国内外共引分析研究现状探析*

朱亮，孟宪学，赵瑞雪，鲜国建，寇远涛
（中国农业科学院农业信息研究所，北京 100081）

摘要：本文选取CNKI数据库、WoS数据库作为数据源，分别检索出以“共引分析”为主题的中外文文献数据，并就作者、期刊、机构等进行了统计分析。此外，以信息可视化软件CiteSpace为工具，绘制出“共引分析”领域的科学知识图谱，并进行解读。从分析结果来看，国际“共引分析”研究前沿包括了应用研究和方法研究两方面内容，主要有科学知识图谱绘制及分析、共引分析流程的改进和优化等；国内“共引分析”研究热点紧跟国际前沿，“科学知识图谱”相关研究占据了其中最重要的地位。

关键词：共引分析；科学知识图谱；CiteSpace；研究前沿
中图分类号：G353
DOI：10.3772/j.issn.1673-2286.2015.04.007

自美国情报学家Small于1973年首次提出“共被引（Co-citation）”概念以来，共引分析作为一种独特的科学计量方法，其理论研究与实践应用得到了广泛开展。共引又称同被引，若两篇文献同时被其他n篇文献所引用，则称这两篇文献具有共引关系，其同被引次数（即共引强度）为n，n越大则意味着这两篇文献的关系越密切。以此为原理，共引分析就是以具有一定学科代表性的一批文章（或著者或期刊）为分析对象，利用聚类分析、多维标度等多元统计分析方法，借助电子计算机，把众多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，使分析对象之间相互关系的格局清晰可辨，在此基础上研究分析对象所代表的学科及文献的结构和特点[1]。当前，根据分析对象的不同，共引分析主要包括文献共引分析、作者共引分析(Author Co-citation Analysis，ACA)、期刊共引分析、主题共引分析等。

1 方法和工具

通常情况下，研读经典文献，学习、了解研究前沿和热点，是全面认识一个学科领域发展脉络和研究现状的两个最直接有效的途径。目前，学科领域经典文献、研究前沿和热点的识别和探测方法有许多，以文献计量方法为主流，但文献计量方法的传统操作方式并不方便和直观，因此，现代图书情报人员尝试借助信息可视化、数据挖掘等技术，探寻一种结果更客观、操作更高效、结果更直观的新方法，在此背景下，科学知识图谱应运而生。CiteSpace是当前应用较为广泛的科学知识图谱的绘制工具，由美国德雷塞尔大学陈超美博士开发，主要用于对特定领域文献（集合）进行计量，以探寻学科领域演化的关键路径及转折点，探测学科领域研究前沿等。CiteSpace分析的数据主要来源于WoS、PubMed、Derwent等国际权威数据库，目前也已支持对CSSCI、CNKI等中文数据的分析，其应用

* 本研究得到国家“十二五”科技支撑计划项目“面向外文科技文献信息的知识组织体系建设与示范应用”子课题“基于STKOS的知识服务应用示范”（编号：2011BAH10B06）资助。

流程包含数据采集和处理、参数功能选择、可视化、图谱解读等四个步骤。根据节点选择的不同，CiteSpace可生成作者、机构、国家合作，术语、关键词共现，文献、作者、期刊共被引，文献耦合等类型丰富的科学知识图谱，以满足不同的应用需求。基于此，本文利用CiteSpace绘制科学知识图谱，实现对“共引分析”领域国际研究前沿及国内研究热点的探测。

2 数据来源

本文中文数据源选择CNKI数据库，采用检索式“主题=共引分析or共被引or同被引”，于2015年1月6日检索中国学术期刊网络出版总库（1980-2014），共检索到651条记录，由于CNKI数据库中的引文数据未完全对外开放，使得下载的文献记录中并不含相应的引文数据，因此，本文在对中文数据进行分析时，主要是利用CiteSpace探寻国内“共引分析”研究热点，而对重要作者、期刊和机构的分析，则是通过统计方法来实现。

外文分析数据来源于WoS数据库，引文数据库选择：Science Citation Index Expanded（SCI-EXPANDED），1995年至今；Conference Proceedings Citation Index - Science（CPCI-S），1995年至今。时间选择为所有年，文献类型选择为：ARTICLE、PROCEEDINGS PAPER、REVIEW，检索式为：主题=“co-citation analysis” or “cocitation analysis”，共检索到299条记录，数据下载日期为2015年1月6日。

3 共引分析国内外研究现状分析

651篇国内论文中共包含945名作者，其中发表论文10篇及以上的作者共6名，包括了邱均平、刘则渊、崔雷、侯剑华等国内图书情报领域的知名专家或学者。299篇外文文献共有514位作者，其中，发文量超过10篇的共有三位，分别是Chen CM、McCain KW和Zhao DZ，发文数量均为11篇，国内外发文量排名前十的作者如表1所示。从国内外作者的学科领域来看，绝大多数均来自图书情报领域，长期从事着科学计量、科技评价、科学知识图谱等方面研究，这也表明“共引分析”在图书情报学研究方法体系中的重要地位。此外，在地理分布上，299篇外文文献分别来自美国（86篇）、中国（含台湾地区）（65篇）、加拿大（20篇）、英国（18篇）、德国（15篇）等30多个国家和地区，值得注意的是，中国的论文数量仅次于美国，排在了第二位，占到了论文总量的21%以上，这说明我国学者已成为国际“共引分析”研究舞台上的一支活跃力量。

表1　国内外“共引分析”研究发文量排名前十作者

排名	国内		国外	
	作者	发文量（篇）	作者	发文量（篇）
1	邱均平	34	Chen CM	11
2	刘则渊	20	McCain KW	11
3	崔　雷	19	Zhao DZ	11
4	侯剑华	18	Boyack KW	8
5	杨思洛	11	Chen TT	8
6	姜春林	10	Strotmann A	8
7	赵蓉英	9	Ding Y	7
8	王贤文	9	Klavans R	7
9	房宏君	9	Lee MR	7
10	侯海燕	8	White HD	7

从发文机构来看，651篇国内论文共来自484家机构，大连理工大学、武汉大学、中国科学院文献情报中心、中国医科大学、南京大学等高校、科研院所的发文量排在了前列。299篇外文文献出自250余家机构，机构发文量排在前三的分别是Drexel大学、Indiana大学、Alberta大学，国内外发文量排名前十的机构如表2所示。由表2可知，在机构类型方面，国内和国际呈现出的特点比较一致，均以高校为主，而且国内发文量排名前两位的大连理工大学、武汉大学也出现在了国际“共引分析”研究重要机构的前十名单中，分别排在发文数的第五位和第六位，这也进一步说明这两所高校不仅是国内“共引分析”研究的领军者，也是国际“共引分析”领域的重要力量。

从文献载体来看，651篇中文文献中的绝大多数都刊载于《情报杂志》、《图书情报工作》、《情报科学》等情报学、图书馆学的核心期刊上，这也是国内“共引分析”研究与应用成果研讨的最主要阵地。299篇外文文献则分布在119种不同的出版物上，载文量排在前两位的Scientometrics和Journal of the American Society for Information Science and Technology，无论是从载文量（分别为74篇和46篇），还是在领域内的被引频次（分别为192和317），都远远超过其他期刊，所以这两者是领域内的国际权威期刊。“共引分析”领域国内外载文量排名前十的出版物如表3所示。

表2 国内外“共引分析”研究发文量排名前十机构

排名	国内		国外	
	机构	发文量（篇）	机构	发文量（篇）
1	大连理工大学	69	Drexel Univ	11
2	武汉大学	66	Indiana Univ	11
3	中国科学院	59	Univ Alberta	11
4	中国医科大学	28	Univ Granada	8
5	南京大学	26	Wuhan Univ	8
6	大连大学	13	Dalian Univ Technol	8
7	天津师范大学	13	SciTech Strategies Inc	7
8	中国科学技术信息研究所	13	Nanyang Technol Univ	7
9	湘潭大学	12	Royal Sch Lib & Informat Sci	7
10	辽宁师范大学	11	SE Missouri State Univ	7

4 共引分析研究前沿与热点分析

4.1 国际共引分析研究前沿分析

研究前沿（Research front）是科学研究中最先进、最新、最有发展潜力的研究主题或方向，通常代表了科学发展的难点、热点以及发展趋势[2]。陈超美将研究前沿定义为一组突现的动态概念和潜在的研究问题，其知识基础是研究前沿概念所在文献的引用文献簇，研究前沿与知识基础相互作用并动态发展[3]。在CiteSpace中，研究前沿是基于从文献题目、摘要、关键词和标识中提取出的突变专业术语而确定的。CiteSpace采用J. Kleinberg于2003年提出的突发监测算法[4]来识别突变专业术语。

首先从299条外文文献的标题、摘要和关键词中提取名词短语（Noun Phrases），再利用CiteSpace内置算法“Detect Bursts”共探测到123个突变专业术语（burst term）。术语类型选择“Burst Terms”，节点类型选择“Term”和“Cited Reference”，时区跨度为1995-2014，时间切片为1年，分别将前、中、后三个时间分区的被引频次、共引频次和共被引率（c,cc,ccv）阈值设置为（2，2，20）、（3，3，20）、（4，3，20）。运行软件，生成突变专业术语共词与文献共被引构成的混合网络图谱，网络中包含259个被引文献节点和23个突变专业术语节点，共有1744条边，图谱的时间线视图（Timeline）如图1所示。

网络中共有9个聚类，分别采用TF*IDF加权算法和对数似然率算法（LLR）从施引文献标题词条中提取出聚类标签，排在第一位的是该方法生成的首选标签，聚类信息见表4。

根据表4对“共引分析”领域国际研究前沿分析如下：

（1）9个聚类中规模最大者为聚类#0，表示的是科学知识图谱研究和应用，包括方法、技术及目的等，

表3 “共引分析”领域国内外载文量排名前十出版物

排名	国内		国外	
	出版物	载文量（篇）	出版物	发文量（篇）
1	情报杂志	51	Scientometrics	34
2	图书情报工作	48	Journal of the American Society for Information Science and Technology	20
3	情报科学	45	Information Processing & Management	19
4	情报理论与实践	35	Journal of Information Science	18
5	现代情报	22	Proceedings of ISSI 2011: The 13th Conference of The International Society for Scientometrics and Informetrics	11
6	中国图书馆学报	18	Annual Review of Information Science and Technology	10
7	科学学研究	17	Journal of The American Society for Information Science	9
8	情报探索	16	ASIST 2002: Proceedings of The 65th ASIST Annual Meeting	9
9	医学信息学杂志	15	European Journal of Information Systems	9
10	图书与情报	12	Journal of The Association for Information Science and Technology	8

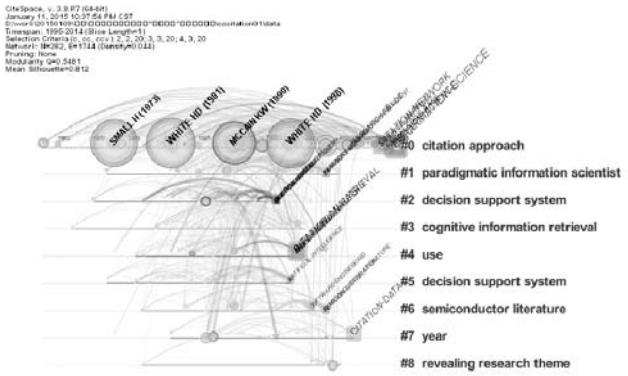

图1 被引文献共引与突变术语共词混合网络知识图谱（时间线视图）

为了证明这一研究前沿的重要性，由表5所示的文献被引突现信息可知，排在前4位的高突现性论文均来自该聚类，包含了陈超美[3,5]、BOYACK KW[6]在2006和2010年发表的三篇科学知识图谱研究重要文献，且其突变时间均为2011-2014，这就从内容和时间上说明了科学知识图谱研究及应用是当前国际“共引分析”领域最新最重要的研究前沿。此外，聚类#8所表示的主题揭示、趋势分析、可视化等内容也都属于该前沿范畴。

（2）聚类#1表示共引分析方法和技术研究，内容包括Pearson相关系数、相似度计算、网络寻址定位、数据挖掘、信息计量（共链分析）等。近些年来，围绕着共引分析流程的优化，国际图书情报学界开展了广泛的讨论。2005年以前，该讨论主要集中在相似系数的选取方面；2006年以后，讨论的内容便扩展到了相似系数矩阵的生成方式方面：是该利用共引矩阵还是引文矩阵。在作者共引分析（ACA）研究及应用方面，针对McCain提出的传统ACA 模式需要大量计算与绘图操作、计算强度高的不足，学者们不断引入网络寻址定位（Pathfinder Network Scaling，PFNETs）、自组织映射（SOM）等新技术对其进行改进，降低了ACA 的计算强度，结果更为可信。

（3）聚类#3和#4表示的是信息检索研究，信息检索是共引分析的一个重要应用领域，相关实践已有不少，其中最主要的方向是基于共引分析产生的文献网络、智力网络等进行检索结果关联展示。聚类#5、#6和#7由于模块较小，且内容分散，难以形成对研究前沿的聚焦，本文就不再具体分析。

4.2 国内共引分析研究热点分析

关键词是一篇文章主题内容的高度概括和凝练，对高频关键词进行分析来确定学科领域研究热点是目前研究人员较常采用的方法之一。本文利用CiteSpace绘制出中文关键词共现网络知识图谱，实现对“共引分

表4 共词与共引混合网络图谱聚类信息

聚类号	包含节点数	聚类标签（TF*IDF）	聚类标签（LLR）
#0	77	citation approach;direct citation; measure;technique;current trend	bibliometric mapping; technique;research front
#1	41	paradigmatic information scientist; pearsons correlation coefficient; data mining;informetric	paradigmatic information scientist; pathfinder network;remapping
#2	35	decision support system; subspecialties;organizational science	decision support system; intellectual development; empirical investigation
#3	32	cognitive information retrieval; economic; conjunct subject	informetric; cognitive information retrieval; bibliometric
#4	31	use; information retrieval user studies; author co-citation analysis	information retrieval studies; author co-citation analysis; intellectual structure
#5	21	decision support system; empirical investigation; reference discipline	empirical investigation; reference discipline;
#6	17	semiconductor literature; software engineering journal; index	semiconductor literature; vocabulary assignment; software engineering journal
#7	16	year;alternative perspective; european conference;citation classic	paradigm; morgan
#8	12	revealing research theme; progressive knowledge domain visualization; trend; tool;	trend;revealing research theme; knowledge management

表5 排名前十的高突现性论文信息表

序号	突现值	论文	突现时间区间	聚类号
1	7.07	Boyack KW，2010，J AM SOC INF SCI TEC，V61，P2389	2011-2014	#0
2	6.93	Chen CM，2006，J AM SOC INF SCI TEC，V57，P359	2011-2014	#0
3	6.4	White HD，2003，J AM SOC INF SCI TEC，V54，P1250	2005-2009	#0
4	6.37	Chen CM，2010，J AM SOC INF SCI TEC，V61，P1386	2011-2014	#0
5	5.15	White HD，1989，ANNU REV INFORM SCI，V24，P119	2000-2003	#1
6	4.55	Zhao DZ，2008，J AM SOC INF SCI TEC，V59，P2070	2010-2014	#0
7	4.45	Ahlgren P，2003，J AM SOC INF SCI TEC，V54，P550	2005-2006	#0
8	4.37	Van Eck NJ，2010，SCIENTOMETRICS，V84，P523	2012-2014	#0
9	4.13	Leydesdorff L，2006，J AM SOC INF SCI TEC，V57，P1616	2008-2009	#0
10	4.09	Zhao DZ，2008，J AM SOC INF SCI TEC，V59，P916	2009-2014	#0

注：高突现性论文是指那些被引频次出现突增的论文，包含突现值和突现时间两个维度，突现性高的论文意味着其在相应的时间区间里受到了格外的关注，一定程度上代表了所在研究领域在相应时间区间的研究前沿。

表6 “共引分析”领域中文关键词出现频次信息（Top20）

序号	关键词	出现频次	所属聚类
1	科学知识图谱	152	4
2	信息可视化	108	4
3	文献计量学	86	2
4	引文分析	79	0
5	Citespace	73	4
6	社会网络分析	66	4
7	作者共被引分析	42	4
8	聚类分析	39	3
9	研究前沿	39	4
10	研究热点	34	4
11	情报科学	26	3
12	共词分析	23	5
13	科学计量学	22	4
14	共现分析	20	4
15	被引分析	15	2
16	知识结构	15	4
17	CSSCI	15	4
18	多维尺度分析	14	3
19	网络计量学	12	2
20	核心著者	11	2

析”领域国内研究热点的探测与分析。

由于中文分析数据中存在着不同词语表达同一概念及用词不规范等现象，如“共引”、“共被引”与“同被引”，“文献计量学”与“文献计量”等，本文首先对网络中的重要节点进行了归并，由于本文的研究对象选取的是“共引分析”领域，为了不对热点分析形成干扰，我们去掉了出现频次（168）排名第一的“共引分析”，得到如下关键词出现频次信息表（表6）。

表6中出现频次超过100的关键词有两个，其中排在第一位的是“科学知识图谱”，其出现频次为152，从数量上远远大于其他关键词，足以体现其在国内“共引分析”领域得到的关注度之高。再结合所属聚类信息来看，表中与其同在一个聚类（聚类#4）中的关键词还有："信息可视化"、"Citespace"、"社会网络分析"、"作者共被引分析"、"研究前沿"、"研究热点"、"科学计量学"、"共现分析"、"知识结构"、"CSSCI"，这些关键词基本涵盖了“科学知识图谱”研究的各个重要环节，包括数据来源、绘制方法和技术、现有系统及工具、应用目的等。因此，可以认为“科学知识图谱”相关研究是当下国内“共引分析”领域最重要的研究热点，其典型的应用流程可概括为：以包含CSSCI在内的中外文数据库作为数据源，利用社会网络分析、共现分析、信息可视化等方法与技术，借助Citespace等系统工具，绘制特定学科领域的科学知识图谱，实现其知识结构的展示、研究前沿和热点的分析。

继续分析表6，可得出国内“共引分析”领域的研究热点还包括：共引分析操作流程中所涉及的聚类分析、多维尺度分析等具体方法研究，以及共引分析的研究及应用范围从传统文献计量学领域向新兴网络计量学领域拓展等。

5 结语

经过近40余年的发展，共引分析已成为国际图书情报领域一种重要且高效的研究方法。本文对CNKI数据库、WoS数据库中收录的以“共引分析”为主题的中外文文献进行了统计分析，并利用CiteSpace软件绘制出共引分析领域的科学知识图谱，实现了其国际研究前沿和国内研究热点的探测。当前，国际“共引分析”研究前沿包括了应用研究和方法研究两方面内容，其中应用研究占据了主流，包括科学知识图谱绘制及分析、关联信息检索等；而在方法本身的研究方面，国际图书情报学界讨论的重点则是共引分析流程的改进和优化。总体来看，国内“共引分析”研究热点紧跟该领域的国际前沿，“科学知识图谱”相关研究占据了其中最重要的地位。此外，本文还对CiteSpace软件的使用流程，以及针对不同应用目的所采取的科学知识图谱分析与解读方式等内容进行了阐述，可为今后同类、相关研究提供思路和方法参考。

参考文献

[1] 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[J]. 情报杂志,1993,12(2):36-42.

[2] 陈仕吉. 科学研究前沿探测方法综述[J].现代图书情报技术,2009(9): 28-33.

[3] Chen C M. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature[J].JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 2006, 57(3): 359-377.

[4] Kleinberg J. Bursty and Hierarchical Structure in Streams[C]// PROCEEDINGS OF THE 8TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. Canada: ACM Press,2002:1-25.

[5] Chen C M. The Structure and Dynamics of Cocitation Clusters: A Multiple-Perspective Cocitation Analysis[J]. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 2010, 61(7):1386-1409.

[6] Boyack K W, Klavans R. Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately?[J]. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 2010, 61(12):2389-2404.

作者简介

朱亮，男，1981年生，助理研究员，在读博士，主要从事文献计量、情报分析研究，E-mail：zhuliang@caas.net.cn。

Analysis of Co-citation Analysis Research at Home and Abroad

ZHU Liang, MENG XianXue, ZHAO RuiXue, KOU YuanTao, XIAN GuoJian
(Agricultural Information Institute of CAAS, Beijing 100081, China)

Abstract: In this paper, We retrieved “co-citation analysis” data records from CNKI and WoS, and made use of statistics method to analyze the important authors, journals, institutions of the literature. In addition, we used the visualization software CiteSpace to draw the scientific knowledge map of co-citation analysis. International research fronts in co-citation analysis field includes application research and methods study, such as mapping knowledge domain and analysis, the improving and optimizing of co-citation analysis processes, etc. Domestic research focus in co-citation analysis field followed the international research fronts, the most important content is the related research of mapping knowledge domain.

Keywords: Co-citation Analysis; Mapping Knowledge Domains; CiteSpace; Research Front

（收稿日期：2015-04-01；编辑：王立学）