

· 科学论坛 ·

学术评价的评价

张春霆*

(天津大学生命科学与工程研究院, 天津 300072)

[摘要] 本文简要介绍了各种学术评价指标,包括基于科研产出、论文引用、引用网络、h-指数和互联网的评价指标,讨论了其优缺点,并最后探讨了同行评议与使用学术指标评价进行评议之间的关系。

[关键词] 论文总数,篇均引用次数,权重引用次数,皇冠指标,特征因子与 SJR 指标,h-指数,同行评议

1 引言

学术评价者学术界之大事也,事关国家科技的良性发展和原始创新能力的提高;也关乎科研人员的入职、职称的提升、资助和奖励的获取等切身利益,不可不察也。2010年6月17日,英国著名刊物 *Nature* 以“评价的评价”为题发表了一篇社论,对该期发表的3篇有关学术评价的文章加以介绍和评论。其中一篇介绍了形形色色的学术评价指标及在世界各地的使用情况^[1]。*Nature* 杂志曾就评价指标的实际使用程度以及与同行评议的关系向世界各地的许多学者和管理者发去了问卷调查。结果有150位研究人员和近30位大学教务长、系主任或其他管理人员反馈了调查问卷,综合后加以公布^[2]。*Nature* 杂志还邀请包括著名的文献(科学)计量学家在内的6名专家就如何改进学术评价指标的使用等学术评价问题发表了看法^[3]。受这些文章的启发,笔者试图用同一个题目结合中国的实际情况,写一篇新的文章,这正是本文的由来,希望对于改进我国学术评价的现状有所参考。本文由两部分组成:一部分介绍各种国际通用的学术评价指标,包括适用于研究群体(大学或研究所等)、刊物和个人的评价指标,并讨论如何克服其缺点;另一部分则探讨同行评议与使用学术指标评价进行评议之间的关系。

2 基于科研产出的评价指标

科学研究,尤其是基础研究其产出的主要形式就是发表学术研究论文。在一般情况下发表论文的数量就成为一项主要的评价指标。当然不排除历史上或现实中某些杰出的学者只发表了极少数的论文却取得了极高的学术成就,在这种情况下再讨论其发表了几篇论文就变得毫无意义。但是在一般情况下这个数字还是有意义的。论文总数这个评价指标在实际计算时存在一个选择什么样的论文来统计的问题。目前国际上存在3大文献检索数据库:Web of Science 或 WoS (Thomson-Reuters 集团);Scopus (Elsevier 集团)和 Google Scholar (Google 集团)。前者从11500种杂志收录论文;中者从16500种刊物收录论文;而后者则包括了比前两者更多的工程学、社会科学和人文科学刊物所发表的论文以及专利^[1]。其中 WoS 又包括 SCIE、SSCI 等7个子库。对于一个给定的研究人员或一所大学(研究所或部门)统计出的论文总数取决于采用哪个数据库。对于自然科学方面的基础研究宜用 SCIE 数据库(约7000种刊物);对于工程技术研究除了用 SCIE 数据库以外,还应参考 Scopus 数据库加以补充;而 Google Scholar 数据库对于两者都有参考价值,而且它是免费的。

* 中国科学院院士。

本文于2010年8月16日收到。

3 基于论文引用次数的评价指标

3.1 篇均引用次数与影响因子

论文的引用具有重要的参考价值,因为它体现了在世界范围内的小同行的署名评议,具有广泛性、公开性、公平性和客观性等优点,历来受到国内外学术界的普遍认可。其中,篇均引用次数是一个极其重要的学术评价指标。统计在一个时间窗口中某刊物所发表论文的总数(分母),以及这些论文在另一个后续的时间窗口中所获得的引用次数(分子),两者之比即为篇均引用次数。若把前时间窗口取为前两年,而把后时间窗口取为今年,则篇均引用次数又称为该刊物在今年的影响因子。显然,某刊物的影响因子,即其篇均引用次数,与在该刊物所发表的某特定论文所获引用次数属于两个不同的概念,两者没有可比性。例如,据 *Nature* 统计,在 2005 年 *Nature* 的 25% 的论文获得的引用次数占总引用次数的 89%^[1]。换言之,另外 75% 的论文只获得了 11% 的引用次数。也就是说即使在 *Nature* 这样的刊物上发表的论文,其中大部分所获得的引用次数相当有限。笔者遗憾地指出,在我国教育界和学术界有一种用刊物影响因子来评价在该刊物所发表论文及其作者学术水平的倾向。有的管理部门甚至要求科研人员必须在影响因子多少以上的刊物上发表论文。这种倾向不仅是十分错误的,因为它背离了科学研究的宗旨;而且也是十分荒唐的,因为它误用了刊物的评价指标。著名文献计量学家 A. van Raan 指出:“如果有一个观点每一位文献计量学家都同意的话,那就是:你们永远不要用刊物的影响因子来评价一篇论文或某个研究人员的学术表现——因为那是一种不可饶恕的大罪(That is a mortal sin)^[2]”。但愿文献计量学家的忠告能引起我国教育界和学术界的警觉。

3.2 考虑作者贡献之不同的权重引用次数

无论是总引用次数或篇均引用次数都是针对论文而言的。如果该论文只有一位作者,那么全部引用次数理应归于作者本人。但是如果有多位作者(这是当前发表论文的普遍形式),国际上三大文献检索数据库的现行做法是把论文的引用次数归于每一位作者,导致了一篇论文的引用次数被反复计算,这是十分错误的。为了解决这一问题,笔者定义了两个概念:针对论文的引用次数和针对作者的引用次数,而后者等于前者乘以作者的贡献权重系数。权重系数是介于 0 和 1 之间的一个实数,可以根据该作者在论文中的排名计算出来。为了计算权重系

数,笔者提出了一个论文荣誉的三分原则:将一篇论文由引用次数获得的荣誉等分为 3 份:第一作者和通讯作者各得 1 份(他们的权重系数均为 1),其他作者的权重系数之和等于 1。而他们的权重系数原则上按作者排名的先后顺序递减,详情可参见文献[4],在这里不再介绍。为了给权重系数的计算提供方便,笔者建立了一个网站,免费提供权重系数和权重引用次数的在线计算,详见:<http://www.wcitation.org/>。

3.3 按学科归一化的引用分数和皇冠指标(Crown indicator)

用引用次数作为评价指标的另一个大缺点是不具有学科可比性。例如,生物、医学学科的引用次数相对较多;而数学、力学学科的相对较少。解决这一问题的方法之一是引入按学科归一化的引用分数。假设一个研究群体在某一学科领域发表了一批论文。首先计算其篇均引用次数;其次再计算世界范围内在同一学科领域和相同情况下的篇均引用次数,两者之比则成为按该学科归一化的引用分数。其值为 1 表示达到了世界平均水平;其值大于 1(小于 1)表示比世界平均水平高(低)。于是归一化的引用分数在不同学科间就有了可比性。若某一研究群体同时从事多学科研究,则还要对所涉及学科求平均。在计算过程中可有不同的变化,于是就有多种不同的归一化方案。其中荷兰 Leiden 大学科学与技术研究中心(CWTS)所提出的皇冠指标就是其中的一种,被作为世界大学排行榜的评价指标使用,详见 <http://www.cwts.nl/ranking/>。但是仔细一想这种做法也有问题。把科学划分为各种领域是一种粗粒化的做法。问题是,粗到多粗为宜?细到多细为好?“如果划分过宽就会抹杀细节;如果划分过细,就会导致毫无用处的结果”^[3],最终使人处于两难的境地。到目前为止,这种指标只应用于对研究群体(大学或研究所等)的评价。

4 基于引用网络(Citation network)的评价指标

如果把每篇论文用平面上一个点(节点)来表示,而它们之间的引用用一条连接两点的线段来表示,则构成一个网络图。引用次数是一正整数,可用实轴上的一个点来表示。而引用网络是一个二维图,应比实轴上的一些点包含更多的信息。令人感兴趣的是互联网(Internet)的众多网页和它们之间的链接关系也可用类似的网络图来表示。这使人联

想到能否用互联网的研究成果来研究引用网络。互联网的网页重要性排序算法的发明是一项重要的成果,其中由美国博士研究生 L. Page 和 S. Brin 研发的 PageRank 算法可能是最重要的之一,它是 Google 搜索引擎的核心。PageRank 算法认为一个网页 A 的重要性取决于链向它的网页的数量和其重要性:链入的网页数越多,A 页面就越重要;与此同时,链向 A 的网页本身越重要,其对 A 网页重要性的贡献就越大,反之亦然。可用一个数来表示网页的重要性,记为 PR。显然,按照这一算法一群网页的 PR 值都耦合在一起。Page 和 Brin 采用了叠代法解决了这一问题,算出了每一个网页的 PR 值,最终给出了网页重要性排序。PageRank 算法的成功使人们想到用该算法对杂志的重要性进行排序。杂志之间通过其论文的引用构成了一个网络。杂志不同于网页,所以应用 PageRank 算法的具体过程是比较复杂的。美国的研究人员在 WoS 的基础上,按照这一思路提出了评价学术刊物的新指标,称为特征因子(Eigenfactor),又细分为特征因子分数(Eigenfactor Score)和论文影响分数(Article Influence Score),详见 <http://www.eigenfactor.org>,这里不再详述。Thomson-Reuters 集团在其 JCR 报告增强版里同时发布特征因子数据,包括特征因子分数和论文影响分数。可以理解,刊物的特征因子与刊物的影响因子呈强正相关,但排序不完全相同。影响因子只考虑引用次数;而特征因子不仅考虑引用次数,而且还考虑引用刊物的重要性,即来自重要刊物的引用其权重就大,反之亦然。从这个意义上来说,特征因子这类指标比影响因子更先进。基于同样的想法但采用 Scopus 数据库,西班牙的研究人员提出了 SJR 指标^[5]。值得指出的是,特征因子和 SJR 指标都是用来评价刊物的而不是用来评价个人的学术表现的。但是在引用网络的基础上发展评价个人学术表现的指标是可行的。意大利学者 Radicchi 等人从 1893—2006 年的 114 年间 *Physical Review* 杂志所发表的 40 多万篇论文以及这些论文所附的 878 多万篇参考文献出发,建立了一个庞大的作者-作者引用网络,然后再采用类似于 PageRank 的算法计算出作者的排序名次^[6]。为了实用,他们建立了一个网站:<http://www.physauthorsrank.org>,提供免费在线排序服务。任何于 2006 年以前在 *Physical Review* 上发表过至少一篇论文的作者皆适用。只要输入姓名,就立即给出该作者的排序名次及显示历年来排序名次变动情况的

直方图,使用十分简便,结果相当合理,建议我国物理学研究人员不妨一试。

社会网络理论中所定义的一些参数在引用网络中可能有意义。其中有一个参数叫中介中心性(Betweenness centrality),它正比于网络中任意两个节点通过被研究节点(对应一篇论文)的最短路径数目;另一个参数叫接近中心性(Closeness centrality),它表示该节点到其他所有节点的最短路径的平均长度。“这些参数在学术评价中有何意义仍不清楚,可能是学科交叉性的一种指标”^[1]。毫无疑问,引用网络是学术评价的重要研究领域,值得重视。

5 基于将论文数与引用次数相结合的评价指标:h-指数^[7]

2005 年美国物理学家 Hirsch 提出一种将论文数与引用次数相结合的新的评价指标,称之为 h-指数^[7]。一个学者的 h-指数为 h 意味着其至少发表了 h 篇论文,同时至少被累积引用了 h^2 次。由于这个学术评价指标的新颖性和简单性,立即在全球范围内引起了广泛的注意,掀起了研究它的热潮。迄今为止,至少发表了 100 多篇研究 h-指数的论文。为了克服其缺点至少提出了十几个 h-指数的变种,又称为 h-型指数。面对这种局面,澳大利亚著名学者 Anne-Wil Harzing 说:“自 2005 年以来提出了一打以上的 h-型指数,就连文献计量学专家也搞不清楚哪些是最好的。结果,大部分科学家坚持还是使用 h-指数,尽管有种种缺点”^[1]。其实这种说法是不对的,h-型指数虽然为数众多,但是它们基本上反映了相同的信息,只是表述方式不同而已。它们之间呈强正相关,这提示用少数一、两个即可。其中 e-指数是笔者提出来的,它被定义为在 h 篇论文所获引用次数中扣除 h^2 之后多余部分的平方根^[8]。简言之,在使用 h-指数的前提下,e-指数补充了其丢失的引用信息,故应该同时使用。美国学者 Dodson 就建议同时使用 h-指数和 e-指数来评价大学教师的学术表现^[9]。天津大学人事部门联合使用 h-指数和 e-指数作为主要评价指标建立了人才评价平台,取得了良好的效果^[10]。在许多 h-型指数中,g-指数是一个重要的指数,它主要反映被 h-指数忽略的高引用论文的引用信息^[11]。在这点上,看,g-指数与 e-指数是相互冗余的。因此,在使用 h-指数的前提下,(h,e)组合或(h,g)组合皆可取得满意的效果。但是 g-指数在若干重要的情况下没有定义^[8]。笔者重新定义了 g-指数,彻底消除了这一缺陷,并推导出一个公式可

以从 h-指数和 e-指数出发快速计算出 g-指数^[12]。意大利学者 A. Bee 开发出一套计算机程序,当用 Firefox 浏览器访问 Google Scholar 时,会自动显示所查询作者的 h-指数、g-指数和 e-指数。该程序应下载到本地计算机中运行,详见 <https://addons.mozilla.org/en-US/firefox/addon/45283/>。

Hirsch 本人认为 h-指数的最大缺陷是没有考虑多作者论文中的作者人数和荣誉分配^[13]。其实,笔者也注意到了这一问题,并提出了用 w-指数来解决它^[14]。所谓 w-指数就是建立在权重引用次数上的 h-指数。Hirsch 在其论文中^[13]引用了文献[4],并提出另一种解决方案;即 h-bar-指数。在使用 h-指数的前提下,当出现多作者论文时,可选用 w-指数或 h-bar-指数。对前者已建立了网络平台(即前已提及的 <http://www.wcitation.org>)提供 w-指数的在线计算;对后者目前尚无此类服务。

6 基于互联网(Internet)的评价指标

随着互联网的普及,越来越多的学术刊物将论文在线发表,变成一份纯电子刊物。另外一些则在保持其纸质印刷的传统形式下,先期在线发表。有一种趋势就是越来越多的论文,尤其是重要的论文在网上先期发表,免费阅读或下载。这就产生了一些新的学术评价指标,如在线阅读次数和下载次数等。有的刊物,如 PLoS ONE,就提供其论文的被阅读次数和下载次数,以及这些次数的总和随时间的变化曲线。例如文献[8]自 2009 年 5 月 5 日发表算起到 2010 年 8 月 15 日为止(本文定稿之日),已经被阅读和下载了共 3494 次。显然,总次数越大就越表示或者这是一篇热门论文;或者这是一篇重要论文。另有一些刊物将论文按阅读次数和下载次数之和排序,排在前面的被赋予“高访问论文”(Highly accessed article)称号;或者将排在前几名的加以公布。可以想象,这些阅读者或下载者很可能成为该论文的潜在引用者,不过这要等 2、3 年之后才能见分晓。在线阅读和下载次数反映了论文的即刻以及长远影响力,这些指标是值得深入研究的。

7 同行评议与使用指标评价之间的关系

同行评议的常用形式之一就是提供个人推荐信。对被评价的群体、个人或一项成果由同行专家提供一封或几封推荐信,对被评价对象进行定性的描述并做出结论。这种做法需要两个先决条件:(1)同行专家必须是真正的同行,即所谓小同行,应

当对被评价内容相当熟悉和了解;(2)评价专家应该认真负责、能够本着“公平、公正和客观”的原则无个人偏见地进行评价。可是有时上述两个条件很难满足。随着科学技术的高度分化,产生了隔行如隔山的现象。对于一个项目、一个成果或某个人的学术表现很难找到真正的小同行来评价。如果只能找一些大同行甚至外行来评价,则往往会导致不公平和不准确的结果。同时由于复杂的人际关系(“互评网”的客观存在),上述第二个条件有时也难以满足。在这种情况下,虽然是干巴巴的、但却是定量的和客观的评价指标可以弥补同行评议之不足。在我国,有的管理者在批评片面使用某些评价指标的弊端后,正尝试用更科学的评价方法,比如,一个院系搞得好不好由国际专家来评估。那么请看国际专家是怎样进行学术评价的。针对 *Nature* 杂志的问题:“你们的院系在做出下列决定时,在某种程度上采用科学表现的评价指标吗?”,问卷调查表明^[2]:在做出雇佣决定、转正(转为终身制副教授)决定、提职称决定和学术表现鉴定方面,回答“*Yes*”的占问卷的 70%左右,而回答“*No*”的只占 10%—20%。而在工资决定和研究经费分配两方面,回答“*Yes*”和“*No*”的大致相同。由此可见,国外的管理者或专家在对许多问题做决定时,学术评价指标起着重要的作用。针对 *Nature* 杂志的另一个问题:“对于评价指标的使用方式,一般说来你满意程度如何?”,约 1/3 的问卷回答“一点也不满意”;超过 1/3 的问卷反馈“不是非常满意”;而将近 1/3 的问卷说“极为满意或非常满意”。在我国,就有管理者不满意评价指标,说他们只看推荐信,不看什么指标。但是清华大学施一公教授一语道破了天机,他说:“我们并不看(被评价者的)论文发表记录,也不告知同行评议者这么做,但实际上评价指标确实有影响,因为同行评议专家会看这些东西。”^[2]。这种情况被描述成:“定性的推荐信有时从后门引入定量的评价指标”^[2]。这说明,在做同行评议时适当参考评价指标可能使其评价更客观和更公正。完全排除评价指标的作用是不可取的。如果找不到适当的同行来评价,就要考虑使用评价指标来评议了,但是完全依靠评价指标也是不可取的。在此情况下,还要尽量收集各种同行的意见,如:在各种引用中引用者如何评价被引用的内容?被评价者是否作为主讲人(Keynote speaker)在重要的国际会议上做过报告?被评价者的文章发表后是否另有专家撰文予以正面评价?如何评价?是否被邀请在重要的专业刊物上发表综述文章?等

等。总之,凡涉及对个人学术表现的评价时,在定性的描述中可引入定量的指标加以佐证;在使用定量的评价指标时可引入引用者的定性的描述加以补充,综合考虑各种因素之后再作结论。最后我要着重指出:学术评价是一个复杂的科学和社会问题,没有最好的答案,只有更好的答案,学术评价工作要永远处在不断的改进之中。

参 考 文 献

- [1] Van Noorden R. A profusion of measures. *Nature*, 2010, 465: 864-866.
- [2] Abbott A et al. Do metrics matter? *Nature*, 2010, 465: 860-862.
- [3] Van Noorden R. How to improve the use of metrics. *Nature*, 2010, 465: 870-872.
- [4] Zhang C T. A proposal for calculating weighted citations based on author rank. *EMBO Rep*, 2009, 10: 415-416.
- [5] Gonzalez-Pereira B et al. A new approach to the metric of journals' scientific prestige: The SJR indicator. *J Informetrics*, 2010, 4: 379-391.
- [6] Radicchi F et al. Diffusion of scientific credits and the ranking of scientists. *Phys Rev E*, 2009, 80: 056103, 1-10.
- [7] Hirsch J E. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*, 2005, 102: 16569-16572.
- [8] Zhang C T. The e-index, complementing the h-index for excess citations. *PLoS ONE*, 2009, 4: 1-4.
- [9] Dodson M D. Citation analysis: Maintenance of h-index and use of e-index. *Biochem Biophys Res Commun*, 2009, 387: 625-626.
- [10] Zhang L et al. Establishment of Paper Assessment System based on Academic Disciplinary Benchmarks. *Scientometrics*, 2010, 84: 421-429.
- [11] Egghe L. Theory and practice of the g-index. *Scientometrics*, 2006, 69: 131-152.
- [12] Zhang C T. Relationship of the h-index, g-index and the e-index. *J Am Soc Inf Sci Tech*, 2010, 61: 625-628.
- [13] Hirsch J E. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, 2010. DOI 10.1007/s 11192-010-0193-9.

ASSESSING ACADEMIC ASSESSMENT

Zhang Chunting

(*Institute of Life Science and Biotechnology, Tianjin University, Tianjin 300072*)

Abstract Various academic assessing indicators, including those based on scientific output, citations, citation network, h-index and Internet, are briefly presented. The relationship between the peer-review and assessments based on various academic assessing indicators is discussed finally.

Key words publication number, citations per publication, weighted citations, crown indicator, eigenfactor and SJR indicator, h-index, peer-review

· 资料 · 信息 ·

我国研究人员发现铁磁石墨烯体系 CT(电荷共轭-时间反演)不变的量子自旋霍尔效应

最近,中国科学院物理研究所孙庆丰和谢心澄研究员在铁磁石墨烯体系中预言了一种新类型的拓扑绝缘体和量子自旋霍尔效应(*PRL*, 104, 066805 (2010))。与由自旋轨道耦合所引起的拓扑绝缘体不同,这类新的拓扑绝缘体与自旋轨道耦合无关,体系也不具有时间反演不变性,但具有 CT 不变性(C 为电子-空穴电荷共轭变换, T 为时间反演变换)。再者,处在平衡时,原来拓扑绝缘体的边界态只携带持续自旋流,而现在的边界态同时携带持续自旋流和持续电流。加电压时,现在的体系可以同时观察到量子自旋霍尔效应和量子霍尔效应。所以,这是一种新类型的拓扑绝缘体。

由于这类体系既有量子自旋霍尔效应也有量子霍尔效应,导致纵向电阻 $R_{14,23}$ 和霍尔电阻 $R_{14,26}$ 都呈现平台结构。另外,他们也研究了这类体系的自旋霍尔阻抗。结果显示自旋霍尔阻抗 RS 也有量子化平台。特别是,这种自旋霍尔阻抗的量子化平台有非常强的抗无序能力,即使无序强度 W 增加到带宽量级,自旋霍尔阻抗的平台还能保持。

以上工作得到了国家自然科学基金和科技部经费的资助。

(中国科学院物理研究所 供稿)