

国外知识图谱研究的文献计量分析

郭颖涛 杨思洛 覃惠迪

(湘潭大学公共管理学院, 湖南湘潭 411105)

摘要: 利用知识可视化软件CiteSpace分析Web of Science中收录的2003—2012年知识图谱研究相关论文, 分别对相关文献的时间分布、期刊分布、学科分布和地区分布, 作者合作网络、机构合作网络以及学科代表人物和经典作品等进行分析, 直观地揭示国外知识图谱研究的发展现状, 并通过关键词共现分析确定国外知识图谱研究热点, 通过主题词突变检测描绘国外知识图谱研究前沿与趋势。

关键词: 知识图谱; 知识可视化; 共被引分析; 合作网络; CiteSpace; 文献计量分析

分类号: G350

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2013.06.014

Visualization Analysis of Foreign Mapping Knowledge Domain Research

Guo Yingtao, Yang Siluo, Qin Huidi

(Department of Public Management, Xiangtan University, Xiangtan 411105)

Abstract: This paper utilizes knowledge visualization software of CiteSpace to analyze thesis related to mapping knowledge domain in Web of Science from 2003 to 2012. It makes a concrete analysis of time distribution, journal distribution, discipline distribution, area distribution as well as co-author network, co-institution network, high-influence authors and literature which reveal development status quo of foreign mapping knowledge domain research intuitively. It detects hot topics of researches by analysis of keyword co-occurrence. It describes research front and trend by term burst detection.

Keywords: mapping knowledge domain, knowledge visualization, co-citation analysis, co-author network, Citespace, visualization analysis

知识图谱是对科学知识以可视化的形式直观展示, 且深入知识内部对其发展进程与结构关系进行系统分析。具体来说, 就是把应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法同计量学引文分析、共现分析等方法结合, 用可视化的图谱形象地展示学科的核心框架、发展历史、前沿领域以及整体知识架构的多学科融合的一种研究方法^[1]。科学知识图谱于20世纪70年代发端于国外。2003年, 美国科学院组织的“知识图谱测绘”大会的召开标志着专家们对知识图谱绘制进入大规模的深入研究阶段。随着学术研究和应用研究的不断发展, 知识图谱理论与方法体系日益成熟, 利用

知识图谱方法来探寻科学研究前沿及发展动态已经成为一种潮流, 广泛应用于人文社会科学与自然科学领域。

经过检索, 国内学术论文偏重于研究国内知识图谱, 对国外知识图谱的研究存在一定欠缺。有鉴于此, 本文利用知识可视化软件CiteSpace对国外知识图谱的研究演进与发展现状进行可视化展示, 并借助文献题录信息统计分析工具SATI^[2]进行文献计量分析。

1 数据来源

对Web of Science (包括SCI-EXPANDED和SSCI)

第一作者简介: 郭颖涛(1986—), 男, 湘潭大学公共管理学院硕士研究生, 研究方向: 信息计量。

基金项目: 国家社科基金项目“中外图书情报学科知识图谱比较研究”(11CTQ015)。

收稿日期: 2013年3月22日。

进行主题检索。检索表达式为：主题= (Science map*) OR 主题= (Bibliometric map*) OR 主题= (Literature map*) OR 主题= (Map* Knowledge Domain) OR 主题= (Scientometrics map*) OR 主题= (informetrics map*) OR 主题= (Science visual*) OR 主题= (Bibliometric visual*) OR 主题= (Literature visual*) OR 主题= (Scientometrics visual*) OR 主题= (informetrics visual*) OR 主题= (coauthor visual*) OR 主题= (coauthor map*) OR 主题= (cit* map*) OR 主题= (cit* visual*) OR 主题= (coupl* visual*) OR 主题= (coupl* map*), 检索时间为2012年12月1日, 学科范围限定在 Information and Library Science, 最终得到800条有效引文数据, 发表年代自2003年1月1日至2012年12月1日。

2 研究的时空分布

2.1 发文量与载文期刊

国外知识图谱研究年度分布曲线见图1。由于数据采集时间的原因, 2012年发表的论文未完全统计。2003—2012年个别年份的发文量出现突降现象, 但整体来讲, 发文量呈现快速增长趋势, 而且

近几年学者越来越多地关注知识图谱在科学研究中的利用价值。

国外知识图谱研究论文主要分布在图书情报学科期刊上, 其中载文量排名前10的高影响力期刊见表1。排名前3位的期刊大量刊载知识图谱研究论文, 一方面说明这3种期刊质量高、认可度高, 另一方也说明知识图谱与科学计量学、信息科学、地理信息科学之间存在着非常密切的联系。Scientometrics (《科学计量学》) 是世界科学计量学的顶级权威刊物, 该期刊是荷兰的一种定量研究科学学、科学交流和科学政策方面的国际期刊, 在计算机领域和图书情报领域影响力很大。Journal of the American Society for Information Science and Technology (《美国信息科学与技术学会杂志》) 主要探讨信息科学的理论与应用, 反映最新进展。International Journal of Geographical Information Science (《国际地理信息科学杂志》) 主要刊载地理信息系统设计和运用方面的文章。

值得一提的是, 10种高影响力期刊中, 英国占据5种, 荷兰2种, 美国3种, 英、荷、美三国对传播知识图谱研究最新成果、提升知识图谱研究水平



图1 国外知识图谱研究年度分布曲线

表1 国外知识图谱研究高影响力期刊

高影响力期刊	载文量	国家
Scientometrics	171	荷兰
Journal of the American Society for Information Science and Technology	131	美国
International Journal of Geographical Information Science	56	英国
Journal of Informetrics	40	荷兰
Information Processing & Management	25	英国
Information Research—An International Electronic Journal	22	英国
Journal of Documentation	19	英国
Journal of Information Science	17	美国
Aslib Proceedings	16	英国
Journal of the Medical Library Association	15	美国

作出了重大贡献。

2.2 研究涉及的学科

由表2可以看出，国外知识图谱研究涉及众多学科，信息科学和图书馆学、计算机科学、地理学、管理学、传播学、医学信息学、教育学等学科都有知识图谱相关研究，其中信息科学和图书馆学、计算机科学、地理学对知识图谱理论、工具、方法、技术的发展和应用起着尤为关键的作用。

管理学、传播学、医学信息学、教育学等其他学科也适时地抓住知识图谱发展成熟的契机，借助知识图谱研究学科研究进展和当前的研究前沿及其对应的知识基础。

2.3 开展研究的国家

国外知识图谱研究的国家见图2，涵盖世界六大洲。整体来讲，北美洲、欧洲的知识图谱研究处于世界领先地位，尤其是美国、英国、荷兰、西班牙等国。

表2 国外知识图谱研究学科发文统计

学科	发文量
Information Science Library Science	800
Computer Science Information Systems	337
Computer Science Interdisciplinary Applications	186
Geography	56
Geography Physical	56
Management	43
Communication	11
Social Sciences Interdisciplinary	10
Medical Informatics	8
Education Educational Research	4
Health Care Sciences Services	3
Computer Science Theory Methods	2
Humanities Multidisciplinary	2

3 研究作者

3.1 影响力

国外知识图谱高影响力作者见表3。发文量排

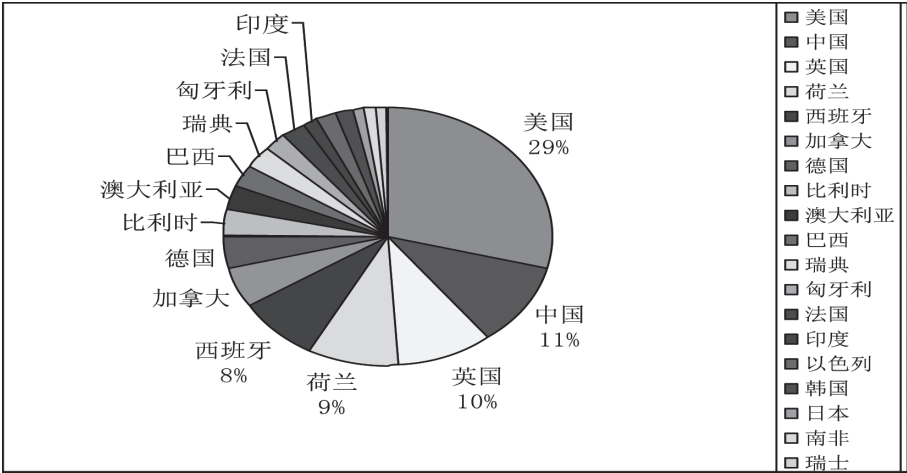


图2 国外知识图谱研究的国家(地区)分布

表3 国外知识图谱研究高影响力作者

高产作者		高被引作者			
作者	发文量	作者	频次	中心度	首次被引年份
Leydesdorff L	41	Leydesdorff L	185	0.27	2002
De Moya-Anegón F	16	Small H	158	0.17	1997
Glanzel W	14	White HD	126	0.08	1989
Boyack Kw	12	Garfield E	116	0.10	1979
Chen Cm	10	Boyack KW	90	0.32	2002
Rafols I	10	Chen CM	89	0.18	2002
Klavans R	9	Mccain KW	84	0.18	1991
Vargas-Quesada B	9	Borner K	74	0.27	2003
Janssens F	8	Glanzel W	72	0.06	2003
De Moor B	8	Price DJD	63	0.30	1965

名前列有 Leydesdorff、Moya-Agegon、Glanzel。Leydesdorff 是知识图谱研究的高产作者，他在信息可视化、社会网络分析、科学计量等诸多方面均有建树。他提出利用双连接图形分析算法可视化知识结构网络，开发了处理共词、合作、耦合、共引等知识单元关系的系列软件，并通过外部可视化软件展示知识关系。

作者被引频次排名前列者有 Leydesdorff、Small 和 White。其中，Leydesdorff 是被引频次最高的作者。Small 吸收 Price 和 Garfield 的科学引文网络思想，创造性地提出了论文“共被引”的概念和共被引分析的方法。White 提出共引分析要与多元统计分析结合起来使用。值得一提的是，华人学者陈超美，其发文量和被引频次均有不错的排名，他开发的 CiteSpace 软件开创了以知识领域为分析单元的可视化综合性学术与应用领域，把对科学前沿的知识计量和知识管理研究推进到以知识图谱与知识可视化为辅决策重要手段的新阶段^[3]。

在中心度方面，Boyack、Price、Leydesdorff 和 Borner 排名前列。Boyack 是中心度最高的作者。高中心度作者在共被引网络中占据着重要地位，中心度的高低在一定程度上代表了该作者影响力的大小。

从首被引年份可以看出作者相关研究的先后时间。Price、Garfield、White 等奠定了知识图谱研究的基础，后继学者 McCain、Small、Boyack、陈超美等继承并发展了相关的理论，拓展了相关的应用领域。

3.2 合作网络

选取被引频次靠前的 50 位作者，使用 SATI 导

出作者共现矩阵，经 Ucinet 可视化可得到研究作者合作网络（图 3）。图 3 中较大节点标识高点度中心性作者。在该网络中，点度中心性等于与一个作者共同出现的关联作者的数量。因此，一个具有高点度中心性的作者必定与许多作者有联系，这也就意味着该作者是整个网络的核心作者。Leydesdorff 与 Moya-Anegon 点度中心性最高，为 7。部分作者点度中心性为 1 或 0。整体来讲，国外知识图谱研究作者合作网络相对稀疏，作者合作研究仍有较大上升空间。

图 3 中节点连线的粗细代表合作次数的多少。Boyack 与 Klavans 合作次数最多，合作论文《Drawing the Backbone of Science》被广泛引用，影响深远。Glanzel、Janssens、De Moor 三人彼此都有紧密合作，他们的研究具有较高的相似性和互补性。

4 研究机构

国外知识图谱研究高发文机构见表 4。这些机构大多是高校且集中分布在欧洲和美国。格拉纳达大学、阿姆斯特丹自由大学、天主教鲁汶大学排名前 3。印第安纳大学和德雷塞尔大学排名紧随其后。可见，欧洲和美国的知识图谱研究走在世界的前列。武汉大学是中国的佼佼者，能够跻身世界前列代表了武汉大学雄厚的科研实力。

选取发文量靠前的 50 个机构，使用 SATI 导出机构共现矩阵，经 Ucinet 可视化可得到国外知识图谱研究机构合作网络（图 4）。图 4 中较大节点标识高点度中心性机构。Georgia Inst Technol 点度中心性最高，为 7。与作者合作网络相似，国外知识图谱

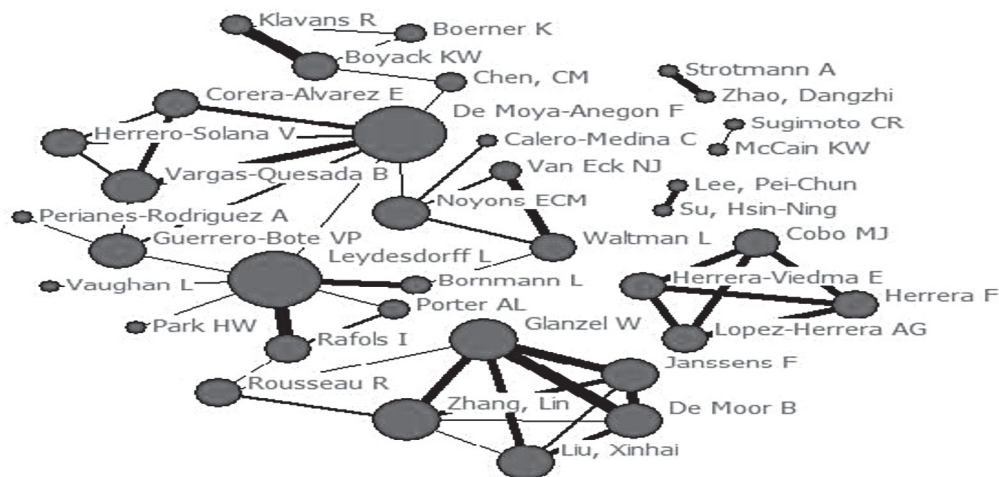


图 3 国外知识图谱研究作者合作网络

表4 国外知识图谱研究高发文机构

机构	国家	发文量
Univ. Granada	西班牙	43
Univ. Amsterdam	荷兰	42
Katholieke Univ. Leuven	比利时	37
Indiana Univ.	美国	24
Hungarian Acad Sci	匈牙利	19
Drexel Univ.	美国	18
Leiden Univ.	荷兰	18
Univ. Sussex	英国	16
Wuhan Univ.	中国	16
SciTech Strategies Inc.	美国	15

研究机构合作网络相对稀疏,机构合作研究仍有较大上升空间。

图4中节点的粗细代表合作次数的多少。Katholieke Univ. Leuven与Hungarian Acad Sci合作次数最多。大连理工大学与Hungarian Acad Sci合作3次。通过高层次的对外交流与合作,大连理工大学已经走在国内知识图谱研究的前列,国际影响力日益增强。

5 被引论文

国外知识图谱研究论文共被引图谱见图5。论文详细信息见表5。图5中较大节点标识了国外知识图谱研究的高影响力论文。1973年,Small在发表的《科学文献的共引分析:两文献间联系的新测度》论文中创造性地提出了文献共被引的概念和方法,定义了共引强度以测量论文之间的共引程度,

认为共引是测量两篇文献相关度的一个新工具,并以1971年的SCI数据库中的粒子物理学为例,绘制了领域论文的共引网络图^[4]。1981年White发表的《作者共被引:知识结构的文献测度》论文被认为是作者共被引分析的又一经典之作。该文通过SSCI数据库,分析了情报学作者共引情况,划分了情报学的学科结构^[5]。1998年,White和McCain发表了《可视化一个学科:1972-1995年信息科学作者共被引分析》。该文通过因子分析、聚类分析可视化展示了1972-1995年信息科学学科机构及其高影响力作者聚类群体^[6]。2005年,Boyack、Klavans和Borner发表了《绘制科学的“骨架”》,通过对自然科学和社会科学期刊论文共被引和聚类分析来可视化展示科学机构^[7]。1989年,Kamada发表了《一种绘制一般无向图的算法》,介绍了一种图布局算法,利用力作用方式,不断调整点的位置达到最小能量完成布局^[8]。2003年,Ahlgren发表了《共被引相似度测量要求:参考皮尔森相关系数》,提出共被引相似度测量需要满足两个自然要求,皮尔森系数并不是相似度测量的最佳选择^[9]。1965年,Price发表在《科学》上的《科学论文的网络》具有最高的中心性,认为科学论文之间的引证和被引证关系以及由此形成的所谓“引证网络”,显示了科学研究前沿的本质特征。1990年,McCain发表了《绘制智力空间中的作者图谱》。该文对作者共引分析方法(ACA)从数据的收集、整理、聚类、可视化、分析结论等进行了全面的总结^[10]。2009年,Leydesdorff发表了《新的期刊评价指标如何添加到

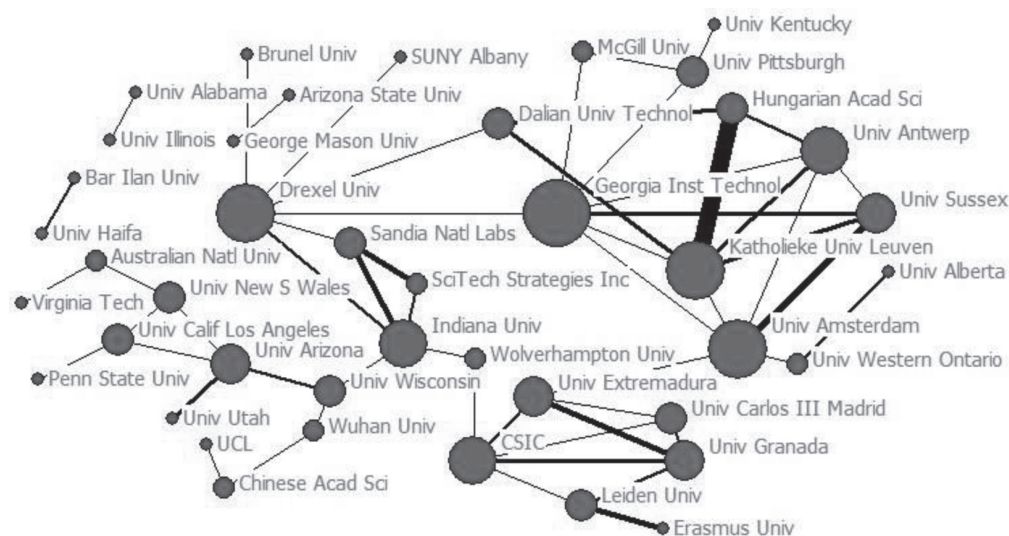


图4 国外知识图谱研究机构合作网络

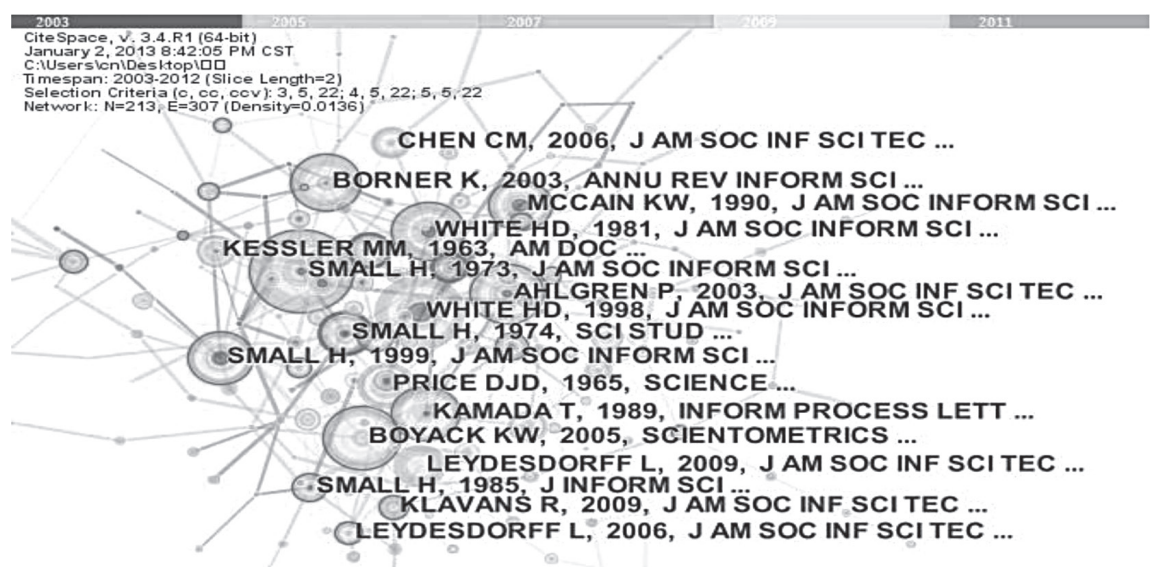


图5 国外知识图谱研究论文共被引图谱

表5 国外知识图谱研究被引论文

被引量	中心度	发表年份	作者	发表刊物
75	0.32	1973	Small H	Journal of the American Society for Information Science And Technology
69	0.17	1998	White HD	Journal of the American Society for Information Science And Technology
56	0.43	1981	White HD	Journal of the American Society for Information Science And Technology
55	0.40	2003	Borner K	Annual Review of Information Science and Technology
55	0.12	2003	Ahlgren P	Journal of the American Society for Information Science And Technology
55	0.45	2005	Boyack KW	Scientometrics
54	0.15	1989	Kamada T	Information Processing Letters
51	0.34	1999	Small H	Journal of the American Society for Information Science And Technology
47	0.24	1990	Mccain KW	Journal of the American Society for Information Science And Technology
46	0.11	1965	Price DJD	Science
38	0.29	2009	Leydesdorff L	Journal of the American Society for Information Science And Technology

文献计量工具箱中》，论述了h指数、PageRank、The Scimago Journal Ranking等新型评价指标各自的优缺点，并与旧评价指标进行了比较^[1]。

从时间和内容层面考虑上述高被引论文，笔者发现，国外知识图谱研究的经典文献集中在引文理论、知识图谱理论与方法、科学计量学评价指标等方面，基础理论研究和方法论研究在继承与创新中走向成熟。

6 论文关键词

6.1 关键词频次

论文的关键词是其内容的浓缩和提炼，关键词的分布频次与特征能显示某领域总体特征、研究内

容之间的内在联系、学术研究的发展脉络与发展方向、学术研究的重点与热点等^[12]。研究关键词共现图谱见图6。在图6中出现频次最高的关键词是科学，出现了173次；其次是网络，出现了73次。其他频次大于30的关键词有技术、指标、信息、地图、知识、可视化、信息科学、共被引、引文、文献计量学、模型、模式、信息检索、知识管理、引文分析等。

知识图谱以科学学为基础，以科学知识为对象，是科学计量学和信息计量学的新发展。借助科学知识图谱，人们可以理顺当代知识大爆炸形成的复杂知识网络。从这个意义上讲，科学知识和知识网络是知识图谱的灵魂所在。随着信息技术的发展

展,信息可视化研究取得重要突破,知识图谱得以进入可视化研究阶段。进入可视化阶段后,知识网络就可以通过可视化知识图谱的方式加以展示。

图谱中出现频次大于20的关键词,大致可归为以下六大方面:一是知识网络,具体关键词包括寻径网络、网络分析、社会网络;二是引文分析,具体关键词包括共词分析、共被引分析、词频分析等;三是信息计量,具体关键词包括指标、文献计量学、科学计量学、h指数;四是可视化,具体关键词包括图表、地图、可视化;五是图书馆学,具体关键词包括知识管理、信息检索、数字图书馆、图书馆学;六是科研合作创新,具体关键词包括交

流、创新、科研合作。这六大方面在一定程度上代表了当前国外知识图谱研究的热点主题。

6.2 关键词突变图谱

把国外知识图谱研究论文近10年的题录数据,利用CiteSpace软件中提供的膨胀词探测技术和算法,通过考察词频的时间分布,将其中频次变化率高的词从大量的主题词中探测出来,并根据词频的变动趋势,而不仅仅是频次的高低,来确定引文研究的前沿领域^[13]。

国外知识图谱研究关键词突变图谱见图7。从图7可以看出,频次变化率较高的一些主题词,一是与引文分析相关的膨胀词,主要有引文矩阵、被

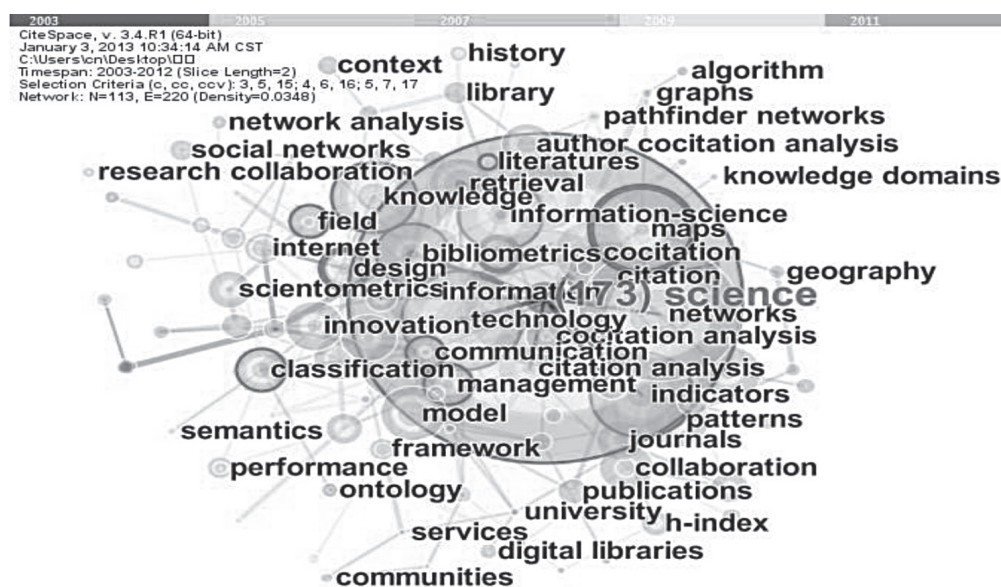


图6 国外知识图谱研究关键词共现图谱



图7 国外知识图谱研究关键词突变图谱

引期刊、科学引文、共被引分析、引文环境、作者共被引分析;二是与知识管理相关的膨胀词,主要有知识型员工、新知识观、知识结构、知识创造;三是与知识图谱技术相关的膨胀词,主要有信息技术、信息检索、科学地图、数据集、认知模型等。由此可见,引文分析、知识管理、知识图谱技术是当前国外知识图谱研究的前沿领域。

通过参阅相关论文,笔者发现,引文环境是当前引文分析领域的一个前沿课题,引文环境的自动提取、多引文环境下摘要的自动生成、不同引文环境下科研评价指标的差异等都是学者关注的焦点。新知识观是当前知识管理领域的一个前沿课题,如何引导用户把更多精力放在理解和解释信息而不是搜索信息上是学者关注的热点。引用路径展示是当前知识图谱技术领域的一个前沿课题,同时使用施引图和被引图可以完整展示引用的来龙去脉,分析比较多个数据集。

7 结论

(1) 国外知识图谱研究整体水平较高,基础理论研究与方法论研究成果突出,不过科研合作圈子还有待拓展。

(2) 国外知识图谱研究集信息科学、图书馆学、计算机科学、地理科学等诸多学科的理论、方法、技术于一身,成熟的知识体系使得知识图谱被广泛应用于自然科学及人文社会科学的应用研究之中。

(3) 国外知识图谱研究人才辈出,涌现出Price、Garfield、White、McCain、Small、Boyack、Glanzel、Leydesdorff、Borner、陈超美等杰出人物。

(4) 格拉纳达大学、阿姆斯特丹自由大学、天主教鲁汶大学是国外知识图谱研究的三大代表性杰出机构。

(5) 国外知识图谱研究的经典文献集中在引文理论、知识图谱理论与方法、科学计量学评价指标等方面,基础理论研究和在方法论研究在继承与创新中走向成熟。

(6) 知识网络、引文分析、信息计量、可视化、图书馆学、科研合作创新是当前国外知识图谱研究的热点主题。

(7) 引文分析、知识管理、知识图谱技术是当前国外知识图谱研究的前沿领域。引文环境是当前引文分析领域的一个前沿课题,新知识观是当前知

识管理领域的一个前沿课题,引用路径展示是当前知识图谱技术领域的一个前沿课题。

参考文献

- [1] 杨思洛,韩瑞珍.知识图谱研究现状及趋势的可视化分析[J].情报资料工作,2012(4):22-28.
- [2] 刘启元,叶鹰.文献题录信息挖掘技术方法及其软件SATI的实现——以中外图书情报学为例[J].信息资源管理学报,2012(1):50-58.
- [3] 刘则渊,陈悦,侯海燕,等.科学知识图谱:方法与应用[M].北京:人民出版社,2008.
- [4] Small H.Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents [J]. Journal of the American Society for Information Science, 1973,24(4):265-269.
- [5] White Howard D, Griffith B. Author Cocitation: A Literature Measure of Intellectual Structure[J]. Journal of the American Society for Information Science, 1981,32(3):163-171.
- [6] Howard D, Katherine W McCain.Visualizing a Discipline: An Author Co-citation Analysis of Information Science, 1972-1995[J]. Journal of the American Society for Information Science, 1998,49(4):327-355.
- [7] Kevin W Boyack,Richard Klavans,Katy Börner.Mapping the Backbone of Science[J]. Scientometrics, 2005,64(3):351-374.
- [8] Tomihisa Kamada,Satoru Kawai. An Algorithm for Drawing General Undirected Graphs[J]. Information Processing letters, 1989(31):7-15.
- [9] Per Ahlgren, Bo Jarneving, Ronald Rousseau. Requirements for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient[J]. Journal of the American Society for Information Science and Technology, 2003,54(6):550-560.
- [10] Katherine W McCain. Mapping Authors in Intellectual Space: A Technical Overview[J]. Journal of the American Society for Information Science, 1990,41(6):433-443.
- [11] Leydesdorff Loet. How Are New Citation-based Journal Indicators Adding to the Bibliometric Toolbox?[J]. Journal of the American Society for Information Science and Technology, 2009,60(7):1327-1336.
- [12] 马费成,张勤.国内外知识管理研究热点——基于词频的统计分析[J].情报学报,2006(2):163-171.
- [13] 栾春娟,侯海燕,王贤文.国际科技政策研究热点与前沿的可视化分析[J].科学学研究,2009,27(2): 240-243.