基于 h 相关指数的科研人员评价案例分析*

□邓红巧 董颖 刘凤琴 秦进红 谢丽斌

摘要 对 h 指数和 g 指数进行比较和研究,重点分析了科研人员 h 指数逐年变化的态势,并以 h 指数逐年变化态势为研究对象,从科学引文索引(SCI)数据库中采集和整理科研人员的 h 指数数据,发现科研人员从初期获得 h 指数后,会经历缓慢的累积,然后会有一段快速增长期,最后会逐渐趋于平缓。

关键词 h 指数 g 指数 时间演化

分类号 G353.1

DOI 10. 16603/j. issn1002—1027. 2016. 05. 019

1 h 指数概述

Hirsch于2005年提出旨在评价科研人员绩效的指标——h 指数。h 指数的计算表述如下:设当某一位科研工作者,他在此之前一共发表了 N 篇科研论文,其被引频次最高时序号是 1,将这些论文按照被引频次的降序排列,当且仅当第 h 篇论文至少获得了 h 次也就是大于或等于被引用次数,其余的所有 N-h 篇论文的引文数全都小于 h 时,那么就称这个 h 值是这位科研工作者的 h 指数^[1]。与个体科研人员发文数量和引文数量等指标相比,很多学者认为 h 指数综合衡量了论文的数量和影响力。h 指数高不仅表明被评价对象发表的论文质量高,而且高质量的论文数量也多,弥补了传统文献计量学单项数量指标只能体现数量的遗憾。

h 指数在被提出之初,主要作为衡量科研人员个人成就的计量指标, Hirsch 以及后续研究者对多个学科的科研人员 h 指数进行了研究。例如物理学,如果在物理行业已经从事科研工作 20 年,他们的 h 指数可以达到 20,就是成功的科学家;能够达到 40,就算是一名杰出的科学家;那些 h 指数能够达到 60 的科学家可以将其视为象牙塔塔尖的人,他们均是首屈一指的大科学家。不同国家、不同学科的 h 指数分布趋势差异较大,对科学家影响力的划

分也不同,例如同等学术时间段生物科学家的 h 指数会稍稍大于物理学家的 h 指数。在美国要想成为一名美国物理学会会员,h 指数的要求最少得在 15至 20 之间,而且美国国家的科学院院士的 h 指数通常应该都要大于或者等于 45。还有人就数学、生物学等学科研究人员的 h 指数分布情况进行研究。这类均属于基于 h 指数的验证性研究[1]。

针对 h 指数的缺点,陆续有研究人员开展基于 h 指数的改进研究,其中既有 g-indice、r-indice 等新 指数的提出,也有 hf 等 h-indice 改进指数^[2]。另外,还有人从作者合作和署名角度来给 h 指数赋予权重。例如 2006 年,Batista 提出根据 h 核内论文的作者数量的均值平分 h 指数^[3],h 核内论文是指被引次数为 h 次及以上的作者论文。但由于每个作者对于论文的投入和贡献都是不均衡的,平分的做法有欠公平,也会在某种程度上妨碍科研人员之间的合作。Egghe 在 2008 年提出根据作者排名来计算论文的引文数量再提取 h 指数^[4]。 2009 年 Schreiber提出修正 h 指数的研究也都是基于类似的想法,根据合作作者数量重新计算论文的被引次数再提取 h 指数^[5]。

^{*} 国家社会科学基金研究项目"数字图书馆动态组合学科服务模式研究"(编号:14BTQ017)和教育部人文社会科学研究规划基金项目"团体导向的数字图书馆评价研究机制"(编号:12YJA870003)及黑龙江省高校图工委"基于 Living library 高校图书馆学科化知识服务创新研究"(编号:2013-B-057)的研究成果之一。

通讯作者:邓红巧,ORCID:0000-0003-0830-1302,dhq1234@126.com。

學圖書館學報

2 h 相关指数和改进指数

除了 h 指数外,陆续有更多简单易于计算的改 进指数被提出,其中 g 指数是在 h 指数提出不久后 于 2006 年 Egghe 提出。论文按被引次数排序后相 对排前的累积被引至少 g²次的最大论文序次 g,亦 即第(g+1)序次论文对应的累积引文数将小于(g+ 1)²。从定义可以看出,g≥h,而按被引量排序靠前 的文章的被引次数越大,g指数越大[4]。设一个科 研人员发表了 N 篇文章,将这些文章按序号排列, 序号为1的为N篇文章的最大被引频次(其实与 h 指数没有区别),此时多出两个计算单位,一个是序 号的平方,另一个是所有累积的被引频次,当某一序 号开始它的平方正好大于或者等于累积的被引频 次,那么这个序号则称为该科研人员的 g 指数,倘若 出现这种特殊情况即序号的平方根不能恰好等于而 是小于对应的累积被引次数,那么则最接近累积被 引次数的序号为g指数。

更多的研究是探索基于作者在论文中的贡献度来改进 h 指数,这类研究主要有两种视角:一是限制论文成果集合,即根据作者署名顺序和署名规则仅提取作为主要贡献人的署名成果,并纳入 h 指数计算范畴。二是考虑全部论文成果集合,但根据作者署名将每篇论文的被引次数赋予权重,将成果的被引次数加权后再重新计算序列提取 h 指数。也就是说,根据每篇论文的作者数量平分引用次数或根据作者排序加权后计算引文次数再提取 h 指数^[5]。

3 h 相关指数和改进指数的案例分析

h 指数及 g 指数等相关指数的提出陆续给科研人员的评价提出更加科学、合理的方法,相关指标也各具方法特性。通过对同一机构科研人员的相关指数进行比较和评价,对指标的评价特点进行分析和总结。

3.1 案例数据描述

选取细胞生物学领域 8 位 h 指数在 25-70 之间的研究人员,根据基本科学指标数据库(Essential Science Indicators, ESI)选定细胞生物学领域,按篇均被引次数从高到低排列,选取 Bishop, JM、Goetzel, EJ、Jan, YN、Mostov, KE、Vale, RD、Verkman, AS、Walter, P、Werb, Z 等 8 位研究人员作为研究样本。

登录 Web of Science 数据库,点击核心文集,机

构扩展并检索机构 Univ Calif San Francisco,年份设置选取 1981-2014年。选取检索类别为细胞生物学(Cell Biology),排序方式为按被引次数降序排列。

3.2 h 指数和 g 指数的比较与分析

按照 h 指数的排序,可以将这些科研人员的贡献排名为 Jan、Bishop、Vale、Verkman、Mostov、Walter、Werb、Goetzel,8 位研究人员 h 指数最低为28,最高为71,差异较大,他们所发表的论文数量分布为102-200。其中论文数量最多的 Walter 的 h 指数为60,论文数量最少的 Mostov 的 h 指数为41。从表1可以看出 Walter 与 Vale 这两位的 h 指数是相同的,但后者的 g 指数高于前者。

表 1 科研人员 h 指数与 g 指数分布表

人名	h指数	g指数	论文数
Bishop, JM	64	94	117
Goetzel, EJ	28	42	130
Jan ,YN	71	105	125
Mostov, KE	41	53	102
Vale, RD	60	95	151
Verkman, AS	46	62	142
Walter, P	60	42	200
Werb,Z	76	53	111

3.3 考虑作者贡献后的数据对比

考虑作者贡献改进 h 指数通常有两种做法,一是将第一作者和通讯作者纳入研究范畴;二是将署名排在前三的成果纳入研究范畴,根据作者署名次序,仅计算排名在前 1、2 或 3 位的论文被引次数后计算 h 指数。本部分从 2 个方面考虑作者贡献改进 h 指数,一是根据每篇论文的合作作者数量均分 h 指数,这种计算方法假设所有论文作者对文献作出了相同的贡献,将文献被引频次均分给所有作者,均分被引频次后计算方法与 h 指数计算相同;二是仅把第一作者署名的成果计算在内,这种计算方法只考虑第一作者的贡献,对科研人员作为第一作者的文献进行 h 指数计算。8 名研究对象的 h 指数、均分 h 指数、第一作者 h 指数差异较大[6-7]。

由于第一作者论文数在全部文献中所占比例不高,平均比例约为1:10,大部分研究人员的第一作者h指数很低,其中7名研究人员第一作者h指数

均在8以下,Vale的第一作者h指数最高,为15,其第一作者论文数量与文献数量比例为1:5。8名研究对象h指数从38—76,均分h指数从13—43,但第一作者h指数除Vale外,数据在1-7变化,差异并不明显。可见,用第一作者h指数衡量研究者的研究成果,对合著论文较多的作者并不适合。

均分 h 指数与 h 指数的平均比例约为 1:2,其中 Bishop 和 Walter 均分 h 指数与 h 指数差异最小,比例约为 2:3。这两位作者的平均作者排名和平均论文人数均为 3人,是 8名研究人员中排名最靠前的。均分 h 指数受论文总数、论文被引频次及平均论文数量影响,在一定程度上能够反映合著论文作者对研究成果的贡献。

作者	第一作 者论 文数	平均 排名	平均作者数	h指数	第一 作者 h 指数	均分 h 指数
Werb,Z	19	4	5	76	7	43
Bishop, JM	11	3	3	64	5	39
Goetzel, EJ	26	3	4	28	6	13
Jan, YN	7	4	4	71	4	35
Mostov, KE	2	4	5	41	2	20
Vale, RD	31	3	4	60	15	33
Verkman, AS	11	4	4	46	5	23
Walter, P	11	3	3	60	3	36

表 2 研究人员文献数据对比

4 考虑时间因素的 h 指数分析

根据研究对象引文数据的时间分布,对 8 位研究人员历年 h 指数进行提取,如图 1 所示。

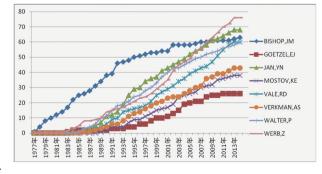


图 1 8 位研究人员历年 h 指数

通过图 1 我们可以看出,Bishop 在 1977 年开始有学术成果,其 h 指数成长曲线从 1977-2001 年呈现快速上升的趋势,2001 年后开始转变为平稳缓慢的增长趋势。

Goetzel 从 1983 年开始有学术成果,其 h 指数成长曲线从 1982—1995 年,呈现一个缓慢增长的态势,1996—2007 年出现快速向上增长的态势,但在2008 年后,又变成平稳缓慢增长。

Jan 是从 1982 年开始有学术成果,在 1985—2012 年,h 指数成长曲线都是呈现出了一个快速向上发展的态势,在 2012 年以后,其曲线变得平稳下来。

Mostov 在 1989 年开始有学术成果,其开始的曲线变化态势有点类似于 Goetzel,但 h 指数成长曲线增长数值要高,且呈现出的是一直持续增长的趋势,在 2012 年后才趋于平稳的发展趋势。

Vale 在 1985 年开始有学术成果,其 h 指数成长曲线在 1985—1997 年间,呈现一个平稳上升的态势,但在 1998—2012 年间,其 h 指数成长曲线则呈现出上升速度非常快的趋势,在 2012 年后,开始趋于平缓发展的状态。

Verkman 在 1986 年开始有学术成就,其 h 指数成长曲线从 1986—2007 年间,呈现一个平稳发展的态势,并在 2007—2008 年间出现一个小幅度快速增长的趋势,但在 2008 年后,其曲线又变成平稳发展的态势。

Walter 在 1984 年开始有学术成果,在 1984—2002 年这段区间里,其 h 指数成长曲线呈现出快速增长的态势,2002—2003 年间,其 h 指数成长曲线暂时持平,但在 2003 年后,P 的 h 指数成长曲线又开始出现平稳快速向上增长的态势。

Werb 在 1981 年开始有学术成就,h 指数成长曲线从 1984 年开始到 2012 年,呈现出一个快速向上增长的态势,通过图 1 可以看出其 h 指数成长曲线非常"陡",其上升势头很快速,并处于高速上升的状态,在 2012 年后,其上升势头逐渐变成缓慢增长的状态。

进一步,以3年为一个时间段,分析每个作者的h指数、均分h指数、第一作者h指数、平均论文人数和平均作者排名。以Werb,Z为例,详细展示他的改进h指数数据。

由图 2 可知, Werb, Z 论文在 1982-2014 年 h 指数、均分 h 指数变化较大。h 指数和均分 h 指数都有了明显提升,其中 h 指数从 3 增长到 76,均分 h 指数从 2 增长到 43。合作论文对 h 指数的提升要比均分 h 指数的提升更为显著。但由于其第一作者论

6 年第5 期

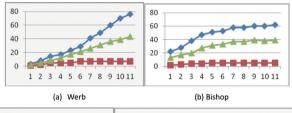
圖

耆

文数有限,第一作者 h 指数无明显差异,与 h 指数及均分 h 指数相比,数据较低。

表 3 Werb, Z 数据

序号	年份	h指数	第一作者	均分	平均	平均
かち	平彻		h指数	h指数	排名	作者数
1	1982 - 1984	3	1	2	3.05	3.14
2	1982-1987	8	2	5	3	3.16
3	1982-1990	14	5	8	2.88	3. 15
4	1982-1993	17	5	12	3. 17	3.38
5	1982-1996	23	5	17	3.32	3.56
6	1982-1999	29	7	21	3.44	3.8
7	1982-2002	41	7	26	3.71	4.08
8	1982-2005	49	7	31	3.93	4.4
9	1982-2008	60	7	36	4.22	4.74
10	1982-2011	70	7	39	4.22	4.75
11	1982-2014	76	7	43	4.37	4.89



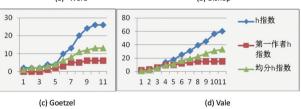


图 2 Werb\Bishop\Goetzel\Vale 相关指数时间趋势

Bishop 论文数据与 Werb, Z 的论文数据比变化相对平滑。第一作者 h 指数变化依然不明显,数据维持在 6 以下。h 指数和均分 h 指数在 3-5 年(即1990-1996 年)提升较显著,h 指数与均分 h 指数都提升了 13,论文成果突出。其他时间段 h 指数与均分 h 指数在缓慢增长。

Goetzel 最初发表论文数量不多,h 指数、均分 h 指数、第一作者 h 指数起点较低。随着其合作论文的发表,各指数有了明显的提升。h 指数从 2 提升为 26,均分 h 指数从 1 提升到 13,第一作者 h 指数从 0 提升到 6。2002 年后,h 指数、均分 h 指数增长显著,h 指数增长尤为突出,由于平均作者数目增多,均分 h 指数增长相对平缓。

Vale 发表文献 h 指数、均分 h 指数均有明显提

高。h 指数由 2 提升到 60,均分 h 指数由 1 提升至 33。与其他研究人员不同的是,其第一作者 h 指数 也有显著提升,2 增长到 15,这与其论文中第一作者 论文比例较高,作者平均论文人数排名较前有关。在 7-11 年(即 2002-2014 年)第一作者 h 指数未出现明显变化,但 h 指数、均分 h 指数增长速率较快。可见,合作论文数目的增多能有效提升 h 指数、均分 h 指数,但对于第一作者 h 指数提升不高。

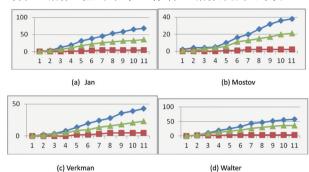


图 3 Jan、Mostov、Verkman、Walter 相关指数时间趋势

四位研究人员 Jan、Mostov、Verkman、Walter的 h 指数、第一作者 h 指数、均分 h 指数的变化与Werb 文献数据变化相似。h 指数、均分 h 指数有了显著提升,Jan、Mostov、Verkman 均分 h 指数在各时间段约为 h 指数的一半,Walter 论文合作人数较少,均分 h 指数与 h 指数比例约为 2:3,差距较小。第一作者 h 指数变化不明显,数值在 5 以下。

5 总结

分析的 8 位研究人员中,其论文大部分为合作 论文,平均论文作者和平均作者排名均为 3-5,第 一作者论文不多。h 指数很高,但不排除其中"挂名 式"合作的可能,单纯用 h 指数进行学术成就评价并 不科学。当 h 指数较大时,提升所需时间也就越长, h 指数较大的科研人员可能在几年都无法提升自己 的 h 指数。在以上 8 组数据中,低起点的科研人员 h 指数增长较快,案例中 Bishop 由于最初 h 指数稍 高,h 指数提升到 60 后,增长速度明显下降,趋于 平缓。

均分 h 指数考虑了合著论文中每个作者的贡献,将文献被引频次均分至每个作者,一定程度上遏制了科研人员挂名合作论文的现象,在现今多作者合作的科研趋势下较单纯采用 h 指数更为科学。在8 名研究人员的文献数据中可以看出,从 1982 —

入學圖書館學都

2014年度,h指数增长很快,均分h指数增长速率稍缓。由于合作论文较多,均分h指数与h指数差距将近一倍。平均作者数量靠前的研究人员,其均分h指数与h指数差距相对较小,比例约为2:3。均分h指数受论文总数、论文被引频次及平均论文数量影响,在一定程度上能够反映合著论文作者对研究成果的贡献。

但同时,在合著论文中不考虑作者的排名,仅均分贡献,虽然方法简单易行,但不同作者在合作论文中付出的贡献差异被忽略了,很少有论文每个作者贡献都是相同的,平均分配无疑有失公平。如本文计算的科研人员论文数据,大部分作者排名在所有论文作者中靠后,平均作者排名与平均作者人数接近,在一定程度上反映出他们在合作论文的贡献比平均值低,得出的均分 h 指数并不客观。

第一作者 h 指数只计算第一作者的文献数据, 能体现科研人员的独立研究水平,避免了论文作弊、 挂名式合作现象对 h 指数的影响。但这种方法忽略 了合著论文中第一作者以外作者对文献的贡献,影 响了作者合作的积极性。

分析的 8 位研究人员文献数据中合作论文占了 很大比例,第一作者论文数与全部文献的平均比例 约 1:10。他们在合作论文中可能作出了不小贡献,作者排名靠前但却无法计入第一作者 h 指数,第 一作者 h 指数大部分很低,其中 7 名研究人员第一 作者 h 指数均在 8 以下。这 8 名科研人员的 h 指 数、均分 h 指数有着不小的差异,但第一作者 h 指数在 1-7 间变化,区分度不高。可见,用第一作者 h 指数衡量研究者的研究成果,对于部分需合作完成的科研科目来说,并不公平,须充分结合其他指数以及作者排名信息。

参考文献

- 1 Hirsch, Jorge E. An index to quantify an individual's scientific research output[J]. PNAS,2005:102(46):16569-16572.
- 2 王梅英,刘雪立. h-指数及其扩展指标的研究进展[J]. 中国科技期刊研究,2011,22(2):184-189.
- 3 Batista, P. D. et al. Is it possible to compare researchers with different scientific interests? [J]. Scientometrics, 2006, 68(1): 179-189.
- 4 Egghe, L. Mathematical theory of the h-and g-index in case of fractional counting of authorship[J]. Journal of the American Society for Information Science and Technology, 2008 (59):1608 -1616.
- 5 Schreiber, M. A case study of the modified hirsch h_m accounting for multiple coauthors[J]. Journal of the American Society for Information Science and Technology, 2009,60(6):1274-1282.
- 6 吕娜. 科研人员 h 指数动态变化趋势分析. 情报理论与实践, 2015.38(5):112-115.
- 7 吕娜. 基于作者主要贡献的 h 指数时间趋势分析. 情报杂志, 2015,34(4):54-58.

作者单位:哈尔滨师范大学图书馆,哈尔滨,150025 收稿日期:2016年4月7日

Case Study of Researcher Evaluation Base on h-index and Related Index

Deng Hongqiao Dong Ying Liu Fengqin Qin Jinhong Xie Libin

Abstract: Based on the research and comparison between h and g index firstly, the paper mainly focuses on analysis of papers of Chinese and foreign scientific research personnel index changes throughout the years, summarizes the regularity of them and discusses relevant advantages and disadvantages. It also makes analysis on the individual researchers focusing on the annual variation of h index trend.

Keywords: h-index; g-index; Time Evolution