



一流大学研究文库
WCU SERIES

大学的新衣？

对基于文献计量学的 科研评价的反思

Bibliometrics and
Research Evaluation

Uses and Abuses

【加】伊夫斯·金格拉斯 (Yves Gingras) 著
刘 莉 董彦邦 王 琪 译校



近十年来，排名 (ranking)、评价 (evaluation)、计量 (metrics)、h 指数 (h-index)、影响因子 (impact factors) 等在高等教育及其研究领域颇为盛行。政府和科研管理人员希望运用定量指标来评价一切，包括普通教师、教授、科研人员、课程和大学。这些术语虽然被频繁使用，但很少有人关注它们的确切含义及其有效性。

上海交通大学出版社
JIAO TONG UNIVERSITY PRESS



一流大学研究文库
WCU SERIES

大学的新衣？

对基于文献计量学的科研评价的反思

Bibliometrics and Research Evaluation
Uses and Abuses

【加】伊夫斯·金格拉斯 (Yves Gingras) 著
刘莉 董彦邦 王琪 译校



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

目前,科研评价正在蓬勃发展。“排名”、“计量”、“h 指数”和“影响因子”在学术界非常流行。政府和科研管理者希望使用定量指标来评价一切,包括对普通教师、教授、课程和大学的评价。文献计量学集出版物和引文于一身,在“卓越研究”的评价工具中占据主导地位。与 17 世纪科研论文首次发表以来使用的同行评价等“主观”和直观评价相比,文献计量指标是一种更有价值的定量测量方法,因此被称为研究水平的“客观”评价标准。伊夫斯·金格拉斯在书中围绕文献计量方法开展了热烈的讨论,探究了大学急于让无效指标影响科研战略的缘故,并在一定的分析维度上提出了建立有效指标的精确标准。

本书将为科研政策制定者、科研管理人员、科研政策研究者以及研究生提供有价值的参考。

图书在版编目(CIP)数据

大学的新衣? ——对基于文献计量学的科研评价的反思 / (加) 伊夫斯·金格拉斯(Yves Gingras)著;刘莉,董彦邦,王琪译校. —上海:上海交通大学出版社, 2019
ISBN 978-7-313-21081-4

I. ①大… II. ①伊… ②刘… ③董… ④王… III. ①文献计量学—研究 IV. ①G250.252

中国版本图书馆 CIP 数据核字(2019)第 050461 号

Les dérives de l'évaluation de la recherche
Du bon usage de la bibliométrie
Yves GINGRAS est l'auteur.
© Editions Raisons d'Agir, 2014

上海市版权局著作权合同登记号:图字 09-2017-514 号

大学的新衣? ——对基于文献计量学的科研评价的反思

著 者: [加] 伊夫斯·金格拉斯(Yves Gingras)	译 校: 刘 莉 董彦邦 王 琪
出版发行: 上海交通大学出版社	地 址: 上海市番禺路 951 号
邮政编码: 200030	电 话: 021-64071208
印 制: 上海天地海设计印刷有限公司	经 销: 全国新华书店
开 本: 787 mm×1092 mm 1/16	印 张: 6.5
字 数: 103 千字	
版 次: 2019 年 5 月第 1 版	印 次: 2019 年 5 月第 1 次印刷
书 号: ISBN 978-7-313-21081-4/G	
定 价: 78.00 元	

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系
联系电话: 021-64366274

近十年来,排名(ranking)、评价(evaluation)、计量(metrics)、h 指数(h-index)、影响因子(impact factors)等词语在高等教育及其研究领域颇为盛行^[1]。政府和科研管理人员希望运用定量指标来评价一切,包括对普通教师、教授、科研人员、课程和大学的评价。这种需求源于 20 世纪 80 年代的“新公共管理”(New Public Management)思想,这一思想使“卓越”、“质量”指标在评价中倍增。这些术语虽然被频繁使用,但很少有人关注它们的确切含义及其有效性^[2]。

在“卓越研究”(research excellence)的评价工具中,文献计量学(bibliometrics)的地位最为显赫。文献计量学作为一个研究领域,通常用于分析一个既定的实体(个人、机构、国家等)发表的科研论文及其被引频次。虽然论文数量提供了一个简单的产出指标(若定义了时间单位,则为生产力指标),但是一篇论文的被引频次往往被认为是对论文质量和科学影响的直观测量。各种数字组合为大学、实验室及科研人员等排名提供了一些指标,这些指标被认为是对研究成果价值的“客观”测量。许多人认为,“客观”测量应该有效地替代 17 世纪以来一直沿用的相对“主观”的同行评价方法。

20 世纪 90 年代,许多科学家发现了文献计量学,并开始普遍使用定量指标来评价科学家的研究活动。同时,科学家们也开始对其应用进行批评,如谴责文献计量学的缺点,揭示简单化指标可能对科学研究动态产生的负面影响等。然而,矛盾的是,科研人员未必了解这些指标的真正含义及其有效性,但他们却往往首先强调自己论文发表期刊的影响因子,并将总被引频次或其 h 指数(后文将进行具体阐述)作为自己的价值体现与成功标志。当成为评价委员会的成员后,这些科研人员往往会毫不犹豫地使用这些相同的指标对其同行进行排名,决定

其是否具有资格获得科研资助。因此,人们不能把20世纪90年代突然爆发的“文献计量热”(bibliometric fever)以及由此引发的令许多大学校长兴奋不已的排名繁荣现象(尽管事实上,正如本书所要讨论的,这些排名没有真正的科学有效性,且很少能够真正测量出想要测量的内容)都归咎于管理人员。

虽然在各种排名和评价中,文献计量数据很少作为唯一的信息,但科学家对文献计量学的发现让人们错误地认为“文献计量学”=“评价”,好像文献计量学在评价活动之外并没有其他的用途。事实上,由于缺乏认真的方法论反思,导致了科研评价中文献计量指标的滥用。尽管科学家在《自然》(*Nature*)、《科学》(*Science*)或者博客中频繁抱怨各种文献计量指标的使用和大学排名,但是到目前为止,大多数批评仅仅强调所谓的文献计量指标的局限性,并想当然地认为它们提供了其他有效指标。批评者很少质疑这些指标的认识论基础:这些指标本身是否有意义?这些指标所测量的是否是它们本应测量的内容?这些指标是否适用于它们应该测量的概念(质量、生产力、影响等),等等。

有关文献计量学在科研评价中的弊端的研究文献非常丰富,甚至略显冗余^[3],这些文献提出的问题往往是在概念与其指标之间界定不清的时候产生的,并没有具体说明指标的有效性和有用性(如果有的话)的标准。评价是不可能完全避免的,关键是理解最常见的文献计量指标的具体属性,并对那些设计不当的指标和使用这些指标可能导致(或已经导致)的意想不到的负面后果进行严厉批评。尽管我们知道这些排名所使用的指标是无效的,而且只有大约15%的国际学生知道这些排名的存在,而这些国际学生中又仅有10%使用大学排名来选择学校^[4]。但是,当得知欧洲大学70%的高等教育管理者承认使用公开的大学排名来决策时,这一任务就更加紧迫了。

本书的主要目的是以简洁的方式阐述文献计量学的基本概念与方法,以证明文献计量学比科研评价的范围广泛得多。科研评价是文献计量学方法相对较新的(和不恰当的)一个应用领域。第一章回顾了文献计量学作为一个研究领域的起源,梳理了文献计量学从图书馆管理(20世纪50、60年代)到科学政策(20世纪70年代)再到科研评价(20世纪80年代)的应用与发展的演变过程。接下来的第二、三、四章则分别讨论了文献计量指标和排名的有效应用、不当应用及滥用情况。

本书首先关注有效应用,第二章对出版物和引文模式的研究,在一定程度上

为分析全球科学随时间推移的动态发展提供了一个独特的工具。一直以来,科学家之间流传着许多关于论文和引文的“神话”,但事实证明,它们缺乏经验基础,只是“神话”而已。第三章证明了科研评价是非常古老的实践活动,伴随 17 世纪中叶科研论文的出现而诞生,又在 20 世纪增加了新的层次,包括对拨款、研究生课程、研究实验室乃至将大学作为一个整体进行评价。这些评价基本上是质性的,并以同行评价委员会的评价为基础,直到 20 世纪 80 年代,文献计量指标才开始被应用于科研评价,以补充定量评价。根据实证主义的“决策”概念,传统的质性方法如今被认为过于主观而无法令人信服,定量评价则被认为比传统质性方法更为客观。

毫无疑问,当越来越多的定量指标被用于高等教育机构与研究者评价时,这些定量指标也越可能被滥用。因此,第三章和第四章分析了界定不清的定量指标的不当应用,即那些使研究动态产生异常的、意料之外的影响的应用,以及那些更糟糕的滥用,如操纵数据,甚至贿赂作者,让其在论文中添加新的署名学校地址,以提升某些机构在世界大学排名中的名次。本书还讨论了文献计量指标的倡导者及批评者很少提及的问题,如为什么所谓的期刊影响因子要保留到小数点后三位?哪些人真正需要了解期刊论文的质量?换言之,本书关注评价的政治性,并试图证明数字的使用是控制科学家并在评价过程中降低其自主性的一种方式。

目前,在蓬勃发展的评价市场中流行着各种各样的文献计量指标。为了更好地进行质量控制,本书最后一章分析了这些指标所用的数据的来源及其局限性,并且在给定的分析范围内提出了建立有效性指标(比如个人、机构或国家)的精确标准。许多大学管理者基于学校的排名去调整他们的实践与政策,反过来,这些实践与政策又是基于对科研质量的不当评价,本书也试图厘清这其中的缘由。总之,我认为,大学急于让无效指标影响其科研战略,事实上是在重演“皇帝的新衣”(The Emperor's New Clothes)这个古老的故事。

本书仅一家之言,并没有对与科研评价相关的诸多方面的计量文献(正如上文所说,冗余和浅显的文献)进行大范围的调查。尾注为读者提供了一些具体文献,但并未对其进行全部引用。对所讨论的问题贡献最大的文献被列为重点文献。对于想要了解更多细节的读者来说,可以通过尾注或者浏览如下期刊,轻松找到合适的(文献)来源,例如《科学计量学》(Scientometrics)、《信息计量学期

刊》(*Journal of Informetrics*)、《信息科学与技术学会会刊》(*Journal of the Association for Information Science and Technology*)、《科学与公共政策》(*Science and Public Policy*)、《科研评价》(*Research Evaluation*)，以及由布莱斯·克罗宁(Blaise Cronin)与凯斯蒂·杉本(Cassidy R. Sugimoto)编写的《超越文献计量学》(*Beyond Bibliometrics*)一书^[5]。

我的同事米歇尔·齐特(Michel Zitt)对本书的法文版进行了长篇评论，他引用了50多篇专业论文，拓展了我在书中讨论的一些问题，使相关内容在本书中得到修订与更新^[6]。本书使我有机会能够清晰地阐明在先前出版的法文版本中尚需阐释的内容。本书面向的读者是科研人员与科研管理者，以及其在工作中与文献计量学打交道的高等教育参与者。我希望本书的简洁风格与基调使其更具可读性。不过，经验丰富的文献计量学专家也许不能在本书中获取更多技术层面的知识(可能除了我提出的用来检测指标有效性的标准)。

齐特在总结其长篇评论时认为我提供了“坚定的愿景，但缺乏细微差别”(committed vision, albeit one without nuances)^[7]。读者会从自身角度判断我所提供的“细微差别”的程度。但是我认为，使读者沉迷于技术内容会使他们错过书中传达的重要信息和问题：在过去50年中，有数以千计的专业论文来讨论文献计量学的基本概念和指标，然而时至今日，为什么科研人员、大学校长、行政人员和管理者仍然对此十分困惑？可能的一种解释是，太多的文献计量学家只关注于计算他们所找到的测量单位(引文、推特、评论、网络连接)的难点，而没有先思考“这些测量的意义是什么”。一位研究者分析了经济合作与发展组织(Organisation for Economic Co-operation and Development, OECD)开展的“国际学生读写能力评价测试项目”(Programme for International Student Assessment, PISA)识字测试后指出：测量结果是答案，但问题是什么呢？^[8]

因此，我希望本书不是在技术层面为某一个具体指标的完善方法而展开讨论，而是通过聚焦评价的基本问题及其对科研的影响，为文献计量评价的使用者与被评价者提供一些概念性工具，帮助他们更合理地评价目前在新兴的科研评价市场上盛行的排名“黑匣子”里的许多指标。

致 谢

本书法文版最初源于 2011 年 6 月我受邀在巴黎法国农业科学院(Institut National de la Recherche Agronomique, INRA)的一场文献计量学演讲。在这次演讲活动中,我注意到与科研评价相关的文献计量学的争论层出不穷且错误百出,这种情况让我确信,是时候对这些争论进行梳理了。我参与过很多文献计量学的评价工作,并且在过去 20 年里也在很多会议上探讨过,所以我认为全面梳理科研评价中人们对文献计量学有效应用、不当应用甚至滥用的情况是非常有价值的^[9]。非常感谢评审本书书稿并提出有益建议的同事们:首先感谢我长期的亲密合作伙伴文森特·拉里维埃(Vincent Larivière)、杰里米·布迪尔(Jérôme Bourdieu)、米歇尔·格罗塞蒂(Michel Grossetti)、约翰·海尔布朗(Johan Heilbron)、卡米尔·里蒙(Camille Limoges)、弗兰克·波普(Franck Poupeau)、维克多·斯托维斯基(Wiktor Stoczkowski)和让-菲利普·华伦(Jean-Philippe Warren)。同时,也感谢让-菲利普·罗比泰尔(Jean-Philippe Robitaille)和皮埃尔·蒙爱昂(Pierre Mongeon)采集数据;感谢宝琳·赫特(Pauline Huet)对尾注的格式进行调整;感谢尤金·加菲尔德(Eugene Garfield)授权本书使用图 2.1 和 2.2;感谢凯蒂·波尔纳(Katy Börner)与凯文·波亚克(Kevin Boyack)授权本书使用图 2.13。当然,我本人会为书中仍可能存在的错误或不当表述承担全部责任。我亲自翻译、修改和更新了本书的英文版本,并在这个过程中参考了法文版的评审者以及麻省理工学院出版社选择的两位匿名读者的评论和建议。还要特别感谢让-弗朗西斯·布兰切特(Jean-François Blanchette),他认为本书的英文版很有意义,并且联系了麻省理工学院出版社;感谢我的同事彼得·柯亨(Peter Keating),他认真地阅读了书稿的最终版;

感谢吉塔·得维·玛纳卡特拉(Gita Devi Manaktala),感谢她在英文版从翻译到出版的过程中出色而高效的支持。最后感谢莫妮卡·杜蒙特(Monique Dumont)编写索引。

目 录

第一章 文献计量学的起源	...1
第二章 文献计量学与科学动态	...10
第三章 科研评价的繁荣发展	...30
第四章 对科研评价的评价	...50
结论 大学的新衣	...72
注释与参考文献	...74
索引	...93

第一章

文献计量学的起源

在开始探讨文献计量学之前,我们首先厘清这一领域的几个术语。科学计量学(*scientometrics*)是对所有学科各种科学活动的测量,这一术语的提出主要归功于物理学家瓦西里·瓦西里耶维奇·纳利莫夫(Vassily Vassilievich Nalimov)(他在1969年出版的一本著作的标题中使用俄语“*naukometriya*”表达了这个概念)^[1]。科学计量学使用的数据包括投入的研发(Research and Development, R&D)经费、参与研发工作的科研人员以及论文、专利等成果的数量。文献计量学(*bibliometrics*)作为一个专业术语,由艾伦·普里查德(Alan Pritchard)在1969年提出,是科学计量学的一个子概念,而且它仅限于对出版物及其属性的分析^[2]。当涉及专利分析的时候,经常使用的是技术计量学(*technometrics*)这一术语。

虽然开展科研评价时,我们倾向于考虑论文,但是也应该考虑其他类型的出版物,至少原则上应该包括图书、博士论文以及所谓“灰色文献”的研究报告等。对这些不同类型出版物或多或少的复杂性分析基本取决于其是否以数据库形式存在。在计算机能够自动处理大量数据之前,文献计量数据一直都是依靠人工收集来完成,收集工作仅包括数量相对较少的出版物(约几十种),且内容与化学、社会学等具体学科相关。随着互联网的出现,网络计量学(*webometrics*)这一术语被提出,它包括对电子出版物的分析和科学出版物的使用。尽管存在这些形式上的区别,但由于对科研产出(无论文献是以电子版还是纸质版的形式)的分析已经占据科学动态研究的中心位置,科学计量学和文献计量学已经快速发展成为可以互换的概念^[3]。

作为期刊收集与管理工具的引文分析

传统文献计量学的诞生通常与 20 世纪 20 年代中期美国统计学家阿尔弗雷德·洛特卡(Alfred Lotka)发表的一篇论文有关,这篇论文阐述了科研人员科学生产力的分布,如今已经成为经典文献。在这篇开创性的论文中,洛特卡建立了以他的名字命名的“洛特卡定律”,即发表 N 篇论文的作者数量 P 与论文数量 N 的平方成反比^[4]。然而,正如贝诺蒂·戈丹(Benoît Godin)所言,事实上,在 20 世纪早期,心理学家们率先在其研究领域分析出版物的数量变化,当时他们已经开展了文献计量学研究,只是没有使用文献计量学这一术语^[5]。为了追溯所属学科的历史,其他领域的科学家也分析了出版物数量随着时间变化的情况^[6]。但与洛特卡的研究不同,这些分析的目标都不是为了寻求一般定律,而仅仅是为了跟踪他们自身专业的发展。

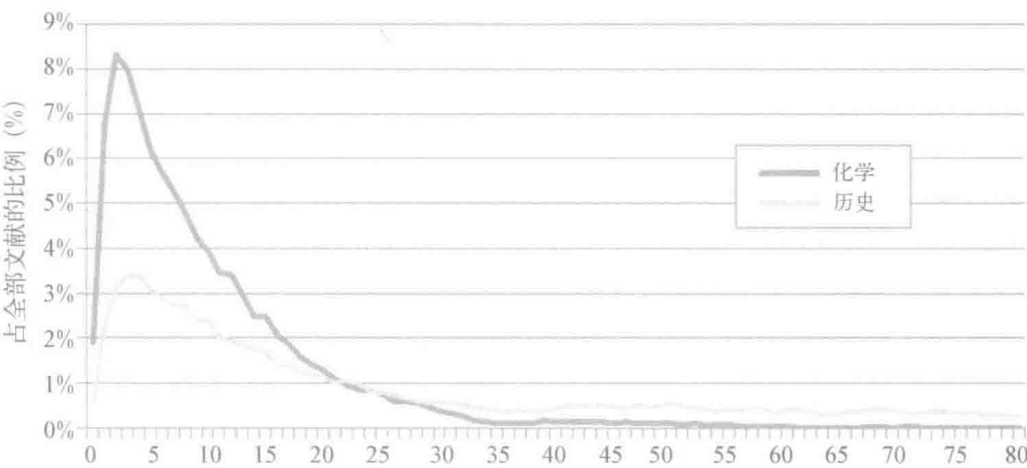
尽管有过一些先例,但是随着 20 世纪 20 年代和 30 年代图书馆期刊馆藏管理的发展,文献计量学研究开始变得更加系统化。随着学术期刊数量和成本的增加,图书馆员开始使用客观方法来选择对科研人员最有用的期刊。在期刊馆藏管理的这种背景下,引文分析开始兴起。图书馆员不仅计算出版物的数量,而且开始查看它们所包含的参考文献(即引文)以衡量论文和期刊对科学家的实际有用程度。通过分析在某一特定学科内或某一特定时期内被引用最多的期刊(起初研究最多的领域是化学),图书馆员能够把那些仍然对研究有用的期刊与那些因极少被引用而被认为过时的期刊区分开来,从而淘汰那些过时的期刊,为有利于解决最新研究问题的期刊腾出空间。

1927 年发表在《科学》上的一项研究表明,与其基于科研人员主观兴趣来选择期刊,不如找到更宽泛、客观的衡量标准来确定这些期刊的实际效用。两位化学家通过分析 1926 年美国化学期刊的主要内容,发现尽管有 247 本不同期刊被引用,但是大部分(超过 130 本)期刊的被引频次小于 3^[7]。随后,另一位化学家强调了德文期刊的流行,指出年轻化学家掌握德语的重要性^[8]。第二次世界大战(以下称二战)带来了很多变化,其中标志性的变化是,美国化学家对德文期刊引用量的急剧下降,这已被 20 世纪 50 年代中期一项对化学工程期刊引用的研究所证明^[9]。

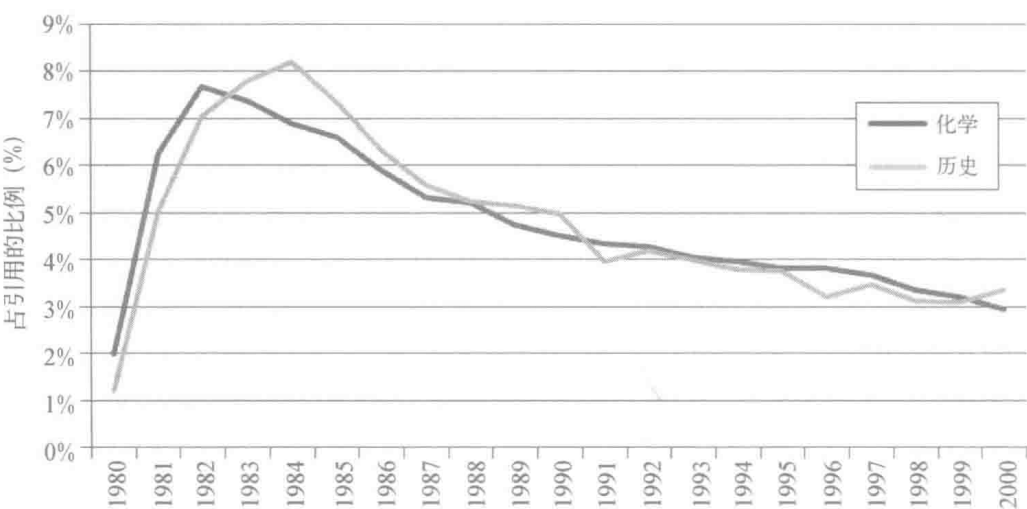
人文与社会科学领域也效仿科学领域的这种做法,基于有限的样本对参考

文献进行类似的描述性分析^[10]。早在 20 世纪 50 年代中期,尤金·加菲尔德(Eugene Garfield)提出“引文索引”(citation index)的很多年前,科学家就已经凭直觉认识到,论文中包含的引文在科学家的研究实践和科学的社会动态方面可以提供有用的信息。事实上,他们已经发现德文期刊被大量引用。由此可以得出两个结论:一是学生必须学习德语;二是一些期刊花费了图书馆宝贵的资金(这些资金本应被用于更好的投资),但是多年来却未被引用。这些结论都是难以通过其他方法得出的。

对不同学科、专业发展的文献计量学分析,导致了这个新的研究领域特有概念的诞生。一篇代表性论文会随着时间的推移而被使用(和被引用)得越来越少,这类似于元素的放射性衰变。为了使这一理念规范化,20 世纪 60 年代初,出现了科学文献半衰期(a half-life)的概念,即通过对论文和期刊中参考文献的年份分布进行分析,发现科学文献半衰期遵循指数衰减曲线规律。以此类推人们可以计算覆盖总参考文献数 50% 的年份来定义半衰期。这个简单的指标为科学文献在某一领域的效用寿命提供了一种测量标准。例如,人们对学科进行比较发现,历史学科比化学学科更经常地引用 20 年前的文献(图 1.1a)。早期研究还表明,在数学学科中,有一半被引用的效用寿命低于 10 年,而在物理学科中,半衰期更短,只有 5 年。这个发现揭示了数学领域的文献通常比物理领域的效用寿命更长^[11]。同样,基于前瞻性的导向,人们可以将引文半衰期(the half-life of citations)定义为达到引文累计总数的一半所需的时间(图 1.1b)。



(a) 发表于 1980 年的化学和历史学科论文中的参考文献的典型时间分布



(b) 化学和历史学科论文自 1980 年发表后被引用的典型时间演变(所有引文经过标准化处理)

图 1.1

科学引文索引：推动文献检索的发展

二战结束后，论文发表数量呈指数型增长。在这样的情况下，科研人员即使在自己的专业领域内也无法追踪新的研究成果。私人企业或学术团体的近期出版物摘要也已经不足以完成此项任务。1946 年，伦敦皇家学会(The Royal Society of London)组织了两次科学信息国际会议，目的是改进现有的科学文献收集、索引和分布方法，并扩展现有的文摘服务^[12]。20 世纪 40 年代末，如何用某种方式对引文进行索引以帮助科学家们更轻松地找到他们所需要的文献，这一问题日益凸显。

在这种背景下，1955 年尤金·加菲尔德提出了一项计划——将科学文献中引用的所有论文都编入索引。该计划所采用的模型以谢巴德援引法(Shepard's Citations)为基础。谢巴德援引法是将所有公开发表的美国法律信息收集起来做成索引，使法律工作者很容易地辨别、提取并援引自己想要的法律信息，也因此创造了法理学。加菲尔德建议弗兰克·谢巴德公司(the Frank Shepard Company)(为律师提供了不可或缺的索引的公司)副总裁威廉·C·阿黛尔(William C. Adair)在《美国文献工作》(*American Documentation*)上发表一篇

论文来描述这一法律工具的运行方式以及它如何适用于科学文献^[13]。这篇论文发表后的几个月里,加菲尔德在《科学》上阐述了该计划的细节,即建立一个科学文献引文电子数据库^[14],目的是通过使用被引论文来帮助确定那些研究同一主题的论文,从而使文献检索更加便利。基本的、潜在的直觉是,对一篇论文的引用表明引文和被引论文之间存在一种概念上的联系。快速找到并引用一篇感兴趣的新论文,就会建立一个相关的参考目录,这似乎是很合理的。例如,阿尔伯特·爱因斯坦(Albert Einstein)1905年发表的关于相对论的论文,该论文确实关注相对论这一主题,而与化学无关。类似地,通过寻找引用自己研究成果的论文,不仅可以找到同一主题的论文,还能找到从事同一主题研究的其他作者,这可能不是仅仅依靠泛读常规的科学期刊能实现的。

加菲尔德深信该计划的实用性及其盈利能力,并于1959年在费城建立了科学信息研究所(Institute for Scientific Information, ISI)。1961年,加菲尔德从美国国家科学基金会(National Science Foundation, NSF)和国立卫生研究院(National Institutes of Health, NIH)获得了30万美元资助,用于研究自动化引文索引(automated citation index)的可行性。1963年,ISI发布了《科学引文索引》(Science Citation Index, SCI),并在接下来的几年里持续扩大其覆盖范围。美国国家科学基金会和国立卫生研究院对该计划的参与表明,在信息呈指数型增长的背景下,科学组织高度重视科学文献的获取问题^[15]。

直到20世纪80年代,SCI才成为评价科学出版物影响力的主要工具。但SCI是在科学文献管理背景下产生的,与科研评价并没有直接联系。20世纪60年代,美国国家科学基金会、国立卫生研究院等组织还没有将科研评价提上议程,更不用说大学或政府了。当时,SCI主要是一种文献检索工具,使科学家们能够在大量出版物中找到他们感兴趣的某一主题的出版物。即使在今天,科睿唯安(Clarivate Analytics)的信息检索平台Web of Science(简称WoS)上的Web of ScienceTM核心合集数据库及其竞争对手爱思唯尔(Elsevier)的Scopus数据库仍提供独特的文献研究工具,可以在特定的主题上快速构建文献目录。值得注意的是,虽然科研人员一直称它们为“索引”,但在20世纪60年代到80年代这一时期,该产品实际上是一本书,按字母顺序排列被引作者的姓名(参见第二章图2.1)。此后,信息技术将这一索引转化为一个电子数据库,包含了可以在互联网上检索到论文的参考文献,但“引文索引”这一术语仍被广泛使用。

文献计量学在科技政策中的应用

英国科学史家德里克·索拉·普赖斯(Derek de Solla Price)赋予了科学计量学第一个理论基础。作为一位经验丰富的物理学家,普赖斯提出将科学作为一种集体现象来分析,应去追踪所有科学家及其论文发表的变化,而不是仅仅关注像爱因斯坦这样极少数的伟大科学家。普赖斯希望能够对科学发展进行定量分析的基础上,创造出所谓的“科学学”(science of science)。1950年,他发表的一项研究表明,大约自1700年以来,科学作为一种系统已经呈指数型增长,每隔15年,期刊和出版物的数量就会翻一番^[16]。尽管精确的增长率会因所使用的方法和数据而有所不同,但是科学的指数型增长已经得到了一些研究的证实^[17]。普赖斯也是第一个使用新SCI数据库进行社会学而不是文献学分析的人。他在1965年发表的一篇有关引文网络的论文中,强调了引文的分布非常不均匀,这种分布遵循一种类似于洛特卡定律(描述科研人员生产力分布)的幂律^[18]。

尽管有图书馆员在引文分析方面所做的工作及普赖斯对科学学的推动,但文献计量学研究仍长期局限于由科学家、图书馆员、社会学家和历史学家组成的小群体中。这些学者都从自身出发研究科学出版物及其所包含的参考文献(即引文)的属性。随着科学政策时代的到来,文献计量学作为学术研究领域在20世纪70年代才真正发展起来。

回想一下,就能理解为什么文献计量学的真正发展会出现在20世纪70年代。1960年下半年,许多国家对科学政策和规划表现出新的兴趣,这成为科学组织的一个转折点。就像20世纪30年代的经济危机和二战后对经济和社会发展指标的制订一样,国际组织对促进研发所需的研究和创新指标进行了思考^[19]。经济合作与发展组织(Organisation for Economic Co-operation and Development,以下简称“经合组织”)因此启动了一系列关于国家科学政策的研究。1962年,经合组织出版了第一版《弗拉斯卡蒂手册》(Frascati Manual),为评价研发活动提供了方便的标准定义^[20]。1965年,普赖斯在《自然》上发表了一篇关于“科学政策的科学基础”(the scientific foundations of science policy)的论文^[21]。

正是在这样的背景下,美国国家科学基金会(如前所述,该基金会帮助加菲尔德创建了SCI)被美国国会授权制订一系列指标来监测和评价美国科学

技术的地位,推动了1972年双年刊《科学指标》(*Science Indicators*)第一卷的出版。1987年,《科学指标》更名为《科学与工程指标》(*Science and Engineering Indicators*),其内容也开始适应政策的变化。在20世纪70年代初,美国国家科学基金会要求弗朗西斯·纳林(Francis Narin)及其公司 Computer Horizons 研究了使用文献计量达到评价目标的可能性。纳林撰写了一份长篇报告,为“可评价的文献计量学”(evaluative bibliometrics)的发展奠定了基础^[22]。

虽然经合组织主要以科学发展的经济愿景为驱动力,但它仍然主要关注投入指标与专利指标(专利指标被用作衡量创新的标准)。美国国家科学基金会更接近科学共同体,寻找出版物及其影响力方面的指标,而这类数据恰恰是SCI所能提供的。因此,除了更常用的科学人力资源和研发投入水平方面有用的数据以外,出版物和引文数据首次被纳入一套科学和创新指标中。20世纪80年代中期开始,美国以外的其他国家也开始使用这些文献计量数据来监测其科学发展水平^[23]。

政府需要能够衡量科学和技术发展水平的指标,从而能够为国家科学和技术政策规划提供必要的信息,这刺激了以前比较分散的研究领域的发展,并促成了1971年《科研政策》(*Research Policy*)期刊的创立。两年后《科学与公共政策》(*Science and Public Policy*)期刊也随之创立。这些期刊尤其致力于分析各国科技发展的社会与经济方面的影响因素。在1974年举行的一次研讨会上,文献计量学的先驱们就“走向科学计量:科学指标的出现”这一主题^[24]进行了讨论。为了给科学指标的发展建立一个连贯模型的目标,来自不同领域的专家从历史学、社会学、经济学等方面对科学计量进行了分析。在此次研讨会上,物理学教授和科学史家杰拉德·霍尔顿(Gerald Holton)提出“科学是否真的能被测量?”这一问题,而斯蒂芬·科尔(Stephen Cole)和乔纳森·科尔(Jonathan Cole)兄弟及其他社会学家对科学领域学科的认知方面进行了讨论。加菲尔德及其同事们则表示,诸如文献耦合(bibliographic coupling)以及共被引分析(co-citation analysis)等量化方法可以作为绘制科学学科和专业的概念结构及其发展路径的工具^[25]。

下一章将讨论这些文献计量方法和指标的不同用途。就目前而言,值得注意的是,文献计量研究的发展推动了科学学科通用标志(期刊、学会和学术团体)的诞生。第一本专门讨论文献计量的期刊是创建于1978年的《科学计量学》。10年后(1987年),一年一度的国际会议召开。1993年“国际科学计量学和信息