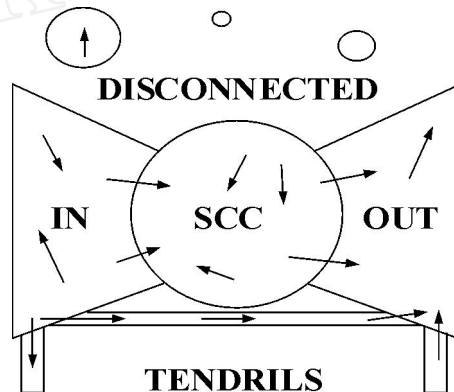


图 1 网络的“蝴蝶结”模型

后来为了强调“SCC”部分的中心作用,“IN”与“SCC”以及“OUT”和“SCC”的紧密联系,Bjorneborn 对“蝴蝶结”模式做了进一步的修改,如一个网站的主页位于“SCC”部分,而其内部的网页则可能位于“OUT”部分。Bjorneborn 还研究了英国大学子网站间的最短链接路径,结果发现计算机科学的网站常常起到连接互不相关学科网站的作用。

与上述的结构分析相对,网络动态分析涉及网络变化的测度、描述和建模。它的一个重大发现就是文献计量学(或其他领域)中的“马太效应”在网络链接中的体现。在网络中,一些网页吸引了千百万的链接,然而也有千百的网页的链接数仅为一个或没有。这种不均衡现象的产生可以归因于:人们往往倾向于将链接指向已有很多链接的网页。当然,没有人会在决定链接哪个网页前数一数各个网页的链接数,但是搜索引擎为人们提供了答案。人们可以通过搜索引擎知道某个网页链接数的多少,因为搜索引擎是根据链接数查找网页,并对它们排序。因此,网页得到的外部链接数越多,则其网络可见性越大。

通过对网页链接分布状况的研究,人们发现“马太效应”适用于某些类型的网页如公司网站主业,而

并不适用如大学网站主页。出现不适用的原因极有可能是第二因素在起作用如对网页的知识准备。两个案例都证明了一点,如果一个网页没能吸引一个链接,那么它就不大可能再吸引到链接了。因此为了吸引更多的链接,网站的建设者应当尽量寻找若干个由外部指向自身的链接。然而在网络计量学中,将网页的链接数做为评价网页内容的指标并不可靠,因为网页较高的链接数可能是由于其过去某一时间有较高的可见度。

最后,动态网络分析的另一类型就是在线信息变化分析。Koehler 从 1996 年开始跟踪一组网页,他发现这些网页内的信息起初有规律地被更换,而随后即时的网页不再会被更换,在一段时间内呈现稳定状态。Koehler 还表示由于专攻不同,不同类型网站网页的保存率会有所不同。随后, Koehler 在 1997 年和 2004 年对 738 个网站进行了调查,发现这期间网站的规模增长了 7 倍,网站某方面的内容也有很大增加如动态网页、站内链接,与之相对应网站的外部链接的增长速度则较慢。此外,研究还发现 1997 年外部链接中,仅有 25 % 延用到了 2004 年。

总而言之,网络是一个复杂的、不断变化的实体,尽管它并不规律,但通过大规模的研究人们仍然可以总结出其明显的变化模式。

4.5 测度 Web 2.0

Web 2.0 是由出版商 Tim O'Reilly 创造的词汇,主要是指面向网民的网站如博客、网络社区,网民在网站上公布信息是它的存在形式。在网民公布信息的大幅增长的背景下,市场智能产业 (market intelligence industry) 得以产生,许多人开始致力于 Web 2.0 测度研究。它们都是基于数据挖掘:因为太多网民在网上通过各种渠道如博客、聊天室、电子公告板以及网络社区等,发布了非正式的观点,那么理论上,从这些信息中发现一些行为模式是有可能的,例如网民对产品或社会事件的反应。为了解决类似问题,一些大公司开发了新的软件,如 IBM 的 Web Fountain,微软的 Pulse。同时,也有人创建或将原公司改组成专业网络智能公司,如 Nielsen BuzzMetrics 和 Market Sentinel。

利用“消费者自组媒体”的一个典型例子就是基于博客的讨论来预测书籍出售模式。然而,这一尝试并不很成功,可能是由于人们常常是在读过书之后才在博客上进行讨论,此时作预测显然太晚了。也有一些商业目的较小的研究项目,如 Gruhl 等人选择分析了博客空间中的几个主题,发现了几个不

同的模式。例如,一些话题只能维持很短一段时间,而有些却可以持续很久,这期间有可能产生或不产生新的话题。一项社会科学性质的研究尝试通过博客和新闻讨论区对大事件的讨论构建出这些事件的追溯时间表 (retrospective timelines),事实证明实现这个目标的可能较小。但相对次要的话题并不受关注,而将它们从时间表中去除掉,那么所有事件的重要方面可能被忽略。

除了利用数据挖掘进行的研究,还有许多旨在描述 Web 2.0 内容和解释 Web 2.0 用户行为的研究。这里将回顾网络社区的研究。早期对 Facebook 的大量研究全面阐明了它的用户行为特征。研究采用 2004 年 2 月至 2006 年 3 月 Facebook 的数据,这段期间 Facebook 是仅为美国大学学生服务的网络社区。研究发现用户使用 Facebook 的时间模式与学习时使用电脑的模式相符,而并非在分散的时间段使用。此外,用户在 Facebook 中联系的朋友多为同学校的校友,并非身在外地大学的朋友。这说明网络社区是现实社会中人际交流的延伸,尽管在技术上 Facebook 完全可以支持新人际关系的产生,它并不会根本上激发新的人际关系,对著名的 MySpace 的定量研究也证明了这一结论。

一项以 MySpace 为研究对象,旨在探索用户特征,并非用户行为的网络计量研究发现大约三分之一的用户访问 MySpace 的周期是一周,他们的平均年龄在 21 岁左右。尽管有研究发现 MySpace 上的好友关系是现实生活中朋友关系的反映,但 MySpace 中男性和女性的朋友中都以女性居多。还有研究发现 MySpace 中的好友在现实中住处的距离大都在 100 米内,仅有少数住在同一城镇中。

市场调研公司已经公布了许多有关 Web 2.0 的统计数据,尽管这些数据的出处不能确定,但有些分析结果也看似合理,由于获取使用权的费用太高,学者们也不大可能利用这些数据。如 HitWise 宣布至 2007 年 12 月, MySpace 已取代 Google 成为全美访问数最大的网站。据报道这些数据来自美国 2 000 000 网络用户,为此 HitWise 和用户的网络服务供应商签署了协议。公布这一分析结果不但提高了 HitWise 的知名度,同时也使网络研究人员对网络有了进一步了解。

5 结论与展望

5.1 文献计量学

文献计量学自 1958 年开始演变,那时文献计量

学还未建成为一门学科,甚至文献计量学方面还没有出现研究群体。而如今,文献计量学成为各图书情报学院的教授课目,并成为世界各地众多科学评价团体采用的核心学科,如荷兰的科学技术研究中心(the Centre for Science and Technology Studies)。以期刊影响为代表的众多文献计量学指标已被广为应用,许多国家都借助文献计量学进行有关政府资助科研项目的重大决策。同时,文献计量学指标也在不断发展中,大部分应用广泛、容易计算的文献计量学指标仍存在很大的缺陷。虽然专家们非常了解这些缺陷,但它们却可能被非专业人士忽略。因此文献计量学专家的一项重要工作就是劝说政策制定者使用高质量的指标,并保证没有按面值(at face value)使用指标。

随着信息来源的增长,可用的信息资源的增多,文献计量学发生了变化。如今,Scopus 和 Google Scholar 已经成为向 Thomson Scientific 发出挑战的两大重要的文献计量数据库。更重要的是由于专利数据库的数字化和索引化的实现,大规模的专利分析已容易得多。这也体现了文献计量研究的商业价值。

最后,由于研究范围的扩大,文献计量学也在发生变化。尤其是如今应用范围广泛的相关性分析为理解学术交流过程和学科结构提供了新方法。它通过期刊间、作者间的引文关系进行分析。可视化后的引文分析更有助于理解学科的学科结构,尤其是对正在兴起并快速发展的重大学科领域,比如纳米技术学和生物技术学。

5.2 网络计量学

现今进行网络计量研究的主要是信息科学和计算机科学方面的专家,当然,他们有着不同的研究动机。在信息科学领域,网络计量学最初集中在类似文献计量学的研究,但现在它已转向更具描述性、面向社会科学的研究方向。随着网络的发展,网络计量技术也极可能进一步发展,并公布更多有价值的描述性发现,商业数据挖掘技术的产生也是可能的。

相对传统的文献计量学,网络计量学有三大优势:(1)网络比 ISI 数据库更加快捷。一般来讲,一个科研项目的运行需要经历几个阶段:投入资金、展开研究、报告成果,最后向期刊投稿。从项目开始到成果公布至少需要 2 年时间。因此,以 ISI 数据库为依据的文献计量研究必然是追溯性的,是以若干年前的现象为研究对象。而基于网络计量的研究可以在网站建立时开始,并在其宣布研究很久前就展开调研。(2)网络包含大量与学术有关的作品,如演讲稿、专

利、数据、软件和大众网站。因此网络计量学更可能收集到有关学术影响力或者相互关系的数据。(3)由于所有的网络用户可以免费使用网上资源,这有可能使无力购买 ISI 数据的人也进行计量分析。

研究发现网络计量学仍有许多缺陷,有些缺陷甚至隐藏在其优势中。首先,与 ISI 出版物不同,网络上的信息没有经过质量控制。因此网络数据往往质量较低,这意味着网络计量学的研究发现多为指示性的,不能做为依据。其次,网络数据并不规范,因此只能从中抽取最简单的数据,区分各类型的信息显得尤为困难。例如,想辨别网上一篇文章的引用者是来自电子期刊还是网络课程并不容易。因此,网络计量的结果往往混合了价值不等的信息。再次,尽管网络数据的时效性很强,但也有可能找不到它的公布日期,因此网络计量的结果大多混合了或新或旧的网络数据。最后,网络数据在某些方面不完善且不够权威。尽管在网上可以免费获得一些学术文章,但大多数的学术文章并不是免费的。此外,有些学者或学术组织建立综合性网站,而有些则没有。这样一来,虽然网络计量可以反映网络中的学术研究情况,但其研究结果也只是对研究活动非常有限的反映。

通过对网络计量学优缺点比较,作者认为网络计量学取代文献计量学的可能性不大,但它在某些方面却很有优势:首先,它可以作为先行研究为后续系统的文献计量研究确定研究领域。其次,由于研究人员的网络发表是其成果公布的重要一方面,因此可以利用网络计量学评价学者在网络公布方面的成就。再次,网络计量学可以进行学科间及人口统计学方面的相关性分析。最后,网络计量学有助于 Web 2.0 以及社会科学知识库、人类学科研究目标的分析。

6 致谢

感谢武汉大学信息管理学院 2008 级出版专业博士研究生安欣同学在此次论文翻译和校对过程中给予的帮助!

作者简介

邱均平 男,武汉大学教授、博导,华中师范大学特聘教授。主要研究方向为信息管理与知识管理、信息计量与科学计量。
杨瑞仙 女,武汉大学信息管理学院情报学专业博士研究生,主要研究方向为网络信息计量学。
陶 雯 女,武汉大学信息管理学院情报学专业硕士研究生。
李雪璐 女,武汉大学信息管理学院情报学专业硕士研究生。