

学术论文的下载频率与被引频率的相关性分析

杜秀杰¹⁾ 赵大良²⁾ 葛赵青¹⁾ 苗凌¹⁾

(1)《西安交通大学学报》编辑部;2)西安交通大学期刊中心:710049,西安)

摘要 为深入研究网络传播对提高学术期刊影响力的作用,从微观层面入手,对一篇学术论文网上下载频率和被引频率的相关性进行研究,给出了相关性的量化计算指标和计算方法。以《西安交通大学学报》的数据为例,对其1994年第1期、2006年第1~3期、2007年第1期发表文章的下载频率和被引频率进行统计,计算出两者的简单线性相关系数,并进行了分析。研究结果表明,网络传播是提升期刊影响力的必要条件和手段,吸引和发表优秀稿件才是提升期刊影响力的充分条件。

关键词 网络传播;下载频率;被引频率;影响因子

Relativity analysis of downloads and citations for scientific papers//DU Xiujie, ZHAO Daliang, GE Zhaoqing, MIAO Ling

Abstract To study the relation between the internet communication and the scientific journal impact, the relativity between the download and citation frequencies of scientific papers is studied. The quantitative index of relativity and its calculated method are presented. The downloads and citations of some papers published in 1994, 2006, 2007 in the *Journal of Xi'an Jiaotong University* are investigated and analyzed, the relativity coefficients are calculated. The results show that the downloads of internet is a necessary condition for the impact of journals; however the quality of published papers is the substantial condition.

Key words internet communication; download frequency; citation frequency; impact factor

First-author's address Editorial Office, Journal of Xi'an Jiaotong University, 710049, Xi'an, China

进入网络时代,信息的网络传播无疑是最重要的方式与途径。网络信息传播的最终目标是为了实现全社会信息共享,满足各个层次、各类公众对信息的需求^[1-2]。学术界也毫不例外。由于网络的发展,科学工作者改变了以往获取科技信息的方式,网络成为其获取新知识、新技能及学术界最新研究成果的主要手段。学术的传播与知识的获取一样离不开网络,科技工作者通过网络获取知识的同时,也渴望自己的研究成果通过网络广泛传播,被学术界关注;但是,在学术界,科技论文的被引频率不高一直是困扰大家的问题^[3-4]。

目前,文章的被引频率和期刊的影响因子已经超出其本质意义,衍生出很多的表征意义,这2个指标与文章或期刊的影响力、文章质量直接关联^[5]。将学术论文的被引频率作为衡量一篇文章学术水平高低、将影响因子作为衡量一种期刊水平高低的观点在学术界并没有取得共识,这中间还存在着许多局限性^[6]。不

可否认,学科研究群体规模的大小,所研究内容受众面的多少,导致这些指标有数十倍的差距,简单地将其放在一起比较是没有什么意义的;但是,对某一专业领域、对某一类别的学术期刊或者对某一种刊物自己纵向相比,被引频率和影响因子还是能反映出其影响力的大小和学术质量的高低的,所以,文章的高引用率一直以来都是期刊界以及作者自己所希望看到的,以至于是大家想尽办法追求的目标之一。

文章的网络浏览量和被引频率是衡量一篇学术论文水平的重要指标。本文试图从微观层面入手,对一篇学术论文的下载频率和被引频率的相关性进行研究,从而揭示出网络传播对文章引用率的影响情况,为扩大学术期刊影响力找到一个正确的方向。

1 网络传播对提高引文频率的作用

鉴于文章被引频率受到期刊界的高度重视,所以,想方设法提高文章的被引率是每个编辑部工作的重中之重。在网络高度发达的今天,网络传播自然是扩大期刊和文章影响力的最主要的手段之一^[7]。从最初的使用电子函件为读者提供孤立的文章信息、使用电子函件列表将文章目录上网,到后来发展起来的网上开放存取,以及目前实施的稿件在线处理平台,这每一步都使我们更加贴近读者,对文章的网络传播都具有重要意义。由IT产业发展起来的各种搜索工具,如百度、google等著名的搜索引擎,使读者想要找到自己所需要的信息变得容易。国内外著名的数据库,如Ei、SCI、ISTP,被认为是最具权威的检索工具,Ei和SCI提供全世界范围内6000多种优秀期刊上的文章检索,ISTP则报道国际上定期召开的1000多个科技会议录,这些几乎涵盖了各个学科、各个专业公认的高水平的科技文献。国内外一些知名网站也为科技工作者提供了丰富多彩的科技信息,如中国期刊网是我国最大的全文现刊数据库,是目前世界上最大的连续动态更新的中国期刊全文数据库,收录1994年至今约7000多种期刊全文,并对其中部分重要刊物回溯至创刊时起。

所有这些为科技工作者打开了一个前所未有的科技的网络世界,改变了学者们传统的获取信息的模式,大大促进了科技信息的网络传播。有人曾经说过,网

络传播的出现极大地改变了人们的信息交流与传播的方式,使人类生活的地球变成了一个小村庄^[8]。拥有这些便利的同时,也带来了一个问题:便捷的网络使所有科技信息几乎同时站到了人们面前,使所有学术论文被引机会又站到了同一起跑线上。在这一有趣的矛盾的情况面前,我们有必要在微观层面上探寻一篇文章的网络传播与被引频率的相关性问题。在该文中文章的网络传播主要用文章的下载频次表征。

2 文章下载频率与被引频率的相关性

通常人们都会认为学术论文的下载频率与被引频率正相关,即文章下载越多,被引的机会也就越多。事实完全如此吗?当人们查找资料时,最先找到的是文章题名,一篇学术论文的题名所能带给人们的信息价值大小是引人注意的关键因素。下载一篇文章后是否引用它,那就要看此文的内在因素,即文章的内在质量。当然,下载频率与被引频率两者之间拥有某些相关性,首先文章研究的内容是否为学科里的热点和难点,这些是相关学科里前沿问题,代表学科最新的研究方向和最高的研究水平。其次是研究领域的大小,一般而言,在学术价值相当的情况下,论题越大,所涉及

的领域越宽,受众也就越多,文章被下载、被引用的也就越多^[9-11]。

文章的下载频率和被引频率是文献计量学评价体系中2个最基本的指标,从统计学角度看它们可构成一个二元体系。人们在研究2个随机变量之间的相互关系或依赖关系时,经常运用数理统计学中的简单线性相关系数来分析,具体公式为

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}。$$

式中: r 为简单线性相关系数; n 为取样数量; x_i 为被引用频率; y_i 为下载频率。

现以《西安交通大学学报》发表文章的数据为例,对文章的下载频率与被引频率的相关性作进一步分析。取1994年第1期(这是能收集到的最早的数据)、2006年第1~3期、2007年第1期,其中发表篇数 n 、每一篇文章的网络下载次数 x_i 和被引用次数 y_i 均列于表1。表中数据来源为中国学术文献网络出版总库中的期刊库2009年2月10日的数据,该数据库是动态的,数据每日更新。

表1 论文网络下载次数和被引用次数的统计与分析

序号	1994年1期		2006年1期		2006年2期		2006年3期		2007年1期	
	x_i	y_i								
1	3	64	9	156	6	159	5	49	0	54
2	4	45	4	60	1	74	3	92	1	84
3	3	29	2	87	0	56	3	69	0	160
4	8	54	2	141	0	14	1	139	1	98
5	2	25	4	230	3	70	0	71	1	84
6	0	9	2	113	0	75	0	84	0	59
7	0	25	4	123	0	109	1	63	0	120
8	0	21	6	89	0	36	1	85	0	174
9	0	20	3	166	0	50	9	196	1	42
10	2	67	1	60	0	48	3	78	0	52
11	4	59	1	129	0	47	4	111	0	52
12	0	14	3	238	1	80	3	73	1	188
13	1	26	38	319	0	94	0	45	1	103
14	43	230	2	108	4	542	1	53	5	102
15	1	12	2	103	0	70	2	88	0	27
16	1	68	2	85	0	90	0	50	2	154
17	0	7	0	124	1	58	2	154	0	208
18	10	72	1	117	0	87	0	114	0	85
19			1	45	3	135	0	40	0	155
20			1	45	0	115	2	105	0	109
21			5	229	3	73	2	61	0	100
22			3	93	0	45	2	69	0	52
23			0	72	2	64	1	73	0	101
24			0	114	3	160	3	120	2	154
25			1	52	8	166	3	174	0	12
26			1	77	1	45	4	81	1	236
27			6	244	1	52	3	36	1	329
28			2	147	1	172	1	204	0	598
29					6	299	1	58	0	153
30					5	71			0	152
31					2	108				
合计	79	783	106	3 566	51	3 264	60	2 635	17	3 997

根据相关性公式,以1994年第1期为例进行计算。该期共发表文章18篇, $n=18$,利用Excel对这些原始数据进行处理,得到

$$\sum_{i=1}^{18} x_i = 79; \sum_{i=1}^{18} y_i = 783; \sum_{i=1}^{18} x_i^2 = 2\ 065;$$

$$\sum_{i=1}^{18} y_i^2 = 79\ 697; \sum_{i=1}^{18} x_i y_i = 11\ 835。$$

根据线性相关系数公式,计算出1994年第1期 r 为0.9476。

采用同样的方法计算出2006年第1~3期相关性系数分别为0.6445、0.5329、0.398,而2007年第1期的相关性系数仅为0.0075,与前面数据差距相当大。由这一组数据看出,文章发表时间长短是影响下载频率和被引频率的关键因素,而且时间因素对两者相关性的影响巨大。2006年1—3月,每过1个月相关性系数下降10%左右,2007年1月两者的相关性系数竟仅为0.0075,下降幅度非常大。

由表1可以看出,在20世纪90年代,网络发展还不很充分,1994年文章的下载次数多为几十次,引用率也不是很高,只有第14篇《平直翅片管换热器传热与阻力特性的实验研究》一文下载230次,引用43次。到2006年以后,文章发表仅1~3a,网上下载次数多为上百次,仅从这一点就可以看出网络传播的力量。但很遗憾的是,引文频率不是简单地与下载频率成正比,引文数量高的还是个别篇,甚至有的文章网上下载次数百次,竟没有一篇引用。其原因应该是在文章的内在质量上,即学术论文的科学意义与价值,研究方法是否新颖、是否具有创造性,研究结果是否对别人的工作具有启发性。表1里2006年第1期中的第13篇《粒子群优化算法的惯性权值递减策略研究》,发表3a多,拥有319次的网上下载和38次的引用次数,2007年第1期中第14篇《一种判断点与多边形关系的快速算法》,发表仅2a,已经有102次的下载和5次引用。这2篇文章的共性均为方法性的研究,尤其是粒子群优化算法显然是近2年研究的热点,大量的文献出现在2007和2008年,《西安交通大学学报》2006年第1期发表这篇文章还是比较早的,所以引用频率比较高。

从以上的分析可以看出:稿件发表时间和稿件学术水平是影响文章被引用的2大核心要素。另外,通过分析发现对《西安交通大学学报》影响因子贡献较大的主要集中在一些重点学科(如能源与动力、机械制造、电气工程、电子与信息工程等)和热点方向(主要有粒子群优化方法、支持向量基方法、混合燃料的特性研究、换热器的传热等)。表1中被引次数 ≥ 5 的均是高引用率的文章,对《西安交通大学学报》影响因子

贡献较大,不言而喻,其作者自然是学报的核心作者。

3 结束语

网络的发展曾经为科技信息的传播带来不可估量的作用,传播的信息量大大增加,信息孤岛减少,传播节奏加快,为提高期刊影响力发挥了很大的作用,尤其是在网络高速发展的初期,这种作用更加明显。在网络发展很充分的今天,这种作用在逐渐降低,但是又不可或缺,借用一个数学名词,网络传播对于提升期刊影响力是一个必要条件,而不是充分条件。目前,要想提高学术论文的被引频率以及期刊的影响因子,仅从网络传播上下工夫是远远不够的,真正起主要作用的还是所发表文章本身的质量和價值。对一种期刊所发表的文章进行深入的引文分析,有助于编辑掌握可以提高期刊影响因子的重要板块、热点问题以及核心作者群,尤其是对于涵盖的学科比较多、所报道的方向比较散乱的大学学报,就显得更加重要。总之,吸引并争取到优秀稿件是提升期刊影响力的必要而充分的条件。

4 参考文献

- [1] 张军华.网络信息传播自由度的影响因素研究[J].高校图书情报论坛,2007,6(2):48-51
- [2] 方锦清.网络科学的理论模型探索及其进展[J].科技导报,2006,24(12):67-72
- [3] 陈广仁,刘元珉.重视科技论文引用率,提高中国科技影响力[J].科技导报,2008,26(5):96-97
- [4] 许书阁,高雪梅,韩晋平,等.影响科技期刊论文被引用率的因素及其提高的途径[C]//第四届中国科技期刊发展论坛文集.北京,2008:818-821
- [5] 张九庆.期刊影响因子的意义及其影响因素和应用原则[J].中国科技期刊研究,2006,17(2):189-191
- [6] 任胜利,王宝庆,严谨,等.应慎重使用期刊的影响因子评价科研成果[J].科学通报,2000,45(2):218-222
- [7] 陈少华.网络环境下学术传播与科技期刊发展研究[C]//第4届中国科技期刊发展论坛文集.北京,2008:449-452
- [8] 董毅.网络传播的特质与取向[J].新媒体在线,2008(3):68-69
- [9] 王楚鸿,陈妙玲.试析影响学术论文他引频次的十大因素[J].科研管理,2007,28(4):44-46
- [10] 戴泳.团队文献高引用率成因探析[J].科技情报开发与经济,2008,18(29):3-5
- [11] 任胜利,李家林,金碧辉,等.我国部分科技期刊参考文献和被引情况统计分析[J].编辑学报,2001,13(5):162-164
(2009-05-22 收稿;2009-07-22 修回)