

情报学视角与社会学视角的 链接分析比较

戴 伟

(南京大学信息管理系,江苏南京 210093)

摘 要:本文在简要介绍了当前 Web 领域内链接分析的研究现状后,重点介绍了情报学视角和社会学视角的链接分析,最后通过比较两个视角研究之异同,提出了深化情报学视角的链接分析研究的建议。

关键词:链接分析;网络信息计量学;社会学;社会网络分析

中图分类号: G353.1 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2008.03.004

Comparison of Link Analysis from Webometric and Sociological Perspectives

Dai Wei

(Department of Information Management, NanJing University, Nanjing 210093)

Abstract: This paper introduces current research status of link analysis in the field of Internet, especially from webometric and sociological perspectives. Through the comparison between similarities and differences of the two perspectives, the author brings forward the personal viewpoint of the further study of informatics in this area.

Keywords: link analysis, webometrics, sociology, social network analysis

1 引 言

从统计物理学来看,网络是一个包含了大量个体及个体之间相互作用的系统。计算机学家与统计物理学家最早通过获取网络链接数据将互联网抽象成复杂网络^[1],探讨整个网络的连通性以及节点与链接的分布模式与增长模式等,研究互联网中类似于人际网络的“小世界现象”。他们

发现互联网不是一个混乱的世界,而是有着自身的特殊规律。计算机学家基于“特定主题位置的链接比全局的链接更加重要”^[2]这样一个假设开发出 HITS 算法,设计搜索引擎进行网络信息检索。同样,Google 也利用网页之间的链接引用结构关系开发了 PageRank 算法而获得成功^[3]。数学家利用图论更加抽象地研究网络链接,将互联网作为超大规模的有向图或无向图对待。传播学家们则关心网络上信息的传播与链接之间有何

第一作者简介:戴伟(1984-),男,湖南汨罗人,研究生,研究方向是网络计量学。

收稿日期:2008年4月12日。

种关系,他们认为链接代表一种信息的流动与传播方式,而传播学中有部分学者已在关注 BLOG 链接^[4]。信息科学在网络链接分析方面的研究主要是从信息计量学的角度,定量分析链接的数量特征与内在规律。社会学将社会网络分析法应用到网络环境中,将整个网络看成是由链接联系起来的人际网络,是现实人际网络的一种扩展。可见,不同学科背景有着不同的研究视角,本文主要对比研究同属于社会科学研究范畴的情报学视角和社会学视角的链接分析。它们的理论认识或许较为容易让彼此接受,研究方法和视角也可能有互通性,这将对深化情报学视角的链接分析研究有所启发。

2 情报学视角

20世纪90年代,万维网的发展对信息计量学研究提出了挑战,促使以科学计量学、文献计量学、情报计量学和技术计量学为主要内容的信息计量学进一步发展成为网络信息计量学^[5]。

2.1 相关理论

“从研究对象、方法、内容和目标等方面来看,网络信息计量学是采用数学乃至统计学等各种定量方法,对网上信息的组织、存贮、分布、传递、相互引证和开发利用等进行定量描述和统计分析,以便揭示其数量特征和内在规律的一门新兴分支学科”^[6]。这一视角的链接分析是“运用网络数据库、数学分析软件等工具,利用数学(主要是统计学和拓扑学)和情报学方法,对网络链接自身属性、链接对象、链接网络等各种对象进行分析,以便揭示其数量特征和内在规律,并用以解决各方面问题的一种研究方法”^[7]。情报学领域自1996年起就关注链接分析研究,并一直受文献引文分析方法的影响与驱动^[8]。

引文分析理论是其主要理论基础。引文分析法是利用各种数学和统计学的方法以及比较、归纳、抽象、概括等逻辑方法,对学术期刊、论文、作者等各种分析对象的引用或被引用现象进行分析,以揭示其数量特征和内在规律的一种文献计量研究方法^[9]。部分学者将网络空间看成类似于

传统引文环境,从而把网页之间的链接关系看作文献之间的引用关系。一个网页/网站被链接的次数(类似被引率和篇均被引频次)越多,说明此网页/网站越具有权威性^[10]。因此,引文分析法同时被扩展用于网络的学术工作评价与网络学术交流模式研究。

2.2 方法与工具

信息计量学者将学术影响因子引入网络,采用引文分析法研究网络链接。采用引文分析方法需要获取网页或网站的链接数以及被链接数等数据,目前主要利用以下两种工具来收集链接数据:商业搜索引擎与专用的 Web Crawler 爬行计算工具。在必要时可以借助手工方式。

2.2.1 商业搜索引擎

目前,商业搜索引擎还是链接分析的主要数据收集工具。支持这一技术的搜索引擎有 AltaVista、AllTheWeb、Google。利用商业搜索引擎有几个问题:一是商业搜索引擎的网络覆盖范围有限,Lawrence 和 Giles 在 1999 年就指出,主要商业搜索引擎的覆盖率只有 16% 左右。使用商业搜索引擎时应当特别说明。二是商业搜索引擎的返回结果在时间与空间上波动性很大^[11-14]。从根本上解决以上问题需要开发专门的网络爬行工具。

2.2.2 专用的 Web Crawler 爬行计算工具

根据链接分析特点专门设计的爬行计算工具可以获得商业搜索引擎所没有的灵活性与可靠性,可以就特定的范围,针对所研究的网络对象在一定时间内完成数据收集。在这方面,国内外所做的工作都很有有限^[15],因此也是研究重点所在。SocSciBot 是一个免费链接分析工具软件(<http://socscibot.wlv.ac.uk>),文献[16]提供其使用参考。

2.3 相关应用

2.3.1 网络影响力评价

这方面国内外已经做了大量实证研究,如文

文献[17]采用5种网络影响因子对中、美两国部分综合实力排名靠前的医院网站的网络影响力进行测度和分析。文献[18]利用链接分析,对大学网站的利用情况和网络影响力在大学评价中的作用进行了研究。Kayvan Kousha 和 Mike Thelwall 认为,通过链接分析评价特定学科的网络影响力具有一定的相关性,然而也有一些问题^[19]。笔者认为以引文分析理论为基础进行网络影响力评价具有一定意义,但由于只有部分链接是实质性链接,应该谨慎使用。链接的识别与分类研究发展将使链接在网络影响力评价方面更加有效。

2.3.2 改进搜索引擎

计算机科学在这方面做了大量工作,包括搜索引擎爬虫程序的改进,以及后期处理结果排序改进等。文献[20-24]对搜索引擎算法做了相关的研究。文献[25-27]研究了链接分析在结果相关度排序方面的应用。

3 社会学视角

社会学视角将网络链接看成是一种社交性的、具有信息传递作用的纽带。通过研究链接各节点之间关系所隐含的信息,比如通过移植社会网络分析法研究网络链接,反映整个网络中不同层面的行动者之间的关系,以及不同网络层次上的文化现象。

3.1 相关理论

在持续不断的发展中,社会网络分析涉及的领域众多。其建立在如下假设基础之上:“在互动的单位之间存在的关系非常重要”,“社会网络理论、模型及应用的基础都是关系数据,关系是社会网络分析理论的基础”^[28]。早在1997年,Garton、Haythornthwaite、Wellman 以及 Jackson 就撰文建议用社会网络分析来理解以计算机为媒介的交流过程^[29,30]。将社会网络分析用于网络链接分析即为链接网络分析,它利用社会网络分析法研究不同层面行动者之间由链接网络表示的各种交流网络的通信交流结构,可以看作是以计算机为媒介的通信交流过程(CMC)研究的一种形

式。

3.2 方法与工具

链接网络分析使用一系列社会网络分析的技术和工具,例如密度、集中度、聚类分析、块模拟、多维标度等。通过从网络中获取的链接数据组成一个基本链接网络数据集,利用社会网络分析的常用方法得出如密度、集中度等有用指标,进而反映这个对象网络的某些特性^[28]。这其中会用到很多专门工具,如 UCINET(http://www.insna.org/INSNA/soft_inf.html) 是社会网络分析的一个国际性专业网站。

3.3 相关应用

国外许多科学家已经开始将超链接网络分析应用于具体的评价研究。Palmer、Bailey 和 Faraj 将超链接网络分析的方法应用于电子商务^[31],但国内这方面的研究论文很少。链接网络分析还用于研究网络上个人之间、组织之间以及国际间的交流,并从另一角度提供研究社会运动的方法。超链接网络分析的一个重要作用是检验中心节点,作为中心网站通常与研究领域内成员保持最多的或最短的链接,中心网站一般起轴心、经纪人或权威性网站的作用。其相关指标有 Bonacich 中心特征向量、Freeman 集中度、Negopy 中心指标。

4 两个视角的比较分析

4.1 相关理论

情报学视角的链接分析主要集中在网络影响因子与搜索引擎研究,节点之间的链接与引文相似,体现了认同评价。因此,在情报学领域的学者将成熟的引文分析方法引入网络链接分析的同时,计算机学者也以此开发出两种重要的搜索引擎算法:PageRank 算法与 HITS 算法。而社会学视角的理论基础是将链接网络理解为一种社会网络,认为网络节点之间的链接是各节点所代表行动者的现实人际关系在网络空间的一种扩展或延伸,采用的研究方法是社会网络分析。

4.2 相关指标与概念

情报学视角的链接分析通常关心总链接数、入链数、出链数、共链数以及共被链数、内部链接数、外部链接数以及最长/平均路径长度等反映研究对象相互作用强度、影响力与重要性的指标。而社会学视角的链接分析主要关心 Bonacich 中心特征向量、Freeman 集中度(包括接受和发出链接的集中度)、接近集中度和两者间集中度、Richards Negopy 中心指标。中心指标表示围绕中心网站组织超链接的范围,密度指示了网络的整体水平,而集中度是其他网站与中心网站链接的比率。

4.3 相关链接分析工具

由于网络信息计量学最早是对商业搜索引擎进行研究,故商业搜索引擎是主要链接分析工具,但随着研究范围不断扩展,已经不能满足相关研究的需求,急需开发专用链接分析工具,包括特定的链接数据采集工具与后期的分析计算工具。社会网络分析方法由于其已有研究工具加以改进便可适应网络环境,已经有一些社会网络分析工具经过改造被前者使用。

中,有很多方面相互交叉重叠,之间的界线甚至难以分清。网络社区发现要利用网络信息计量学中链接分析的相关算法,同时运用社会学视角中链接网络分析的相关假设进行聚类分析,融合两个不同视角的研究成果,这或许体现了学科发展趋势。笔者也发现社会网络分析中心度概念与引文分析方法中的影响因子、引文分析中引用耦合现象研究与社会网络分析法的凝聚子群分析(同时其方法也与搜索引擎中使用的聚类算法相似)都具有一定的相似性。社会网络分析中的这些概念产生于社会学目的与视角,反映了对链接不同的认识,也许可引入到情报学视角的链接分析中,以深化其研究。

Callon 等人的行动者网络理论(ANT)使社会学研究从单纯关注人与人之间的关系延伸到关注社会网络中的物(新技术、信息系统等),将物看作社会网络中的行动者,将人与物之间的相互关系当作社会网络的一部分,考虑受技术产物的影响。在这一点上,情报学正发生着相反的变化:很多学者在研究物(信息)的网络链接的同时更多地关注人(信息生产者),研究网络资源节点所代表的人之间的关系。情报学研究的最终目的是使信息资源符合人的需要,已有学者研究网络信息系统的用户接受模型。情报学在网络链接分析时可以借鉴社会学的理论假设,考虑人对链接网络的影响与接受情况,对于研究链接生成动机、分布规律,网络学术群体分布与交流,改进搜索引擎用户体验等众多方面都有启发作用。

因此,笔者认为,情报学领域的链接分析不能停滞在引文分析理论基础之上,而应该积极引入社会网络分析方法,关注信息网络,同时关注与信息网络相关的人的网络,深入发掘网络信息资源的潜在规律,寻找学科创新与发展的突破口。

表1 情报学视角与社会学视角的链接分析比较

视角	基础理论	相关指标与概念	链接分析工具
情报学视角	将链接看成与传统引文相似的一种引用关系,体现了认同与评价	总链接数、入链数、出链数、共链数以及共被链数、内部链接数、外部链接数以及最长/平均路径长度等	商业搜索引擎、少量专门网络爬虫与改造社会网络分析工具、手工
社会学视角	将链接看作人际关系在网络空间的一种扩展或延伸	Bonacich 中心特征向量、Freeman 集中度、Richards Negopy 中心指标等	多种专门网络爬虫、手工与计算机辅助

5 结语

随着网络链接研究的深入,从两个不同领域内发展起来的研究团体之间的沟通会越来越多。从前面的讨论中可以看出,在两个视角研究

参考文献

[1] 吴金闪,狄增如.从统计物理学看复杂网络研究[J].物理学进展,2004,24(1):18-46.
 [2] Flake, G. W., Lawrence, S., Giles, C. L., Coetzee, F. M. Self-organization and identification of Web communities [J/OL]. [2008-03-28]. <http://clgiles.ist.psu.edu/papers/computer2002-web-communities.pdf>.

- [3] Brin, S., Page, L. The anatomy of a large scale hyper-textual Web search engine [J/OL]. [2008-03-28]. <http://infolab.stanford.edu/pub/papers/google.pdf>.
- [4] 麦尚文, 丁玲华, 张印平. 博客日志: 一种新的网络传播方式——从传播学角度看 blog 的勃兴[J]. 新闻界, 2003(6): 37-38.
- [5] 邱均平, 张洋. 网络信息计量学综述[J]. 高校图书馆工作, 2005, 25(1): 1-12.
- [6] 邱均平. 信息计量学(一): 信息计量学的兴起和发展[J]. 情报理论与实践, 2000(1): 75-80.
- [7] 董江山, 胡吉祥, 邱均平. 链接分析法及其应用[J]. 情报科学, 2004, 22(9): 1081-1084, 1099.
- [8] Han Woo Park, Mike Thelwall. Hyperlink Analyses of the World Wide Web: A Review[J/OL]. [2008-03-28]. <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1083-6101.2003.tb00223.x>.
- [9] 邓秀林. 基于引文分析法的期刊栏目的学术影响力评价[J]. 现代情报, 2007, 27(5): 154-155.
- [10] 杜友桃, 何琳. 基于引文分析法的 Web 超链接分析新进展[J]. 情报探索, 2006(9): 29-32.
- [11] Bar-Ilan, J. Search engine results over time - A case study on search engine stability[EB/OL]. [2008-03-28]. <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>.
- [12] Mettrop, W., Nieuwenhuysen, P. Internet search engines - fluctuations in document accessibility[J/OL]. [2008-03-28]. <http://www.vub.ac.be/BIBLIO/nieuwenhuysen/presentations/2001-j-doc-mettrop-nieuwenhuysen/2001-mettrop-nieuwenhuysen.pdf>.
- [13] Rousseau, R. Daily time series of common single word searches in AltaVista and Northern Light[EB/OL]. [2008-03-28]. <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>.
- [14] Snyder, H., Rosenbaum, H. Can search engines be used for Web-link analysis? A critical review[J/OL]. [2008-03-28]. <http://publ.ac.uk/archive/journals/jdocumentation/v55n0499.htm#2can>.
- [15] 邱均平, 李江. 链接分析工具——博客链接索引(BSI)的功能与应用[J]. 图书情报工作, 2007, 51(9): 25-28, 108.
- [16] Thelwall, M., Stuart, D. Web crawling ethics revisited: Cost, privacy and denial of service[J/OL]. [2008-03-28]. http://www.scit.wlv.ac.uk/~cm1993/papers/Web_Crawling_Ethics_preprint.doc.
- [17] 朱雷. 中美两国医院网站网络影响力指标对比评测研究[J]. 现代图书情报技术, 2006(3): 64-67, 81.
- [18] 段宇锋, 邱均平. 网络信息计量学研究(V)——链接分析在大学评价中的应用研究[J]. 情报学报, 2005, 24(6): 735-741.
- [19] Kayvan Kousha, Mike Thelwall. Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis[J/OL]. [2008-03-28]. <http://eprints.rclis.org/archive/00006416/01/google.pdf>.
- [20] 王卫玲, 刘培玉, 刘克非. 改进的 Web 链接主题提取算法[J]. 计算机工程与设计, 2007, 28(2): 294-296.
- [21] 宋玲玲, 李村合. 基于链接结构分析的 Web 信息检索方法研究[J]. 现代情报, 2007, 27(2): 133-135, 137.
- [22] 黄德才, 戚华春, 钱能. 基于主题相似度模型的 TS-PageRank 算法[J]. 小型微型计算机系统, 2007, 28(3): 510-514.
- [23] 钱功伟, 倪林, MIAO Yuan, 等. 基于网页链接和内容分析的改进 PageRank 算法[J]. 计算机工程与应用, 2007, 43(21): 160-164.
- [24] 张丽. PageRank 算法的改进[J]. 科学技术与工程, 2007, 7(5): 673-677.
- [25] 刘菁菁, 林鸿飞, 赵晶. 基于 PageRank 和锚文本的网页排序研究[J]. 计算机工程与应用, 2007, 43(10): 170-173.
- [26] 王秀平, 李治柱. 一种改进的搜索引擎网页排序算法[J]. 计算机科学, 2004, 31(9): 17-19.
- [27] 郑煜, 钱榕. 一个基于链接分析的相关度排序算法及其在专题搜索引擎中应用[J]. 计算机应用与软件, 2007, 24(7): 54-55, 65. <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1083-6101.2000.tb00342.x>. Journal of Computer-Mediated Communication, 2000, 5(3).
- [28] 刘军. 社会网络分析导论[M]. 北京: 社会科学文献出版社, 2004.
- [29] Garton, L., Haythornthwaite, C., Wellman, B. Studying online social networks[EB/OL]. [2008-03-28]. <http://jcmc.indiana.edu/vol3/issue1/garton.html>.
- [30] Jackson, M. H. Assessing the structure of communication on the world wide web[EB/OL]. [2008-03-28]. <http://jcmc.indiana.edu/vol3/issue1/jackson.html>.
- [31] Palmer J W, Bailey J P, Faraj S. The Role of Intermediaries in the Development of Trust on the WWW: The Use and Prominence of Trusted Third Parties and Privacy Statements[J/OL]. [2008-03-28]. <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1083-6101.2000.tb00342.x>.