

从JoI期刊载文窥探计量学主题演化与学科交叉

吴亚平, 陈迪, 王继民

(北京大学信息管理系, 北京 100871)

摘要: 国际顶级期刊载文的主题演化和学科交叉很大程度上反应了一个学科的发展趋势与学科内涵。文章选取计量学领域国际顶级期刊Journal of Informetrics, 从关键词演化和参考文献的学科分类对其历年载文进行分析, 结果显示: 2007年以来计量学的三大研究主题为“文献计量”、“h指数”、“引文分析”, 在大数据的背景下“主题挖掘”的研究热度在攀升。计量学与多个学科间存在学科交叉渗透, 与计算机分支学科“亲缘”关系最近, 在与众学科的交叉“贸易”中, 计量学都表现为包容性强但影响力略差的特点, 在未来的学科交叉大趋势下, 计量学要加强自身独立性和竞争力, 为更多的学科带来影响。

关键词: JoI; 计量学; 主题演化; 学科交叉

中图分类号: G250.252

DOI: 10.3772/j.issn.1673-2286.2015.08.013

1 引言

主题演化是学科研究热点的动态演变过程, 通常采用分析文献中能够表征主题的特征项之间的关联关系来分析。在发现主题的基础上, 从横向角度对学科不同发展阶段上文献集合的内容进行比较, 找出主题演化的过程和未来发展趋势。对主题演化进行研究可以揭示学科知识发展变化及其相互作用的特征和规律, 找出学科主题演化的动力, 为学术工作者指明未来研究的热点。现今已经有多位学者在主题演化及其方法上开展了有特色的研究。王莉亚^[1]系统地总结主题演化中基于信息熵的方法、基于词频统计的方法、基于突发词共词分析法以及引文分析法等主要方法, 并指出基于词频的分析是具有代表性的一类方法。词频分析法是利用能够解释或表达文献核心内容的关键词或主题词在某一研究领域文献中出现的频次高低来确定该领域研究热点和发展动向的文献计量方法^[2]。论文在很大程度上可以承载一个学科的发展, 而关键词作为论文必要的组成部分是一篇论文主题的浓缩, 反映了论文核心要义, 对关键词的计量分析是对文献计量分

析的有效替代。邱均平等^[3]通过国内外核心期刊的关键词统计总结了情报学研究的10大领域。黄小燕等人^[4]通过情报领域论文关键词进行了统计分析得出1999年至2003年的情报学研究热点。学科交叉的研究一直是国内外学者的研究重点, 学科发展过程中与其它相关学科相互渗透影响, 学者通过借鉴其它学科的研究方法来寻求新的研究视角, 从而解决本学科在发展过程中的问题与瓶颈。引用关系直接代表着知识的借鉴和交流, 在研究学科交叉中有显著应用。冯雪梅等人^[5]对图书情报学8种核心期刊的引文数据进行定量分析, 分析了情报学的相关学科, 测度学科之间的融合程度, 并研究了情报学与相关学科的发展变化。张金柱等人^[6]通过参考文献的学科分类统计分析, 得出了图书情报领域的学科交叉特征, 并指出利用参考文献的学科分类分析学科交叉性的方法可以扩展到其他领域。在借鉴已有研究的基础上, 本文选取计量学高影响因子期刊“Journal of Informetrics”(简称JoI), 对主题演化和学科交叉维度进行分析, 以期通过对该期刊载文的分析, 窥得计量学领域的主题演化特征和学科交叉情况。

2 数据和方法

JoI是情报学刊载定量研究的国际著名期刊, 主要研究领域包括文献计量学、科学计量学、网络计量学和替代计量学等。虽然JoI期刊2007年才创办, 但其在Journal Citation Reports (JCR) 2013年图书情报领域期刊的影响因子为3.58, 排名第4位^[7]。本研究的文献来源于Web of Knowledge数据库中的子库(SCIE, SSCI, A&HCI, CPCI-S, CPCI-SSH, CCRE, IC.)收录的文献数据, 检索出2007年以来(至2015年4月30日)JoI刊载的所有论文共496篇。

本文将关键词演变情况映射到主题演变, 描绘出期刊中刊载文献的主题演化过程和各主题文献的数量分布, 然后对期刊所刊载文献的引用与被引情况进行追溯和统计分析, 基于参考文献的学科分类, 分析计量学学科与其它学科的交叉关系, 从定量角度看计量学与其它学科之间的相互影响。

3 结果与分析

3.1 基于关键词的主题演化分析

将JoI载文的关键词聚类成多个研究主题, 以词频热力图来展示关键主题的研究热度变化, 探究期刊的研究热点, 并以此管窥计量学的研究热点。通过数据字段抽取, 496篇文献共得到1868个关键词, 篇均3.77个。对关键词数据集做如下处理: 第一步, 指定关联规则, 将同个词的不同表示方式合并, 如h-Index、Hirsch index、h-index、h index、Hirsch's index均归为“h-index”, 经过此步骤后共得到关键词163个; 第二

步, 将关键词映射到若干主题, 如将合著、共词、共引、文献耦合归入社会网络-引文分析分类下, 将聚类、分类归入主题挖掘分类下。经过此步骤共得出14个主题类别; 第三步, 列出各个主题下每年的关键词频次变化情况, 结果如图1所示。

从图1可以看出, 在三大计量学中文献计量学随着时间的推移研究热度不断上升, 网络计量学集中出现在2010年, 替代计量学保持在一个不温不火的水平, 这也在于替代计量学的理念备受吹捧, 但是目前的资源开放程度以及本身的定位不清晰, 不好做实证研究。随着替代计量学的提出, 社会化媒体的概念得到了较为广泛的关注, 这也表明, 计量学的数据源不再局限于传统的文献著作, 而是将视野拓宽到更为广阔的社会化媒体中, 例如在评价学者影响力和学术期刊影响力时, 也采用了如Twitter中的转发数、评论数、点赞数之类的社交媒体数据指标^[8]。这也表明计量学开始向更全面、更科学的方向发展。

主题挖掘的研究热度波动攀升, 随着大数据时代的到来, 在科研大数据环境下, 主题挖掘成为一个重要发展方向。主题识别中所涉及到的聚类、分类的算法日益成熟并且不断推陈出新, 产生了基于关键词分析的潜在语义分析等新方法, 使得对文献主题的识别更加准确。

引文分析是文献计量中的重要部分, 而社会网络分析与引文分析密不可分, 社会网络分析的方法和思想为引文分析提供了基础。另外关于共词网络、合著网络的研究也越来越多, 值得提出的是近两年有关零被引现象的研究热度有所增加, 2013年出现了3次, 2015年仅在5月份就出现了4次, 可以预见零被引现象受到了更多的关注。与此同时, 随着数据量的增大以及网络复杂度的增加, 可视化技术愈加体现出其价值, 尤其在

序号	关键词/年份	2007	2008	2009	2010	2011	2012	2013	2014	2015	总数
1	文献计量	3	3	2	9	8	9	13	16	5	68
2	网络计量	0	0	0	4	0	2	1	0	0	7
3	替代计量	1	2	2	1	2	4	2	1	2	17
4	社会化媒体	0	0	0	0	1	1	1	1	1	5
5	基础理论	7	2	0	3	4	3	5	4	2	30
6	主题挖掘	3	0	4	4	6	4	7	9	5	42
7	引文分析(社会网络)	5	6	2	5	5	9	13	7	10	62
8	可视化	2	0	0	4	5	1	0	2	0	14
9	影响因子	1	4	0	2	0	1	5	2	2	17
10	评价指标	0	0	2	1	1	1	3	0	0	8
11	同行评议	3	0	0	2	4	3	1	1	0	14
12	g指数	0	6	2	11	1	1	3	3	1	28
13	h指数	7	14	5	13	8	15	23	7	4	96
14	其他指数	1	1	0	1	1	1	3	2	0	10

图1 JoI主题热力图

2010-2011年引发的讨论最多,讨论的范围涉及呈现方法、可视化工具、布局算法等。

随着期刊、学者的增多,如何对期刊影响能力和学者学术水平进行科学的评价成为计量学领域热议的又一大话题。关于评价指标的选取,学者们进行了较多的探索研究,主要集中在09-13年,其中同行评议作为一种传统的方法在11年得到了最多的讨论和反思。而作为弥补同行评议不足之处的基于发文质量和数量的h指数、g指数等评价指标也得到了较多的研究。然而h指数也不是完美无缺的,其在评价学术论文和学者影响力上的局限性也促发了更多指数的提出,如近几年提出的NSP指数、success指数等。

综上可以将JoI近年来的研究主题分为以下几大方面:计量、引文网络、主题挖掘和科学评价,这也是未来主要的发展方向。根据以上分析可以预见这四方面的发展趋势如下:计量方面,网络计量学的研究高峰已过,目前热度下降,文献计量学热度攀升,替代计量学引起了广泛关注,但在数据、方法等各方面需要突破;三大领域中,文献计量仍然为目前的核心。引文网络中,社会网络分析方法是其中的重要方法,通过网络分析作者、期刊、机构之间的关系,在展示方面更注重借助可视化技术,从而更好地展示数据。主题挖掘热度攀升,在科学大数据的时代,主题挖掘在知识发现、主题追踪、前言预测方面的重要性日益凸显,并有着巨大发展潜力。评价指标方面,传统的同行评议方法越来越少被采用,取而代之的是基于引文的评价指标,主要的指数包括h指数、g指数等,然而由于它们自身的不足,如h指数完全忽视最高被引论文、不关心高被引论文的具体次数、完全忽视低被引论文等,这些不足促使很多新指数的产生,如success index,可以预见随着学者的不断努力,更科学的指数将会被提出。

3.2 学科交叉现象分析

一个学科的发展和其它学科有着紧密的联系,科技的进步使得学科间的界限被打破,学科间的相互渗透、影响已成为一种潮流和趋势。从期刊论文间的互引关系可以分析一个期刊的学科交叉特征,从而反应一个学科的学科交叉特征。参照WOS中的学科分类体系,截至2015年4月30日,除本学科外,JoI共引用其它学科8062次,共被其它学科引用3299次,从整体来看,引用次数达被引次数的两倍多。

引用关系代表此期刊对其它学科的参考度和依赖度,体现了其它学科向计量学的影响和渗透,引用TOP15学科如表1所示。可以看出JoI引用了很多计算机科学跨学科应用、计算机科学信息系统、多学科科学、管理学、经济学、社会学等学科的新方法充实自己的理论、技术、方法等,其中引用最多的是计算机科学下的两个分支学科。这表明,计量学的发展和进步与计算机科学有很大关系,吸纳了计算机科学中的数据挖掘算法、聚类分类算法、可视化方法等算法和思想。

表1 引用次数前15名的学科及频次表

RANK	学科	频次
1	COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS	1525
2	COMPUTER SCIENCE INFORMATION SYSTEMS	1382
3	MULTIDISCIPLINARY SCIENCES	534
4	MANAGEMENT	503
5	ECONOMICS	316
6	PLANNING DEVELOPMENT	284
7	STATISTICS PROBABILITY	215
8	BUSINESS	196
9	PHYSICS MULTIDISCIPLINARY	152
10	MEDICINE GENERAL INTERNAL	139
11	SOCIAL SCIENCES MATHEMATICAL METHODS	134
12	COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	122
13	HISTORY PHILOSOPHY OF SCIENCE	114
14	SOCIOLOGY	114
15	MATHEMATICS INTERDISCIPLINARY APPLICATIONS	104

被引关系表示计量学的理论/技术等被其它学科引用,代表计量学对其它学科的渗透程度,被引TOP15的学科及频次如表2所示。数据表明计量学的内容多被计算机跨学科应用、计算机科学信息系统、多学科科学、管理学、人工智能等学科参考。其中JoI论文被引用次数最多的学科为计算机学科,这表明计量学对计算机学科的影响和渗透也非常显著,计量学也有许多计算机学者积极参与研究,除此之外,计量学的方法被用在多个学科领域分析其研究前沿、新型趋势等。

表2 被引次数前15名的学科及频次表

RANK	学科	频次
1	COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS	957
2	COMPUTER SCIENCE INFORMATION SYSTEMS	682
3	MULTIDISCIPLINARY SCIENCES	185
4	MANAGEMENT	138
5	COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	75
6	BUSINESS	66
7	COMPUTER SCIENCE THEORY METHODS	60
8	OPERATIONS RESEARCHMANAGEMENT SCIENCE	60
9	PHYSICS MULTIDISCIPLINARY	55
10	ENGINEERING ELECTRICAL ELECTRONIC	49
11	PLANNING DEVELOPMENT	47
12	ENGINEERING INDUSTRIAL	45
13	EDUCATION EDUCATIONAL RESEARCH	43
14	MATHEMATICS APPLIED	42
15	SOCIAL SCIENCES INTERDISCIPLINARY	41

被引和引用体现的互引关系可以体现出学科间的“亲密”程度,从被引和引用排名看,和JoI发生的被引和引用均排在前15名的学科为: 计算机科学跨学科应用、计算机科学信息系统、多学科科学、管理学、商科、规划发展、跨学科物理学、人工智能,这体现了计量学与上述学科间的亲密程度较高。如: 物理学中的复杂网络算法和指标被计量学借鉴,计量学中的研究前沿识别也为物理学提供了较好的科研服务。

我们使用IMPORT(进口)表示引用其它学科的论文数,EXPORT(出口)表示被其它学科引用的论文数^[9]。基于期刊的IMPORT/EXPORT指数代表着期刊的学科交叉“贸易”特征,当IMPORT/EXPORT指数值>1时,代表这个期刊吸纳其它学科的内容多于被其它学科吸纳的内容,对其它学科的依赖性较强,独立性较差;当IMPORT/EXPORT指数值<1时,代表这个期刊吸纳其它学科的内容少于被其它学科吸纳的内容,对其它学科的影响较大,生命力较强。将JoI论文中每篇论文的引用和被引论文所属的学科进行汇总统计,选取JoI引用或被引次数大于45的学科,即双方发生学科交叉现象显

著的学科,计算得出IMPORT/EXPORT指数,如图2所示。

从图2中可以看出JoI的IMPORT/EXPORT指数在很多学科都大于1,处于不平衡的状态,尤其是科学历史哲学学科最为明显,IMPORT/EXPORT指数达到了57,即JoI引用科学哲学史学科的论文数与期刊中的论文被科学哲学史学科引用的论文数的比值为57,引用远大于被引。这一方面表明JoI与多个学科存在交集,包容性较强;另一方面也表明期刊吸纳了很多科学哲学史、社会学、概率统计、计算机软件工程等学科的内容,对其它学科的依赖性较高、独立性较差,需要更多地提升自己对其它学科的影响力。

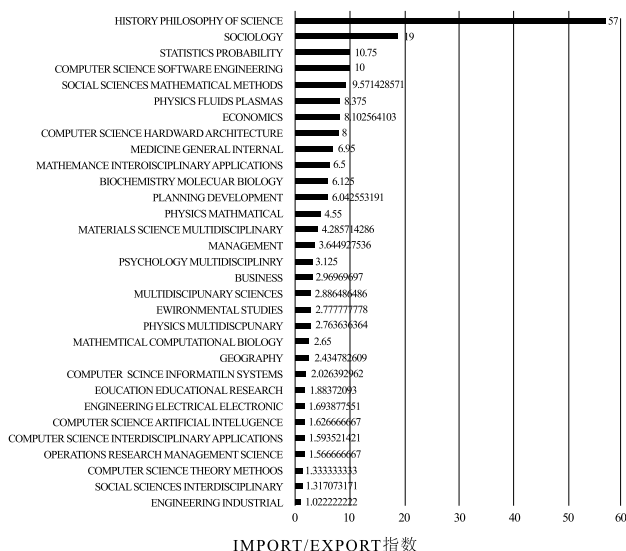


图2 JoI期刊IMPORT/EXPORT指数分布图

4 结论

本文通过定量的方法对计量学领域国际顶级权威期刊JoI的历年载文数据,进行了主题演化分析和学科交叉分析,从一定程度上反应了计量学近年来的发展脉络,主要结论如下:

(1) 近些年计量学的研究内容主要分布在文献计量学、网络计量学、替代计量学、社会化媒体、主题挖掘、引文分析、可视化研究、影响因子与评价指标等范围内。整体来看,文献计量、引文分析、h指数是计量学研究的三大领域,除了这三大领域之外,新的热点领域如“主题挖掘”正在悄然升起,这一方面体现了计量学在大数据环境的积极应对,一方面也与计算机科学、数

学、物理学中的算法、挖掘模型等方法对计量学的渗透相关,也表明学科的交叉与主题的演化有一定的关系。

(2) 计量学学科交叉现状表明,计量学与多个学科之间存在相互渗透和融合。JoI与多个学科存在交集,与众多学科协同发展,其中与计算机分支学科的“亲缘”关系最近。然而由互引现象反映出的学科交叉“贸易”情况显示,其它学科对计量学的渗透远远大于计量学对其它学科的渗透,这一方面表明计量学的学科包容性较强;另一方面也表明计量学对其它学科的依赖性较高、独立性较差。预计将来,学科交叉趋势会更明显,学科交叉现象中体现的“贸易”问题需要引起注意,计量学本身的独立性和竞争力需要再加强,这样才会在学科交叉的大趋势下,保持自己的核心竞争力,向更多的学科渗透。

参考文献

- [1] 王莉亚. 主题演化研究进展[J].情报探索,2014(4):29-32.
- [2] 邱均平,温芳芳. 近五年来图书情报学研究热点与前沿的可视化分析——基于13种高影响力外文源刊的计量研究[J].中国图书馆学报, 2011(02):51-60.
- [3] 邱均平,赵蓉英,侯经川. 2002年国内外情报学发展动向研究[J]. 情报学报,2003(5):512-516.
- [4] 黄小燕. 情报领域研究热点透视——情报领域论文关键词词频分析(1999-2003) [J]. 图书与情报, 2005(6):82-84.
- [5] 冯雪梅,邓小昭. 论情报学的相关学科及发展[J]. 情报杂志, 2008, 27(2):96-98.
- [6] 张金柱,韩涛,王小梅. 利用参考文献的学科分类分析图书情报领域的学科交叉性[J].图书情报工作,2013(1):108-111.
- [7] JCR最新图书情报领域期刊因子排名[EB/OL].[2015-06-02].http://admin-apps.webofknowledge.com/JCR/JCR?RQ=LIST_SUMMARY_JOURNAL.html.
- [8] Thelwall M, Haustein S, Larivière V, et al. Do Altmetrics Work? Twitter and Ten Other Social Web Services[J]. Plos One, 2013, 8(5): e64841.
- [9] Lariviere V.A Bibliometric Chronicling of Library and Information Science's First Hundred Years[J].Journal of the American society for information science and technology, 2012, 63(5):997-1016.

作者简介

吴亚平,女,1991年生,北京大学信息管理系硕士生,研究方向:Web挖掘、科学评价等。

陈迪,女,1992年生,北京大学信息管理系硕士生,研究方向:Web挖掘、科学评价等。

王继民,男,1966年生,北京大学信息管理系教授,研究方向:搜索引擎与Web数据挖掘、科学评价等,通讯作者,E-mail: wjm@pku.edu.cn。

Research on Topic Evolution and Interdisciplinary of Informetrics Based on JoI Papers

WU YaPing, CHEN Di, WANG JiMin

(Department of Information Management, Peking University, Beijing 100871, China)

Abstract: The topic evolution and interdisciplinary of the top international journal papers reflects a subject's corresponding trends. This paper selected the top international journal of Informetrics field——Journal of Informetrics, analyzed the key words evolution and the references subject categories of its papers. The results show that: "bibliometric", "h index", "citation analysis" are three main topics of Informetrics from 2007; the heat of "topic mining" is rising under the background of big data. The interdisciplinary exists between Informetrics and multiple disciplines, Informetrics has the close "genetic" relationship with Computer science. The majority of interdisciplinary "trade" shows that, Informetrics has less impact than its tolerance. Informetrics should strengthen its independence and competitiveness to bring more penetration and impact to more subjects in the future.

Keywords: JoI; Informetrics; Topic Evolution; Interdisciplinary

(收稿日期: 2015-07-15; 编辑: 王立学)