

国内网络信息计量学研究的 作者共被引分析

邓中华 郑 曦

(南京大学信息管理系, 江苏南京 210093)

摘 要: 本文利用统计分析软件 SPSS11.0 的因子分析、聚类分析以及多维尺度分析, 对 CSSCI 中收录的 2000-2005 年网络信息计量学的 109 篇被引文献进行了作者共被引分析, 归纳了国内网络信息计量学领域的作者群体, 指出当前国内主要集中于对网络信息计量学理论、方法和应用的研究。

关键词: 网络信息计量学; ACA; 因子分析; 聚类分析; 多维尺度分析

中图分类号: G353.1 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2008.03.008

ACA of Chinese Webometrics

Deng Zhonghua, Zheng Xi

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

Abstract: Using the factor analysis, clustering analysis and multi-dimension analysis of the SPSS11.0, this paper makes an Author Co-citation Analysis to the 109 cited articles of Chinese Webometrics from CSSCI and from 2000 to 2005, pointing out that Chinese Webometrics is now mainly concentrating on it's theory, method and application.

Keywords: Webometrics, ACA, factor analysis, clustering analysis, multi-dimension analysis

1 关于 ACA

ACA (Author Co-citation Analysis) 的全称是作者共被引分析, 是共被引分析法中的一种。常见的共被引分析法主要有 DCA (Documents Co-citation Analysis, 即文献共被引法) 和 ACA。这两种方法的基本原理相同, 即分析两篇文献或两位作者同时被第三篇文献或作者引用的关系, 两篇文献

或两位作者的同被引次数越高, 二者之间的相关性越大^[1]。本文拟利用 ACA 方法来分析国内网络信息计量学的研究状况。具体做法是统计出两位作者同时被第三篇文献引用的次数, 得出作者共被引矩阵, 在此基础上利用 SPSS11.0 (Statistical Product and Service Solutions, 统计产品与服务解决方案) 软件进行作者共被引矩阵的标准化和多元统计分析, 从而形成不同研究方向的作者的共被引可视化聚类结果。

第一作者简介: 邓中华 (1984 -), 女, 硕士研究生, 研究方向是网络信息资源管理。

收稿日期: 2008 年 4 月 12 日。

2 数据来源

随着网络技术和现代信息技术的发展,在传统“三计学”——文献计量学、科学计量学和信息计量学的基础上,1997年,丹麦学者 T. C. Almind 等在《Informetric Analyses on the World Wide Web: Methodological Approaches to Webometrics》中首次提出“Webometrics”^[2],吸引了众多研究者的参与,国内常将其译为网络信息计量学或网络计量学。关于网络信息计量学的研究,国外特别是欧洲,不仅开展得较早,研究领域也比较广泛,与之相比,国内的研究还远远不够。那么,国内网络信息计量学的研究现况如何呢?本文以 PM = 网络 * 计量学为检索表达式,在 CSSCI(即“中文社会科学引文索引”)中检索到 2000-2005 年被引文献 109 篇,于 2001-2007 年总被引 240 次。

3 高被引次数的 ACA 分析

3.1 统计并建立作者共被引矩阵

先统计出国内网络信息计量学各个作者(本文取第一作者)被引用文献的总篇数(当然,同一作者可能有多篇文献被引用),亦即作者被引次数,并按被引次数由高到低排序(表 1),本文选取被引次数大于或等于 3 的作者共 17 个。这些作者是国内网络信息计量学相关研究中被引次数较高的,他们之间的共被引关系在很大程度上可以体现我国网络信息计量学的研究状况。

然后,在 Excel 中内嵌的程序设计语言 VBA(Visual Basic for Applications)^[3] 自编统计分析程序两两统计他们被同一篇文献引用的次数,建立 17×17 的作者共被引矩阵(表 2)。该矩阵是对称矩阵,其中主对角线上的数据定义为缺省值,非主对角线上的数据表示任意两个作者共同被同一篇文献引用的次数。

表 1 作者被引次数列表

姓名	被引次数	姓名	被引次数
邱均平	39	谈大军	2
吴华香	29	王知津	2
徐久龄	25	姚贾虹	2
金岩	17	张翠玲	2
马建华	17	赵春英	2
李长忠	14	蔡明月	1
夏旭	14	陈亦伟	1
杨涛	10	成波	1
龚立群	8	李长玲	1
霍丽萍	8	李长志	1
齐艳红	8	李伟超	1
邹菲	6	刘远翔	1
牟华	5	沈峰	1
段宇锋	4	徐家坤	1
樊春林	3	张立肖	1
苏云	3	张晓雁	1
谢奇	3	张洋	1
邓湘琳	2	邹勇	1
苏新宁	2		

表 2 17 个高被引作者的共被引矩阵

	邱均平	吴华香	徐久龄	金岩	马建华	李长忠	夏旭	杨涛	龚立群	霍丽萍	齐艳红	邹菲	牟华	段宇锋	樊春林	苏云	谢奇
邱均平		7	9	4	5	7	0	3	4	3	0	1	2	4	2	3	3
吴华香	7		12	12	9	9	7	4	2	3	2	2	2	0	1	1	0
徐久龄	9	12		5	10	7	3	2	4	3	0	3	1	0	2	3	0
金岩	4	12	5		2	5	5	1	0	3	3	3	1	0	0	0	0
马建华	5	9	10	2		6	0	1	4	1	0	0	2	0	3	3	0
李长忠	7	9	7	5	6		3	1	2	4	1	1	2	0	2	1	0
夏旭	0	7	3	5	0	3		0	1	2	3	1	2	0	1	0	0
杨涛	3	4	2	1	1	1	0		2	1	0	0	0	1	0	0	1
龚立群	4	2	4	0	4	2	1	2		0	1	0	1	0	3	2	0
霍丽萍	3	3	3	3	1	4	2	1	0		0	2	2	0	0	0	0
齐艳红	0	2	0	3	0	1	3	0	1	0		0	0	0	1	0	0
邹菲	1	2	3	3	0	1	1	0	0	2	0		1	0	0	0	0
牟华	2	2	1	1	2	2	2	0	1	2	0	1		0	1	1	0
段宇锋	4	0	0	0	0	0	0	1	0	0	0	0	0		0	0	4
樊春林	2	1	2	0	3	2	1	0	3	0	1	0	1	0		2	0
苏云	3	1	3	0	3	1	0	0	2	0	0	0	1	0	2		0
谢奇	3	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	

表 3 作者共被引的相关矩阵(局部)

	邱均平	吴华香	徐久龄	金岩	马建华	李长忠	夏旭
邱均平		1	0.587205	0.72248	0.425535	0.87151	0.703111
吴华香	0.587205		1	0.751774	0.786989	0.609989	0.857393
徐久龄	0.72248	0.751774		1	0.693744	0.882985	0.931728
金岩	0.425535	0.786989	0.693744		1	0.606826	0.779169
马建华	0.87151	0.609989	0.882985	0.606826		1	0.812326
李长忠	0.703111	0.857393	0.931728	0.779169	0.812326		1
夏旭	0.398265	0.561209	0.406343	0.866936	0.517313	0.575005	
杨涛	0.743503	0.403803	0.727809	0.622273	0.713831	0.752789	0.4155
龚立群	0.515178	0.43998	0.611564	0.194234	0.729096	0.567951	-0.109
霍丽萍	0.519411	0.786205	0.627655	0.731764	0.569005	0.760531	0.5745

3.2 作者共被引矩阵的标准化——相关矩阵

进行 ACA 分析,只有高被引作者的共被引矩阵是不够的,我们需要在此基础上知道高被引作者之间的相似性,也就是将作者共被引矩阵标准

化,即利用 SPSS11.0 中相关分析的 Bivariate 过程将共词矩阵转化为相关矩阵^[4](表 3)。

3.3 作者共被引的多元统计分析

将表 3 所示相关矩阵的数据导入 SPSS11.0, 运用不同的分析过程可以进行多元统计分析, 具体包括因子分析、聚类分析和多维尺度分析。

3.3.1 因子分析

因子分析^[4](有时也称作主成份分析),是最常用的数据简化方法,需要将特征值准则和碎石检验准则相结合,通过因子分析将众多的观测变量浓缩成少数几个因子。将表 3 的相关矩阵的数据导入 SPSS11.0 进行因子分析即可得到主成份列表(表 4),表中列出了所有的主成份,它们按特征值从大到小排列,特征值大于 1 的前 3 个主成份解释了总变异的 93.838%。在此基础上建立碎

石图(图 1),图中自第 4 个观测变量开始曲线变平,故取最大因子数为 3。

结合表 4 和图 1,我们认为在接下来所进行的聚类分析和多位尺度分析中将本文高被引作者分为 3 类较合理。

3.3.2 聚类分析

聚类分析有快速聚类法、系统聚类法等^[4]。文采用系统聚类法,是一种比较常用并且有效的聚类方法,关于作者共被引的系统聚类分析原理与共词分析类似。首先将每个高被引作者看作一类,然后把距离最近的两类合并,再重新计算类与类之间的距离,之后再把距离最近的两类合并,依此类推,直至将所有的高被引作者归为一类。将表 3 的相关矩阵数据导入 SPSS11.0 进行系统聚类并加以优化整理,就可以得出聚类龙骨图(图 2)。

表 4 因子列表

主成分	初始Eigen值			提取的主成分值		
	特征根	解释总变量的百分比	累积解释变量的百分比	特征根	解释总变量的百分比	累积解释变量百分比
1	9.4	55.292	55.292	9.4	55.292	55.292
2	5.002	29.426	84.718	5.002	29.426	84.718
3	1.55	9.12	93.838	1.55	9.12	93.838
4	0.488	2.873	96.711			
5	0.339	1.997	98.708			
6	0.116	0.683	99.391			
7	0.022	0.133	99.524			

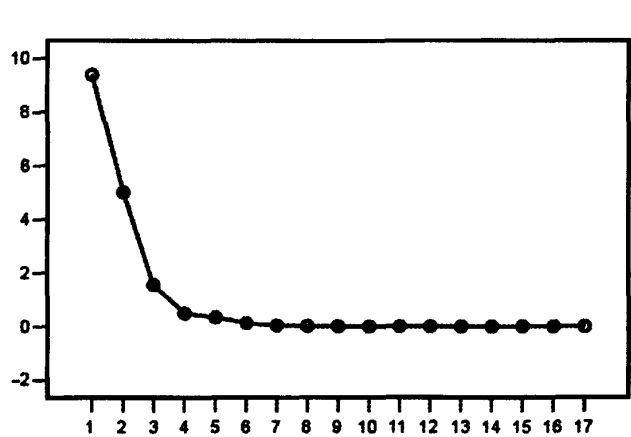


图 1 碎石图

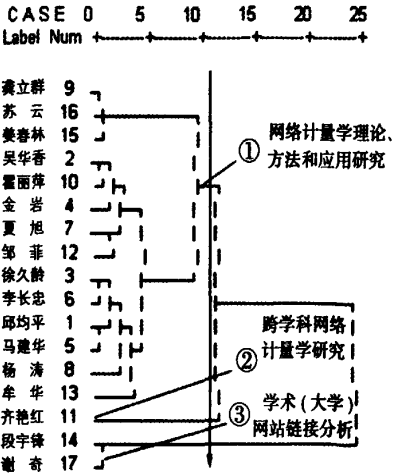


图 2 共被引作者聚类龙骨图

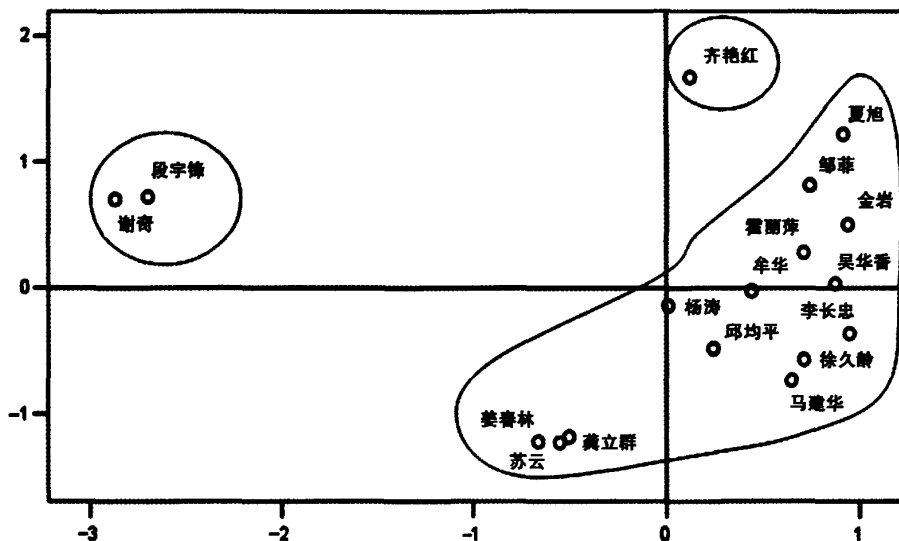


图3 共被引作者多维尺度图

3.3.3 多维尺度分析

多维尺度分析利用平面距离来反映关键词之间的相似程度,它通过低维空间(本文是二维空间)展现作者之间的联系^[4]。根据表3的相关矩阵,利用SPSS11.0进行多维尺度分析,并结合图2的聚类结果,得出多维尺度图(图3)。在图3中,有高度相似性的作者聚集到一起,并且越居中的作者与其他作者之间的联系越多,在该领域中处于核心地位,反之则越独立。从图3我们可以看到,在聚集作者最多的一类中,邱均平、吴华香、徐久龄等居于中心位置(表1),他们也是被引次数居高的几位作者。由此可见,这几位作者处于该领域的核心地位。

在多维尺度分析中,Stress和RSQ分别是信度和效果估计值。一般而言,Stress < 0.2, RSQ > 0.8时,结果是可信有效的,并且Stress值越小,RSQ值越大^[4]。在本文所做的多元统计分析中,Stress = 0.04759 < 0.2, RSQ = 0.99211 > 0.8。这表明本文的分析结果有很强的适合度,能大致反映出我国网络信息计量学研究者之间的关联及领域发展状况。

4 结果分析

在本文所做的ACA分析中,因子分析给出

分类参考值,在此基础上聚类分析图将作者分为三大类,但反映不出作者之间的关联程度,多维尺度分析图可以直观地反映作者之间的关联程度,但不能明确界定类别。综合因子分析、聚类分析龙骨图和多维尺度分析图,同时参考本文分析所用样本的叙词及摘要,可以得出以下分析结果。

(1)当前国内网络信息计量学的研究基本上可分为3类。在图2所示的聚类龙骨图中,将①命名为网络信息计量学理论、方法和应用研究,将②命名为跨学科网络信息计量学研究,将③命名为学术(大学)网站链接分析。在图2和图3中,可以清楚地看到目前国内网络信息计量学的研究侧重于①,即网络信息计量学的理论、方法及应用,具体我们将在下文第(4)点中详细阐述。同时,对于③的研究也占有一席之地,它其实可归属于①,但因为链接分析是与网络信息计量学相互独立而又互有交叉的研究领域,因此将网络信息计量学视角的学术网站链接分析单独作为一类。另外,网络信息计量学的发展吸引了跨专业领域的参与,如图2中②所示。

(2)我们将②命名为跨学科网络信息计量学研究(如图2和图3所示),一方面是因为作者齐艳红来自跨院系图书馆,另一方面其关于网络信息计量学的文献《网络计量学的一种Internet分布式聚类分析软件》^[5],多为偏重于技术类文献

引用,如王曰芬等的《文献计量与内容分析综合应用软件的开发与实验》^[6]一文。

(3) 将③命名为学术(大学)网站链接分析研究(如图2和图3所示),是因为该类中的两个代表作者段宇锋、谢奇的论文侧重于大学网站的链接分析研究,如段宇锋的《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》^[7]就是以链接分析为基础,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探索。同时,引用他们的文献也多是基于链接分析进行大学网站评价研究的,如王丽伟等在《基于链接的网络计量指标与大学科研得分相关性研究》^[8]一文中就同时引用了这两位作者关于学术网站链接分析的3篇文献。

(4) 图2和图3所示的①,即网络信息计量学理论、方法和应用研究是国内网络信息计量学的研究之重。该类(如图2和图3所示)聚集了国内网络信息计量学研究的大部分学者,包括徐久龄、李长忠、邱均平、马建华、杨涛、牟华、吴华香、霍丽萍、金岩、夏旭、邹非、龚立群、姜春林、苏云。网络信息计量学理论研究主要论述有关网络信息计量学的新概念、起源、产生的意义及与传统文献计量学的关系等,方法研究主要讨论诸如链接分析法及其计量指标、工具等,而应用研究探讨其原理、方法、指标、工具等在各个领域的应用。具体来说,邱均平的《网络信息计量学综述》^[9]对网络信息计量学作了一个综述,包括网络信息计量学的产生、定义、内容和工具等基础理论问题,并对其研究现状作了全面、详细的介绍。而引用该文的文献也对网络信息计量学的应用等方面进行了一系列探讨,如朱良兵等的《网络信息计量学及其应用》^[10]。吴华香的《网络计量学:互联网上的文献计量学》^[11]一文,认为网络计量学是文献计量学的拓展,同时介绍了网络计量学在互联网上进行信息分析与测量的标准方法。吴华香在其另一篇文献《链接分析法——网络计量学方法初探》^[12]

中,指出链接分析法是网络计量学研究方法之一,利用因特网搜索引擎数据库为网络计量研究提供数据,可对网络信息进行链接计量分析。还有从实证角度探讨网络信息计量学理论、方法和应用的,如杨涛的《网络信息计量学实证研究:对国内20个大学网站的分析》^[13],从中国大陆20所大学网站的链接数入手,结合网络影响因子等指标,对各类链接数与我国较有影响的两所大学排名的相关关系进行测度。

参考文献

- [1] 马费成,等.我国情报学研究分析:以ACA为方法[J].情报学报,2006,25(3):259-268.
- [2] T. Almind, P. Ingwersen. Informetric Analyses on the World Wide Web: Methodological Approaches to "Webometrics" [J]. Journal of Documentation, 1997(53)404-426.
- [3] 杜茂康,等. Excel与数据处理(第2版)[M]. 北京:电子工业出版社,2005.
- [4] 张文彤,等. 世界优秀统计工具 SPSS11.0 统计分析教程(高级篇)[M]. 北京:北京希望电子出版社,2002.
- [5] 齐艳红. 网络计量学的一种 Internet 分布式聚类分析软件[J]. 情报科学,2003,21(10):1069-1072.
- [6] 王曰芬,等. 文献计量与内容分析综合应用软件的开发与实验[J]. 图书情报工作,2005,49(6):24-28.
- [7] 段宇锋. 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(6):735-741.
- [8] 王丽伟,等. 基于链接的网络计量指标与大学科研得分相关性研究[J]. 情报学报,2007,26(3):384-390.
- [9] 邱均平. 网络计量学综述[J]. 高校图书馆工作,2005,25(1):1-12.
- [10] 朱良兵,等. 网络计量学及其应用[J]. 中国信息导报,2005(12):38-40.
- [11] 吴华香. 网络计量学:互联网上的文献计量学[J]. 图书馆杂志,2001(1):33-36.
- [12] 吴华香. 链接分析法——网络计量学方法初探[J]. 情报科学,2002(1):71-73.
- [13] 杨涛. 网络计量学实证研究:对国内20个大学网站的分析[J]. 图书情报工作,2003(9):61-66.