

定量学术评价体系的构成和应用

宋黎明,冯振卿

(南京医科大学图书馆,江苏 南京 210029)

摘要:学术评价体系不仅是科研经费获取、科研绩效评价的主要手段之一,同时也作为重要动力激发了学者和科研人员的学术生命力。文章从定量学术评价体系的构成和应用上阐述了近年来评价指标的进展,就评价指标如权威因子、被引半衰期、H指数及其变体、F1000指标重点予以介绍,并结合南京医科大学具体情况进行分析。

关键词:学术评价指标;定量分析;同行评议

中图分类号:G644.4

文献标识码:A

文章编号:1671-0479(2014)03-238-005

doi:10.7655/NYDXBSS20140317

作为国家科研管理工作的重要内容之一,学术评价先后经历了行政评价、同行评价、指标量化评价和国际科学计量评价四个阶段。随着我国科技文献量的日益增长,对文献所属地区、机构、课题组或个人、出版物进行分析评价的迫切需求已将学术评价推到了一个前所未有的高度。学术评价体系还成为了科研经费获取、科研绩效评价的主要手段之一,同时也作为重要动力激发了学者和科研人员的学术生命力。基于定性标准的评价体系过于笼统粗放,面对日益增长的海量学术文献已无法满足学术评价的要求,定量标准的引入则弥补了这一缺憾,随着数据量的急剧攀升,其严谨精确,以质量、创新和贡献为导向的优势也日益凸显出来。

学术期刊文献的定量研究目前已然是学术评价体系的主要构成,其客观地展示了高校、科研机构乃至一个国家和地区的科研学术水平和实力,得到了世界范围内各国科学界的认可。但应注意的是,定量研究需与定性标准结合,并关注文献的实际应用价值和社会效益^[1]。

一、文献定量指标

定量指标包括影响因子、被引频次、即年指标、被引半衰期、H指数等。汤森路透集团旗下的《期刊引证报告》无疑是定量指标集大成者,特别是对于南京医科大学等以自然科学研究为主的机构。

(一)总体介绍

总文献量、总引文量、总下载量、被引频次以及影响因子等由于简明便于理解,目前这些指标仍为我们进行学术评价所常用。此外,由上述指标直接衍生出的指标还有:①平均引文量,为期刊当年发表论文的总引文量与当年总文献量之比,是评价科研人员吸收利用文献能力的指标;②即年指标或即时指数,为期刊当年发表论文的被引频次与当年总文献量之比,是表征期刊即时反应速率的指标;③即年下载率,为期刊当年发表论文的总下载量与当年出版并上网的总文献量之比,是评价期刊在网络环境下传播效率的指标;④5年影响因子,为期刊前5年发表论文在统计当年的被引频次与期刊前5年总文献量之比,是测度期刊有用性和显示度的指标,而且也是测度期刊的学术水平乃至论文质量的重要指标;⑤5年零被引论文率,为期刊前5年发表论文在统计当年零次被引的文献量与期刊前5年总文献量之比,是反映期刊发表低质量论文的数量指标;⑥普赖斯指数,为期刊当年发表论文的近5年引文量与总引文量之比,是测度期刊老化程度的数量指标;⑦学科扩散指标,为统计源期刊范围内,引用该刊的期刊数与学科内全部期刊量之比;⑧学科影响指标,为期刊所在学科内,引用该刊的期刊数与学科内全部期刊量之比,这两者是从学科延伸和影响的角度来评价期刊;⑨扩散因子,为期刊当年每被引100次所涉

收稿日期:2014-02-25

作者简介:宋黎明(1972-),男,河南叶县人,馆员,主要研究科技文献的计量分析。

及的期刊数,是评估期刊影响力的学术指标,显示总被引频次扩散的范围。上述指标都仅着眼于期刊文献的来源或引用、下载的某一方面,难以全面、综合、整体地来评价期刊文献和科研人员,但仍有其存在的必要性。

就我校而言,2007~2012年所发表 SCIE 期刊论文在文献量和被引频次上都有明显增长,2007年文献量和3年被引频次分别为374篇、1357次,而2011年结果分别为998篇、5450次。2012年文献

量更较上年激增50%,截至2014年1月底施引文献和被引频次已达4279篇、4711次,篇均被引3.15次。此6年内即年指标由0.19渐升至0.45,说明我校科研人员对新出版文献能够迅速予以关注,密切追踪学科发展。从5年零被引论文率看,2007年为26.2%,之后逐步降至20%以下,按照文献引用周期理论,2009年前后文献到今年就达到引用高峰,对比其与之前结果可以得出文献质量逐步提高的结论(表1、2)。

表1 南京医科大学 SCIE 论文 2007~2012 年主要定量指标统计

项目	2007年	2008年	2009年	2010年	2011年	2012年
总文献量(篇)	374	557	660	778	998	1497
3年被引频次(次)	1357	2522	3011	4327	5450	4711
3年施引文献(篇)	1134	2306	2591	3642	4728	4279
篇均被引频次(次)	3.63	4.53	4.56	5.56	5.46	3.15
即年指标	0.19	0.34	0.33	0.38	0.39	0.45
5年零被引论文率(%)	26.2	21.5	16.7	19.2	21.1	32.4

数据收集截至2014年1月31日。

表2 南京医科大学 SCIE 论文 2003~2013 年定量指标统计

年份	总文献量(篇)	被引频次(次)	篇均被引频次(次)	施引文献量(篇)	H ₁ 指数
2003~2012	5662	42795	7.56	34073	60
2008~2012	4490	30745	6.85	24643	48
2012	1497	4711	3.15	4279	21
2013	1809	1395	0.77	1185	10

数据收集截至2014年1月31日。

(二)权威因子

基于期刊引用网络 Prestige Rank 算法所得到的权威因子,综合考虑了引用次数和引用质量来计算期刊分数,因弥补了被引频次和影响因子单纯计算次数而不考虑引用重要性区别的弊端,所以更能客观地反映期刊权威性而非流行性^[2]。因为临床医学类期刊较多,研究人员也多,按论文被引频次排名自然优于基础类期刊;而权威因子则不然,期刊之间的引用不再是同等对待,重要期刊的一次引用比次要期刊的一次引用被赋予了更高的权重,它可部分消除因为学科规模和性质不同造成的被引机会不同的问题。因此可认为权威因子更多地是一个总量的测度,它提供的期刊评价结果和被引频次具有很强的相关性,唯计算较为繁琐。该指标已纳入《中国科技期刊引证报告》收录范围。此类期刊评价指标还包括基于 Page Rank 算法得出的 Eigenfactor,并实现了自引数据排除,已为汤森路透集团旗下的《期刊引证报告》所采用。

(三)被引半衰期

被引半衰期指期刊在统计当年的被引频次中,较新一半被引论文发表的时间跨度。被引半衰期与

引用半衰期虽都是测度期刊老化速度的一种反向评价指标,半衰期越短说明文献老化速度越快。但引用半衰期长,说明期刊本身引用的参考文献年代久远,测度的是其他文献的老化速度;而被引半衰期长,则说明其自身的影响较长远,测度的是文献自身的老化速度,较符合现今医学科技发展的趋势。随着获取文献信息的途径和方式发生了根本变革,虚拟图书馆和网络期刊已逐渐成为获取期刊文献和信息的主要手段,被引半衰期作为文献计量的指标已有些落伍,可取而代之的是与之接近的下载量半衰期。下载量半衰期指期刊在统计当年的总下载量中,较新一半是距离当年多长一段时间内累计达到的,数值往往短于被引半衰期。下载量半衰期长,说明被下载文献老化速度较慢,文献时效性较差,被读者关注的速度较慢。当前读者往往喜欢下载最新的权威期刊文献,并以占有文献为目的,使得文献下载高峰较以往提前并迅速老化,此点与引用文献需要吸收消化和存在反复引用的情形有所不同。

(四)H 指数

H 指数则与上述期刊评价指标有所不同,其作为一项累积性指标,可用于评价科研人员的个人绩

效。将被评价者论文按被引频次从高到低排列并核对,直到某项序号大于其被引频次,则该序号减去1就是H指数。H指数将论文数量与引文数量紧密结合,概念简单且易于计算,同时对科学家长期累积的研究成果的影响力也能提供强有力的评估。H指数还可以和高等院校的办学资源、教学水平、科学研究以及学校声誉等指标相结合,从不同视角共同对高等院校的重要性和影响力作出全面评价。但H指数也有其自身固有的缺陷:①缺乏灵敏度,多数被评价者出现指标数值几年不变的呆滞局面;②缺乏区分度,多数被评价者处于较低水平的H指数范围且数值等同;③缺乏波动性,被评价者绩效评价的指标数值只会升高或维持,不会下降。为了试图解决H指数的某些缺点,众多H型指数或H指数变体,包括R指数、AR指数、G指数、HG指数等又被提出^[3-4]。

R指数为H指数划定的绩效核内论文被引频次的平方根,A指数为H指数绩效核内论文被引频次的平均值,而AR指数为H核内各篇论文的年均被引频次总和的平方根。因为H指数只给出了论文数量与被引频次数量相交叉的阈值,而没有对阈值内绩效核的不同强度作进一步的测定。R指数与A指数则对H指数无力染指的H核进行了强度测量,解决了H指数对绩效核内的变化缺乏灵敏度和区分度的问题。AR指数是在R指数的基础上引入“论文发表年限”这一参数,将论文的被引频次与论文的发表年限两个参数构成“论文年均被引频次”这一因变量,对H指数形成一种辅助性的上升和下降机制。H指数与R指数、H指数与AR指数以及R指数与AR指数之间都存在高度相关性,鉴于科研绩效评价需要最大程度的精确性,因此建议将H指数与R指数配对,或H指数与AR指数配对使用^[5]。

G指数是一种基于学者以往贡献的测度指标,将论文按被引频次由高到低排序,将序号平方、被引频次按序号层层累加,当序号平方等于累计被引频次时,该序号即为G指数。如序号平方不是恰好等于而是小于对应的累计被引频次,则最接近累计被引频次的序号为G指数。作为H指数的衍生指数,G指数弥补了H指数不能很好反映高被引论文的缺陷。HG指数为被评价者H指数与G指数的几何平均数,HG指数既考虑到了高被引论文,又显著降低了高被引论文的影响,可以在被评价者的主要高质量论文和高被引论文之间达到一种平衡,为科研绩效评价提供了一种更为合理的方法^[6]。

当H指数用于评价科研机构时,如果按照个人H指数的计算方法计算,所得机构全部论文数据的H指数即为 H_1 指数,这是目前较为常用的方法;如果分别计算机构每位研究人员的个人H指数后再计算出的机构H指数,即为 H_2 指数。机构 H_1 指数重在测度机构论文表现出来的水平,在机构 H_1 指数的计算中,如果某人H指数明显高于其他研究人员,那么个人H指数有可能很大程度上决定机构 H_1 指数;而 H_2 指数从拥有高水平研究人员的角度来评价机构水平,主要揭示机构研究人员个人H指数的差异。机构 H_2 指数为机构评价提供了一种新的思路,尽管其对中低端水平的研究机构仍缺乏区分度^[7]。我校SCIE论文的2008~2012年总 H_1 指数为48,略逊于首都医大和协和医大,各年 H_1 指数分别为32、30、31、25、19。个人H指数一般达到30以上即为杰出科学家,我校沈洪兵教授和沙家豪教授的个人H指数分别为33、17,已达到很高的水平(表3)。

(五)影响因子

表3 南京医科大学 SCIE 论文 2008~2012 年 H_1 指数与其他医学院校的横向对比

年份	南京医科大学	天津医科大学	南方医科大学	首都医科大学	中国医科大学	协和医科大学
2008~2012	48	45	35	50	40	59
2008	32	26	25	31	30	38
2010	31	29	25	28	25	34
2012	19	16	13	17	16	22
个人 H 指数	33(Shen)	22(Yao)	14(Hou)	21(Li)	18(Wang)	28(Zhang)

数据收集截至2014年1月31日。

影响因子在近年的发展过程中形成了两个新指标,即复合影响因子和综合影响因子。复合影响因子是以期刊综合统计源文献、博硕士学位论文统计源文献、会议论文统计源文献为复合统计源文献计算,而综合影响因子是以科技类期刊及人文社会科学类期刊综合统计源文献计算,但这

两者均为被评价期刊前两年发表的被引文献在统计当年的被引用总次数与该期刊在前两年内发表的被引文献总量之比。因为复合影响因子在统计源的选取上引入博硕士学位论文和会议论文,其统计数据也囊括了博硕士学位论文和会议论文的引文数据,使得计算结果能更全面地反映

期刊在学术研究和高级人才培养方面的影响力,从而一定程度上克服了传统影响因子和综合影响因子仅计算期刊数据而带来的固有缺陷,如恶意自引或自引同盟、低水平引用、无关引用、引用不全面等^[8]。

此外,还有一些评价指标相继被提出,如质量指数、引证强度、合作系数等。国内外学者提出的计算现有多种评价指标权重的综合评价体系也在不断探索和研究中。

从上述定量指标来看,我校近十年(特别是近五年)在肿瘤学、药理与药剂学、生物化学与分子生物学等多方面都处于国内独立设置医药院校的前列,

沉寂文献(即发表后从未被引用的文献)的比例也处于同类院校的较低水平,并且在某些学科的全球最顶尖刊物上还能发表高质量论文,这一趋势必将随着校院领导的重视、支持和顶尖人才的培养、引进而日趋明显(表4)。不足之处在于Q1区最顶尖期刊中我校文献量(合计2066篇,占总数5662篇的36.5%)仍偏少,且会议摘要比重较高,皮肤病学此点表现最为明显(119篇SCIE文献中仅有2篇原著论文)。投稿期刊倾向性明显,如《PLoS One》发文章增长迅速,2003~2012年累计已达165篇,2013年更是高达创纪录的214篇,远超第2位的《Chinese Medical Journal》,这是值得深思的。

表4 我校SCIE论文2003~2012年与其他医学院校的横向对比 (篇)

项目	南京医科大学	天津医科大学	南方医科大学	首都医科大学	中国医科大学	协和医科大学
学科						
肿瘤学	756(253)	622(192)	333(110)	281(91)	657(143)	864(307)
药理与药剂学	563(137)	196(34)	210(38)	294(58)	306(49)	1126(105)
生物化学与分子生物学	500(114)	257(56)	299(62)	385(57)	430(73)	926(182)
遗传学	218(77)	66(35)	62(20)	82(30)	119(30)	298(134)
毒理学	142(52)	19(7)	25(5)	54(11)	112(28)	64(10)
生殖生物学	94(49)	2(0)	42(29)	15(6)	18(8)	57(31)
期刊						
Nature Genetics	15	0	0	6	11	20
J. Invest. Dermatol.	119	0	0	4	18	5
PloS One	165	54	72	131	74	166
Chinese Medical Journal	153	88	39	597	122	480
文献总和	5662	3027	2874	5903	4365	9151
零被引文献率(%)	24	25	25	30	25	21

括号内数字为Q1区SCIE期刊文献量。数据收集截至2014年1月31日。

二、同行评议指标

同行评议一直以来也被国内外公认为评价论文质量和学术影响力的较好方法。但当同时对不同学科、不同研究领域中的大量的研究机构或个体进行比较或评价时,评价组织者有时难以恰当地选择出符合被评价对象实际情况的评议专家;或者参与评议的同行专家在知识结构、学术水平、学术倾向、学术态度等各个方面存在差异;或者评价组织者、同行评议专家与被评价对象之间可能存在复杂的社会、学术利益关系,这些因素的存在使得真正意义上的、可确保评价结果具有科学性、客观性和公正性的同行评议往往难以实现。

基于此点,一种新的学术评价指标——F1000论文评价体系目前仅针对生物学和医学研究领域,遴选出的各专业顶级专家和学者采取随机推荐论文的方式,按优秀(10分)、必读(8分)、推荐(6分)对论文给予评价^[9]。其中优秀是指里程碑式的论文,必读是指对一般读者均有益的论文,推荐是指论文本

身专业性较强,可能仅对某一学科或主题感兴趣的研究者有用的论文。这3个等级的论文对应的附加分值分别为3、2、1。F1000论文因子的计算方法为选取一个最高评分,加上其他所有评价的附加分总和。经典论文一般能达到30分以上的高分。目前我校被收录F1000论文已达10篇以上。

在F1000论文因子的基础上衍生出的期刊因子可以用来评价期刊绩效水平。由于论文推荐范围主要囿于Medline数据库,与影响因子并无关联,所以许多非SCI论文(不含综述等)和期刊也能得到公正客观的评价。入选期刊需12个月内有24篇论文被Medline数据库收录且至少1篇是F1000论文,通过该期刊年度F1000论文因子之和与F1000论文数量占总论文量的比重来计算期刊因子。利用此指标可以快速找到学科内重点期刊,并及时跟踪研究方法和成果。

三、结论

从文献被引来看,被引频次、影响因子、引用期

刊数、即年指标、他引率、被引半衰期、权威因子、扩散因子、学科扩散指标、学科影响指标等凸显了读者使用和重视文献的程度,以及期刊在学术交流中的地位和作用,其对期刊影响力的客观评价作用是毋庸置疑的。从期刊来源来看,总文献量、总引文量、总下载量、平均引文量、平均作者数、地区分布数、机构分布数、海外论文比、基金论文比、文献选出率和引用半衰期等指标所描述的学术水平、编辑状况和学术交流程度也具有客观性和全面性,适宜作为基础指标来使用。H 指数及 H 型指数则将文献被引与来源文献指标紧密结合,使两者优势互补,克服了单项指标所固有的缺憾。从现有 H 指数与 H 型指数的发展来看,其完全具有发展成为下一代核心评价指标的潜质^[10]。除实证分析外,数学模型也已成为支撑和优化 H 指数与 H 型指数的研究方向。

总之,评价期刊的适时性看影响因子,而期刊的经典性或综合性与专业性看被引半衰期;评价期刊的稳定性要看更长时间段的 5 年影响因子;评价作者个人论文的质量要看引用率和 H 指数等。此外,对于因为学科差异、期刊出版差异、文献类别差异、自引等因素而产生的影响也应该考虑在内。目前,影响因子仍是学术期刊影响力和显示度的重要评价标准之一,尽管 Q1 区 SCIE 期刊(各学科内 IF 排名前 25% 的最优期刊)影响因子的可操作性很高,但以期刊影响因子的高低来衡量论文水平的高低还是有缺陷的,因为期刊的影响力并不代表期刊所刊发的所有论文的影响力。用于学术期刊评价时,H 指数不仅显示出传统文献计量指标所不具有的特点,而且能与影响因子等指标相结合,使得期刊评价更具合理性和科学性。所以,对我校学术水平的评价不能只局限于文献量和被引频次等少数几个指标,而

是要综合多方面因素予以权重和权衡,必要时可进行荟萃分析来解释看似矛盾的数据结果。

国家教育部 2013 年已明确提出要改变分类评价实施不到位、开放评价和长效评价机制不健全的弊端,进一步深化完善学术评价机制,充分发挥其在高等教育的宏观管理和资源配置中的重要作用。可见,学术评价是一项长期、复杂而又非常必要的活动。

参考文献

- [1] 郭红梅,谢娟. 学术论文评价指标的发展过程[J]. 科技管理研究,2011,31(9):65-67
- [2] 苏成,潘云涛,马峥. 权威因子:一个新的期刊评价指标[J]. 编辑学报,2010,22(4):369-373
- [3] 曾纪洁. H 指数及其扩展指数用于机构评价研究[J]. 图书情报工作,2011,55(10):65-68
- [4] 周志峰. 国内 h 指数和 h 型指数研究文献综述与计量分析[J]. 情报杂志,2009,28(6):52-57
- [5] 金碧辉,Ronald R. R 指数、AR 指数:h 指数功能扩展的补充指标[J]. 科学观察,2007,2(3):1-8
- [6] 吴雷,孙莹莹. 基于 h 指数和 g 指数的高等学校学术表现评价应用研究[J]. 经济研究导刊,2013(17):245-247
- [7] 杨立平,岳婷,杨立英. h2 指数用于机构评价的理论分析与初步应用[J]. 图书情报工作,2010(16):18-21
- [8] 伍军红. 复合影响因子与期刊影响力评价[J]. 编辑学报,2011,23(6):552-554
- [9] 宋秀芳,吴鸣,王甲卫. Faculty of 1000 在生物和医学研究中的应用——以 MicroRNA 研究为例[J]. 医学信息学杂志,2013,34(2):45-50
- [10] 许新军. 《中文核心期刊要目总览》评价指标的缺失与建议[J]. 情报杂志,2013,32(10):89-92

Constitution and application of quantitative academic evaluation system

Song Liming, Feng Zhenqing

(Library of Nanjing Medical University, Nanjing 210029, China)

Abstract: The academic evaluation system is not only one of the major ways for acquisition of scientific research funds or assessment of achievement in scientific research, but also the important driving force for scholars and researchers to arouse their academic vitality. Here, the recent progress of academic evaluation system is summarized centering around its constitution and application, dwelling on indicators such as authority factor, cited half-life, H index and H-type indices, F1000 index, and further analysis were performed basing on the specific conditions of our university.

Key words: academic evaluation system; quantitative analysis; peer evaluation