

● 齐世杰^{1,2}, 郑军卫¹

(1. 中国科学院兰州文献情报中心, 甘肃 兰州 730000; 2. 中国科学院大学, 北京 100049)

科技论文定量评价方法研究进展*

摘要: [目的/意义] 科技论文评价是科研活动评价的一项重要内容, 论文评价方法的客观性和准确性对科研绩效评估有至关重要的作用。Web 2.0 时代的到来与在线科研环境的新形势, 为论文定量评价方法的革新创造了机遇。[方法/过程] 通过梳理国内外科技论文评价相关文献, 总结分析了科技论文定量评价的主要方法, 包括传统文献计量评价法、网络分析法、基于 Altmetrics 评价法和综合评价方法, 并对未来科技论文评价的发展趋势与探索方向做出展望。[结果/结论] 基于科技论文交流过程的综合评价、多源异构数据聚合的科技论文评价是科技论文定量评价的发展方向, 在这种趋势下, 科研人员需结合语境界定和分析读者行为动机, 并大力推进单篇文献质量与影响力评价系统建设。

关键词: 科技论文评价; 定量评价; 评价方法; 评价指标; 补充计量学

Research Progress of Quantitative Evaluation Methods of Scientific Papers

Abstract: [Purpose/significance] Evaluation of scientific papers is an important part of the evaluation of scientific research activities. The objectivity and accuracy of the evaluation methods have a crucial influence on the evaluation of scientific research performance. The arrival of the Web 2.0 era and the new situation of the online research environment have promoted the development of the quantitative evaluation methods of scientific papers. [Method/process] By combing the relevant literatures at home and abroad, this paper summarizes and analyzes the main methods of quantitative evaluation of scientific papers, including traditional bibliometric evaluation method, network analysis method, evaluation method based on altmetrics and comprehensive evaluation method, and makes a prospect for the development trend and exploration direction of future scientific and technical papers evaluation. [Result/conclusion] The comprehensive evaluation based on the communication process of scientific papers and the evaluation of scientific papers of multi-source heterogeneous data aggregation is the development direction of quantitative evaluation of scientific papers. Under this trend, researchers should define and analyze readers' behavior motive according to contexts and promote the construction of quality and influence evaluation system on single paper.

Keywords: scientific paper evaluation; quantitative evaluation; evaluation method; evaluation index; altmetrics

科技论文是科研活动中形成的知识形态成果, 也是科学研究成果的重要表现形式, 能够从一定角度反映科研绩效水平。科技论文的评价是科学评价活动的重点, 它关系到国家科技创新能力、科技成果转移转化能力的提高, 有助于促进国家科技的良性发展及对科研工作有效管理。

目前, 科技论文定量评价主要以引文分析为主, 它在论文评价中发挥着重要作用, 但其缺陷也被广大学者诟病, 如引文分析的时滞性、评价方法的单一性等。随着数字出版技术的迅速发展和开放获取 (OA) 的普及, 科研交流环境发生了巨大变化, 一种新的计量评价方法——Altmetrics (补充计量学) 应运而生, 给科技论文定量评价带来了新的视角。因此, 本文从传统文献计量评价方

法、基于网络分析的评价方法、基于 Altmetrics 的评价方法、综合评价方法 4 个方面梳理了论文定量评价方法现状并探讨了未来发展趋势, 以期为后人改进评价指标或构建新形势下的评价体系提供一定的理论依据。

1 传统文献计量评价方法

传统文献计量学的评价方法主要是通过论文的被引频次、发表期刊的影响因子、h 指数等外在指标来衡量论文的质量。期刊影响因子和被引频次是论文质量评价中两个最重要的指标。

1.1 基于影响因子与被引频次的评价

1955 年, 美国科技信息研究所 (ISI) 的 Garfield 提出了论文被引频次指标^[1], 1963 年 ISI 出版《科学引文索引》并发布了期刊影响因子, 并于 1975 年起每年发布上一年度的《期刊引证报告》。期刊影响因子通常是指期刊前两年发表论文在统计当年中被引用总次数除以该刊前两

* 本文为中国科学院“文献情报和期刊出版领域引进优秀人才计划”的研究成果, 项目编号: 传播字 [2014] 10 号。

年发表的论文总数。

被引频次和影响因子的提出在论文定量评价中具有里程碑意义,为了使其更好地发挥作用,大量学者针对其时滞性、偏态分布、不同学科引用差异等不足进行了研究和修订。

值得注意的是, Schubert^[2]和 Bornmann^[3]相继提出了相对指标和百分位法。相对指标是将被引频次、期刊影响因子除以某学科的平均值、某期刊的平均值以及某年的平均值,消除不同论文的固有差异,以达到跨学科、跨期刊、不同发表时间的论文的可比性,使得评价更加准确和客观。百分位法是根据被引频次的偏态分布规律对其分段处理及标准化的方法。该方法被美国国家科学基金会(National Science Foundation)和ESI数据库(Essential Science Indicators)借鉴使用,将学科论文被引频次标准划分为多个等级,进行等级内的比较与评价。至今,这两个指标仍在被不断改进与完善,因此才能被科研评价界沿用至今。

1.2 h 指数评价法

2005年, Hirsch提出了测度学者影响力的指标——h指数,之后, Schubert^[4]将h指数应用到论文评价中,提出用参考文献的h指数代替作者的h指数以评价论文质量,即若文献P的参考文献中有n篇论著每篇至少被引用了n次,那么n即论著P的h指数。

值得一提的是,国内学者邱均平^[5]基于h指数和g指数提出论文质量指数PQI(Paper Quality Index),公式为:

$$PQI_{ij} = \frac{IF_{ij}}{IF} + \frac{TC_{ij}}{TC}$$

式中, IF_{ij} 表示论文所在期刊i在j年的影响因子; \overline{IF} 表示论文所在期刊j年的平均影响因子; TC_{ij} 代表论文在j年总被引频次; \overline{TC} 表示论文所在期刊j年的平均被引频次。PQI论文质量指数消除了学科差异和论文年龄差异,但是因为计算过程中用到影响因子,故该指标只能计算发表两年以后的论文,存在一定的时滞性。

1.3 其他方法

一些学者基于被引频次提出了新的论文评估指标。例如, Bornmann^[6]提出了引用速度指标(Citation Speed Index),指的是论文从发表至第一次被引的时间间隔,间隔越短则论文的影响力越高。Kosmulski^[7]提出“成功论文”(Successful Paper)概念,指被引次数超过该论文参考文献数量的论文。Hu Xiaojun等^[8]提出基于文章参考文献数量和引用数的CC指数、RC指数。徐芳等^[9]提出了同论文数(EPN)指标,从作者对论文投入的角度,测量学术论文中包含的等同工作量,实现对论文质量的评价。

2 基于网络分析方法

随着Web时代的到来,互联网已然成为科研工作中文献来源与发表的重要途径,社会网络、复杂网络与引文分析的碰撞给论文评价带来了新思路。基于引文网络分析方法主要包括类PageRank算法和社会网络分析方法。

2.1 类PageRank算法

早在1976年, Pinski^[10]就提出对有声望的期刊引用应赋予更高权重,而链接分析法的出现使得这一想法得以通过计算机技术规模实现并应用于实践,其中具有代表性的方法是PageRank算法。

1998年, Google提出PageRank算法来标识网页的等级和重要性,其基本假设是:每个到页面的链接都是对该页的一次投票,被链接的越多意味着该网页越重要,同时网页的重要性可以通过链接传递^[11]。在引文分析中,主要是通过引文网络中节点(论文)的入链(被引用)与出链(引用)计算与分析,对论文的被引频次进行加权,通过引文质量评估论文的质量。

为了让PageRank更好地适用于引文网络,学者通过引入虚拟节点、时间因素权重、衰减时间因素等参数,对PageRank算法进行改进与完善,出现了PaperRank、CiteRank、PrestigeRank等诸多类PageRank算法^[12-14],打破了将所有引用行为等权设置的传统方式。重要的是,一些学者根据引文网络的分层和嵌套等结构特征探索出更合理、更细致的权重分配方式。王凌峰^[15]、杜朝东^[16]以年为单位将论文引文网络分为多个层级,利用变异系数法对各级赋予不同权重。Sidiropoulos^[17]则认为直接引用和近期引用对论文影响更大,由此提出SCEAS Rank指标。不仅如此,学者Fragkiadaki^[18]还将路径深度作为影响因素,提出基于多层网络的f-Value指标。

2.2 社会网络分析方法

除了类PageRank算法,社会网络分析方法也常用来评价论文的质量,能够根据节点在网络中所处位置评价论文的重要性,主要包括度中心性、中介中心性、结构洞约束系数等指标。

度中心性通常用来测度节点在引文网络中核心性,被引频次越多节点中心性越高,在引文网络所处位置也越重要。中介中心性和结构洞都是从流量角度对节点测度的指标,判定在知识流通中起到枢纽作用的节点^[19]。节点的中介中心度越高,表明节点承上启下的控制能力越强。结构洞的存在可以在节点间起到连接作用,节点的结构洞系数越低,证明其越具有多领域的知识背景。科研人员通常借助社会网络分析工具,如ucinet, MultiNet, Gephi等实

现对网络的分析。例如,姜磊^[20]利用社会网络工具绘制引文网络图谱,并选取度中心性、中介中心性等社会网络评价指标,结合传统评价方法集成了单篇学术论文评价体系。

此外,社会网络分析中表示邻近节点间关系、网络结构的理论与方法,如 Bonacich 权利指数,被学者借鉴并应用到引文网络中。相关研究有:纪雪梅^[21]通过计算被引文献的权利指数评估单篇论文的质量。高霞^[22]基于社会网络分析中结构声望 (Structural Prestige) 理论,构建了用于评价基础研究影响力的新指标——欢迎度、临近声望、横权重 3 个结构声望指标。Chaomei^[23]提出用改变边界跨越的知识网络结构程度表征一篇论文的学术水平。

综上所述,基于网络的分析方法与分析工具更加多样化,评估更加细致化。相较于传统文献计量指标,具有以下优势:①根据引文网络结构,对被引频次分层级赋权,使评价更加真实和客观。②基于知识流通的评价,有助于发现研究领域起到枢纽作用的重要论文与潜在的创新论文。③通过社会网络工具能够方便地消除自引的影响,弥补了传统评价方法的缺陷。

3 基于 Altmetrics 的评价方法

Web 2.0 时代的到来催生了在线科研的新环境。2010 年, Priem 等^[24]正式提出 Altmetrics,侧重研究论文在网络交互活动中产生的影响力。Altmetrics 指标包括获取、传播、利用和引用 4 个维度,如表 1 所示。

表 1 Altmetrics 部分指标分类及来源

维度	指标	部分指标来源
传播	点击量 (Clicked)	PLoS
	浏览量 (Viewed)	PLoS HTML usage, PLoS PDF usage
	下载量 (Download)	Dryad
	被分享 (Shared)	Figshares
	被推荐 (Recommended)	F1000 recommendation
获取	喜欢 (Liked)	Youtube
	书签 (Bookmarked)	Mendeley, CiteULike
利用	被提及 (Mentions)	PLoS search
	评论 (Comments)	Youtube
	微博	Blog posts
	被讨论 (Discussed)	Facebook, Twitter
引用	学术引用	Web of Science, Scopus
	非学术引用	Wikipedia, Nature blogs

Altmetrics 指标将传统文献计量指标与社交媒体新指标结合,一方面能够对学术成果进行影响力的补充性评估,另一方面又扩充了对非科研人员关注、下载文献等行为分析的大众评审。国外已有诸多实践案例,并获得英国《自然》等国际著名期刊的认可。国内方面,西安交通大学将 WoS 和 PlumX 与机构知识库整合并使用,但其他机

构仍持谨慎态度。

相较于传统方法与引文网络分析法, Altmetrics 指标涵盖的平台更广泛:除网络数据库外,也将社交网络、文献管理工具等新型开放平台纳入其中。补充计量指标聚焦于科学交流与学术成果传播的过程,更能提供实时更新、多种形式的文献评价数据。但是, Altmetrics 也存在一些不足: Altmetrics 工具平台数据源参差不齐;数据质量标准、使用规范及方案仍有待进一步完善;其指标应用于论文评价体系缺乏大规模有效数据验证。

4 综合指标评价方法

科技论文的质量与早期被引频次、参考文献数目、跨学科参考文献数目、高被引及跨学科引证文献数目等指标有密切关系^[25]。因此,在进行论文评价时,通常会将论文本身属性,如作者、机构、参考文献数量、项目基金支持等作为影响因素加入到评价体系中,构成论文的综合指标评价方法,避免指标单一造成评价结果的偏差。

国内外学者进行了大量相关研究,如:苏学^[26]构建了包含期刊影响因子、论文被引情况、论文基金资金情况等因素分别对期刊论文定量评价体系进行探讨。林德明等^[27]提出论文内部指标、外部指标和社会网络分析指标相结合的论文评价体系,内部指标包括:参考文献数量、基金资助和作者合作规模等。Cho 等^[28]提出了开源环境下的论文评价综合指标,包括:引文量、点击量、论文年龄、自引、可获得渠道数目、论文页数和论文类型,构建了 OAR 模型。

综合指标评价方法融合了多资源与多角度的评价指标,考虑了论文的内在和外在属性,在一定程度上使评价方法之间优势互补,使评价结果更具客观性、准确性和全面性。

5 科技论文定量评价发展方向

科技论文影响力评价是科研评价的重要组成部分,评价标准因评价主体、评价需求和评价目的的不同有所差异,目前尚未形成统一的科技论文影响力评价的指标体系,科技论文质量与影响力评价平台建设更是鲜有问津,据此本文认为科技论文定量评价有以下未来发展趋势及探索方向。

1) 基于论文交流过程的科技论文综合评价。随着整个科研环境的开放、情报研究技术的发展以及学科间的相互渗透与融合,论文定量评价方法更加多样化、细致化和过程化,未来更是朝着综合利用多种方法进行多维度、全方位评价的方向发展。一方面,科研人员需探索基于引文分析等传统定量评价方法在大数据环境下的改进与迁移,

促使科技论文定量评价方法日臻完善；另一方面，要注重评价指标的相对化与权重设置的科学合理，利用计算机技术与网络分析工具进一步规范评价指标的设立、评价方法的选取及评价过程的精确与完整，以更精准、可靠的评价结果为广大学者以及科研评价工作服务。

2) 多源异构数据聚合的论文定量评价。目前，论文评价数据来源广泛且结构复杂。除了文献引文数据外，在线科研以及开放平台数据，如社交媒体指标等也成为论文定量评价方法的重要组成部分。为了确保论文评价结果的可靠性及有效性，在评价指标确立之前需对数据源进行标准化规范与整合，以保证论文评价指标数据来源的可信度。建立多来源数据的整合系统主要包括：①网络平台论文评价数据源的标识符认证与转化系统，统一各个数据源间论文的标识符，确保评价对象的唯一性。②网络平台数据采集与分类系统，采集各个网络平台的数据并根据其数据来源属性分类，以便评价指标的选取。③数据源的语境分析与判别系统，根据数据源的可靠性，分析读者使用行为的语境，以判别数据源的有效性与数据的权威性。④数据源的实时监测系统，实时监测并跟踪数据源的累积态势。

3) 结合语境界定和分析读者行为动机。无论是学术引用指标还是收藏、提及等使用行为指标，都是根据读者对文章的反馈进行的计量与分析。当读者对一篇论文分享、标签和引用时，很大程度上是对论文的认同，也可能是通过分享供大家讨论辨析，但也可能是持反对意见。囿于读者使用行为及动机的分类标准尚未确立和统一，相关研究虽多却几乎无可比性，也不能判定其结果的有效性。因此，通过语境界定读者对论文引用、使用等行为动机是首要且必要的。随着网络技术的发展与开放获取环境的影响，探究读者行为动机并非难事，首先应在引用和使用动机的分析和归纳上形成统一的认识和理解，由先前研究中引文分析的视角转向基于读者的语境视角进行研究之后，利用整合思维实现大数据环境下引用和使用行为及动机的规范划分，才能从根源上实现评价的真实性，促使论文评价方法进行多维度全方位整合，最终朝着体系化方向发展。

4) 大力推进单篇文献质量与影响力评价系统建设。科技论文评价伴随计量学的发展已有 60 余年历史，评价模式从依赖于核心期刊遴选逐渐转向单篇论文的全方位评价，但相关理论研究仍未能应用至单篇文献影响力评价的实践平台建设，国内单篇文献评价平台屈指可数，且相关指标也都依赖被引频次与影响因子。平台建设应以收录开放获取文献的网络数据库为基础，以能反映文献的短期、动态影响力为目标，同时选择相应 Altmetrics 指标与

工具，融合对文献的定性评价（如专家评审或软同行评议），建立在 Web 2.0 环境下的单篇文献评价平台，推动科研评价的发展。

6 结束语

本文从传统文献计量评价法、引文网络分析法、Altmetrics 评价法和综合评价法 4 个方面对科技论文定量评价方法进行梳理，发现科技论文定量评价日益趋向细致化和过程化，同时具有评价数据规模变大、指标数据来源广泛、评价指标多样化、评价内容全面化的发展特点。随着大数据时代情报学定量方法和计量工具与科学环境的相互促进与发展，科技论文定量评价发展趋势将是基于论文交流过程的科技论文综合评价以及多源异构数据聚合的定量评价，科研人员需从论文评价数据的根源出发，结合各研究领域的新思维和新方法，界定和分析语境下读者行为动机，引进先进的计算机辅助技术，大力推进文献质量与影响力评价系统的开发，在加强理论探究的同时需将研究成果应用于实际，发挥科技论文定量评价在各个层面的真实效用，促进我国科研事业的发展，从而提高我国在大数据时代 Web 2.0 环境下的科技竞争力。□

参考文献

- [1] GARFIELD E. Citation indexes for science: a new dimension in documentation through association of ideas [J]. Science, 1955, 122 (3159): 108-111.
- [2] SCHUBERT A, BRAUN T. Cross-field normalization of scientometric indicators [J]. Scientometrics, 1996, 36 (3): 311-324.
- [3] BORNMANN L. Towards an ideal method of measuring research performance: some comments to the Ophhof and Leydesdorff (2010) paper [J]. Journal of Informetrics, 2010, 4 (3): 441-443.
- [4] SCHUBERT A. Using the h-index for assessing single publications [J]. Scientometrics, 2009, 78 (3): 559-565.
- [5] 邱均平, 马瑞敏, 程妮. 利用 SCI 进行科研工作者成果评价的新探索 [J]. 中国图书馆学报, 2007, 33 (4): 11-16.
- [6] BORNMANN L, DANIEL H D. The citation speed index: a useful bibliometric indicator to add to the h, index [J]. Journal of Informetrics, 2010, 4 (3): 444-446.
- [7] KOSMULSKI M. Successful papers: a new idea in evaluation of scientific output [J]. Journal of Informetrics, 2011, 5 (3): 481-485.
- [8] HU X, ROUSSEAU R, CHEN J. A new approach for measuring the value of patents based on structural indicators for ego patent citation networks [J]. Journal of the Association for Infor-

- mation Science and Technology, 2012, 63 (9): 1834-1842.
- [9] 徐芳, 刘文斌, 李晓轩. 等同论文数 (EPN): 学术论文质量评估的新指标 [J]. 科研管理, 2011, 32 (7): 150-156.
- [10] 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.
- [11] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine [J]. Computer Networks and ISDN Systems, 1998, 30 (1): 107-117.
- [12] 郑美莺, 梁飞豹, 梁嘉嘉. 单篇论文评价方法——Paper-Rank 算法 [J]. 科技与出版, 2016 (7).
- [13] WALKER D, XIE H, YAN K K, et al. Ranking scientific publications using a model of network traffic [J]. Journal of Statistical Mechanics Theory & Experiment, 2006, 6 (6).
- [14] SU C, PAN Y T, ZHEN Y N, et al. PrestigeRank: a new evaluation method for papers and journals [J]. Journal of Informetrics, 2011, 5 (1): 1-13.
- [15] 王凌峰, 张泽玺. 基于引文网络的单篇论文新评价指标: Hi 指标 [J]. 情报理论与实践, 2012, 35 (11): 52-55.
- [16] 杜朝东, 胡军. 网络文献中单篇科技论文的学术影响力分析 [J]. 情报理论与实践, 2013, 36 (1): 43-47.
- [17] SIDIROPOULOS A, MANOLOPOULOS Y. A citation-based system to assist prize awarding [J]. Acm Sigmod Record, 2005, 34 (4): 54-60.
- [18] FRAGKIADAKI E, EVANGELIDIS G, SAMARAS N, et al. f-Value: measuring an article's scientific impact [J]. Scientometrics, 2011, 86 (3): 671-686.
- [19] 宋歌. 网络分析方法在引文分析中的整合研究 [J]. 中国图书馆学报, 2011, 37 (4): 106-114.
- [20] 姜磊. 单篇学术论文的评价指标体系研究 [D]. 大连: 大连理工大学, 2014.
- [21] 纪雪梅, 李长玲, 许海云. 基于权力指数的引文网络分析方法探讨 [J]. 图书情报工作, 2009, 53 (24): 111-114.
- [22] 高霞, 官建成. 结构声望视角下的基础研究影响力评价: 以 Hirsch 指数为例 [C]. 全国科学学理论与学科建设暨科学技术学两委联合年会, 2012.
- [23] CHEN C. Predictive effects of structural variation on citation counts [J]. Journal of the American Society for Information Science & Technology, 2012, 63 (3): 431-449.
- [24] 刘春丽. Altmetrics: 从理论假说、术语提出到内涵的重新界定 [J]. 图书情报工作, 2015 (6): 82-89.
- [25] MCNAMARA D, WONG P, CHRISTEN P, et al. Predicting high impact academic papers using citation network features [C] // Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2013: 14-25.
- [26] 苏学. 期刊论文学术水平定量评价指标体系的初步设计 [J]. 情报探索, 2010 (5): 7-9.
- [27] 林德明, 姜磊. 科技论文评价体系研究 [J]. 科学学与科学技术管理, 2012, 33 (10): 11-17.
- [28] CHO S R. New evaluation indexes for articles and authors' academic achievements based on open access resources [J]. Scientometrics, 2008, 77 (1): 91-112.
- 作者简介:** 齐世杰 (ORCID: 0000-0001-8757-7037), 女, 1992 年生, 硕士生。研究方向: 情报理论与方法。
郑军卫 (ORCID: 0000-0001-7390-5757), 男, 1973 年生, 博士, 研究员, 硕士生导师。研究方向: 情报理论与方法。通讯作者。
- 作者贡献声明:** 齐世杰, 文献调研与分析, 撰写论文。
郑军卫, 提出研究命题与思路, 并修订最终版本。
- 录用日期:** 2017-05-09

(上接第 122 页)

- [9] 谢丽星, 周明, 孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取 [J]. 中文信息学报, 2012, 26 (1): 73-83.
- [10] 唐晓波, 向坤. 基于 LDA 模型和微博热度的热点挖掘 [J]. 图书情报工作, 2014, 58 (5): 58-63.
- [11] SAVOY J. Text representation strategies: an example with the state of the union addresses [J]. Journal of the Association for Information Science and Technology, 2016, 67 (8): 1858-1870.
- [12] LEE N, KIM E, KWON O. A real-time combination of text-mining methods for flexible movie recommendation in human robot interaction [C] // International Conference on Platform Technology and Service, 2016: 1-6.
- [13] POSCH L, PANAHIAZAR M, DUMONTIER M, et al. Predicting structured metadata from unstructured metadata [J]. Database the Journal of Biological Databases & Curation, 2016 (7): baw080.
- [14] 寇宛秋, 李芳. 基于种子词汇的话题标签抽取研究 [J]. 中文信息学报, 2013, 27 (5): 114-121.
- [15] 陈兴蜀, 高悦, 江浩, 等. 基于 OLDA 的热点话题演化跟踪模型 [J]. 华南理工大学学报: 自然科学版, 2016, 44 (5): 130-136.
- 作者简介:** 陈斌, 男, 1992 年生, 硕士生。马静, 女, 1966 年生, 博士, 教授。通讯作者。
- 作者贡献声明:** 陈斌, 提出研究思路, 设计研究方案, 采集、清洗、分析数据以及进行实验和论文的起草。马静, 指导研究方向、实验方案设计, 把握文章结构以及最终版本修订。
- 录用日期:** 2017-04-10