

文献计量学发展演进与研究前沿的知识图谱探析*

赵蓉英 许丽敏

摘 要 以 ISI Web of knowledge 收录的文献题录作为数据样本,利用动态网络分析的信息可视化技术和工具——CiteSpace II 对所采集的数据进行文献共引分析和聚类分析,绘制出文献计量学研究发展演进的知识图谱,揭示了文献计量学领域的代表人物和代表文献;展现了文献计量学研究的热点领域,其中以科学引文和研究产出等为代表;呈现了文献计量学的发展趋势:①网络和网络计量的研究日益成为文献计量学的研究前沿;②目前有关指标的研究是文献计量学研究前沿,且以指标的应用为主要研究前沿;③文献计量学呈现与其他学科特别是医学相结合的趋势。研究表明,文献计量学研究日益呈现应用化、综合化和网络化的态势。图 6。表 2。参考文献 14。

关键词 文献计量学 CiteSpace II 知识图谱 研究前沿 研究热点

分类号 G350

ABSTRACT The authors aim to use newly developed information visualization technology — CiteSpace II to conduct con-citation and cluster analysis in which the data came from ISI Web of knowledge. The authors got the reference co-citation map to identify the dominant research fields and representative scientists and hot topics in which the science citation and research outputs are classical. The authors also point out future frontiers of bibliometrics: 1) webliometrics; 2) research of indexes, especially their application; 3) integration with other disciplines, especially the trend of combining medicine. Research shows that study on bibliometrics is becoming more and more applicable, comprehensive and networked. 6 figs. 2 tabs. 14 refs.

KEY WORDS Bibliometrics. CiteSpace II. Knowledge map. Research front. Research focus.

CLASS NUMBER G350

1 引言

早在 20 世纪初,人们已经开始对文献进行定量研究,但是当时文献计量学并没有作为一门独立的学科而存在。直到 1969 年,英国著名情报学家阿伦·普理查德(Alan Pritchard)首次提出用术语“Bibliometrics”取代“统计目录学”的名称。这一术语的出现标志着文献计量学的正式诞生^[1]。

从人们开始对文献进行定量研究到现在,经历了将近一百年的历史。在这将近一百年的时间内,文献计量学由萌芽走向了成熟。文献计量学无论是理论还是应用都得到充分的研

究,取得很大的成就,诞生出一系列的代表人物和代表作品。然而随着网络和网络相关技术的发展,文献计量学的研究发生了巨大变化,其研究的热点领域也有所改变。

本文旨在对文献计量学的代表人物、代表作品、研究热点与前沿进行定量考察和可视化分析,即通过词频分析方法确定文献计量学研究的热点领域;通过检测词频变动趋势显著的主题词确定文献计量学研究的前沿领域和发展趋势;并利用信息可视化软件 CiteSpace II,绘制出文献计量学代表人物、代表作品、热点领域与研究前沿的知识图谱,形象地展示文献计量学

* 本文为国家社会科学基金资助项目“社会网络分析在企业知识共享中的应用研究”(项目编号:09BTQ022)的研究成果之一。

可以追溯到1917年。在这一初始阶段,文献计量学的研究比较分散,规模不大,涉及的主题也不多,大多是进行某一领域的文献统计的基础工作。尽管如此,这一阶段的研究却拉开了文献计量学的序幕,具有重要的开创性意义。在这一阶段出现的主要研究者与其作品有:

(1)科尔与《科学进展》。由图1我们发现,1917年以科尔(COLE F. J.)为第一作者发表在《科学进展》(Science Progress)上的《比较解剖学的历史——对文献的统计分析》(The history of comparative anatomy——A statistical analysis of the literature)是文献计量学研究的最早期作品。这是现代意义上的第一项较完备的科技史计量研究。它将1543—1860年间欧洲各国关于动物解剖学方面的出版物(6436件)进行了统计,并据此绘制了出版物数量的时间分布曲线,从该曲线可较为清楚地看出比较解剖学的发展进程;根据该曲线比较和分析了欧洲各国在此期间对比较解剖学的贡献及不同时期的各种研究、论文和研究者对解剖学发展的影响。该曲线与后来的科学发展指数增长规律曲线十分吻合^[3]。这次分析不但使他们系统地了解了比较解剖学在这一时期的研究重点及发展变化,而且掌握了这一时期的文献在世界各国的分布情况。更为重要的是用事实说明了文献统计的方法及意义,大大开阔了人们的眼界,为文献计量学研究开创了一条崭新的路子。

(2)洛特卡与《科学生产率的频次分布》。从图1可知,萌芽时期的文献还有1926年洛特卡(LOTKA A. J.)发表在《华盛顿科学院会刊》(Journal of the Washington academy of Science)杂志上的《科学生产率的频次分布》(The frequency distribution of scientific productivity)。这篇文献是这一时期文献计量学研究的又一重大成果。文中,洛特卡(LOTKA A. J.)从科学文献作者与其撰写的论文的纷乱现象中首先发现了“平方反比”的数量关系,提出了至今被人称为“经典”的洛特卡定律,从而成为文献计量学中最早的一个著名定律,为文献计量学的诞生奠定了一定的基础^[4]。由CiteSpace II导出的数据,发现该篇文献在所有文献中的被引频次为第三。这充分地说明了该篇文献在文献计量学研究领域的地位。

(3)格鲁斯与《大学图书馆与化学教育》。另外由CiteSpace II导出的数据可知,发表的较早的文献还有格鲁斯(GROSS P. L. K.)等人1927年发表在《科学》(Science)的《大学图书馆与化学教育》(College libraries and chemical education)。他们统计了化学专业的一些期刊的论文的参考文献并进行了分析,得出了化学教育方面的核心期刊^[5]。这是文献学史上第一次引文分析,为文献计量学后期的研究奠定了基础。

从上面的文献,我们看到在这一时期文献研究人员首创文献统计方法,并在一些学科领域进行了文献计量分析的大胆尝试,取得了一定的成果。这些研究都为文献计量学的诞生与后期的发展奠定了基础。

3.2 文献计量学的奠定时期(1934—1960年)

由文献计量学领域代表人物、代表作品时间序列图谱,可以看出这一时期的研究比较注重理论与规律的发现。同时我们发现,这一时期的研究者在文献统计调查的基础上,尝试地发掘出文献的规律,归纳出具有普遍指导意义的理论性结论,并取得了许多重要的成果。从图1中,我们看到这一时期的成果主要有以下几个方面:

(1)文献计量学的基本规律。文献计量学的大部分规律都是在这时期提出的,其中有著名的文献计量学的三大基本定律中的布拉德福定律以及齐普夫定律。

由图1看到,1934年英国著名文献学家S. C. 布拉德福在《工程》(Engineering)周刊的“图书与文献”栏目上发表了一篇题为《专业学科的情报源》(Sources of Information on scientific subjects)的论文。作者在做了大量的统计研究工作以后,发现尽管学科不同,但它们的论文在相应的期刊中有着同样的分布规律,即 $n_1:n_2:n_3=1:a_2:a_3$,这就是著名的布拉德福定律^[6]。这一定律的确立,奠定了文献计量学的理论基础,对文献计量学的形成和发展起了极其重要的作用,是文献计量学史上具有划时代意义的里程碑。

同时我们也看到了由美国学者G. K. 齐普夫1949年首次出版的名为《人类行为与最省力法则——人类生态学引论》的专著。这部专著引用

了大量的数据和事实,对“最省力法则”作了精辟的论述,这一法则较好地解释了齐普夫定律的内在成因和机制,是齐氏定律的理论基础。

(2)文献计量学的其他规律的研究。在这一阶段,除了文献计量学基本规律的研究以外,还有一些其他规律的研究。如文献引用规律、文献增长规律以及文献老化规律。对这些规律的研究在这一时期都取得了可喜的进展。

在图1中我们清晰地看到由美国情报学家加菲尔德(E. Garfield)1955年在《科学》上发表的《科学引文索引:一个新维度的文献》(Citation Indexes for Science: A new dimension in documentation through association of ideas)一文。在该篇论文中,作者系统地提出了用引文索引检索科技文献的新方法,从而打破了分类法和主题法在检索方法中的垄断地位,打开了从引文角度来研究文献及科学发展动态的新领域,为文献

计量学的发展开辟了道路^[7]。
综上所述,在这一时期,文献计量学的很多定律或规律的研究都取得了很大的发展,这些定律或规律共同构成和奠定了文献计量学的理论基础,为文献计量学后期的研究提供了很好的理论指导。

3.3 文献计量学的全面发展与分化时期(1960年至今)

由图1我们可知,这一时期文献计量学已由狭隘的理论研究发展到了广阔的应用研究,同时涉及的领域和主题也越来越多。在这一时期涌现了一大批优秀的研究者和经典文献。在图1中,我们可以看到12个比较大的节点,这些节点代表作品的详细信息如表1所示,其中有9个节点是这一时期的作品。其中主要的作品有:

表1 文献计量学领域的代表作品信息统计(被引频次≥40的节点信息)

序号	被引频次	节点作者	发表年份	发表刊物
1	101	HIRSCH J. E.	2005	P NATL ACAD SCI USA
2	67	SEGLEN P. O.	1997	BRIT MED J
3	58	GARFIELD E.	1955	SCIENCE
4	53	LOTKA A. J.	1926	J WASHINGTON ACADEMY
5	45	SMALL H.	1973	J AM SOC INFORM SCI
6	43	MOED H. F.	2005	CITATION ANAL RES EV
7	43	GARFIELD E.	1996	BRIT MED J
8	41	WHITE H. D.	1989	ANNU REV INFORM SCI
9	41	GARFIELD E.	1972	SCIENCE
10	41	GARFIELD E.	2006	JAMA - J AM MED ASSOC
11	40	VANRAAN A. F. J.	2006	SCIENTOMETRICS

(1)《一个量化个人研究产出的指标》。由图1与表1可知,被引用频次最多的作品是2005年美国加州大学圣迭哥分校的物理学家HIRSCH J. E.教授发表于《美国科学院院刊》(P NATL ACAD SCI USA)的《一个量化个人研究产出的指标》(An index to quantify an individual's scientific research output)。利用CiteSpace II的Google scholar链接,导出该文的被引频次为101次,可见该文章在文献计量学中有重要的影响力。该篇文章首次提出一个量化个人研究产出的指标—H指数,对H指数

进行了定义,并给出了这一指数的计算方法。他将H指数定义为当且仅当某科学家发表的N_p篇论文中有H篇论文每篇至少获得了H次的引文数,其余的N_p-H篇论文中各篇论文的引文数都≤H时,此H值就是该科学家的H指数^[8]。
(2)《为什么期刊的影响因子不应该用于评价研究成果》。SEGLEN P. O. 1997年发表于《英国医学杂志》(BRIT MED J)的《为什么期刊的影响因子不应该用于评价研究成果》(Why the impact factor of journals should not be used for e-

valuating research) 被引频次为 67 次,位于第二。文中描述了影响因子(IF)在科研评价研究中的局限性,分析了影响因子不能用于评价研究的原因。他认为,期刊影响因子隐藏了文章引用率的不同,一种期刊中部分文献的被引频次是另外一部分文献的 10 倍;定义期刊的影响因子时只考虑技术性问题,而没有考虑到文章的科学质量;高影响因子可能会包含迅速增长的基础研究^[9]。该篇文献为人们有效地使用影响因子提供了一定的指导。

(3)《科学文献的共被引:两篇文献之间的一种新的测度方法》。1973 年由美国费城科学信息研究所的 H. Small 撰写并发表于《美国信息科学技术学会学报》的《科学文献的共被引:两篇文献之间的一种新的测度方法》(Cocitation in the scientific literature: A new measure of the relationship between two documents)一文,是文献计量学领域的另一篇重要文献,其被引频次为 45。文中对“共被引”(co-citation)作出定义,指出共被引文献网络可以对某个具体的科学学科进行分析,被引文献集群对科学构成提供了一种新的研究思路^[10]。

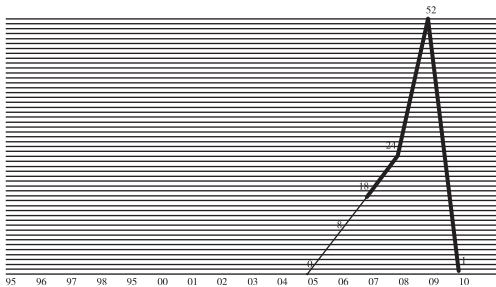


图2 HIRSCH J. E. 2005 年
发表文献的被引情况统计

由图 1 和表 1 可知,文献计量学在这一阶段的研究在进一步完善基础理论之上,较侧重于文献计量学的应用以及指标的研究。同时我们也看到文献计量学在这一阶段,特别是到 20 世纪 80 年代之后,开始分化出一些新的学科和研究方向。20 世纪 80 年代,信息计量学从文献计量学中分离出来,成为一门独立的学科,到目前已经发展得比较成熟。20 世纪 90 年代中期,文献计量学开始出现对网络信息的研究,并提出了“网

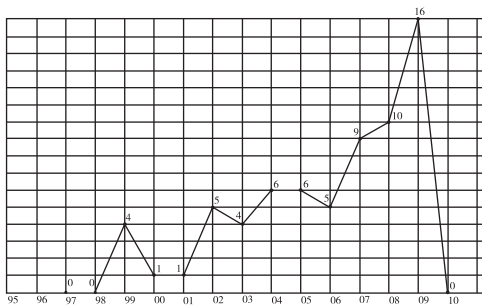


图3 SEGLEN P. O. 1997 年
发表文献被引情况统计

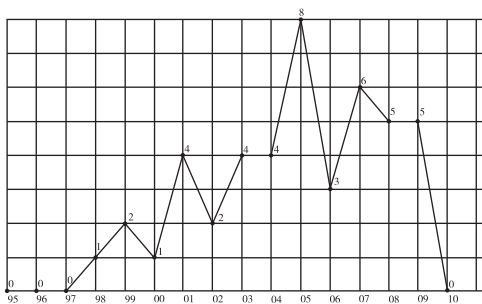


图4 SMALL H. 1973 年
发表文献的被引情况统计

络计量学”的名词。到目前,网络计量学也成为一门独立的学科,而从文献计量学中脱离出来。

如今,文献计量学经历了由萌芽、发展到成熟与分化阶段。它作为一门学科,无论是理论还是应用都取得了很多成果,已经得到广大学者的认可。随着社会的发展,在不同的阶段,文献计量学产生了一些适应当时社会的分支学科。随着这些分支学科的不断发展,到一定阶段便从文献计量学中脱离出来。当然,文献计量学在以后的发展中还会出现一些新的学科分支,并分离出一些新的学科。这是社会的需要,同时也是文献计量学学科的需要。

4 文献计量学研究热点与前沿的知识图谱

4.1 文献计量学研究热点的知识图谱

利用文献题录中的关键词,并借助 Citespace

图5中的名词短语是从数据样本中的每一篇文章中提取出来的,出现频次较多的名词短语也可以在一定程度上代表文献计量学的研究热点。由图5中看到,出现频次较多的名词短语有文献计量分析、影响因子、科学引文索引(science citation index)、文献计量指标、引文分析等,

同样将 ISI Web of knowledge 收录的 1995—2010 年的 “Bibliometric * ” 相关文献数据 (共 1459 篇) 输入到 CiteSpace II 软件中, 并利用 CiteSpace II 软件中提供的膨胀词探测 (burst detection) 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 (burst term) 从大量的主题词中探测出来, 依靠词频的变动趋势, 而不仅仅是频次的高低, 来确定文献计量学的前沿领域和发展趋势。同时选择适当阈值, 运行 CiteSpace II 软件。通过运行 CiteSpace II 软件, 一共探测出 656 个膨胀词, 得到文献计量学研究前沿与发展趋势图 (见图 6)。

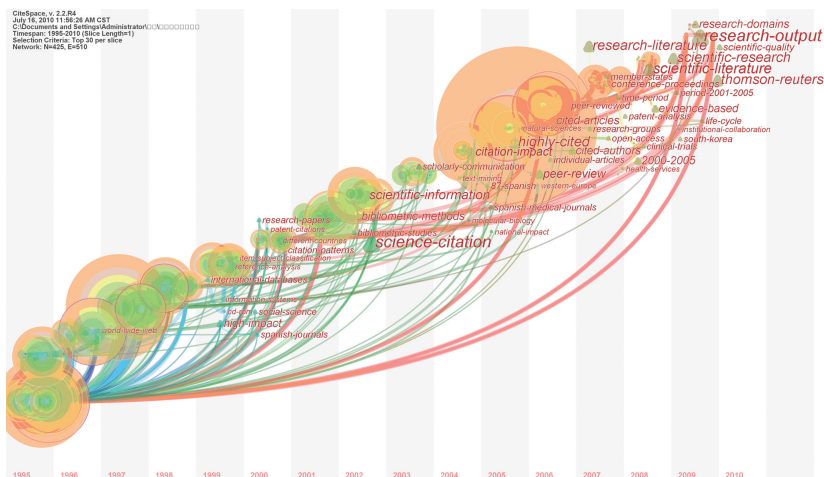


图6 文献计量学研究前沿与发展趋势

由图6看出,主题词变化率较高的词有科学引文 (science-citation) 和研究产出 (research-output),说明科学引文和研究产出的研究是近年来重要前沿领域之一。特别是科研产出,其位于文献计量学研究前沿与发展趋势图的前端,可见科研产出是近些年文献计量学重要的前沿研究。随着科学技术与推动社会进步和发展所起作用的日益加强,人们愈来愈清楚地意识到科学技术与经济发展、国际竞争力等的关系也愈来愈密切。而某一科研机构或学科的研究产出成为人们主要关注的对象。论文是研究机构科研产出的重要表现形式,使用文献计量学的原理及方法,对研究机构所发表的论文进行评价,可以充分反映出科研机构的产出情况。这就使得利用文献计量学的方法对科研产出的研究成为文献计量学的研究前沿。因此利用文献计量学的方法对某一特定的研究机构或国家进行测量与评价是文献计量学重要的前沿研究。

此外,科学文献 (scientific-literature)、汤姆森-路透公司 (tomson-reuters)、科学研究 (scientific-research)、文献研究 (research-literature)、高被引 (highly-cited)、科学信息 (scientific-information)、同行评议 (peer-review)、引文影响 (citation-impact)、研究域 (research-domains)、高影响力 (high-impact)、循证医学 (evidence-based)、被引作者 (cited-authors)、被引作品 (cited-articles)、文献计量学研究方法 (bibliometric-methods)等也是文献计量学领域的研究前沿。有关影响力和科学的研究可以代表文献计量学的前沿领域。目前无论是对影响力计算的研究还是对提高影响力方法的研究,均是文献计量学的研究前沿。有关科学的研究也如此,无论是对科学文献还是科学的研究都是文献计量学领域的专家们关注的对象,是文献计量学的研究前沿。需要指出的是,循证医学也日益成为文献计量学研究的前沿,可见文献计量学研究的内容跨度很大,不仅涉及到普通的情报与图书馆学,还涉及到医学领域,并成为其研究的前沿。

另外,我们在文献计量学研究前沿与发展趋势图中看到若干与网络有关的膨胀词,如万维网 (world-wide-web)、网络站点 (web-site(s))、

网络爬虫 (web-crawling)、网络引用 web-citation (s)、web-based-survey等。虽然有关网络膨胀词出现的频次并不是太多,但是这些膨胀词出现的个数较多,可见有关网络的研究也是文献计量学重要的前沿领域。同时我们在文献计量学研究前沿与发展趋势图中还发现一些与医学、健康有关的膨胀词,如循证医学、公共健康研究 (public-health-research)、临床实验 (clinical-trials)、医学信息 (medical-informatics)等。可见有关医学的研究也是文献计量学重要的前沿领域。

5 结论

本文对从 ISI Web of knowledge 中检索到的与“Bibliometric*”相关的1459篇文献题录进行分析,得到以下结论:

第一,文献计量学从产生到现在,经历了由萌芽到成熟的发展过程。在这一时期,诞生了多部经典著作,推动了文献计量学的迅猛发展;出现了一批杰出的学科带头人和骨干力量,形成了一支优秀的研究队伍,为文献计量学的发展做出了杰出的贡献。文中通过动态网络分析的信息可视化工具,绘制出文献计量学领域的文献共被引图谱,追溯到文献计量学的思想源泉以及文献计量学领域的经典著作。最早的文献计量学研究始萌于20世纪初,一般认为1917年以 COLE F. J. 为第一作者发表在《科学进展》(Science Progress)上的《比较解剖学的历史——对文献的统计分析》一文是文献统计研究的起点,是文献计量学研究的萌芽之作。在文献计量学的发展过程中出现了一系列的经典著作,其中以 HIRSCH J. E. 发表的《一个量化个人研究产出的指标》、SEGLEN P. O. 发表的《为什么期刊的影响因子不应该被用于评价研究成果》等为代表,这些文献大部分主要侧重于对文献计量学领域的指标的研究。

第二,随着信息技术的发展以及信息的网络化,文献计量学研究的主要领域发生了很大的变化。文中借助文献题录中的关键词和名词短语,并通过 CiteSpace II 信息可视化软件,利用词频分析方法,确定了文献计量学研究的若干

热点领域,如文献计量学、科学、影响力、期刊、指标、文献计量分析法、引文分析、影响因子、科学引文索引等,这些研究热点大多是针对文献计量学理论、方法和应用的研究。

第三,通过 CiteSpace II 信息可视化软件的主题词探测技术,探测出了若干文献计量学研究的前沿领域和发展趋势,其中以科学引文和研究产出等为代表。由探测出的主题词,我们发掘出文献计量学呈现三大发展趋势:①网络和网络计量的研究日益成为文献计量学的研究前沿;②文献计量学内容体系中的指标,从提出到现在一直都是文献计量学研究的热点,目前有关指标的研究是文献计量学研究前沿,且以指标的应用为主要的研究前沿;③文献计量学呈现与其他学科相结合的趋势,特别是医学。通过以上研究表明,文献计量学研究日益呈现应用化、综合化和网络化的态势。

参考文献:

- [1] 邱均平.文献计量学[M].北京:科学技术文献出版社,1988.
- [2] 邱均平.信息计量学[M].武汉:武汉大学出版社,2007.
- [3] COLE F J, Hulme E W. The history of comparative anatomy—A statistical analysis of the literature [J]. Science Progress, 1917 (11): 578-596.
- [4] Lotka A J. The frequency distribution of scientific productivity[J]. Journal of the Washington Academy of Sciences, 1926,16(12):317-323.
- [5] Gross P L K, Gross E M. College libraries and chemical education [J]. Science, 1927, 66 (1713):385-389.
- [6] Bradford S C. Sources of information on specific subjects[J]. Engineering, 1934, 137: 85-86.
- [7] Garfield E. Citation Indexes for Science: A new dimension in documentation through association of ideas [J]. Science, 1955, 122 (3159): 108-111.
- [8] Hirsch J E. An index to quantify an individual's scientific research output[J]. The National Academy of Sciences of the USA, 2005, 102 (46): 16569-16572.
- [9] Seglen O. Why the impact factor of journals should not be used for evaluating research [J]. BMJ, 1997,314(7079):497-502.
- [10] 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 and Technology, 1973,24(4): 265-269.
- [11] 栾春娟,赵呈刚.基于SCI的基因操作技术国际前沿分析[J].技术与创新管理,2009,30(1):11-13.
- [12] Chen C. Generalised similarity analysis and pathfinder network scaling[J]. Journal of the American Society for Information and Technology, 1998,10(2):107-128.
- [13] 陈超美. CiteSpace II:科学文献中新趋势与新动态的识别与可视化[J].情报学报,2009(3): 401-421.
- [14] 刘则渊,陈悦,侯海燕.科学知识图谱的方法与应用[M].北京:人民出版社,2008.

赵蓉英 武汉大学信息管理学院教授、博士生导师,武汉大学科学评价研究中心副主任。通讯地址:湖北省武汉珞珈山。邮编:430072。

许丽敏 武汉大学信息管理学院硕士研究生。通讯地址同上。

(收稿日期:2010-02-08;修回日期:2010-03-14)