国外共引分析研究进展及发展趋势

The Research Progress and Trends of Cocitation Analysis in Foreign Countries

耿海英　肖仙桃

（中国科学院国家科学图书馆兰州分馆　兰州　730000）

摘　要　共引分析是一种重要的引文分析方法。阐述了国外两类突出的共引分析法的发展过程：以 Small 为代表的文献共引分析研究和以 White 为代表的作者共引分析研究。并指出共引分析中存在的问题及其发展趋势。

关键词　引文分析　共引分析　研究进展

引文分析是研究科学文献之间引证关系的一种科学计量方法。所谓引文分析（Citation Analysis），就是利用各种数学及统计学的方法和比较、归纳、抽象、概括等逻辑方法，对科技期刊、论文、著作等各种分析对象的引用或被引用现象进行分析研究，以便揭示其数量特征和内在规律，达到评价、预测科学发展趋势的目的[1]。引文分析中最具影响力的就是共引分析方法。自从 1973 年 Small 提出共引分析的概念后，共引分析已成为一种潜在多产的分析方法，它不仅可以用来揭示科学结构的发展现状乃至变化情况，还可以用来进行研究前沿分析、领域分析、科研评价等，进而为宏观科技决策提供先行支持，为科技规划与评估提供基础[2]。

共引（Cocitation）又称同被引，即两篇文献同时被后来的一篇或多篇文献所引用，同时把共同引用这两篇文献的文献数称为共引强度[1]，共引强度越大这两篇文献关系越密切。共引分析就是以具有一定学科代表性的一批文献为分析对象，利用聚类分析、多维尺度等多元统计分析方法，借助计算机，把众多的分析对象之间错综复杂的共引网状关系简化为数目相对较少的若干类群之间的关系并直观地表示出来，在此基础上分析研究分析对象所代表的学科及文献的结构和特点[3]。同传统的学者个人归纳、访谈调查等主观分类方法相比，共引分析最大的优势是它的客观性、分类原则的科学性和数据的有效性[4]。引文分析的分析单位可以是论文、作者或期刊，因此除了文献共引，还可将共引推广至作者共引和期刊共引。

1　国外共引分析发展历程

纵观国外共引分析研究的发展历程，主要有两个系列：以 Small 为代表的以文献为分析单位所做的文献共引分析研究和以 White 为代表的以作者为分析单位所作的作者共引分析研究。期刊共引分析是在文献共引分析的基础上发展起来的，研究相对较少，故此文只对文献共引和作者共引的发展情况作了阐述。

1.1　*文献共引分析*　文献共引分析研究的始祖是 ISI 首席科学家 Henry Small。Small 的共引理论是基于共引可以反映文献主题内容方面的相似性，及对共引关系的测度可以作为揭示科学结构的一种有效方法这样的假设而提出的。此后，Small 一直致力于文献共引分析的理论与实践研究，所以，Small 的研究历程也即是文献共引分析研究的发展历程。

早在 1972 年，Small 受库恩范式理论的启发，通过对文献编码，记录每篇文献中的作者、关键词、从属关系、分类标题、参考文献等，尝试通过科学映射方法描述核物理学的科学结构及其随时间的演变历程。1972 年，Small 加入 ISI 后将研究重点转向参考文献的共现研究，并于 1973 年提出了共引概念[5]。Small 认为高被引文献可能有着非同寻常的重要性，代表了特定的发现、方法，或者是引用作者所共同认可的概念；高被引文献间的强共引链也常常成为人们关注的焦点。1974 年，Small 和 Drexel 大学的情报学教授 Griffith 着手分析一个季度的 SCI 文件，目的就是利用共引方法把整个科学的结构整合在一个图中显现出来。

为了通过共引分析确定研究领域，首先需要确定一个阈值来选择高被引文献，即选择被引频次高于某数值的文献作为分析对象，其中最关键的是要从各学科均匀地进行高被引文献的取样。但不同学科在引文数量上有很大差异，这样必然会导致科学结构分析结果失真，Small 为此提出了用改进的指标分数引文量（fractional citation counting）来选择文献，就是每一篇引文都用引用它的来源文献的引文长度进行加权，以此来平衡学科结构[6]。同时针对不同学科引文率不同，还引入了可变水平聚类方法（Variable level clustering）和以类聚类（clustering of clusters）的反复聚类方法[7]。为了得到一个映射图，还需以一定的方式将已聚类的文献显现在一个二维结构中以便能够反映出文献间的相关程度，这就需要降维技术，为此 Small 借用了 Joseph Kruskal 发展的多维尺度（Multidimensional scaling，MDS）技术。

实践应用方面，Small 等人首先开发了基于共引理论的单机系统 SCI－Map 来描绘科学文献间的结构[8]；通过连续时间内共引聚类图的历时比较，反映科学结构的变化；从不同学科间的共引关系中寻找某一学科到另一学科的可通路径，从而描述知识结构；基于 ISI 数据将共引聚类用于科学研究前沿分析。

1.2　*作者共引分析*（Author Cocitation Analysis ACA）　作者共引分析（ACA）起源于美国费城的 Drexel 大学，是以作者而不是文献为分析单位。1981 年 White 和 Griffith 合作发表的《作者共引：科学结构的文献测量方法》一文开创了 ACA 的先河，该文通过对

作者简介：耿海英，女，1980 年生，硕士研究生，研究方向为文献计量学理论及应用；肖仙桃，女，1965 年生，研究馆员，硕士生导师。

1972～1979年39位情报学家的共引情况描绘了他们在学科中的位置和情报学的学科结构[9]。描述学科结构是White开发ACA的初衷。ACA假定两个作者的作品同时被后继的作品引用则表明这两个作者之间有联系，共同被引用的次数越多，他们之间关系就越紧密。一组相关作者的共引频次模式分析能揭示出作者间突出的链接，并能揭示他们各自或共同代表的主题领域。此后的20多年里，ACA一直以映射图的方式用于探究科学和学术内部的专业知识结构、揭示文献的影响力、探讨学科范式等。这期间的研究主要有：1989年，White和McCain将情报学分为两个主要领域——文献计量学(包括引文分析)和情报检索[10]；1998年，White和McCain再次采用ACA技术，通过对1972～1995年24年间的一些代表性的文献特征(作者共引数目)归纳总结情报学领域的结构特征和24年来的发展情况，并将结论可视化[11]。

1990年，McCain对作者共引技术进行综述，将ACA的程序规整为选择作者、检索共引频次、生成共引矩阵、转化为Pearson相关系数矩阵、多元分析和解释结果等几个步骤[12]，人们称其为传统的ACA模式。但传统ACA需要大量的计算与绘图操作。分析者首先必须通过各种来源确定能够覆盖一个学科各个分支的作者集合，进而通过分析程序，并依赖于支持因子分析、多维尺度和聚类分析的统计工具(如SPSS)，通过多维图观察相似性而形成簇，同时借助统计方法确定作者的重要性[13]。此工作流程中烦琐的数据搜集、计算中存在的矩阵对角线值设定问题，特别是Pearson相关系数的计算增加了ACA的复杂性，这些都严重阻碍了ACA的广泛应用[14]。为此人们开始寻找新的技术方法替代传统的Pearson相关系数。网络寻址定位(Pathfinder Network Scaling，PFNETs)就是人们作出的尝试之一。PFNETs起源于认知心理学语义关系研究，已被情报学家广泛用于文献检索界面的主题索引词显示。White于2003年采用PFNETs对1998年的同一数据进行了第二次分析，得到了更为准确可靠的分析结果。比起传统的ACA，PFNETs可直接产生于原始数据矩阵，而不需要再将原始矩阵转化为Pearson相关系数矩阵，减少了ACA的计算强度，结果更为可信。

ACA有两个目的，一个就是前面提到的通过共被引作者概括一个领域的知识结构；另一个就是为了共被引作者检索的目的，将映射图转换为可视化的信息检索界面(VIRIs)。1981年，White就提出用作者共引分析方法改进检索策略来辅助联机检索，提高检索效率；1989年，White又做了关于共被引作者检索方面的研究报告；2000年，White又提出可以将ACA应用拓展到主题检索。但是由于技术等方面的原因，终未能实现。而信息可视化的迅速发展为ACA用于信息检索提供了可能。1997年美国肯塔基大学的Linxia首先尝试将Kohonen的自组织映射技术(self－organizing map，SOM)用于共引矩阵，并在2000年生成一个将情报学家聚到几个主题域的图谱[15]。随后，White带领由Linxia和Jan Buzydlowski组成的研究小组开展了实时环境下ACA绘图及主题检索研究，利用Dialog和SCI的数据，开发出了AuthorLink检索系统。AuthorLink也因此成为用实时共引映射图实现检索重要数据库的开创者。利用AuthorLink进行检索时，输入一个作者名，用户从该系统得到的不仅是一个作者的信息，而是与该作者高频共引的24位作者，以及基于共引强度以图的形式展示的作者间相互关系。

2 共引分析中存在的问题及其发展趋势

2.1 *共引分析中存在的问题* 共引分析自诞生以来的30多年里，其理论与方法逐渐成熟，应用实践范围不断扩大，分析结果的客观系统性，使得其已成为一种可靠实用的情报研究方法。但共引分析作为一种分析方法，除了存在引文分析所固有的缺陷外，自身还存在一些问题，主要有以下几方面：a.数据搜集过程烦琐且费时，搜集好的数据还需要转化成统计工具或可视化工具所需要的形式，还没有专门的软件工具能够将此过程程序化。b.相似性的计算问题。传统的共引分析中，Small主要采用Salon余弦测度或Jaccard系数测度文献间的相似性，而White主要用Pearson相关系数进行作者间的相似性计算。相似性计算方法众多，哪种方法测度更准确可靠，ACA分析中的原始矩阵是否还需转换为Pearson相关矩阵，这些还都有待商榷。c.共引分析的数据源一般都来自ISI，但ISI中只对文献的第一作者进行标引，为了方便起见，传统的ACA都是针对第一作者进行的共引分析研究，但随着合著文献的日益增多，这种第一作者的共引分析无疑会使分析结果在一定程度上失真。d.用共引分析进行科学前沿和热点分析，由于分析时时对高被引的论文进行聚类，而一些新出现的研究领域，因为太新可能在分析时还未被高被引，因此，分析结果可能会漏掉一些研究前沿领域。

2.2 *共引分析的发展趋势* a.综合多种分析方法。例如研究前沿和热点分析时，将共引分析结果和文献耦合、共词聚类、词频统计等方法的分析结果加以比较分析；揭示科学结构时，将共引分析与共词分析相结合，分析结果会更准确可靠。b.不断融入新的技术。由最初借用多维尺度技术进行降维，到现在用PFNETS替代Pearson相关系数，引入自组织映射(Self－Organization Map，SOM)技术、潜在语义索引(Latent Semantic Indexing，LSI)技术等。随着各种技术的发展，共引分析中不断融入其他学科新的技术，真可谓吸众家之长为我所用。c.扩展至网络结构研究。网络环境中，站点的链接关系类似于文献的引用关系，因此可以将共引分析方法移植到网络站点共引研究或称其为网页共链分析(Web Colink Analysis，WCA)，反映网络本身的结构和网络中知识的结构。d.不断探究共引分析中的一些细节问题。这其中包括相似性计算方法的优化，如何对合著者进行所有作者的共引分析等。

共引分析作为一种情报研究方法，以其卓有成效的科学结构描述以及信息检索方面应用的新突破，使得其备受各国研究人员的青睐。随着信息技术的发展和人们探索的不断深入，相信共引分析会更加成熟，从而为科学决策者、各级部门管理者和科研工作者提供有效的决策支持。

参考文献

1 庞景安.科学计量研究方法论.北京：科学技术文献出版社，2002

2 王建芳，冷伏海.共引分析理论与实践进展.中国图书馆学报，2006；(1)

3 赵党志.共引分析——研究学科及其文献结构和特点的一种有效方法.情报杂志，1993；(5)

4 刘林青.作品共被引分析与科学地图的绘制.科学学研究，2005；(2)

5 Small H. Paradigms, Citations, and Maps of Science: A Personal History. Journal of the American Society for Information Science and Technology, 2003；(5)

(下转第72页)

表3 西安理工大学信息化评价表

		e_1	e_2	e_3	e_4	e_5
U_1	U_{11}	1	0	0	0	0
	U_{12}	0.2	0.7	0.1	0	0
	U_{13}	0.4	0.5	0.1	0	0
	U_{14}	0.7	0.3	0	0	0
	U_{15}	0.6	0.3	0.1	0.1	0
	U_{16}	0.2	0.3	0.4	0.1	0
	U_{17}	1	0	0	0	0
	U_{18}	0	0	0.7	0.3	0
	U_{19}	0.8	0.2	0	0	0
	U_{110}	0.7	0.3	0	0	0
U_2	U_{21}	0.3	0.4	0.3	0	0
	U_{22}	0.3	0.5	0.2	0	0
	U_{23}	0.5	0.5	0	0	0
	U_{24}	0.4	0.4	0.2	0	0
	U_{25}	0.8	0.2	0	0	0
	U_{26}	0	0.5	0.5	0	0
U_3	U_{31}	0.6	0.4	0	0	0
	U_{32}	0.8	0.2	0	0	0
	U_{33}	0.3	0.6	0.1	0	0
U_4	U_{41}	0	0.4	0.5	0	0
	U_{42}	0	0.3	0.7	0	0
	U_{43}	0.9	0.1	0	0	0
U_5	U_{51}	0.4	0.5	0.1	0	0
	U_{52}	0.7	0.3	0	0	0
	U_{53}	0.6	0.3	0.1	0	0
U_6	U_{61}	0.8	0.2	0	0	0
	U_{62}	0.6	0.3	0.1	0	0
	U_{63}	0.3	0.5	0.2	0	0
U_7	U_{71}	0.8	0.2	0	0	0
	U_{72}	0.9	0.1	0	0	0
	U_{73}	0.4	0.4	0.2	0	0
	U_{74}	0.1	0.6	0.3	0	0
U_8	U_{81}	0	0.5	0.4	0	0
	U_{82}	0	0.5	0.5	0	0

再结合表1所确定的指标权重和表3的评价值，计算各二级指标和一级指标得分结果，见表4。

表4 各级指标模糊评判计算结果

一级指标	二级指标							
U	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8
80.47	85.1	82.25	86.52	79.35	85.38	84.12	84.82	74.70

由表4可得，西安理工大学教育信息化建设的综合得分为80.47。其中，基础设施建设得分为85.1，信息资源建设得分为82.25，信息技术教育得分为86.52，信息化应用得分为79.35，应用效果得分85.38，综合管理得分为84.12，人才队伍得分为84.82，经费投入得分为74.7。

计算结果显示，西安理工大学教育信息化建设整体上处于较好水平。其中，基础设施建设、信息技术教育、应用效果等处于较好偏上的水平；由于资金不足，经费投入和信息化应用处于一般水平；信息资源建设也处于较好偏下水平。因此，从该评价结果可以提出以下信息化发展对策：该校应进行信息化建设规划，并多渠道筹措资金，加大生均经费投入力度，以便进一步促进信息化的发展，同时应加强校园一卡通、远程教育建设，以提高信息化应用水平；另外还应重视信息资源建设，提供丰富的信息资源以满足师生的需求。

4 结 论

本文以我国现有的大学教育信息化测度指标为基础，吸取现有研究成果之所长，对我国大学教育信息化测度指标进行了设计修正，减少了定性指标的数量，相应增加定量指标，将一些总量指标变成了具有可比性的指标，删除了一些较难量化的指标及数据较难获得的指标，并增加一些替代指标以保持指标体系的完整。本文还对该指标体系的应用进行了个案研究，采用层次分析法确定了指标体系中各级指标的权重，可以为利用该指标体系进行的大学信息化实证研究提供参考。本文以笔者所在学校研究个案，运用模糊综合评价对西安理工大学的信息化水平进行了实证测度，测度结果与该校实际基本相符，测度方法可为相关研究借鉴。

参 考 文 献

1 赵书民.机遇与挑战:我国高等教育信息化的探讨.日本问题研究,2003;(4)

2 张 忠,张 萍.教育信息化的现状及对策.西南民族大学学报(人文社科版),2004;(10)

3 刘 文,葛敬民.国内外信息化水平测度理论研究比较.情报理论与实践,2004;(2)

4 朱桂娟.教育信息化水平测评方法研究.网络与信息化,2004;(5)

5 邵晋蓉,王桂香.宁夏教育信息化指标体系建设构想.现代情报,2004;(4)

6 闫 慧.教育信息化测度指标体系的设计.情报杂志,2004;(7)

7 刘军跃,徐 刚,黄伟九.高等教育信息化评价指标体系探讨.高教探讨,2004;(3)

8 陈爱娟,任晓燕.我国高等教育信息化测度研究理论综述.情报科学,2006;(9)

9 翁 佳,郑建明.我国信息化水平测度方法研究述评.情报杂志,2006;(5)

(责编:京梅)

(上接第69页)

6 Small H,Sweeney E. Clustering the Science Citaiton Index Using Co-Citation: 1 A Comparison of Methods. Scientometrics,1985;(3-6)

7 Small H,Sweeney E,Greenlee E. Clustering the Science Citaiton Index Using Co-Citation:2 Mapping Science. Scientometrics,1985;(5-6)

8 Small H. A SCI-Map Case Study: Building a Map of AIDS Research. Scientometrics,1994;(1)

9 White H D,Griffith B C. Author Co-Citation:A Literature Measure of IntellecTual Structure. Journal of The American Society for Information Science,1981;(3)

10 White H D, McCain K W. Bibliometrics. Annual Review of Information Science and Technology,1989;(24)

11 White H D, McCain K W. Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995. Journal of the American Society for Information Science,1998;(4)

12 McCain K W. Mapping Authors in Intellectual Space:A Technical Overview. Journal of the American Society for Information Science,1990;(6)

13 White H D. Pathfinder Networks and Author Cocitation Analysis: a Remapping of Paradigmatic Information Scientist. Journal of the American Society for Information Science and Technology,2003;(5)

14 宋丽萍,徐引篪.基于可视化的作者同被引技术的发展.情报学报,2005;(2)

15 Lin X. Map Displays for Information Retrieval. Journal of the American Society for Information Science,1997;(1)

(责编:阳)