

# 期刊下载频次和被引频次的相关性研究

## ——复合H指数视角

朱雯 陈荣 刘颖

(华东理工大学科技信息研究所, 上海 200237)

**摘要:** 利用中国知网引文数据库, 以理、工、农、医、人文、经济、社会科学大类中3门学科在2011—2015年核心期刊的下载频次和被引频次为基础, 通过构建期刊下载和被引的H指数、复合H指数等指标, 分析不同学科期刊2011—2015年下载频次和被引频次复合H指数的相关性系数变化情况及其原因。结果发现, 2011—2015年理、工、农、医类期刊下载频次与被引频次的相关性高于经济、人文、社会科学类期刊; 除文学及临床医学学科外, 理、工、农、经济、社会科学同一大类中学科间的复合H指数间相关性系数大小可以互相推测, 但是各年相关性变化情况不一致; 在期刊评价中, 不同时间、不同学科下载频次和被引频次需赋予不同权重。

**关键词:** 下载频次; 被引频次; 复合H指数; 期刊评价

**中图分类号:** G351; G237.5

**DOI:** 10.3772/j.issn.1673-2286.2018.10.004

在纸质媒介时代, 期刊的被引频次、影响因子<sup>[1-2]</sup>等引文索引指标在评价期刊学术影响力方面发挥了重要作用, 但随着互联网的普及和数字技术的高速发展, 文献资源的传播方式越来越多样化。在Web 2.0时代, 若将单一的被引指标、影响因子作为衡量一本期刊的学术影响力必然存在一定的局限性<sup>[3]</sup> (如滞后性等问题), 并有学者指出利用被引指标评价期刊论文只能表现出原文影响力的30%<sup>[4]</sup>。为改善单一指标的缺点, 学者们开始研究下载频次<sup>[5-7]</sup>及社交媒体使用量<sup>[8-10]</sup>对期刊评价的影响, 而下载频次具有较强的时效性, 可以在短时间内反映论文价值<sup>[11]</sup>, 并且能在一定程度上反映已被使用但未被最终引用论文的价值<sup>[12]</sup>, 因此下载频次有望成为一个新的考量科学论文传播和影响力的有用的工具和指标<sup>[4]</sup>。

目前, 下载频次与被引频次作为期刊学术影响力的重要指标, 其相关性研究受到学者的广泛关注, 主要集中在多学科单篇论文的被引频次和下载频次的相关性或者单一学科单年期刊的被引频次和下载频次的相关性<sup>[13-17]</sup>, 多使用基于篇均被引和篇均下载的数据处理方法。但由于统计方法、学科及分析角度等的差异性和

局限性, 仅利用下载量和被引量简单平均的方法研究相关性, 容易出现数据稳定性差等问题。学者们对两者之间的关系观点不一, 并且较少有论文以期刊为单位探讨期刊的下载频次与被引频次的相关性在不同年份的变化情况。

由于复合H指数大幅提高了H指数的区分度<sup>[18]</sup>, 不仅能反映出期刊研究的创新性, 还兼顾了文章数量、质量及高被引论文对即时学术影响力的影响。因此本文通过构建期刊的复合下载H指数 ( $S_d$ 指数) 和复合被引H指数 ( $S_c$ 指数) 等指标, 探讨不同学科2011—2015年下载频次与被引频次复合H指数的相关性系数变化情况及其原因, 为建立综合、完善的期刊评价指标体系奠定基础。

## 1 研究方法设计

(1) 被引H指数 ( $H_c$ 指数)。根据H指数定义<sup>[19]</sup>, 对期刊当年发表的n篇论文组合而言, 在统计年有h篇论文每篇至少被引用h次, 剩下的n-h篇论文每篇被引用的次数都小于h次, 则h为该刊当年被引H指数。如某

期刊2011年共有500篇文章,在2017年有20篇论文每篇至少被引用20次,剩下的480篇论文每篇被引用的次数都小于20次,则该刊2011年被引H指数为20。

(2) 下载H指数 ( $H_M$ 指数)。对期刊当年发表的n篇论文组合而言,在统计年有h篇论文每篇至少下载h次,剩下的n-h篇论文每篇下载的次数都小于h次,则h为该刊当年下载H指数。如某期刊2011年共有500篇文章,在2017年有312篇论文每篇至少下载312次,剩下的188篇论文每篇下载次数都小于312次,则该刊2011年下载H指数为312。

(3) 被引G指数 ( $G_C$ 指数)。根据G指数定义<sup>[20]</sup>,对期刊当年发表的n篇论文组合而言,在统计年,论文按被引次数降序排列,并按照降序将每篇论文的序号平方,同时按照同样的顺序将每篇论文被引次数逐一累加,当前g篇论文累计被引次数大于或者等于第g+1篇论文的序号平方时,则g就为该刊当年被引G指数。

(4) 下载G指数 ( $G_M$ 指数)。对期刊当年发表的n篇论文组合而言,在统计年论文按下下载次数降序排列,并按照降序将每篇论文的序号平方,同时按照同样的顺序将每篇论文下载次数逐一累加,当前g篇论文累计下载次数大于或者等于第g+1篇论文的序号平方时,则g就为该刊当年下载G指数。

(5) 被引R指数 ( $R_C$ 指数)。根据R指数定义<sup>[21]</sup>,对期刊当年发表的n篇论文组合而言,在统计年有h篇论文每篇至少被引用h次,剩下的n-h篇论文每篇被引用的次数都小于h次,则h篇论文的被引次数总和的平方根为该刊当年被引R指数。

(6) 下载R指数 ( $R_M$ 指数)。对期刊当年发表的n篇论文组合而言,在统计年有h篇论文每篇至少下载h次,剩下的n-h篇论文每篇下载的次数都小于h次,则h篇论文的下下载次数总和的平方根为该刊当年下载R指数。

(7) 复合被引H指数 ( $S_C$ 指数)。某刊当年被引H指数÷所有期刊样本当年总被引H指数+某刊当年被引G指数÷所有期刊样本当年总被引G指数+某刊当年被引R指数÷所有期刊样本当年总被引R指数=复合被引H指数。

$$S_C(x) = \frac{H_C(x)}{\sum_1^n H_C(n)} + \frac{G_C(x)}{\sum_1^n G_C(n)} + \frac{R_C(x)}{\sum_1^n R_C(n)} \quad (1)$$

其中,  $S_C(x)$  表示在统计年第x本期刊发表当年的复合被引H指数,  $H_C(x)$  表示第x本期刊当年的被引H

指数,  $H_C(n)$  表示第n本期刊当年的被引H指数 (n为整数),  $G_C(x)$  表示第x本期刊当年的被引G指数,  $G_C(n)$  表示第n本期刊的当年的被引G指数,  $R_C(x)$  表示第x本期刊当年的被引R指数,  $R_C(n)$  表示第n本期刊的当年的被引R指数。

(8) 复合下载H指数 ( $S_M$ 指数)。某刊当年下载H指数÷所有期刊样本当年总下载H指数+某刊当年下载G指数÷所有期刊样本当年总下载G指数+某刊下载R指数÷所有期刊样本当年总下载R指数=复合下载H指数。

$$S_M(x) = \frac{H_M(x)}{\sum_1^n H_M(n)} + \frac{G_M(x)}{\sum_1^n G_M(n)} + \frac{R_M(x)}{\sum_1^n R_M(n)} \quad (2)$$

其中,  $S_M(x)$  表示在统计年第x本期刊发表当年的复合下载H指数,  $H_M(x)$  表示第x本期刊当年的下载H指数,  $H_M(n)$  表示第n本期刊的当年的下载H指数 (n为整数),  $G_M(x)$  表示第x本期刊当年的下载G指数,  $G_M(n)$  表示第n本期刊当年的下载G指数,  $R_M(x)$  表示第x本期刊当年的下载R指数,  $R_M(n)$  表示第n本期刊当年的下载R指数。

## 2 实证研究

本文以中国知网的《中国学术期刊网络出版总库》及《中国引文数据库》为数据源,以核心期刊为例,从理、工、农、医、经济、人文、社会科学中各选择3门有代表性的学科(见表1)。为保证数据的完整性,每门学科的期刊按照综合影响因子排序,选取排名前5名和后5名共210本期刊,收集每本期刊在2011—2015年每篇论文的下下载次数和被引频次,检索时间截至2017年8月1日。本文随后将论文题名、下载次数和被引频次记录组成每本期刊每年记录的3个特征项,将其汇总到Excel中,运用Excel进一步整理筛选,共计306 320篇期刊论文纳入统计分析。

表1 选取的学科

大 类	学 科
工业技术	电子通信技术、环境科学、仪器仪表工业
农业科学	林业、渔业、农学
医药、卫生	基础医学、临床医学、药学
自然科学	化学、数学、物理
文化、教育、历史	文化理论/新闻事业、图书情报、文学
经济	贸易经济、综合性经济科学、经济学
哲学、社会学、政治、法律	法律、民族学、综合性人文社会科学

2.1  $S_C$ 指数与 $S_M$ 指数相关性概况

本文利用SPSS 22.0对样本数据进行K-S检验,结

果显示 $S_C$ 指数与 $S_M$ 指数并非所有数据符合正态分布,因此选取Spearman系数计算指标间的相关性系数,结果见表2。

表2 2011—2015年各学科 $S_C$ 指数与 $S_M$ 指数相关性概况

大 类	学 科	2011年相关性系数 (Sig.)	2012年相关性系数 (Sig.)	2013年相关性系数 (Sig.)	2014年相关性系数 (Sig.)	2015年相关性系数 (Sig.)
工业技术	电子通信技术	0.770 (0.009)	0.818 (0.004)	0.806 (0.005)	0.721 (0.019)	0.503 (0.138)
	环境科学	0.855 (0.002)	0.770 (0.009)	0.891 (0.001)	0.952 (<0.001)	0.721 (0.019)
	机器仪表工业	0.903 (<0.001)	0.855 (0.002)	0.964 (<0.001)	0.988 (<0.001)	0.891 (0.001)
农业科学	林业	0.939 (<0.001)	0.939 (<0.001)	0.964 (<0.001)	0.952 (<0.001)	0.976 (<0.001)
	渔业	0.661 (0.038)	0.539 (0.108)	0.515 (0.128)	0.321 (0.365)	0.782 (0.008)
	农学	0.673 (0.033)	0.697 (0.025)	0.794 (0.006)	0.758 (0.011)	0.806 (0.005)
医药、卫生	基础医学	0.697 (0.025)	0.806 (0.022)	0.794 (0.006)	0.806 (0.005)	0.644 (0.044)
	临床医学	0.770 (0.009)	0.564 (0.09)	0.370 (0.293)	0.394 (0.26)	0.316 (0.374)
	药学	0.818 (0.004)	0.709 (0.022)	0.745 (0.013)	0.758 (0.011)	0.915 (<0.001)
自然科学	化学	0.903 (<0.001)	0.855 (0.002)	0.879 (0.001)	0.915 (<0.001)	0.915 (<0.001)
	数学	0.661 (0.038)	0.673 (0.033)	0.794 (0.006)	0.576 (0.082)	0.418 (0.229)
	物理	0.818 (0.004)	0.77 (0.009)	0.705 (0.023)	0.867 (0.001)	0.926 (<0.001)
文化、教育、历史	文化理论/新闻事业	0.430 (0.214)	0.673 (0.033)	0.539 (0.108)	0.721 (0.019)	0.620 (0.056)
	图书情报	0.515 (0.128)	0.297 (0.405)	0.576 (0.082)	0.030 (0.934)	0.212 (0.556)
	文学	0.866 (0.001)	0.866 (0.001)	0.844 (0.002)	0.997 (<0.001)	0.883 (0.001)
经济	贸易经济	0.333 (0.347)	0.200 (0.580)	-0.030 (0.934)	0.212 (0.556)	0.624 (0.054)
	综合性经济科学	0.467 (0.174)	0.491 (0.150)	0.418 (0.229)	0.576 (0.082)	0.770 (0.009)
	经济学	0.600 (0.067)	0.721 (0.019)	0.867 (0.001)	0.903 (<0.001)	0.818 (0.004)
哲学、社会学、政治、法律	法律	0.309 (0.385)	0.321 (0.365)	0.406 (0.244)	0.394 (0.260)	0.527 (0.117)
	民族学	0.515 (0.128)	0.418 (0.229)	0.685 (0.029)	0.584 (0.077)	0.552 (0.098)
	综合性人文社会科学	0.224 (0.533)	0.370 (0.293)	0.794 (0.006)	0.927 (<0.001)	0.818 (0.004)

①整体而言,除文学学科及临床医学学科外,2011—2015年理、工、农、医大类中的其他学科的相关性系数普遍高于人文、经济、社会科学,总体相关性保持在中等相关与强相关之间;而2011—2015年人文、经济、社会科学大类中的其他学科的相关性处于中等程度相关或弱相关。②除文学及临床医学学科外,同一大类中各学科之间 $S_C$ 指数与 $S_M$ 指数相关性系数相差不大。③除2013年贸易经济学科外,2011—2015年其他学科 $S_C$ 指数与 $S_M$ 指数呈正相关。

分析原因:①学科间自身存在较大差异,导致理、工、农、医类与人文、经济、社会科学类相关性系数差异大。理、工、农、医类的期刊论文被下载后得到有效引用,导致理、工、农、医大类中学科的 $S_C$ 指数与 $S_M$ 指数相关性强,其相关性系数大于经济、人文、社会科学学科。②同一大类中学科期刊发展具有一定的相似性,

学科之间交叉程度高,这可能是同一大类中各学科之间 $S_C$ 指数与 $S_M$ 指数相关性系数相差不大的主要原因。

③统计每年期刊的低被引(被引频次小于4)论文数及每年期刊参考外文文献比(总参考外文文献数÷当年期刊总发文量)发现,贸易经济学科2013年发表期刊的低被引论文数高于其他年份,这可能是导致呈现负相关或无相关的主要原因;而临床医学期刊参考外文期刊的数量高于大类中其他学科,这可能是导致相关性系数低于大类中其他学科的主要原因。④进一步探讨文学学科,首先,依据各学科 $S_C$ 指数与 $S_M$ 指数方差发现文学学科2011—2015年各数据波动较小,数据的稳定性较高;其次,文学期刊对现代中国的文化传播起着重要作用,本文所选取的文学学科样本期中论文大部分是研究中国文学,学者参考中文文献的比例较高,导致文学学科的期刊论文下载后得到有效引用。

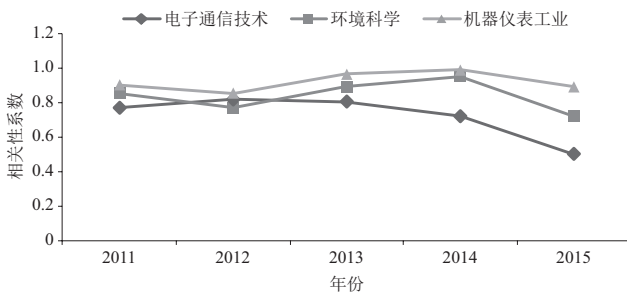
## 2.2 2011—2015年 $S_C$ 指数与 $S_M$ 指数相关性变化情况

本文将上述各学科 $S_C$ 指数与 $S_M$ 指数的相关性系数放入Excel中绘制折线图。

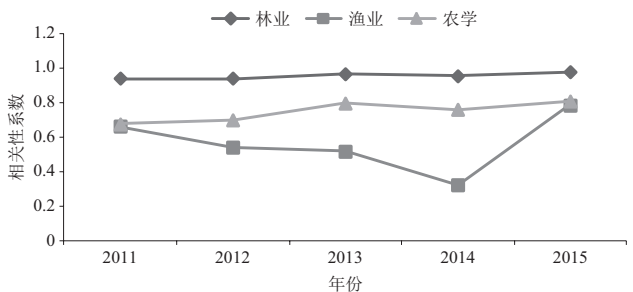
由图1可知：①除林业、化学学科外，大部分学科2011—2015年相关性系数变化幅度较大；并且2011—2015年同一大类中的3门学科相关性系数变化情况不一致。②从整体来看各学科相关性系数变化在2013年或者2014年出现转折，如仪器仪表工业学科在2014年出现下降趋势，医药卫生类中3门学科在2013年出现上

升趋势等。③农业科学类、经济类及哲学、社会学、政治、法律类中的3门学科2011—2015年相关性系数基本呈现波动式上升趋势，工业技术类、文化教育历史类中的3门学科2011—2015年相关性系数基本呈现波动式下降趋势。

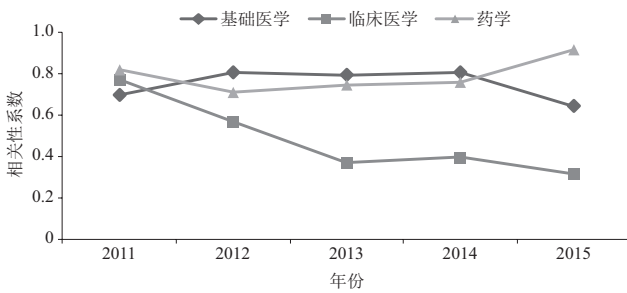
分析原因：①由于论文从下载到引用一般需要2年或者更多的时间<sup>[22]</sup>，期刊论文的下载频次和被引频次在2—3年前等发生波动，进而导致各学科 $S_C$ 指数与 $S_M$ 指数相关性系数发展趋势在2013年或者2014年发生变化。②虽然同一大类中学科存在相似性，但各个学科期刊的发展又有着其特性并且学者们的下载和引用行



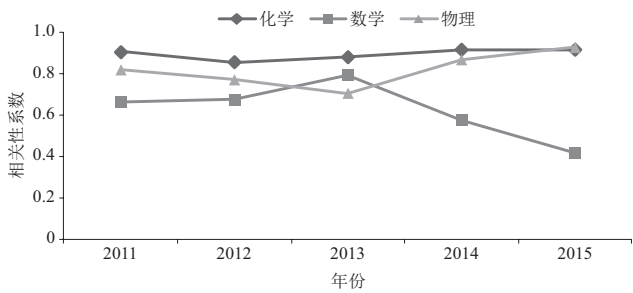
(1) 工业技术类学科 $S_C$ 和 $S_M$ 指数相关性



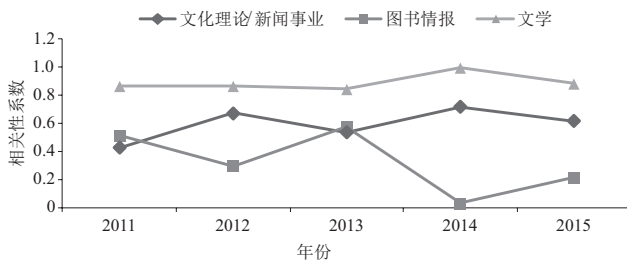
(2) 农业科学类学科 $S_C$ 和 $S_M$ 指数相关性



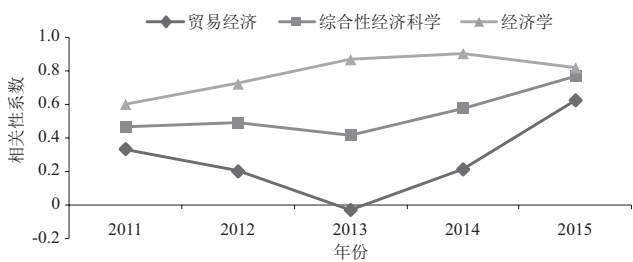
(3) 医药卫生类学科 $S_C$ 和 $S_M$ 指数相关性



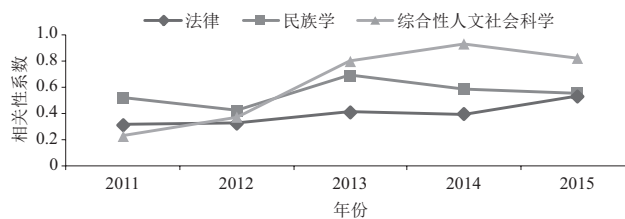
(4) 自然科学类学科 $S_C$ 和 $S_M$ 指数相关性



(5) 文化教育历史类学科 $S_C$ 和 $S_M$ 指数相关性



(6) 经济类学科 $S_C$ 和 $S_M$ 指数相关性



(7) 哲学、社会学、政治、法律类 $S_C$ 和 $S_M$ 指数相关性

图1 2011—2015年各学科 $S_C$ 指数与 $S_M$ 指数相关性系数变化情况

为不确定性高,这可能是同一大类中的3门学科相关性系数变化情况不一致的主要原因。③从各学科期刊论文主题发现,林业、化学学科期刊论文多为基础理论研究,而其他学科实证研究居多,这可能导致这2门学科的 $S_C$ 指数与 $S_M$ 指数的相关性系数变化幅度不大,而其他学科相关性系数变化幅度大。④工业技术类、文化教育历史类近2—3年的期刊论文下载后没有得到有效引用,导致相关性系数下降,而农业科学类、经济类及哲学、社会学、政治、法律类学者从近2—3年文章着手,发现当前研究热点,期刊论文下载并得到了及时引用,导致 $S_C$ 指数与 $S_M$ 指数的相关性系数上升。

由上述分析发现,一方面,虽然同一大类中各学科之间 $S_C$ 指数与 $S_M$ 指数相关性系数相差不大,但各年相关性系数变化情况不一,不能通过同一大类中的学科预测其他学科 $S_C$ 指数与 $S_M$ 指数的相关性系数变化情

况;另一方面,不同学科中 $S_C$ 指数与 $S_M$ 指数相关性系数每年变化情况不同,在期刊评价中应考虑不同时间下载频次和被引频次的关系。

## 2.3 $S_C$ 指数与 $S_M$ 指数相关性系数变异程度分析

变异程度分析是用来观测总体数据的波动程度的,一般通过变异系数来表示,变异系数为标准差与均值的比,变异系数越大,说明这组数据中的平均数的代表性差。本文进一步分析不同学科2011—2015年的“ $S_C$ 指数与 $S_M$ 指数相关性系数”变异程度,旨在从客观上探析2011—2015年 $S_C$ 指数与 $S_M$ 指数相关性的波动概况,结果见表3。

表3 2011—2015年各学科 $S_C$ 指数与 $S_M$ 指数相关性系数变异程度

学 科	变异系数	学 科	变异系数	学 科	变异系数
电子通信技术	17.82%	临床医学	38.40%	文学	6.82%
环境科学	11.07%	药学	10.22%	贸易经济	89.06%
机器仪表工业	5.93%	化学	2.91%	综合性经济科学	25.44%
林业	1.69%	数学	22.28%	经济学	15.67%
渔业	30.60%	物理	10.44%	法律	22.26%
农学	7.87%	文化理论/新闻事业	19.28%	民族学	17.70%
基础医学	9.95%	图书情报	68.54%	综合性人文社会科学	49.37%

可以看出环境科学、机器仪表工业、林业、农学、基础医学、药学、化学、物理、文化理论/新闻事业、文学、经济学和民族学学科外,其他学科的 $S_C$ 指数与 $S_M$ 指数相关性系数变异程度较大,由此说明,2011—2015年不同学科中下载频次和被引频次相关性系数处于波动中,不能使用统一的标准(如仅使用单一指标下载频次)进行期刊评价。

结合图1和表3的分析发现:一方面,在期刊评价中需要将期刊论文发表时间列入评价指标,不同时间下载频次和被引频次应赋予不一样的权重值;另一方面,环境科学、机器仪表工业、林业、农学、基础医学、药学、化学、物理、文化理论/新闻事业、文学、经济学和民族学学科中已知某一年的下载频次和被引频次的相关性系数,一般情况下可以预测未来4年内下载频次和被引频次的相关性。

## 3 结论

本文以理、工、农、医、人文、经济、社会科学大类中3门学科在2011—2015年核心期刊论文的下載频次和被引频次为基础,构建期刊的复合被引H指数( $S_C$ 指数)和复合下载H指数( $S_M$ 指数)指标,探讨2011—2015年不同学科中的 $S_C$ 指数与 $S_M$ 指数相关性系数变化情况及其原因,得出如下结论。

(1)除贸易经济学科2013年外,期刊的下载频次与被引频次一般具有正相关性。说明下载频次指标在某种程度上与被引频次指标所反映出的信息一般具有一致性。

(2)2011—2015年理、工、农、医类期刊的下载频次和被引频次的相关性高于经济、人文、社会科学类。说明2011—2015年理、工、农、医类期刊的利用效率高于经济、人文、社会科学类期刊。导致这一情况的原因

可能是学科之间的差异及不同样本期刊的研究主题,如理论基础类的期刊论文下载频次和被引频次均高,而实证研究类的期刊论文下载频次偏高、被引频次偏低。

(3) 除文学及临床医学学科外,同一大类中的不同学科间的期刊下载频次和被引频次的相关性系数相似,但2011—2015年同一大类中的不同学科间的期刊下载频次和被引频次的相关性变化情况不一。说明除文学及临床医学学科外,理、工、农、经济、社会科学同一大类中的学科之间的复合H指数相关性可以互相推测,但相关性变化情况不能互相推测,导致这一情况可能的原因是同一大类中的学科间的交叉程度高,存在一定的相似性。

(4) 2011—2015年不同学科间的下载频次和被引频次的相关性系数波动较大。说明在期刊评价中需要将期刊论文发表时间列入评价指标,不同时间下载频次和被引频次所赋权重不同。如农业科学类的评价近4—6年的期刊时下载频次所赋权重偏大,工业技术类学科近2—3年的期刊时下载频次所赋权重偏大,而经济类及哲学、社会学、政治、法律类学科进行期刊评价时下载频次和被引频次需要相互参照,所赋权重值相差不大。

(5) 2011—2015年环境科学、仪器仪表工业、林业、农学、基础医学、药学、化学、物理、文化理论/新闻事业、文学、经济学和民族学学科在下载频次和被引频次的相关性系数变异程度小。说明上述学科期刊在已知某一年下载频次和被引频次的相关性,在一定程度上可以预测未来4年内的期刊下载频次和被引频次的相关性概况。

最后需要特别指出的是,在多学科比较中,法律学科复合H指数之间的相关性偏低,数据源仅是导致相关性系数偏低的一种原因,下一步将深入研究法律学科的学者引用行为、研究主题等对复合H指数的影响。

## 参考文献

- [1] GARFIELD E, SHER I H. New factors in the evaluation of scientific literature through citation indexing [J]. *Journal of the Association for Information Science & Technology*, 2010, 14 (3): 195-201.
- [2] PINSKI G, NARIN F. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics [J]. *Information Processing &*

*Management*, 1976, 12 (5): 297-312.

- [3] 丁佐奇, 郑晓南. 期刊影响因子、论文被引证次数与学术质量评价的矛盾分析 [J]. *中国科技期刊研究*, 2009, 20 (2): 286-288.
- [4] O'LEARY D E. The relationship between citations and number of downloads in Decision Support Systems [J]. *Decision Support Systems*, 2008, 45 (4): 972-980.
- [5] 许新军. 基于下载量的期刊半衰期实证研究 [J]. *情报杂志*, 2014, 33 (6): 117-121.
- [6] 许新军. h<sub>d</sub>指数评价期刊网络传播力的探讨 [J]. *中国科技期刊研究*, 2013, 24 (1): 83-87.
- [7] 朱玉奴. 学术期刊的网络传播及影响力研究——以国内23种图书情报类期刊为例 [J]. *情报杂志*, 2012, 31 (11): 71-76.
- [8] 张洋, 郎林芳. Mendeley阅读数用于期刊评价的适用性分析 [J]. *情报杂志*, 2018, 37 (2): 142-146, 103.
- [9] HUANG W, WANG P, WU Q. A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals [J]. *Plos One*, 2018, 13 (4): e0194962.
- [10] GRUNDY Q, DUNN A G, BOURGEOIS F T, et al. Prevalence of disclosed conflicts of interest in biomedical research and associations with journal impact factors and Altmetric scores [J]. *Jama the Journal of the American Medical Association*, 2018, 319 (4): 408-409.
- [11] 张小强. 期刊下载频次与被引频次及影响因子相关性——以中国知网CSCD与CHSSCD刊物为样本的计量分析 [J]. *情报理论与实践*, 2011, 34 (8): 36-40.
- [12] 陆伟, 钱坤, 唐祥彬. 文献下载频次与被引频次的相关性研究——以图书情报领域为例 [J]. *情报科学*, 2016, 34 (1): 3-8.
- [13] 杜秀杰, 赵大良, 葛赵青, 等. 学术论文的下载频率与被引频率的相关性分析 [J]. *编辑学报*, 2009, 21 (6): 551-553.
- [14] 丁佐奇, 郑晓南, 吴晓明. 科技论文被引频次与下载频次的相关性分析 [J]. *中国科技期刊研究*, 2010, 21 (4): 467-470.
- [15] GUERRERO-BOTE V P, MOYA-ANEGÓN F. Relationship between downloads and citations at journal and paper levels, and the influence of language [J]. *Scientometrics*, 2014, 101 (2): 1043-1065.
- [16] VAUGHAN L, TANG J, YANG R. Investigating disciplinary differences in the relationships between citations and downloads [J]. *Scientometrics*, 2017, 111 (3): 1533-1545.
- [17] HAUKE J, KOSSOWSKI T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data [J]. *Quaestiones Geographicae*, 2011, 30 (2): 87-93.
- [18] 刘颖. 基于衍生h指数期刊评价指标构建与实证研究 [J]. *图书*

- 情报工作, 2011, 55 (14): 144-148.
- [19] HIRSCH J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102 (46): 16569-16572.
- [20] MACROBERTS M H, MACROBERTS B R. Problems of citation analysis: A study of uncited and seldom-cited influences [J]. Journal of the American Society for Information Science and Technology, 2010, 61 (1): 1-12.
- [21] EGGHE L. Theory and practice of the G-Index [J]. Scientometrics, 2006 (1): 131-152.
- [22] 赵一权, 王振民, 熊文炳, 等. 科学论文的下載与引用关系研究: 以ACM数字图书馆为例 [J]. 中国科技期刊研究, 2014, 25 (6): 818-823.

## 作者简介

朱雯, 女, 1994年生, 硕士研究生, 研究方向: 计量评价, E-mail: zhuwen5678@163.com。

陈荣, 女, 1969年生, 硕士, 副研究馆员, 硕士生导师, 研究方向: 信息计量、专利分析。

刘颖, 女, 1978年生, 博士, 副研究馆员, 研究方向: 文献计量。

### Relationship between Citations and the Number of Downloads of Journals: Based on Compound H-Index

ZHU Wen CHEN Rong LIU Ying

(Institute of Science and Technology Information of East China University of Science and Technology, Shanghai 200237, China)

**Abstract:** The research is using the CNKI and is based on the citations and the number of downloads of the 2011—2015 core journals of the three disciplines of the science, engineering, agriculture, humanities, economics and social sciences. Then the research builds the journal's download and cited H-index, compound H-index and other indicators. The research is aim to analyse the relationship and trends between compound H-index of different journals in different disciplines from 2011 to 2015. It is found that the correlation between the numbers of downloads and citations of science, technology, agricultural and medicine journals in 2011—2015 is higher than that of economic humanities and social science journals. In addition to literature and clinical medicine, the correlation of compound H-index between disciplines in the same category of science, industry, agriculture, economics and social science can be inferred from each other, but the correlation changes in different years are inconsistent. When evaluating journals, the frequency of downloads and citations of different subjects are different at different times.

**Keywords:** Citations; Number of Downloads; Compound H-Index; Journal Evaluation

(收稿日期: 2018-09-28)