

期刊编委比非编委论文作者能更早 探测出研究前沿吗*

张丽华¹ 曲建升²

(1.山西财经大学 太原 030006; 2.中国科学院兰州文献情报中心 兰州 730000)

摘要 [目的/意义]期刊编委是“科学的守门人”,能够把握期刊的办刊方向,并合理选择哪篇论文更适合在本刊出版。因此,期刊编委理应对非编委的论文作者对本领域的前沿动态有较早的察觉和高度的敏感。基于此,提出了期刊编委能否比非编委的论文作者更早探测到研究前沿的问题。[方法/过程]以科学计量学领域为研究对象,构建了不重复的期刊编委数据集和非编委论文作者数据集,通过比较两个数据集中重复前沿所占的比例来判断究竟是期刊编委较早探测到研究前沿还是非编委论文作者较早探测到前沿。[结果/结论]大多数情况下,期刊编委比非编委论文作者能够较早地探测到同一个研究前沿。在极少数情况下,期刊编委能够与非编委论文作者同时探测到同一个研究前沿。

关键词 期刊编委 非编委论文作者 研究前沿 研究前沿探测 重复前沿 科学计量学

中图分类号 G350 文献标识码 A 文章编号 1002-1965(2017)08-0113-07

引用格式 张丽华,曲建升.期刊编委比非编委论文作者能更早探测出研究前沿吗[J].情报杂志,2017,36(8):113-119.

DOI 10.3969/j.issn.1002-1965.2017.08.020

Could Journal Editors Detect Research front Earlier than Paper Authors Who Don't Hold the Post of Journal Editor?

Zhang Lihua¹ Qu Jiansheng²

(1. Shanxi University of Finance & Economics, Taiyuan 030006;

2. Lanzhou Library, Chinese Academy of Sciences, Lanzhou 730000)

Abstract [Purpose/Significance] Journal editor is a "gatekeeper" of science, who can grasp the general direction of the journal and choose reasonably which paper is more suitable to be published in this journal. Therefore, journal editors should be highly sensitive to research front and can detect research front earlier than paper authors who don't hold the post of journal editor. Based on this hypothesis, this paper proposed a question of whether journal editors could detect research front earlier than paper authors who don't hold the post of journal editor or not. [Method/Processes] This paper takes scientometrics as research field, constructing two separate datasets: journal editors dataset and dataset of paper authors who don't hold the post of journal editor. According to the proportion of repetitive research front which detected early in journal editors dataset or paper authors who don't hold the post of journal editor dataset, the paper judges whether journal editors could detect research front earlier than paper authors who don't hold the post of journal editor or not. [Result/Conclusion] In most cases, the journal editors can detect research front earlier than paper authors who don't hold the post of journal editor. In rare cases, the journal editors and paper authors who don't hold the post of journal editor could detect research front at the same time.

Key words journal editor paper authors who don't hold the post of journal editor research front detection of research front the same research front scientometrics

收稿日期: 2017-03-07 修回日期: 2017-04-01

基金项目: 国家自然科学基金面上项目“气候变化科学成果集成研究范式及其实现平台研究”(编号: 41671535)研究成果之一; ISTIC-ELSEVIER 期刊评价研究中心开放基金项目“中国学者的会议论文产出影响力及其对科研人员排名的影响研究——基于 Scopus 数据库的分析”研究成果之一。

作者简介: 张丽华 (ORCID: 0000-0003-3415-9370), 女, 1986 年生, 博士, 讲师, 研究方向: 科学计量学; 曲建升 (ORCID: 0000-0002-2806-3447), 男, 1973 年生, 博士, 研究员, 研究方向: 气候政策分析、集成研究。

0 引言

作为“科学的守门人”^[1],期刊编委最主要的目标是把握期刊的办刊方向,产生高质量期刊。在不同时期,期刊编委的职能有一些不同。传统上,期刊编委的主要职能是决定哪篇文章更适合出版。在过去的二三十年,该功能发生了一些变化。随着匿名评审过程的普及,期刊编委又增加了一项新的职能:搜寻并评估审稿人。因此,从一定程度上来说期刊编委和审稿人的工作都是进行同行评议^[2-3]。随着影响因子等科学计量学指标的出现,期刊编委的工作又进行了小幅调整:编委互相竞争以吸引最好的论文,即最可能被引用的论文^[4]。

期刊编委有能力重塑期刊的编辑过程并调整期刊的出版政策。鉴于编委在期刊发展中的重要性,只有那些能够成功为本刊吸引高质量论文并在建设和维持期刊声望方面有足够自信并获得本刊同事信任的人才有能力担任期刊编委。也就是说,担任期刊编委是一项巨大的荣誉,能够反映一个人在专业领域获得同行认可的突出贡献^[5]。

可以认为,担任期刊编委的均是各学科领域具有较高学术水平的人员,是各学科领域的学术权威和科研“带头人”。他们能够站在领域发展的最前沿,引领本学科的发展方向。在期刊的编辑出版工作中,除了期刊编委,还存在一个重要的角色:论文作者。论文作者希望将自己的科研成果发表在高质量的期刊中,并通过被引用来提高科研成果的学术影响。而期刊编委则希望尽可能吸引最好的论文来提高本刊的声望。在这个过程中,期刊编委的专业知识以及对本学科领域发展的前瞻性、引领性起到重要作用。因此,期刊编委理应比论文作者更早地觉察到研究前沿,并以此为依据进行期刊的选稿工作。那么,与论文作者相比,期刊编委真的能够更早觉察到研究前沿吗?

为了解决这个问题,本文提出了基于期刊编委探测到的研究前沿与基于论文作者探测到的研究前沿出现时间的先后性比较问题。在分别构建期刊编委所著论文的数据集(set_1)和该刊非编委的论文作者所著论文的数据集(set_2)的基础上,通过比较两个数据集中前沿主题出现的先后顺序,回答期刊编委是否能够比非编委的论文作者更早察觉到研究前沿的问题。

1 研究方法

为了比较基于期刊编委探测到的研究前沿与基于非编委论文作者探测到的研究前沿的时间先后性,涉及到以下几方面的问题:

a. 什么是研究前沿? 如何探测研究前沿?

Price^[1]最早提出“研究前沿(research front)”的概念,用它来描述研究领域的动态本质。Price认为期刊文献可能是由两种有着不同半衰期的文献组成——经典文献和过渡文献,过渡文献实质上对应于研究前沿^[15]。其它学者对研究前沿的内涵界定基本与Price的认识一致。但对于研究前沿的具体表征方式,不同学者有不同的认识。归纳起来大致分为以下3种:①将一组高被引文献定义为研究前沿,如Price、Small^[6]、Braam^[7]的定义;②将一组施引文献定义为研究前沿,如Persson^[8]、Morris^[9]的定义;③将突现的概念或主题定义为研究前沿,如陈超美^[10]的定义。

目前从科学计量学角度探测研究前沿的方法主要包括基于引文的方法^[9,11-17]、基于主题词的方法^[18-21]、基于引文和主题词的复合方法^[7,22,23]以及基于文献计量学指标的方法^[24-26]。此外,近年来还出现了一些新方法,如社区结构探测方法^[17,27]、离群数据探测方法^[28-29]以及直接聚类探测方法^[30]等。

本文选择较为成熟的共词聚类方法作为研究前沿探测方法,并认为基于共词矩阵聚类获得的研究主题即是研究前沿。具体操作步骤为:①抽取作者关键词。作者关键词能够较准确的揭示论文的主要研究内容。为了尽量减少不同作者用词习惯、使用偏好的影响,需对作者关键词的同义、近义等现象进行人工清洗。②构建作者关键词共现矩阵。选择满足一定频次要求的高频关键词,利用TDA构建不同时间窗的作者关键词共现矩阵。③聚类及可视化。在作者关键词共现矩阵的基础上,利用Ucinet、Netdraw和VOS viewer进行聚类及可视化。首先将共词矩阵依次导入Ucinet、Netdraw中生成.net文件,然后将.net文件导入VOS viewer中进行可视化,获得关键词聚类图谱。

b. 如何判断 set_1 和 set_2 数据集中相同或不同时间窗内的研究前沿是同一个?

由上文可知,研究前沿是由一组相互关联的词汇(即词汇簇)构成。判断两个词汇簇是否是同一个研究前沿,可以根据不同词汇簇中相同关键词所占的比例来判断。如果两个词汇簇中相同的关键词数量越多,那么这两个词汇簇越有可能是同一个研究前沿。但由于词汇簇在不同的时间窗可能存在分化、合并、融合、新增、消失等现象,经过综合考虑,本文认为:如果两个词汇簇中有30%及以上的词汇重复,那么这两个词汇簇有可能是同一个前沿。

此外,在计算重复关键词所占的比例时,还需要考虑不同词汇簇的大小。假设2个较小词汇簇的重复关键词数量等于2个较大词汇簇的重复关键词数量,那么2个较小词汇簇是同一个研究前沿的可能性更高。为了平衡词汇簇大小的影响,本文采用的计算公式为:

$$P_{com} = \frac{N_{com}}{(N_1 + N_2) / 2} \quad (1)$$

其中, P_{com} 为 C_1 和 C_2 两个研究前沿中重复关键词所占的比例, N_{com} 为 2 个前沿中重复的关键词数量, N_1 和 N_2 分别是 C_1 和 C_2 前沿中包含的关键词数量。若 $P_{com} \geq 30\%$ 则称 C_1 和 C_2 为重复前沿。

c. 研究前沿的出现时间怎么确定?

研究前沿由一组关键词组成, 那么研究前沿的出现时间则应该根据前沿内部包含的关键词的出现时间来确定。每个关键词均可能在不同的年份出现。找出每个关键词最早出现的年份, 并以所有关键词最早出现年份的平均值来表示研究前沿的出现时间。某个研究前沿 C 的出现时间计算公式为:

$$C = \frac{\sum_{i=1}^n T_i}{n} \quad i = 1, 2, \dots, n \quad (2)$$

其中 n 是前沿 C 中包含的关键词数量, T_i 表示第 i 个关键词在统计时间段内最早出现的年份。

d. 如何综合判断 set_1 和 set_2 数据集中研究前沿出现时间的先后性?

在相同的统计时间段内, set_1 和 set_2 中均包含多个研究前沿, 且可能包含多个既存在于 set_1 中又存在于 set_2 中的重复前沿。在这种情况下, 如何从整体上比较 set_1 和 set_2 数据集中研究前沿出现时间的先后性?

对于某一个重复前沿来说, 其在 set_1 和 set_2 数据集中的出现时间可能存在 3 种情况: ① set_1 中的出现时间较早, 说明期刊编委能够较早探测到这个前沿; ② set_2 中的出现时间较早, 说明非编委论文作者能够较早探测到这个前沿; ③ set_1 和 set_2 中的出现时间相同, 说明期刊编委和非编委论文作者能够同时探测到这个前沿。

对于 set_1 和 set_2 中的若干个研究前沿来说, 如果在 set_1 中较早出现的重复前沿所占的比例越高的话, 说明期刊编委越有可能比非编委论文作者更早探测到研究前沿。相反, 如果在 set_2 中较早出现的重复前沿所占的比例越高的话, 说明非编委论文作者越有可能比期刊编委更早探测到研究前沿。如果 set_1 和 set_2 中重复前沿所占的比例相同, 则说明期刊编委和非编委论文作者能够同时探测到领域内的研究前沿。本文所采用的计算公式为:

$$P_{ce} = \frac{N_{ce_com}}{N_{c_com}} \times 100\% \quad (3)$$

$$P_{cl} = \frac{N_{cl_com}}{N_{c_com}} \times 100\% \quad (4)$$

$$P_{eq} = \frac{N_{eq}}{N_{c_com}} \times 100\% \quad (5)$$

其中, N_{c_com} 指统计时间内 set_1 和 set_2 中重复前沿

的数量。 N_{ce_com} 和 N_{cl_com} 分别指较早出现在 set_1 和 set_2 中重复前沿的数量。 N_{eq} 指同一时间出现在 set_1 和 set_2 中重复前沿的数量。 P_{ce} 和 P_{cl} 分别指 set_1 和 set_2 中较早出现的重复前沿所占的比例, P_{eq} 指在 set_1 和 set_2 中同一年出现的重复前沿所占的比例。

2 数据来源

2.1 研究领域 本文选择“科学计量学”作为实证分析领域。以科学计量学领域中两个重要的核心期刊《Journal of Informetrics》和《Scientometrics》为研究对象, 比较 2007-2014 年间这两种期刊中基于期刊编委探测到的研究前沿与基于非编委论文作者探测到的研究前沿的先后性。这两个期刊中共包括 41 名编委, 编委列表见表 1-2。其中, H. Small、L. Egghe 以及 L. Leydesdorff 等 8 名学者同时担任两个期刊的编委工作。

表 1 《Journal of Informetrics》的编委列表

序号	姓名	机构	国别
1	L. Waltman	莱顿大学科学技术研究中心(CTWS)	荷兰
2	L.Egghe	哈塞尔特大学	比利时
3	V. Larivière	魁北克大学蒙特利尔分校	加拿大
4	G. Abramo	意大利国家研究委员会(CNR)	意大利
5	P.Ahlgren	瑞典皇家理工学院(KTH)	瑞典
6	J. Bar-Ilan	巴伊兰大学	以色列
7	J.Bollen	印第安纳大学	美国
8	L.Bormmann	马克斯·普朗克学会	德国
9	K. Boyack	美国科学技术策略公司	美国
10	Q. Burrell	天主教鲁汶大学	比利时
11	Y. Ding	印第安纳大学	美国
12	W. Glanzel	天主教鲁汶大学	比利时
13	R. Guns	安特卫普大学	比利时
14	P.Ingwesen	皇家图书情报学院	丹麦
15	L.Leydesdorff	阿姆斯特丹大学	荷兰
16	T. Marchant	根特大学	比利时
17	O.Persson	优密欧大学	瑞典
18	F. Radicchi	印第安纳大学	美国
19	I. Rafols	瓦伦西亚理工大学 萨塞克斯大学	西班牙 英国
20	S. Redner	波士顿大学	美国
21	R.Rousseau	天主教鲁汶大学	比利时
22	J. Ruiz-Castillo	马德里卡洛斯三世大学	西班牙
23	J. Schneider	奥胡斯大学	丹麦
24	H. Small	美国科学技术策略公司	美国
25	M. Thelwall	胡弗汉顿大学	英国
26	N.Van Eck	莱顿大学	荷兰
27	A.F.I.VanRaen	莱顿大学	荷兰
28	P.Vinkler	匈牙利科学院	匈牙利
29	D. Wolfram	威斯康辛大学	美国
30	叶鹰	南京大学	中国

Journal of Informetrics Editorial Board. [2016-10-25]. <http://www.journals.elsevier.com/journal-of-informetrics/editorial-board>.

表 2 《Scientometrics》的编委列表

序号	姓名	机构	国别
1	W. Glänzel	天主教鲁汶大学	比利时
2	T. Braun	匈牙利科学院	匈牙利
3	A. Schubert	匈牙利科学院	匈牙利
4	E. Garfield	汤森·路透	美国
5	S. Heeffler	天主教鲁汶大学	比利时
6	L. Egghe	哈塞尔特大学	比利时
7	P. Ingwersen	皇家图书情报学院	丹麦
8	L. Leydesdorff	阿姆斯特丹大学	荷兰
9	B. Martin	萨塞克斯大学	英国
10	K. W. McCain	德雷塞尔大学	美国
11	H. F. Moed	罗马大学	罗马
12	F. Narin	美国知识产权咨询公司	美国
13	O. Persson	优密欧大学	瑞典
14	R. Rousseau	天主教鲁汶大学	比利时
15	H. Small	汤森·路透	美国
16	A. F. J. Van Raan	莱顿大学	荷兰
17	P. Vinkler	匈牙利科学院	匈牙利
18	H. D. White	德雷塞尔大学	美国
19	M. Zitt	法国农业科学研究院	法国

EDITORIAL BOARD of Scientometrics . [2016-10-25]. <http://www.springer.com/computer/database+management+%26+information+retrieval/journal/11192?detailsPage=editorialBoard>

2.2 检索方法 选择国际上公认的获取学术信息的权威科学文献数据库 Web of Science 作为数据源。检索时间为 2016 年 11 月 15 日。

在检索期刊编委所著论文时所采用的检索方法是:①访问 41 名编委的个人主页。梳理编委的工作履历、教育经历和研究兴趣(专长);②查看编委的个人出版物列表。收集编委姓名的不同表达方式,并随机查看若干篇文献的全文,核对编委姓名的不同书写格式;③在 Web of Science 数据库中,将时间限定为 2007-2014 年,采用作者途径进行检索,对检索结果去重、人工剔除不相关文献之后获得相关记录 1271 条。

在检索论文作者所著论文时所采用的检索方法是:将期刊来源分别限定为 Journal of Informetrics 和 Scientometrics,并将时间限定为 2007-2014 年。对获得的数据集进行合并、去重之后获得相关记录 2308 条。

2.3 数据清洗方法 本文中涉及的数据清洗主要包括人工剔除不相关文献与清洗作者关键词两部分。剔除期刊编委数据集中不相关文献所采取的清洗方法包括:①剔除非编委所著论文。②剔除非本领域的论文。

剔除论文作者数据集中不相关文献的主要目的是剔除期刊编委所发表的论文。由于一些期刊编委会将论文发表在自己担任编委的期刊中,因此,为了让两个数据集具有可比性,应该在论文作者数据集中剔除期刊编委所发表的论文。由此构成了本文的分析用数据集:期刊编委数据集 set_1 中包含 1271 条记录,非编委

论文作者数据集 set_2 中包含 1 806 条记录。

对作者关键词的清洗内容包括:①一般清洗。所采用的工具为 TDA。利用 TDA 的列表清理(List Cleanup)功能,对关键词中存在的拼写错误、单复数变化、美式/英式拼写进行清洗;②去停用词。使用 TDA 的停用词叙词表完成。③人工清洗。包括合并同义词,合并简称与全称,人工剔除无区分意义或必须与其它词组配才有意义的词,如 study, use, method 等。

3 研究结果

3.1 期刊编委数据集(set_1)和非编委论文作者数据集(set_2)中发文量对比 在 2007-2014 年间,期刊编委数据集(set_1)和非编委论文作者数据集(set_2)中发文量的年变化曲线如图 1 所示。从图 1 中可以观察到:①期刊编委的总发文量低于非编委论文作者的总发文量。在 2007-2014 年间,期刊编委的发文量始终少于非编委论文作者的发文量。这与期刊编委的总数要远远少于非编委论文作者有关。 set_1 中包括 718 名作者,而 set_2 中包括 2832 名作者。②期刊编委的平均发文量高于非编委论文作者的平均发文量。 set_1 中所有作者的平均发文量约为 1.8 篇, set_2 中所有作者的平均发文量约为 0.6 篇。显然期刊编委的平均发文量要高于非编委论文作者的平均发文量。③期刊编委发文量与非编委论文作者发文量差距越来越大。这可以从两条曲线的走势中得到验证。在 2007-2008 年间,两条曲线几乎平行。虽然期刊编委数量与非编委论文作者数量差距较大,但两者发表的论文总数基本持平,说明这一时期是期刊编委非常高产的时期。在 2009-2010 年间,两条曲线开始逐渐分离,经过 2011 年的短暂靠近之后,两条曲线的分离趋势越来越显著。说明在 2012 年之后,期刊编委的高产特征已经越来越不显著。

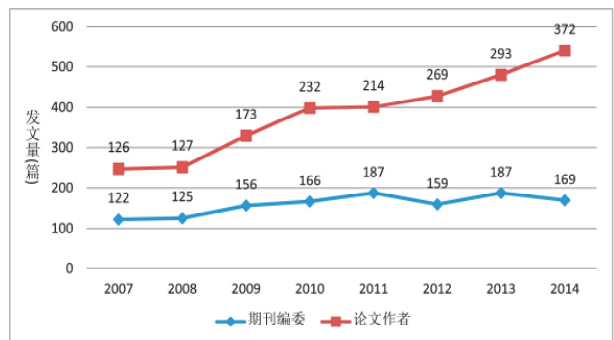


图 1 2007-2016 年 set_1 和 set_2 中发文量的年变化曲线

3.2 期刊编委数据集(set_1)和非编委论文作者数据集(set_2)中研究前沿对比 将 2007-2014 年的统计时间段,以 2 个为 1 个时间窗,共划分为 2007-2008 年,2009-2010 年,2011-2012 年,以及 2013-2014 年 4 个

时间窗。经过共词聚类之后, 获得 set_1 和 set_2 中不同时间窗的研究前沿。不同时间窗中包含的研究前沿数量与每个研究前沿中包含的作者关键词数量统计结果见表 3。

表 3 不同数据集中研究前沿数量与规模的对比

	Set ₁	Set ₂
2007-2008	111 (7) , 112 (6) , 113 (5) , 114 (5) , 115 (2)	211 (3) , 212 (2)
2009-2010	121 (8) , 122 (7) , 123 (6) , 124 (5) , 125 (2)	221 (13) , 222 (8) , 223 (7) , 224 (7) , 225 (3) , 226 (2)
2011-2012	131 (13) , 132 (9) , 133 (8) , 134 (8) , 135 (6) , 136 (5) , 137 (4) , 138 (3) , 139 (2)	231 (22) , 232 (15) , 233 (13) , 234 (12) , 235 (10) , 236 (10) , 237 (10) , 238 (9) , 239 (3) , 2310 (1)
2013-2014	141 (11) , 142 (10) , 143 (8) , 144 (7) , 145 (7) , 146 (7) , 147 (5)	241 (18) , 242 (14) , 243 (13) , 244 (11) , 245 (11) , 246 (10) , 247 (10) , 248 (10) , 249 (10) , 2410 (9) , 2411 (7) , 2412 (7) , 2413 (6) , 2414 (6)

在表 3 中, set_1 和 set_2 中的数字表示研究前沿编号。该编号由 3 位或 4 位数字组成, 其中第 1 位数字表示数据集。1 表示 set_1 即期刊编委数据集, 2 表示 set_2 即非编委论文作者数据集。第 2 位数字表示不同的时间窗, 1 表示 2007-2008 年, 2 表示 2009-2010 年, 3 表示 2011-2012 年, 4 表示 2013-2014 年。第 3 位和第 4 位数字表示在特定时间窗内的研究前沿序号。括号中的数字代表研究前沿中包含的关键词数量。例如 2411(7) 即表示在 set_2 数据集中 2013-2014 年的第 11 个研究前沿中包含 7 个作者关键词。

set_1 和 set_2 中不同时间窗的研究前沿对比结果如下:

① set_1 中研究前沿呈现数量少但主题集中的特点, 而 set_2 中研究前沿呈现数量多但主题相对较分散的特点。在期刊编委数据集 set_1 中, 学者们关注的重点内容从 h 指数及其衍生指数(2007-2010 年) 发展到文献计量学分析(2009-2014 年)、引文分析(2009-2010 年)、大学评估(2009-2014 年) 等内容。除了上述重点研究内容, set_1 数据集中还包括科学地图、同行评议、高被引论文等研究主题。但总的来说, 期刊编委的研究兴趣相对集中。在非编委论文作者数据集 set_2 中, 学者们关注的重点除了集中在文献计量学、引文分析、h 指数、期刊影响因子、合作分析等方面之外, 还包括了大量的分散主题, 如政策分析、知识结构、数据库、PubMed、经济等内容。造成这种差异的原因可能是由于期刊编委的人数固定, 相应的研究方向确定, 研究主题稳定, 而非编委论文作者的人数并不固定, 研究人员可能来自多个不同的研究领域, 从而使得研究主题变得不稳定。

② 在某些主题上, 期刊编委和非编委论文作者的

研究兴趣相同。比如文献计量学分析和引文分析。文献计量学分析是科学计量学的首要任务。引文分析是科学计量学的一个主要分支。自 E. Garfield 创立科学引文索引以来, 引文分析一直是科学计量学学者热衷的研究点。据评价, 引文分析是当代科学计量学和信息计量学最杰出的贡献^[31]。

③ 从某些主题来看, 期刊编委比非编委论文作者的关注时间更早。比如 h 指数。在 2007-2008 年间, 期刊编委已经注意到 h 指数及其相关的衍生指数, 如 g 指数、r 指数等, 而非编委论文作者仅仅关注到 h 指数本身, 并未将重点放在 h 指数的衍生指数方面。2009-2010 年间, 期刊编委与非编委论文作者的关注重点依然如此。

从上述分析中可知, 期刊编委的研究主题稳定, 重点突出, 并能够比非编委论文作者更早地关注到一些前沿主题, 这能够从一定程度上反映期刊编委的确是学科领域发展的引领者与前瞻者这一论断。

3.3 期刊编委数据集(set_1) 和非编委论文作者数据集(set_2) 中研究前沿出现时间对比 为了验证期刊编委是否能够较早地比非编委论文作者探测到研究前沿, 本文中采用的研究思路是首先寻找期刊编委数据集(set_1) 和非编委论文作者数据集(set_2) 中的重复前沿, 然后比较这些重复前沿在 set_1 和 set_2 中的出现时间, 分别计算较早出现在 set_1 和 set_2 中的重复前沿的比例, 以及同一时间出现在 set_1 和 set_2 中的重复前沿的比例, 据此判断究竟是期刊编委先探测到前沿还是非编委论文作者先探测到前沿, 抑或是期刊编委和非编委论文作者能够同时探测到前沿。

将 set_1 和 set_2 中的研究前沿两两比较, 使用两个前沿中包含的重复关键词的比例来确定是否是同一个前沿。研究前沿的两两比较结果如表 4 所示。

表 4 set_1 和 set_2 中的研究前沿两两比较结果 (因篇幅有限, 此处仅截取部分数据)

	211	212	221	222	223	224
111	0.0	0.0	0.0	0.3	0.1	0.1
112	0.0	0.3	0.3	0.0	0.0	0.3
113	0.3	0.0	0.0	0.3	0.0	0.0
114	0.5	0.0	0.0	0.2	0.0	0.0
115	0.0	0.0	0.0	0.0	0.2	0.0
121	0.2	0.0	0.1	0.0	0.1	0.0
122	0.0	0.2	0.3	0.1	0.1	0.0
123	0.0	0.3	0.0	0.1	0.2	0.2
124	0.3	0.0	0.0	0.3	0.0	0.0
125	0.4	0.0	0.0	0.2	0.0	0.0
131	0.0	0.1	0.4	0.0	0.0	0.0
132	0.2	0.0	0.0	0.0	0.1	0.0
133	0.2	0.0	0.0	0.3	0.0	0.0
134	0.0	0.2	0.1	0.3	0.0	0.1
135	0.0	0.0	0.1	0.0	0.2	0.0
136	0.0	0.0	0.1	0.0	0.0	0.0

续表 4 set₁ 和 set₂ 中的研究前沿两两比较结果
(因篇幅有限, 此处仅截取部分数据)

	211	212	221	222	223	224
137	0.0	0.0	0.0	0.0	0.2	0.0
138	0.3	0.0	0.1	0.2	0.0	0.0
139	0.0	0.0	0.0	0.0	0.0	0.0
141	0.0	0.2	0.2	0.0	0.1	0.0
142	0.3	0.0	0.0	0.3	0.2	0.0
143	0.0	0.0	0.2	0.0	0.1	0.0
144	0.0	0.2	0.1	0.1	0.0	0.1
145	0.2	0.0	0.0	0.1	0.0	0.0
146	0.0	0.0	0.2	0.0	0.1	0.0
147	0.0	0.0	0.0	0.0	0.0	0.0
151	0.4	0.0	0.1	0.2	0.2	0.0
152	0.0	0.2	0.2	0.0	0.0	0.0
153	0.0	0.2	0.0	0.3	0.0	0.1

本文中将两个前沿中重复关键词占比大于等于 0.3 的前沿定义为重复前沿。set₁ 和 set₂ 中共包含 24 个重复前沿。重复前沿在不同数据集中的出现时间如表 5 所示。

表 5 重复前沿在不同数据集中的出现时间

序号	Set ₁	出现时间	Set ₂	出现时间
1	114	2007.00	211	2007.00
2	125	2009.00	211	2007.00
3	138	2011.00	211	2007.00
4	142	2013.00	211	2007.00
5	151	2015.07	211	2007.00
6	112	2007.00	221	2009.54
7	122	2009.43	221	2009.54
8	131	2011.00	221	2009.54
9	111	2007.00	222	2009.75
10	113	2007.00	222	2009.75
11	124	2009.20	222	2009.75
12	142	2013.00	222	2009.75
13	153	2015.00	222	2009.75
14	112	2007.00	224	2009.57
15	141	2013.27	233	2011.46
16	114	2007.00	237	2011.40
17	125	2009.00	239	2011.00
18	136	2011.20	241	2013.00
19	122	2009.43	244	2013.00
20	122	2009.43	249	2013.10
21	133	2011.00	2411	2013.14
22	135	2011.00	2412	2013.00
23	132	2011.11	2414	2013.00
24	142	2013.00	2414	2013.00

通过比较在 set₁ 数据集和 set₂ 中较早出现的重复前沿的数量, 并计算 P_{ce} 、 P_{cl} 和 P_{eq} 的值, 能够说明期刊编委与非编委论文作者谁能够更早地探测到前沿。经过计算, P_{ce} 、 P_{cl} 和 P_{eq} 的值分别是 0.62、0.31 和 0.07, 即期刊编委数据集中较早出现的重复前沿所占比例为 62%, 非编委论文作者数据集中较早出现的重复前沿所占比例为 31%, 而期刊编委和论文作者集中同一年出现的重复前沿所占的比例为 7%。也就是说, 大多数情况下, 期刊编委比非编委论文作者能够较早地探测到同一个研究前沿。在极少数情况下, 期刊编委能

够与非编委论文作者同时探测到同一个研究前沿。

4 总结

本文写作的初衷是考虑到期刊编委是“科学的守门人”, 能够把握期刊的办刊质量, 并合理选择哪篇论文更适合在本刊出版, 那么期刊编委理应比非编委的论文作者具有更高的专业知识与学术水平, 同时能够比非编委的论文作者对本领域的前沿动态有较早的察觉和高度的敏感, 由此才能高瞻远瞩地引领本学科领域的发展。

本文中提出了期刊编委是否能够比非编委的论文作者更早探测到研究前沿的问题。为了解决这个问题, 首先构建了两个不重复的数据集: 期刊编委数据集 set₁ 和非编委论文作者数据集 set₂, 然后采用相同的共词聚类方法获得 set₁ 和 set₂ 中不同时间窗的研究前沿, 接下来对 set₁ 和 set₂ 中所有的前沿进行两两比较, 通过计算重复关键词的比例来寻找重复前沿, 最后通过计算较早出现在 set₁ 和 set₂ 中重复前沿的比例以及同时出现在 set₁ 和 set₂ 中重复前沿的比例来判断究竟是期刊编委较早探测到研究前沿还是非编委论文作者较早探测到前沿。研究结果表明: 大多数情况下, 期刊编委比非编委论文作者能够较早地探测到同一个研究前沿。在极少数情况下, 期刊编委能够与非编委论文作者同时探测到同一个研究前沿。

由于篇幅和时间有限, 本文的研究尚存在以下不足:

①在判断 1 篇论文属于 set₁ 还是 set₂ 集合时, 本文所采用的标准为: 如果该篇论文中至少包含 41 名期刊编委中的一位, 则判断该篇论文属于 set₁, 否则属于 set₂。在这个过程中, 并没有考虑 1 篇论文中不同作者的贡献问题。第 1 作者或通讯作者与末位作者对该篇论文的贡献肯定不同。这是下一步研究中应该解决的问题。

②在判断两个研究前沿是否是同一个时, 本文通过比较两个前沿中重复关键词所占的比例来判断。但在研究前沿中, 不同关键词根据其重要性是具有不同的权重的。本文中并未考虑到研究前沿中不同关键词权重对研究结论的影响。这也是进一步研究中需要考虑的问题。

③本文选择了科学计量学领域进行实证研究, 结果发现多数情况下期刊编委能够比非编委论文作者更早地探测到研究前沿。该结论仅适用于科学计量学领域, 至于其它领域中该结论是否适用还有待于更多的验证。

参考文献

[1] Crane D. The gatekeepers of science: Some factors affecting the

- selection of articles of scientific journals [J]. *American Sociologist* ,1967(32) : 195-201.
- [2] Bakanic V ,Mcp hail C ,Simon R J.The manuscript review and decision-making process [J]. *American Sociological Review* ,1987 (52) : 631-642.
- [3] Coupe T.What do we know about ourselves? On the economics of economics [J]. *Kyklos* ,2004(57) : 197-215.
- [4] Baccini A ,Barabesi L.Interlocking editorship.A network analysis of the links between economic journals [J]. *Scientometrics* ,2010 , 82(2) : 365-389.
- [5] Kaufman G G.Rankings of finance departments by faculty representation on editorial boards of professional journals: A note [J]. *The Journal of Finance* ,1984(39) : 1189-1197.
- [6] 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.
- [7] Braam R R ,Moed H F ,Vanraan A F J.Mapping of science by combined co-citation and word analysis .I. Structural Aspects [J]. *Science Technology & Human Values* ,1988 ,13(1-2) : 97-98.
- [8] Persson O.The intellectual base and research fronts of JASIS 1986-1990 [J]. *Journal of the American Society for Information Science* ,1994 ,45(1) : 31-38.
- [9] Morris S A ,Yen G ,Wu Z ,et al.Time line visualization of research fronts [J]. *Journal of the American Society for Information Science and Technology* ,2003 ,54(5) : 413-422.
- [10] Chen C.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.
- [11] 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: 265-269.
- [12] Garfield E.Research fronts [J]. *Current Contents* ,1994(41) : 3-7.
- [13] Persson O.The intellectual base and research fronts of JASIS 1986-1990 [J]. *Journal of the American Society for Information Science* ,1994 ,45(1) : 31-38.
- [14] White H D ,McCain K W.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.
- [15] Garfield E.Historiographic mapping of knowledge domains literature [J]. *Journal of Information Science* ,2004 ,30(2) : 119-145.
- [16] Klavans R ,Boyack K W.Identifying a better measure of relatedness for mapping science [J]. *Journal of the American Society for Information Science and Technology* ,2006 ,57(2) : 251-263.
- [17] Shibata N ,Kajikawa Y ,Takeda Y ,et al.Detecting emerging research fronts based on topological measures in citation networks of scientific publications [J]. *Technovation* ,2008 ,28(11) : 758-775.
- [18] Kleinberg J.Bursty and hierarchical structure in streams [J]. *Data Mining and Knowledge Discovery* ,2002 ,7(4) : 373-397.
- [19] Callon M ,Law J ,Rip A.Mapping the dynamics of science and technology: *Sociology of science in the real world* [M]. London: Palgrave Macmillan Ltd ,1986.
- [20] Kostoff R N.Citation analysis cross-field normalization: A new paradigm [J]. *Scientometrics* ,1997 ,39(3) : 225-230.
- [21] Blei D M ,Ng A Y ,Jordan M I.Latent dirichlet allocation [J]. *Journal of Machine Learning Research* ,2003 ,3(4-5) : 993-1022.
- [22] Van Den Besselaar P ,Heimeriks G.Mapping research topics using word-reference co-occurrences: A method and an exploratory case study [J]. *Scientometrics* ,2006 ,68(3) : 377-393.
- [23] Janssens F ,Glanzel W ,De Moor B.A hybrid mapping of information science [J]. *Scientometrics* ,2008 ,75(3) : 607-631.
- [24] Chen C M.Measuring the movement of a research paradigm [J]. *Visualization and Data Analysis* ,2005 ,5669: 63-76.
- [25] Small H G.Co-citation model of a scientific specialty-longitudinal-study of collagen research [J]. *Social Studies of Science* , 1977 ,7(2) : 139-166.
- [26] Tu Y-N ,Seng J-L.Indices of novelty for emerging topic detection [J]. *Information Processing & Management* ,2012 ,48(2) : 303-325.
- [27] Newman M E J ,Girvan M.Finding and evaluating community structure in networks [J]. *Physical Review E* ,2004 ,69(2) .
- [28] 张英杰.科技领域前沿计量探测方法研究 [D].北京: 中国科学院大学 ,2011.
- [29] 王莉亚.基于离群数据的主题演化研究 [D].北京: 中国科学院大学 ,2012.
- [30] Hartigan J A.Direct clustering of a data matrix [J]. *Journal of the American Statistical Association* ,1972 ,67(337) : 123-129.
- [31] 邱均平 ,宋艳辉.引文分析领域研究热点前沿与高频作者的二维时空分析 [J]. *图书情报知识* ,2011(6) : 18-24.

(责编: 王平军; 校对: 王 菊)