

# 权威因子：一个新的期刊评价指标\*

苏成 潘云涛 马峥 袁军鹏 郭红 俞征鹿

中国科学技术信息研究所情报方法研究中心, 100038 北京

**摘要** 2009年版的中国科技期刊引证报告(核心版)新提出了一个权威因子指标,权威因子的计算采用的是 PrestigeRank 算法, PrestigeRank 算法是在著名搜索引擎算法 PageRank 算法的基础上修改优化而成的, PrestigeRank 算法专门针对期刊引用网络缺失严重提出了合理的解决方案,在权威因子的计算中,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要被赋予更高的权重。利用 PrestigeRank 算法计算 2008 年的中国科技论文与引文数据库(CSTPCD)的所有期刊的权威因子,并对权威因子与总被引频次进行了比较研究,发现总被引频次在一定程度上是流行性测度,而权威因子在更大程度上是权威性测度。我们也发现权威因子能部分消除因为学科规模和性质不同造成的被引机会的不同的影响。**关键词** PrestigeRank; 引文分析; 期刊评价; PageRank; 权威因子; 影响因子

**Authority factor: a new indicator for journal evaluation**// SU Cheng, PAN Yuntao, MA Zheng, YUAN Junpeng, GUO Hong, YU Zhenglu

**Abstract** The China S&T Journal Citation Report (CORE) promulgated by ISTIC in 2009 used a new journal evaluation indicator, authority factor, the calculation of which is based on the PrestigeRank algorithm, which is rooted in PageRank algorithm. PrestigeRank algorithm has found a solution suitable for cases where there are missing data in the database citing network. We made use of PrestigeRank to give the ranking of all journals in the Chinese Scientific and Technology Papers and Citation Database (CSTPCD) in 2008. We compared PrestigeRank result with citation ranking. We found that PrestigeRank could reflect journal's authority favorably. We also found that authority factor could partially eliminate the problem of the different cited opportunities because of the different size and nature of the subject.

**Key words** PrestigeRank; citation analysis; journal evaluation; PageRank; authority factor; impact factor

**Authors' address** Research Center for Information Science Methodology, Institute of Scientific and Technical Information of China, 100038, Beijing, China

在目前的期刊评价中,比较有影响的指标有总被引频次和影响因子,但这 2 个指标均存在一个缺点:即它们只是单纯考虑被引次数的多少,并不区分不同引用之间的重要性区别<sup>[1-2]</sup>。而搜索引擎算法 PageRank 可以有效解决这个问题,因为在 PageRank 算法中,不

但考虑一个网页被链接的次数,还考虑这些链接的重要性区别,来自重要的权威性高的网页的链接被赋予更高的权重<sup>[3]</sup>。J. Bollen 于 2006 年利用 PageRank 算法对 ISI 的期刊引证数据进行了期刊排序研究<sup>[4]</sup>, SCImago Research Group 利用 PageRank 算法提出了基于 SCOPUS 数据库的 SCImago Journal Rank (SJR), ISI 2008 年公布的 Journal Citation Report 也加入了一个新的期刊评价指标 Eigenfactor, 它的计算基于 PageRank 算法,并排除了期刊自引,苏成分别在 PageRank、HITS 算法的基础上提出新的期刊排序算法,并利用它进行了中国科技期刊的实证研究<sup>[5-7]</sup>。这些研究的一个共同点就是综合考虑引用的数量和质量从而计算期刊分数,他们区分了不同引用之间的重要性差别,比只考虑被引用次数而言无疑更客观合理些。但是我们也发现在前述这些研究中还存在一个普遍的问题:在构建期刊引用网络时并未考虑系统缺失部分。在期刊引用网络中,期刊上刊载论文的引用文献类型是多种类的,既有期刊论文,也有会议论文、书籍、标准、专利、网络信息等,目前还没有一个数据库能涵盖所有的这些文献类型,缺失比较严重,如果在构建期刊引用网络矩阵时不考虑系统缺失部分的内容就可能会造成结果的偏差。

针对以上问题,本研究在 PageRank 算法的基础上提出一种新的基于期刊引用网络的 PrestigeRank 算法,此算法针对目前数据库中期刊引用网络缺失严重的情况提出了解决方案。

## 1 数据和研究方法

**1.1 数据** 在本研究中,我们选取中国科技论文与引文数据库(CSTPCD)2008 年所有收录期刊共 1 868 种,所有这些期刊平均被引次数 804 次,大于 1 000 次共有 447 种(见表 1),这些期刊总的参考文献量为 5 251 613 条,平均每刊的参考文献量 2 728 条<sup>[8]</sup>,在这 5 251 613 条参考文献量中,刊登在这 1 868 种期刊上的有 1 501 139 条,所占百分比为 28.58%,其他不是刊登在这 1 868 种期刊的参考文献有 3 750 474 条,所占百分比 71.42%,统计数据告诉我们在期刊引用网络中,其缺失部分所占比例是十分大的,所以我们在处理期刊引用网络时对其缺失部分必须有所考虑。

\* 国家科技支撑计划课题(2006BAH03B05);国家自然科学基金项目(70973118);中国科学技术信息研究所科研项目预研基金

表1 期刊引用网络的基本情况

总被引频次	期刊数	所占百分比/%
0 ~ 500	929	49.73
501 ~ 1 000	492	26.34
1 001 ~ 2 000	303	16.22
2 001 ~ 3 000	80	4.28
3 001 ~ 4 000	32	1.71
4 001 ~ 5 000	19	1.02
5 001 ~ 6 000	6	0.32
>6 000	7	0.37
合计	1 868	100.00

**1.2 基于期刊引用网络 PrestigeRank 算法** 笔者于2009年提出过一个基于期刊引用网络的修正PageRank算法用于期刊评价<sup>[6]</sup>,该算法的缺点是没有对指向期刊引用网络之外的缺失部分进行特别处理,只是简单地把缺失部分删除。后来笔者提出了一个基于论文引用网络的 PrestigeRank 算法,并证明了在论文引用网络层面要比从期刊引用网络层面评价期刊更为合理<sup>[9]</sup>,但是因为国内目前数据库收录年限均比较短,CSTPCD 收录了1989年以来中国1 000多种科技期刊就算是时限比较长的了,所以在目前数据库收录年限较短的情况下要构建所有这些论文的引用矩阵基本上是不可能,另外因为参考文献著录不规范或者错误,会造成好多参考文献与原始论文无法链接,也会造成我们构建论文引用矩阵的误差。而在期刊引用层面,在2008年涉及的期刊数量只有1 868种,在经过我们人工的仔细核对与规范后,可以保证构建的期刊引证矩阵的准确性,其计算量也比论文引用矩阵要小很多,因此我们用基于期刊引用网络的 PrestigeRank 算法来计算期刊的权威因子。

图1是一个期刊引用网络的示意,每个节点代表一个期刊,节点间带箭头的线条代表期刊间的引用关系,箭头指向被引用期刊,指向自身的线条表示期刊的自引,线条的粗细表示期刊间的引用次数的多少,对于缺失的数据,我们用一个“虚拟点”表示,这个“虚拟点”代表所有未在系统中包含的参考文献,“虚拟点”与整个期刊引用网络相互引用,“虚拟点”的权重被期刊引用网络中的所有期刊平分。

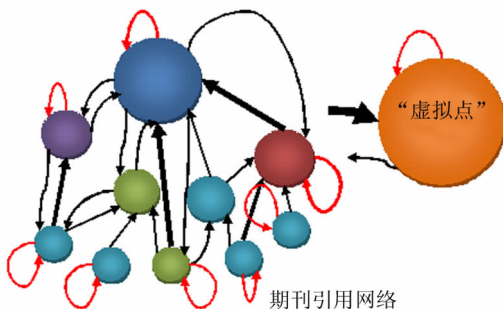


图1 期刊引用网络缺失部分的处理

对于期刊引用网络来说,其矩阵与论文引用网络矩阵存在差别,在期刊引用网络中,期刊之间互相引用很常见,但论文引用网络肯定是单向的,只能是晚发表的论文引用早发表的论文;并且期刊引用网络中存在自引,而论文引用网络中是不存在自引的;另外2种期刊间的引用往往是很多次,而2篇论文间的引用只是1次。因此作者在文献[11]中提出的基于论文引用网络的 PrestigeRank 算法不能直接用于期刊引用网络,必须针对期刊引用网络矩阵作相应的改变。

对于一个包含了  $n$  个期刊的期刊引用网络,在引入“虚拟点”之后,其期刊引用矩阵  $M$  为

$$\begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1j} & \cdots & M_{1n} & M_{1(n+1)} \\ M_{21} & M_{22} & \cdots & M_{2j} & \cdots & M_{2n} & M_{2(n+1)} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ M_{i1} & M_{i2} & \cdots & M_{ij} & \cdots & M_{in} & M_{i(n+1)} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ M_{n1} & M_{n2} & \cdots & M_{nj} & \cdots & M_{nn} & M_{n(n+1)} \\ M_{(n+1)1} & M_{(n+1)2} & \cdots & M_{(n+1)j} & \cdots & M_{(n+1)n} & M_{(n+1)(n+1)} \end{bmatrix} \quad (1)$$

其中:

$$M_{ji} = \begin{cases} t/r, & \text{如果期刊 } i \text{ 引用期刊 } j \text{ 共 } t \text{ 次,期刊 } i \text{ 的} \\ & \text{参考文献量为 } r, \quad (i \leq n, j \leq n); \\ 0, & \text{如果期刊 } i \text{ 没有引用期刊 } j \end{cases}$$

$$M_{i(n+1)} = n - \sum_{j=1}^n M_{ij} \quad (i \leq n);$$

$$M_{(n+1)j} = \frac{1}{n+1} \quad (j \leq n+1).$$

这样,笔者用引入了“虚拟点”的期刊引用网络矩阵  $M$  代替文献[9]中的论文引用网络,不但解决了期刊引用网络与论文引用网络的差异问题,也解决了期刊引用网络的缺失问题,那么基于期刊引用网络的 PrestigeRank 算法的公式<sup>[10]</sup>为

$$\pi^{(k+1)T} = \pi^{(k)T} (\alpha M + (\alpha a + (1-\alpha)e) \frac{1}{n e^T}) \quad (2)$$

在原始的 PageRank 算法中,  $\alpha$  的值一般定为 0.85,然而, P. Chen 等的研究表明:在论文引用网络中  $\alpha$  的值为 0.5 更为合理<sup>[11]</sup>,所以在式(2)中我们把  $\alpha$  的值为 0.5。

Bollen 提出,单纯考虑数量更多地是表示一个期刊的流行性,而利用 PageRank 算法进行的期刊排序更多地是表明一个期刊的权威性<sup>[1]</sup>,因为 PrestigeRank 算法源自 PageRank 算法,不但考虑了被引次数的数量,还区分了不同引用的重要性,在一定程度上我们可以认为它能够更好地反映期刊的权威性,所以我们可以把 PrestigeRank 算法得出的值称为“权威因子”。

**1.3 引入“虚拟点”对排序结果的影响** 假设有一个系统含有如图2所示的一个期刊引用网络,共有9种期刊,期刊间的箭头指向表示期刊间的引用关系,指向自身的表示自引,线条粗细表示引用次数的多少,比如期刊 $j_1$ 有指向期刊 $j_5$ 的箭头,表示期刊 $j_1$ 引用了期刊 $j_5$ 。期刊 $j_2$ 、期刊 $j_5$ 、期刊 $j_6$ 、期刊 $j_7$ 、期刊 $j_9$ 均有未包含在系统中的参考文献,按照 PrestigeRank 算法,我们可以引入期刊 $j_{10}$ 这个虚拟点代表未包含在系统中的参考文献。

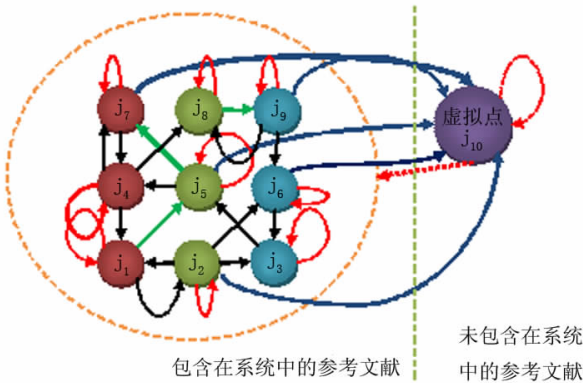


图2 期刊引用网络示例

图2所示引入了“虚拟点”的期刊引用网络如表2所示:表中所示值为这行的期刊引用这列期刊的次数,比如第1行表示 $j_1$ 自引2次, $j_1$ 引用 $j_2$ 、 $j_5$ 分别为1、2次。那么表格每行求和就得期刊的参考文献数,表格每列求和就得期刊的被引次数。

表2 示例期刊引用网络的引用与被引情况

	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$	$j_6$	$j_7$	$j_8$	$j_9$	$j_{10}$	参考文献量
$j_1$	2	1			2						5
$j_2$	1	2	1			1				3	8
$j_3$			1		1						2
$j_4$	1			3			1	1			6
$j_5$				1	2		2			3	8
$j_6$		1				1				1	3
$j_7$				1			2			4	7
$j_8$								1	2		3
$j_9$						1		1	1	2	5
$j_{10}$	1	1	1	1	1	1	1	1	1	1	10
被引次数	5	4	4	6	6	4	6	4	4	14	57

如果不引入虚拟点10,根据表2与PageRank算法权重分配的方法,图2所示期刊引用网络的矩阵为

$$P = \begin{bmatrix} 2/5 & 1/5 & 0 & 0 & 2/5 & 0 & 0 & 0 & 0 & 0 \\ 1/5 & 2/5 & 1/5 & 0 & 0 & 1/5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/6 & 0 & 0 & 3/6 & 0 & 0 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 2/5 & 0 & 2/5 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 & 0 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \end{bmatrix} \quad (3)$$

如果引入虚拟点论文10,根据表2与公式(1),则期刊引用网络矩阵为

$$\begin{bmatrix} 2/5 & 1/5 & 0 & 0 & 2/5 & 0 & 0 & 0 & 0 & 0 \\ 1/8 & 2/8 & 1/8 & 0 & 0 & 1/8 & 0 & 0 & 0 & 3/8 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/6 & 0 & 0 & 3/6 & 0 & 0 & 1/6 & 1/6 & 0 & 0 \\ 0 & 0 & 0 & 1/8 & 2/8 & 0 & 2/8 & 0 & 0 & 3/8 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/7 & 0 & 0 & 2/7 & 0 & 0 & 4/7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/5 & 0 & 1/5 & 1/5 & 2/5 \\ 1/10 & 1/10 & 1/10 & 1/10 & 1/10 & 1/10 & 1/10 & 1/10 & 1/10 & 1/10 \end{bmatrix} \quad (4)$$

式(3)和式(4)所示矩阵的行与列依次代表期刊 $j_1$ 至期刊 $j_9$ (或 $j_{10}$ ),其中的值代表这行的期刊引用这列的期刊的次数除以这行的期刊参考文献数。比如式(4)中矩阵第2行 $[1/8 \ 2/8 \ 1/8 \ 0 \ 0 \ 1/8 \ 0 \ 0 \ 0 \ 3/8]$ 表示 $j_2$ 引用了 $j_1$ 、 $j_2$ 、 $j_3$ 、 $j_6$ 及虚拟期刊 $j_{10}$ 的次数分别为1、2、1、1、3,其权重被这5个期刊分配,各自占得权重比例为 $1/8$ 、 $2/8$ 、 $1/8$ 、 $1/8$ 和 $3/8$ 。

利用式(2)以及式(3)、式(4), $\alpha$ 取0.5,可以分别求出不包含虚拟点以及包含虚拟点的每个期刊权威因子值(见表3)。

表3 被引次数、无虚拟点和有虚拟点 PrestigeRank 算法的期刊排序结果

期刊	有虚拟点	排序	无虚拟点	排序	被引次数	排序
$j_1$	0.088 0	7	0.092 3	8	4	4
$j_2$	0.076 0	9	0.081 0	9	3	5
$j_3$	0.102 5	2	0.121 1	4	3	5
$j_4$	0.095 5	4	0.122 1	3	5	1
$j_5$	0.115 4	1	0.130 3	2	5	1
$j_6$	0.086 7	8	0.108 6	5	3	5
$j_7$	0.093 4	5	0.137 7	1	5	1
$j_8$	0.090 5	6	0.100 2	7	3	5
$j_9$	0.097 7	3	0.106 8	6	3	5
$j_{10}$	0.154 3					

从表3可以看出,在被引次数排序中,其只是简单计数,不能区分不同引用重要性差别,比如 $j_3$ 和 $j_6$ ,被引次数均为3次,其排名均为第5。而在PrestigeRank算法中, $j_3$ 排名第2, $j_6$ 排名第8。下面我们具体分析 $j_3$ 和 $j_6$ 的被引情况: $j_3$ 分别被 $j_2$ 、 $j_3$ 、 $j_6$ 引用1次, $j_6$ 分别被 $j_2$ 、 $j_6$ 、 $j_9$ 引用1次, $j_3$ 和 $j_6$ 的被引有2次是来自同样的刊 $j_2$ 、 $j_6$ ,差别就在于 $j_3$ 被 $j_3$ 自引1次,而 $j_6$ 被 $j_9$ 引用1次,从表3可以看出, $j_3$ 和 $j_9$ 均被引3次,但 $j_3$ 只有2个对外引用,而 $j_9$ 具有5个对外引用,所以 $j_3$ 得到来自 $j_3$ 的 $1/2$ 权重,而 $j_6$ 只得到 $j_9$ 的 $1/5$ 权重,所以 $j_3$ 的排名比 $j_6$ 的排名靠前。从这个例子可以看出,PrestigeRank算法可以区分不同引用间重要性的差异。

结果显示,无虚拟点与有虚拟点的结果存在差异,特别是那些具有较大比例参考文献不在系统的期刊引用的期刊,比如说  $j_4$  和  $j_7$  分别被  $j_4$ 、 $j_5$  和  $j_7$  引用,据表 2 以及式(3)、式(4)可以得知: $j_4$  没有不在系统的引用,加入虚拟点与否其权重的分配没有变化; $j_5$  共有 8 个对外引用,其中 3 个指向了系统外  $j_{10}$ ,如果不加入虚拟点, $j_4$  和  $j_7$  可以分别获得来自  $j_5$  权重的  $1/5$  和  $2/5$ ,加入虚拟点后  $j_4$  和  $j_7$  可以获得它们应该得到的来自  $j_5$  权重的  $1/8$  和  $2/8$ ; $j_7$  共有 7 个对外引用,其中 4 个指向系统外  $j_{10}$ ,如果不加入虚拟点, $j_4$  和  $j_7$  可以获得来自  $j_7$  权重的  $1/3$  和  $2/3$ ,而加入虚拟点后  $j_4$  和  $j_7$  可以获得它们应该得到的来自  $j_7$  的  $1/7$  和  $2/7$ 。所以不引入虚拟点时  $j_4$  和  $j_7$  其权重被不恰当地增加了,表现在排序结果上就是不加入虚拟点  $j_4$  和  $j_7$  的排名要比加入虚拟点的排名靠前,从表 3 得知在不加入虚拟点的排序中, $j_4$ 、 $j_7$  分别排第 3、第 1 位,而在加入虚拟点的排序中, $j_4$ 、 $j_7$  分别排第 4、第 5 位,以上分析表明虚拟点的加入能更加客观反映期刊重要性排序。

## 2 结果

**2.1 权威因子与总被引频次** 从表 4、图 3 可以看出,期刊权威因子与期刊的总被引频次正相关,Spearman 相关系数为 0.930,两者具有强相关关系。权威因子与总被引频次的相关系数要比权威因子与影响因子的相关系数要高。因此可以认为权威因子更多的是一个总量的测度,它提供的结果和总被引频次具有很强的相关性。

表 4 期刊权威因子、总被引频次、影响因子间的 Spearman 相关系数

比较项	Spearman 相关系数
权威因子 vs 总被引频次	0.930*
权威因子 vs 影响因子	0.696*
总被引频次 vs 影响因子	0.706*

\* Correlation is significant at the 0.01 level (2-tailed)

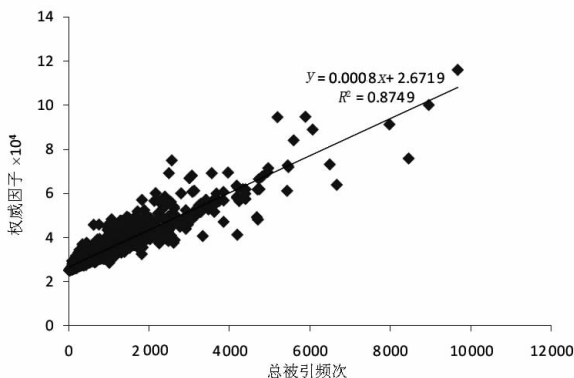


图 3 期刊权威因子与期刊的总被引频次散点分布

但是,权威因子的排序结果与总被引频次排序结果之间也存在差异,从表 5 中可以看出,总的来说权威因子前 10 位的期刊其总被引频次也较高,但有一个例外就是《石油勘探与开发》,其总被引频次只有 2 561,按总被引频次排序为第 91 名,而按权威因子排序排名为第 9。而权威因子排名第 10 的《中国组织工程研究与临床康复》其总被引频次达到了 6 494,远远高于《石油勘探与开发》的 2 561。那么,为什么会存在这个差异呢?表 6 和表 7 分别列出了引用《石油勘探与开发》与《中国组织工程研究与临床康复》的期刊的情况:包括这些期刊各自引用《石油勘探与开发》或《中国组织工程研究与临床康复》的次数,这些期刊的参考文献总量,以及这些期刊的权威因子值。尽管引用《中国组织工程研究与临床康复》的期刊数量和总的引用次数均大大多于引用《石油勘探与开发》的相应数量,但是根据 PrestigeRank 算法我们可以得知,期刊的权威因子值由以下因素决定:期刊的被引次数多少;引用这个期刊的其他期刊的重要性高低;引用它的期刊的权值被平分的状况(即引用它的期刊的参考文献量的多寡)。比如在表 6 中,A 表示一些期刊引用《石油勘探与开发》的次数,B 表示这些期刊的参考文献总量,那么  $A/B$  表示引用《石油勘探与开发》的这些期刊的权值被《石油勘探与开发》分配的比例,C 表示这些期刊的权威因子,它可以反映这些期刊的重要性高低。把所有引用《石油勘探与开发》期刊的  $A'C/B$  累加就大致可以反映《石油勘探与开发》得到的权值。表 7 与表 6 类似。从这 2 表可以发现:所有引用《石油勘探与开发》期刊的  $A'C/B$  累加为 0.000 588 356,大于《中国组织工程研究与临床康复》的 0.000 566 869;所以,尽管《石油勘探与开发》的总被引频次远低于《中国组织工程研究与临床康复》的总被引频次,但其权威因子却比《中国组织工程研究与临床康复》要高。

表 5 权威因子前 10 位的期刊

刊名	权威因子	总被引频次	影响因子
中国电机工程学报	0.001 160 50	9 678	1.557
生态学报	0.001 001 40	8 956	1.669
中华医院感染学杂志	0.000 948 26	5 887	1.402
岩石力学与工程学报	0.000 946 27	5 192	1.448
应用生态学报	0.000 913 38	7 979	1.632
电力系统自动化	0.000 890 08	6 066	1.540
科学通报	0.000 841 39	5 595	0.866
物理学报	0.000 759 27	8 457	1.426
石油勘探与开发	0.000 750 52	2 561	1.904
中国组织工程研究与临床康复	0.000 732 26	6 494	0.843

表6 引用《石油勘探与开发》的期刊分布情况

刊名	引用次数 A	参考文献量 B	权威因子 C	AC/B
石油勘探与开发	600	2 169	0.000 750 52	0.000 207 613
油气地质与采收率	122	2 467	0.000 424 10	2.097 29E-05
大庆石油地质与开发	114	2 363	0.000 448 18	2.162 19E-05
西南石油学院学报	107	3 336	0.000 354 09	1.135 72E-05
∴	∴	∴	∴	∴
总计	2 561			0.000 588 356

表7 引用《中国组织工程研究与临床康复》的期刊分布情况

刊名	引用次数 A	参考文献量 B	权威因子 C	AC/B
中国组织工程研究与临床康复	1 082	49 587	0.000 732 26	1.597 81E-05
中国康复医学杂志	135	5 899	0.000 369 94	8.466 15E-06
中国康复理论与实践	123	6 230	0.000 347 82	6.867 08E-06
现代护理	112	14 709	0.000 377 37	2.873 45E-06
∴	∴	∴	∴	∴
总计	6 494			0.000 566 869

表8中列出了权威因子前100名与总被引频次前100名的期刊的学科分布情况,从表中可以看出,在权威因子前100名期刊中工程技术、地学领域期刊明显要比总被引频次前100名要多,而医学期刊则明显要少,这说明用权威因子排序比较有利于地学、工程技术等学科,对医学则不利。文献计量学研究表明,因为学科规模和性质不同,不同学科领域论文的引用机会是不同,医学领域因为其期刊数量多,从事的相关科研人员多,相应的被引的机会也较大,而工程技术领域期刊因为应用性较强,相应的被引的机会较小。如果按总被引频次来比较期刊,则医学等领域要占据优势,但是如果按权威因子排序,前100名中医学期刊明显减少,工程技术、地学类期刊明显增多,因此我们可以说权威因子能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。

表8 权威因子和总被引频次前100期刊的学科分布情况

学科领域	权威因子前100刊数	总被引频次前100刊数
地学	13	5
工程技术	32	23
化学	3	3
农学	8	9
生物学	7	7
物理	5	6
医学	31	46
综合类	1	1
总计	100	100

### 3 讨论与结论

用总被引频次、影响因子用来评价期刊,固然有其容易理解和结果易于获取的优点,但由于只重数量而不区分不同引用间重要性程度的差异,因此并不合理,

总被引频次多、影响因子大在很大程度上只反映期刊的流行程度,并不能完全反映期刊的权威性。权威因子正好弥补了传统的总被引频次、影响因子只是单纯计算次数而不考虑引用重要性区别的弊端,综合考虑引用次数和引用质量,更能客观地反映期刊权威性。

PrestigeRank 算法解决了期刊引用网络缺失严重的问题,在目前数据库系统收录范围有限的情况下提出了期刊评价可行的方法。

权威因子与总被引频次相关性较高,但是它能发现一些并不流行(总被引频次相对较少)但权威性较高(权威因子相对较高)的期刊。权威因子能部分消除因为学科规模和性质不同造成的被引机会的不同的问题。

与被引次数、影响因子相比,权威因子的缺点是较为抽象,计算量较大。

### 4 参考文献

- [1] Gualberto Buela-Casal. Assessing the Quality of Articles and Scientific Journals: Proposal for Weighted Impact Factor [J]. Psychology in Spain, 2004, 8(1): 60-76
- [2] Maslov S, Redner S. Promise and pitfalls of extending Google's PageRank algorithm to citation networks[J]. Journal of Neuroscience, 2008, 28(44): 11103-11105
- [3] Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual web search engine[J]. Computer Networks and ISDN Systems, 1998, 30(1-7):107-117
- [4] Bollen J, Rodriguez MA, Van de Sompel H. Journal status [J]. Scientometrics, 2006, 69(3): 669-687
- [5] Su Cheng, Pan YunTao, Yuan JunPeng, et al. PageRank, HITS and Impact factor for Journal Ranking[C]. 2009 WRI World Congress on Computer Science and Information Engineering. IEEE Computer Society, 2009:285-290
- [6] 苏成, 潘云涛, 袁军鹏, 等. 基于 PageRank 的期刊评价研究[J]. 中国科技期刊研究, 2009, 20(4): 614-617
- [7] 苏成, 潘云涛, 袁军鹏, 等. 基于 HITS 算法的期刊评价研究[J]. 编辑学报, 2009, 21(4): 366-369
- [8] 潘云涛, 马崢, 苏成, 等. 2009 年版中国科技期刊引证报告:核心版[M]. 北京: 科学技术文献出版社, 2009
- [9] Su Cheng, Pan YunTao, Zhen YanNing, et al. PrestigeRank: A new evaluation method for papers and journals[J]. Journal of Informetrics, 2010, doi:10.1016/j.joi.2010.03.011( In press)
- [10] Langville A, Meyer C. Google's PageRank and Beyond: The science of search engine rankings [M]. Princeton University Press, 2006
- [11] Chen P, Xie H, Maslov S, et al. Finding scientific gems with Google's PageRank algorithm[J]. J Informet, 2007(1): 8-15  
(2010-03-25 收稿;2010-05-09 修回)