

·信息分析与科学评价·

开栏语:今天,我们已经进入大数据时代,数据在科学研究中发挥着越来越重要的作用。科学研究的主导范式也由逻辑驱动、实验驱动转向了数据驱动的研究范式。数据驱动研究范式就是以数据为基础,进行数据处理、数据分析,建立数据间的关系,发现新生的事物,探索科学的规律,为科技决策提供依据。20世纪上半叶诞生的文献计量学已经逐步拓展到信息计量学、科学计量学、网络计量学。科学研究的发展变化、科学技术的不断进步,为定量分析开拓了广阔的前景,带来了无限的机遇,各类信息分析的方法和工具层出不穷,各种评价分析指标更是纷纷诞生。本刊开办“信息分析与科学评价”栏目,就是为信息分析与科学评价领域学者提供一个很好的平台,并期待大家重视计量学理论和方法研究,并以数据为基础,利用新兴技术与工具,进行深入的信息分析,积极开展科学评价实践活动,为学界推出更多、更具意义的优秀成果。

本栏目将关注文献计量学、信息计量学、网络计量学等理论研究成果,欢迎对新兴评价指标及评价体系的探索和研究,重视各类信息分析与科学评价的实践成果,我们将竭尽全力打造这一平台,为繁荣我国信息分析与科学评价领域和学者们共同努力,力尽我们的绵薄之力。

文献计量学与科学评价中有关问题思考

苏新宁

(南京大学信息管理学院 江苏南京 210093)

摘要:文章主要探讨了文献计量学三大定律对科学评价的指导作用,指出了这些定律的某些方面对当前分析研究存在的不适合性,还对目前评价实践中常用的几个计量指标进行了分析,指出了照搬这些指标存在的问题,提出了有效地解决方法,最后给出了在实际评价操作中应注意的相关问题。

关键词:文献计量学 科学评价 布拉德福定律 齐普夫定律 洛特卡定律 文献计量指标

中图分类号: G250.252

文献标识码: A

文章编号: 1003-6938(2013)01-0079-05

Study on Bibliometrics and Related Problems in Scientific Evaluation

Abstract This paper mainly discusses how the three laws of Bibliometrics guide scientific evaluation, and points out that some aspects of these laws are not suitable to the current analysis study. Several measurement indicators used widely in evaluation practice are analyzed as well. This paper not only points out that the existing problems from copying these indicators but also puts forward the effective solution. Finally, this paper emphasizes the relevant problems in the actual evaluation operation.

Keywords bibliometrics; scientific evaluation; Bradford's Law; Zipf's Law; Lotka's Law; bibliometrics index

1 引言

文献计量学与科学评价是图书情报学科十分活跃的研究领域,它不仅促进着图书情报学科领域的快速发展,对科学学学科也是很好的补充,而且对我国科学研究、科学评价、科研管理与科研规划等都发挥了显著作用。然而,由于长期以来我们采取的文献计量学理论或评价指

标大多来源于数十年前,是否适应当前的科学评价也没有过多的去分析评判,致使我们的实际分析可能出现问题,甚至可能出现与实际情况的背离。

多年的信息分析与科学评价领域的实践,笔者深深地体会到,过去的一些理论、指标等对当今的科学评价依然有着指导意义,但不能完全照搬,应当根据当前的科学发展和文献出版状况进行必要的分析,合理使用,有效借

鉴,甚至进行改造,使分析评价结果更加客观、合理和科学。因此本文将对文献计量学定律以及一些常用评价指标作一分析,阐述笔者对这些问题的思考及看法。

2 文献计量学定律应用中的思考

文献计量学三大定律是信息分析与定量评价的理论基础,在文献计量学领域发挥着重要指导和引领作用。文献计量学植根于三大定律肥沃土壤,获得蓬勃发展。但这些理论已提出近百年,是否适合于当今“信息爆炸”、文献激增的环境?这是文献计量学领域十分关心的问题。

2.1 布拉德福定律的思考

布拉德福定律 (Bradford's Law or Bradford's Distribution) 是英国文献学家、化学家布拉德福先生于 20 世纪 30 年代创立,主要描述了文献分散规律的经验定律。布拉德福定律认为,如果按科技期刊刊载某专业论文的数量多少,以倒序排列,则可以分出一个核心区 and 相应的几个区,每个区的论文数量相等,则核心区期刊数量和相继区期刊数量成 $1:n:n^2 \cdots$ 的关系 ($n>1$)^[1]。布拉德福定律用专业文章数量的多少确定各专业期刊核心区的概念,为我们早期确定学科核心期刊起着很重要的作用。

但是,由于 80 年前的期刊远没有像今天这样专业化,多数期刊是作为综合科学类期刊出现的,没有专业偏向或侧重。所以,布拉德福定律确定学科期刊的核心区具有科学依据。但在今天这样,期刊多采用专业化出版的情况下,单纯采用布拉德福定律确定核心期刊区已经不能适应当今期刊的发展。例如,笔者对图书情报近百种期刊的载文进行了统计,按期刊刊载图书情报专业论文数量的排列,确实符合布氏分布,但大多数数学界认为的核心期刊却不在此列。例如,《中国图书馆学报》、《情报学报》、《大学图书馆学报》等均不在核心区。因此,完全照抄布氏定律遴选核心期刊显然已经不适合了。

但是不是说布拉德福定律就不适合当今的文献计量规律,回答是否定的。虽然依赖刊载专业论文的数量,不能确定核心期刊的分布,但布氏思想对当今选择核心期刊仍有很大的指导作用,如果我们换一个角度,把图书情报期刊的影响因子从大小排序,并依据布拉德福定律分区,发现每个区影响因子之和相等的话,其期刊数量呈现出布拉德福比例,更加可喜的是,处在核心区的期刊与多个期刊评价系统的核心期刊目录基本相符,这说明,当我们换个角度利用布拉德福定律的话,不仅能为我们确定

核心期刊提供依据,还为我们划分期刊等级提供了非常有价值的参考。因此,我们不能盲目的抄用布拉德福定律,但可以根据布氏思想寻找规律,更科学的应用布拉德福定律。

2.2 齐普夫定律的思考

齐普夫定律 (Zipf's Law) 是美国学者 G.K. 齐普夫于 20 世纪 40 年代提出的关于词汇在文献中出现频次的分布规律,也称词频分布定律、最省力法则、齐氏分布定律。齐普夫定律的主要思想:如果把文献中词的频率 (f) 从高至低递减排序,并用自然数给这些词编上等级序号 (r), 则有 $fr=C$ (C 为常数)^{[1]166-167}。这个定律与 80/20 规律有异曲同工之效,它提示我们在进行评价分析过程中可以花费较小的力气取得较好的效果,以及如何选取更加科学的指标和有效参数获得理想的结果。

齐普夫的词频分布规律的思想为我们分析学科研究热点提供了一个非常好的思路,我们可以通过文章标引的关键词次数分析各学科研究热点,也可以根据关键词各年出现的次数和排序变化预测研究趋势。许多学者利用这一规律和方法进行过相关研究,并获得了较好的效果^[2-4]。但根据研究结果我们发现,单纯关键词频率统计方法,还不能确切体现研究热点,一些较为泛指的词出现的频率会很高,如“数字图书馆”、“信息资源建设”、“检索技术”等,从这些关键词很难观察到具体的研究热点和趋势,如果能够将与这些高频词经常在同一文献共同出现的关键词进行统计排列,则更能反映出研究的热点和趋势。

在进行学术评价中,涉及许多评价指标,一个评价体系看似指标越全似乎越好。其实,齐普夫定律中最省力思想告诉我们,从效率出发并不是指标越多就越好,大量的指标数据采集将会耗费更多的时间和精力,而大量的评价研究是有时间要求的。因此,建立高效的评价体系应当分析所有指标在评价体系中的作用,选择影响大的、易获取的指标,对在评价中产生微小作用的一些指标并且需花费大量精力的指标可以舍弃。

2.3 洛特卡定律思考

洛特卡定律 (Lotka's Law) 揭示了科学研究人员数量与所著文献数量之间的关系,认为发表两篇论文的作者数量是发表 1 篇论文作者数量的 $1/4$,发表 3 篇论文的作者数量是发表 1 篇论文作者数量的 $1/9$,如此类推,发表 n 篇论文的作者数量是发表 1 篇论文作者数量的 $1/n^2$ ^{[1]149}。洛特卡定律又称倒平方定律,描述了科学生产率的经验

规律,是美国学者 A.J.洛特卡于 20 世纪 20 年代提出的。

洛特卡定律为我们评价核心学者提供了一条思路,即发文越多越可能成为核心学者。这在十多年前可能是评价核心学者一种非常有效的途径,但在追求量化考核学者绩效的今天,量化考核学术业绩的政策促进了大量学术垃圾横空而出,抄袭事件举不胜举。所以,这种定量的方法确定核心作者,已经失去了原有价值。正如,有学者呼吁“学术评价:超越量化模式”^[5],希望能采取代表作制度等有效方法取代量化模式。当然,我们可以换个角度应用洛特卡定律,例如,采取作者成果的被引情况来考量作者的影响或核心学者,如果这种考量能够排除自引和小团体引用,则可能更有价值。

今天,学界考察科研工作者的科研绩效又推出了 H 指数系列方法^[6-7],主要是通过被引来分析个体学者的学术影响力,如果我们将 H 指数和被引次数以及发文量结合考察学者,可能就会得到较为准确的核心学者群,还能够产生出学者的科研绩效效率。

3 评价指标应用中的思考

科学评价的实施需要建立评价体系,评价体系有一系列指标和方法构成。长期以来我们主要使用国际惯用指标进行学术评价,但这些指标是否适合当今的学术资源发展规律?是否适应我国当前的学术研究环境?这是我们应当思考的问题。笔者下文将对定量评价中几个常用指标进行一定评述。

3.1 影响因子的思考

影响因子是上个世纪 70 年代由加菲尔德创建,主要用来评价期刊的学术影响。它的计算方法:若计算期刊某年的影响因子,则为该刊此年的前两年所发文章在这一年的篇均被引率。即,前两年所发文章的数量为分母,前两年所发文章在计算当年被引次数做分子,计算所得之值即为该刊这一年的影响因子。影响因子诞生于 40 年多前,主要用于反映期刊近期的学术影响,早期主要用于评价自然科学与工学期刊,现在所有期刊均采用这一指标作为期刊的主要评价指标是否合适,笔者给出以下思考。

第一,当前的影响因子的计算对各学科期刊都是一个统一的公式,没有体现出学科间的差异,尤其是人文科学,它和自然科学的成果引用的时间关系是完全不一样的,自然科学的引用绝大多数来自于近期发表的文章,随着时间的推移,较早发表的文章其引用价值大大降低,所

以,由近两年文章被引计算影响因子尚具有一定的合理性。但人文科学则不一样,如历史、文学、语言学等,可能几十年前的研究成果,对今天的研究同样具有重要价值,并且还会大量引用。所以,仍用前两年文章被引计算人文科学期刊的影响因子是否合理,也值得商榷。

第二,影响因子前两年为计算区间是否适合。我们知道,如今一篇文章从撰写到发表,一般需要经过一年以上(主要在杂志社积压时间较长),也就是说,文章很难在发表后的一年以内被引用。这一现象促使我们质疑影响因子前两年的区间是否适合?笔者曾经做过统计,社会科学期刊的文章发表后被引用的高峰值在发表后的第三年^[8],那么,仍然用前两年的计算方法就有悖于影响因子表达的初衷,应当对学科各年份被引比例做一统计分析,选择恰当的年份进行影响因子的计算区间。

第三,影响因子按年份计算,这对在一年中不同月份发表的文章在第二年的被引机会(计算在影响因子中的被引机会)大不相同,如,12 月份发表的文章,几乎在发表后的第三年才可能被引用。这就导致同年内,不同时期发表的文章在影响因子的计算上存在很大差异。解决这种差异,应当采取将影响因子的按年度计算的方法改变为按月份的计算方法,这种改变虽然使影响因子的计算方法变得复杂,但这在当前普遍用计算机计算处理的环境下,并不是件困难的事情。

3.2 即年指数的思考

即年指数是指期刊当年发表的文章在当年被引用的比例。该指标主要用于反映期刊对学术领域的及时反应速率,也体现了期刊对当前研究热点的关注度。正是由于这样的作用,所以该指标在许多期刊评价体系中得到使用。但是这一指标在我国是否完全适用,是否需要改进,也将引起我们思考。

正如前文所述,一篇文章发表后一年内很难有其他文章引用它,这主要是发文的时滞原因。我们曾做过一个统计,在 CSSCI 中,当年发表的文章在当年有被引的期刊种数只占 CSSCI 所有有被引论文的期刊种数的 42%,如果除去被引用 1~2 次的期刊,那么剩下的有引用的期刊只占不到 20%的比例^[8],这样的结果是不能有效地反映该指标的作用的,该指标评价人文社会科学期刊必须要进行改进。

从另一个角度我们发出提问,即年指数对评价自然科学期刊可能有很重要的功能,对社会科学期刊也能起

一定作用的话,那么,对评价人文科学类期刊是否具有价值则值得考证,如果这一指标对人文科学期刊没有评价意义,则是否需要改正我们的人文科学期刊评价标准。

3.3 网络即年下载率的思考

网络即年下载率指标是在网络资源普及的环境下产生的对期刊评价的一个新指标,这一指标可以了解到所收录期刊中文章被读者阅读的情况,这是过去传统期刊、传统评价指标所无法完成的。可以说,这一指标从一个独特的角度反映了期刊在论文传播中所起到的作用,是对其他指标一个很好的补充。

但是该指标在使用过程中还是存在一定的不足。例如,在统计数据库中,如果没有收录该期刊,则该期刊该项指标则为0。也可能有的评价系统将该期刊在其他数据库中的下载率作为补充,但不可否认的是,这个补充的指标是不对称的,他们的用户群是不一样的,可能产生的差别非常之大,势必影响该期刊在这一指标上的劣势,造成评价结果的不合理性或缺乏公正性。也有的期刊在任何数据库中都没有收录,这将导致该刊这一指标值的空缺而无法判定此刊。

3.4 其他相关指标的思考

在人文社会科学期刊评价中,许多系统采用二次文献转载指标,毫无疑问,有的二次文献转载期刊转载数量少,学科也不是非常全,有的学科很少有文章被转载,这势必影响这些学科在这一指标上对期刊的鉴别性。

还有些期刊考虑到知识产权问题不允许二次文献的全文转载,这也影响了这些期刊在二次文献转载指标上的缺损,也影响了这些期刊被全面的评价。

有些学者在评价期刊时,认为新设立指标体现了评价的创新性,例如,期刊H指数评价期刊层出不穷,但在使用H指数评价期刊时要特别注意所产生的期刊H指数是否具有分辨力,一般来说,小数据下的H指数不能说明问题,H指数需要大容量数据的支撑。

4 评价操作中需要注意的问题

以文献计量的方法进行科学评价主要依赖二次文献数据,在操作过程中,除了需要选择正确的评价指标,进行大量的数据处理,以及统计计算工作。在具体操作中还需注意一些相关问题。

4.1 人为制造数据的排除

十年前当人们对学术评价过程或指标(如期刊评价)

尚未理解时,这时的定量评价结果可以说是客观、公正,结果也是真实的反映了评价对象的实际情况。但目前大家都了解了评价指标的计算方法,一些不正之风也由此而出。有的期刊为了提高自己的被引数量,采取大量的自引、互惠引用、联盟引用等手段,我们应当在评价过程中剔除这些人为因素产生的虚假数据。

学术期刊中的自引本无可厚非,因为在学者的研究中会自然的引用相关文章,作者一般注重相关主题的文章,而不会刻意选择期刊。这种自然行为,会造成期刊一定的自引,但这种自引超过一定比例,可能就会存在问题。在SCI中,自引率超过20%就可能被剔除来源期刊。虽然我们不可能采取一票否决的剔除手段,但我们可以根据同类期刊的平均自引率,对高自引期刊削减被引次数,让被引回归自然。

对互惠引用或联盟引用(两个期刊人为的相互引用或多个期刊人为的循环引用),我们可以建立期刊引用网络,用一定的阈值呈现期刊间高引用网络,然后将网络呈现数据与过去数据进行比较,发现巨变情况,就很可能存在互惠引用和联盟引用。我们曾对CSSCI来源期刊建立了引用网络,发现了许多互惠引用和联盟引用,经过调查并得到证实。所以,在评价过程中当发现了这些人为制造的引用,可采取剔除手段,将自引、互惠引用和联盟引用产生的数据剔除掉。

4.2 选择指标的考证

在评价过程中,我们应当对每一个指标进行实际数据的考证,确定这些指标是否适合评价。如,当我们选择了影响因子作为评价期刊的指标后,应分析所计算的年代区间的论文被引是否占据被引总量的大部分,如果某学科论文在选取的时间区段间被引数量尚不足全部被引数量的一半,则证明采用这一区段计算影响因子不能体现该学科期刊水平的真实情况,应当进行修正。例如,SCI已经发现原有影响因子的计算方法存在一定不足时,即增加了5年区间计算影响因子的算法。

在进行科学评价中,经常会采用一些新的评价指标,在采用前同样需要进行多方面数据调研和分析。例如,当学界采用H指数评价学者的学术影响力获得认可后,一些学者就将其引入多个评价领域,例如将H指数引入评价期刊^[9-10],这些引入作为探索性研究应得到鼓励,但真正用于期刊评价体系,并用于选择核心期刊就需要仔细鉴别。也就是说,这项指标是否对期刊的优劣有分辨率,

这种分辨是否将优质期刊呈现出来,如果新指标确实可用再纳入评价体系。如,SCI在近几年就引入了期刊特征因子指标评价期刊和论文^[11],并获得了很好的效果。

4.3 显性评价与数据的一致性

文献计量学评价科学研究是一种定量评价方法,既然是定量评价就应当把所有计算结果显性公布,提供公开查询。这种数据公开可以使每个参评对象了解自己的优劣势所在,也保证了评价的公开、公正。在定量评价中切忌暗箱操作,只有将统计数据全部呈现给公众,才能得到公众的信任,也便于公众从数据中发现问题,完善评价体系。

另外,有的评价系统自己没有数据,完全依赖其他机构提供数据。这时,评价机构应当对数据进行严格的检验,对缺损数据要在相同的环境下补缺,不可随意性补充数据,否则难以保证数据的一致性。在使用多方数据源的评价系统,要注意各方数据的协调性,切忌互相矛盾数据进入同一评价系统。

4.4 弱化排名探索科学特征

借助定量分析方法不仅仅是给分析对象排座次,更重要的是借助定量分析的方法探索科学研究内在关系。因此,我们在进行科学评价的过程中,尽可能淡化排序功能,注重借助文献计量学原理深度挖掘科学研究领域、研究主题之间的关系,建立科学文献间的联系关系。通过对大量的数据分析,探寻数据之中存在的科学规律,发现科学研究源流,以及呈现各学科研究领域的研究特征等。也就是说,以文献计量学为理论基础,将分析评价的成果直接服务于科学研究。

5 结语

长期以来的信息分析与科学评价工作,使我们深深体会到,进行科学评价可以促使我们对科学研究状况的了解,可以更有效地进行科学研究,我们切不可把科学评价和科学研究对立起来。但不科学的评价也会造成一些负面效应,我们必须选择恰当的评价指标,采用科学的评

价方法,摒弃不合理因素,使科学评价工作健康发展,真正做到促进科学研究的蓬勃发展。

参考文献:

- [1]《中国情报学百科全书》编委会.中国情报学百科全书[Z].北京:中国大百科全书出版社,2010.
- [2]郑江淮等.中国经济学研究热点与趋势分析——基于CSSCI的分析[J].重庆大学学报(社会科学版),2009,15(1):64-68.
- [3]何日取.基于CSSCI的社会学研究热点与趋势分析(2005—2006)[J].学海,2009,(2):185-189.
- [4]苏新宁.图书馆、情报与文献学研究热点与趋势分析(2000—2004)——基于CSSCI的分析[J].情报学报,2007,26(3):373-383.
- [5]陈洪捷,沈文钦.学术评价:超越量化模式[N].光明日报,2012-12-18(15).
- [6]金碧辉,Rousseau Ronald. R指数、AR指数:h指数功能扩展的补充指标[J].科学观察,2007,(3).
- [7]张学梅.用h指数对我国图书情报学界作者进行评价[J].图书情报工作,2007,(8).
- [8]苏新宁.构建人文社会科学学术期刊评价体系[J].东岳论丛,2008,(1):35-42.
- [9]杨建林.哲学期刊的h指数分析——基于CSSCI(1998—2007年度)数据[J].西南民族大学学报(人文社科版),2009,(5):64-67.
- [10]邓三鸿,金莹.图书馆、情报与文献学CSSCI来源期刊的h指数分析——基于CSSCI(1998—2007年度)数据[J].西南民族大学学报(人文社科版),2009,(6):95-98.
- [11]任胜利.特征因子(Eigenfactor):基于印证网络分析期刊和论文的重要性[J].中国科技期刊研究,2009,20(3):416-418.

作者简介:苏新宁(1955-),男,南京大学信息管理学院教授,博士生导师,研究方向:信息分析与科学评价,信息智能处理与检索。