

普莱斯和他的定律及模型

陈关荣

(香港城市大学)

【一】

当今的学者们对科学引文索引 SCI (Science Citation Index) 爱恨交加。它是由美国情报学家和科学计量学家尤金·加菲尔德 (Eugene E. Garfield, 1925-2017) 于 1955 年引进的。当年他在《科学》杂志上发表了题为“为科学设立引文索引指标” (Citation indexes for science) 的文章，第一次提出了引文索引 (SCI) 的构想，旨在提供一种文献计量工具来帮助科学家寻找和识别感兴趣的专业文献。加菲尔德于 1960 年创办了美国科学信息研究所 ISI (Institute for Scientific Information)，1963 年建立了全球 SCI 数据库，1964 年起开始公布 SCI 年度数据，随后在 1973 年开始公布社会科学引文索引 SSCI 年度数据和在 1978 年开始公布艺术与人文引文索引 A&HCI 年度数据。他生前担任汤森路透科技集团 (Thomson Reuters Corporation) 的终身名誉董事长。

国际科学计量学和信息计量学学会 (International Society for Scientometrics and Infometrics) 于 1984 年设立了“德里克·德·索拉·普莱斯纪念奖” (Derek de Solla Price Memorial Medal) 并把第一个奖章授予加菲尔德，表彰他对情报学和计量学做出的卓越贡献。

次年，加菲尔德在《科学计量学》 (Scientometrics) 杂志为纪念这位普莱斯 (1922 年 1 月 22 日-1983 年 9 月 3 日) 而写的题为“向德里克·约翰·德·索拉·普莱斯致敬”的文章中说：“只要我们生活和工作在这正在成长的领域中，我们不仅会怀念德里克，还会每天都想到他的影响。从这个意义上说，他已经成为永恒。”

在这篇致敬文章中，加菲尔德评论道：“‘科学论文网络’可能是普莱斯最具影响力的信息科学领域的论文。它试图描绘世界科学文献网络，并证明了世界论文引文的模式，为科学前沿研究提供了参数。最初让普莱斯成为世界上最引人注目的科学家之一的工作是他的著作‘巴比伦以来的科学’。1960 年，他被任命为耶鲁大学新创立的科学史系主任。他关于科学史的转折点作了五次讲座，定义了现代科学的基本特征。这个就职系列演说的内容后来以‘巴比伦以来的科学’一书出版。”

这段话刻画了普莱斯的两个学术背景：他是一位科学计量学专家，也是一位科学史专家。



图 1 科学计量学和科学史专家普莱斯

【二】

普莱斯出生于英国伦敦东郊的 Leyton 镇，父亲菲利普·普莱斯（Philip Price）是个裁缝，母亲范妮·德·索拉（Fanny de Solla）是名歌手。

1938 年，普莱斯中学毕业后到西南 Essex 技术学院物理实验室从事了一段时间的助理工作，之后进入伦敦大学修读物理和数学。在伦敦大学，他于 1942 年获学士学位、1946 年获实验物理博士学位。在研究生阶段，他发表了几篇论文和一项光学高温计专利。随后，他获得美国联邦基金会的助研金前往美国，在匹兹堡和普林斯顿工作了一年，于 1947 年返回英国。同年，他在哥本哈根与 Ellen Hjorth（1925-1995）结了婚，两人后来育有两子一女。

1948 年，普莱斯到了新加坡 University of Malaya 的 Raffles College 工作，担任应用数学教师。在那里，有两件事对他后来的学术发展产生了深远影响。一是他遇到了英国剑桥大学的海军历史学家诺斯科特·帕金森（C. Northcote Parkinson, 1909-1993）。帕金森激发了普莱斯对历史的兴趣和热情。二是普莱斯在大学图书馆里负责《英国皇家学会哲学汇刊》的整理工作。当他按 1665 至 1850 年的时间顺序排列会刊时，注意到了它们按年份堆砌起来的高度随着时间呈指数式增长。普莱斯后来回忆道：“我把它按时间顺序整齐地堆放在卧室内的墙边，.....我注意到它们的高度形成一条漂亮的指数曲线.....”。他于是产生了科学发展可能也是指数式增长的想法并在多年以后设计了一个数学模型来刻画这种增长规律。

普莱斯在新加坡工作三年后便返回英国。他决定改变自己的研究方向，来到剑桥大学修读科学史博士学位。在剑桥大学 Peterhouse 图书馆，他意外地发现了一篇用中古英语写成的手稿“行星赤道”（The Equatorie of the Planetis），描述了一种天文用的行星定位仪（Equatorium）。于是他以之为题材，在 1954 年完成了博士学位论文，并在翌年出版了一本关于这个行星定位仪的科学史专著。普莱斯以为该行星赤道手稿是属于“英国文学之父”杰弗里·乔叟（Geoffrey Chaucer, 约前 1400-1340）的，因为乔叟还

写过另一篇天文星盘论的文章。但经后人考证，该手稿其实是英国古天文学家约翰·韦斯特维克（John Westwyk，约前 1400-1350）留下的。

在剑桥攻读博士学位期间，普莱斯遇见了中国科学史学家李约瑟（Noel Joseph T. M. Needham, 1900-1995）。由于普莱斯在行星赤道方面的研究工作，李约瑟邀请他参加了一个关于中世纪中国天文钟的调研项目。这次合作让李约瑟，王铃和普莱斯三人联名在 1960 年出版了《天钟》（Heavenly Clockwork）一书，其主要内容后来收录在李约瑟的巨著《中国科学技术史》中。王铃（1917-1994）是江苏南通人，当年在剑桥留学历史专业，是李约瑟早期中国科技史研究和著述的主要助手。1968 年，普莱斯还发表了一篇关于希腊雅典水力钟的文章。

普莱斯获得第二个博士学位后，于 1957 年移居到美国。他开始时在华盛顿史密森学会（Smithsonian Institution）当科学史顾问，参与协助建立美国国家历史和技术博物馆。他后来到普林斯顿高等研究院（IAS）工作至 1959 年，然后到耶鲁大学历史系任职教授，1960 年担任系主任及大学 Peabody 自然历史博物馆馆长。1962 年，普莱斯晋升为 Avalon 讲座教授。他在耶鲁大学工作直到 1983 年在伦敦朋友家因突发心脏病去世，享年仅 61 岁。



图 2 普莱斯在剑桥大学（1954）

【三】

普莱斯在 1950 年代开始研究神秘的“安提基特拉机械装置”（Antikythera mechanism）。这台机器是 1900 年由潜水员从希腊克里特（Crete）岛西北部的安提基特拉海域一艘沉船的残骸中打捞出来的，后来一直收藏在雅典的国家考古博物馆里。它结构复杂，功能未知，许多研究人员通过各种技术手段分析过它，结论众说纷纭。普莱斯在 1959 年和 1974 年先后发表过两篇关于该机械装置的论文，指出它是一台关于行星轨道的计算机，制造时间可追溯到公元前 65-80 年之间。1959 年，普莱斯在《科学美国人》发表了一篇封面文章“一台古希腊计算机”（An Ancient Greek Computer），详尽地描述了并尝试初步还原出这台复杂的机器。他写道：这台机器“在海底下两千年被腐蚀得摇摇欲坠了，但它的表盘、齿轮和铭牌依然向历史学家展

示着神秘的疑团。……它或许会改变我们对古希腊科学的诸多猜测”。他指出：“据我们所知关于希腊化时代（Hellenistic age）的科学技术，便应该认为这样的设备是不可能存在的”。他认为，那是历史上第一个“High-Tech”例子。1974年，普莱斯出版了论文及专著“来自希腊人的齿轮：安提基特拉机械装置——公元前80年的日历计算机”，把它和牛津大学博物馆里1221-1222年出土的伊斯兰齿轮日历装置作了比较。普莱斯始终借用“计算机”这一称谓，意指该机器就像一台计算机，使用旋转部件的连续运动来模拟特定的天文和历法中时间周期关系。在这之前，1961年普莱斯还出版了一本极具影响的科学史书《巴比伦以来的科学》。他的科学史研究指出，巴比伦人是第一代“程序员”，他们的天文数学读起来就像今天计算机程序的打印输出一样，折射出巴比伦古代技术的复杂性和先进性。



图3 普莱斯和安提基特拉机械装置模型（1982）

【四】

1963年，普莱斯出版了一本有口皆碑的计量科学科普著作《小科学，大科学》（Little Science, Big Science）。这本书一直是他所有论著中被引次数最高的。

1965年是普莱斯学术创新成果最为丰富也最备受学术界关注的一年。从普莱斯的学术论文被引数分布图可以看出，他用自己的实践印证了在《小科学，大科学》里说的一句话：“任何一个年轻的科学家，在回顾他正常人生中从现在开始到职业生涯结束的经历时，将会发现他所完成的全部科学工作的80%到90%都是在他眼前发生的，而只有10%到20%在这个时段之前发生。”

Derek J. de Solla Price's Published Works

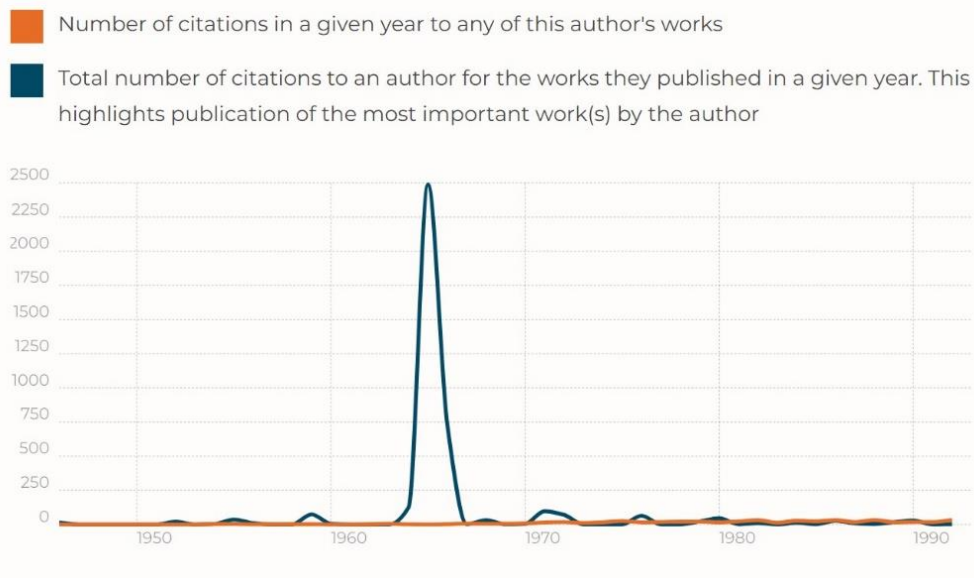


图4 普莱斯学术论文被引数分布图（高峰在1965年）

1965年，普莱斯在伦敦皇家学会作了一个题为“科学政策的科学基础”的讲座，阐述了他关于“科学学”（Science of Science）的理念。同年，该报告在《自然》杂志以同样标题正式发表。

普莱斯的科学学和计量科学的基本思想来自于后来以他命名的平方根定律，即“普莱斯定律”。该定律说：每份工作的一半是由参与工作的总人数的平方根数目的人去完成的。比如有25位作者发表了100篇论文，则其中50篇论文是由5位作者参与撰写的，这个5就是25的平方根。这种比例也可以用来反映其它一些社会现象。

当然，普莱斯并非研究这种作者人数与其出版文献数量之间关系的第一人。统计学家阿尔弗雷德·洛特卡（Alfred J. Lotka, 1880-1949）很早就研究过这类问题并提出了一个“洛特卡定律”：发表 n 篇论文的作者数约等于 n 的倒平方数。也就是说，发表 n 篇论文的作者总数，是发表1篇论文作者总数的 n 平方分之一。例如写2篇论文的作者数量约为写1篇论文作者数量的 $1/4$ ，写3篇论文作者数量约为写一篇论文作者数量的 $1/9$ ，如此类推。还可以估算出，只发表一篇论文的作者总数约占全部作者总数的60.79%。

常言道，没有最早，只有更早。哲学家让-雅克·卢梭（Jean-Jacques Rousseau, 1712-1778）在一个世纪之前就已经指出了：任何产品集合 n 中，品质优良的产品数量约等于 n 的平方根。

现代许多学术工作和文献的计算都基于洛特卡定律和普莱斯定律。这两个定律与意大利工程师和经济学家维尔弗雷多·帕累托（Vilfredo F. D. Pareto, 1848-1923）提出的“80-20法则”类似。该法则说：通常80%的贡献是由20%的人做出的。洛特卡定律和普莱斯定律也被比喻为马太（也称马修）效应（Matthew Effect），其典出自《马太福音》。

音》：“凡有的，还要加给他，让他有余；凡没有的，连他拥有的也要夺去”。这很生动地描述了今天我们常说的“富者越富，穷者越穷”的经济规律和社会现象。

1934年，英国数学家塞缪尔·布拉德福德（Samuel C. Bradford, 1878-1948）描述了一种模式，称为“布拉德福德定律”，用来估计在科学期刊中搜索参考文献的回报呈指数式递减。比如一个领域的期刊按文章数量分为三组，每组大约占有所有文章总数的三分之一，那么每组的期刊数量将以 $1:n:n^2$ 成比例，即少量的杂志发表大量的文章而大量的杂志只发表很少的文章。

1949年，哈佛大学语言学家乔治·齐夫（George K. Zipf, 1902-1950）也发表了一个基于大量数据的统计实验定律，称为“齐夫定律”（Zipf Law）：在自然语言的语料库里，一个单词出现的频率与它在频率表里的排名成反比。具体地说，频率最高的单词出现的频率大约是出现频率第二位的单词的2倍，而出现频率第二位的单词则是出现频率第四位的单词的2倍，如此类推，于是全部结果服从一个幂律分布。

事实上，类似的计量研究报告还能从古老的文献堆里找到，例如有英国统计学家乔治·尤尔（George U. Yule, 1871-1951）的离散幂律分布。也就是说，类似的幂律分布规律随处可见——至少前面提到的学者们认为是这样。这种普遍的社会现象受到了广泛的关注但也引来许多争议。1968年，美国社会学家罗伯特·莫顿（Robert K. Merton, 1910-2003）用马太效应去描述这类社会现象时指出：“相对于那些不知名的研究者，声名显赫的科学家通常得到更多的声望，即使他们的成就是相似的。同样地，在一个项目里，声誉通常都被给予那些已经出名的研究者。例如，一个奖项几乎总是授予最资深的研究者，即使全部工作都是由一个研究生完成的。”

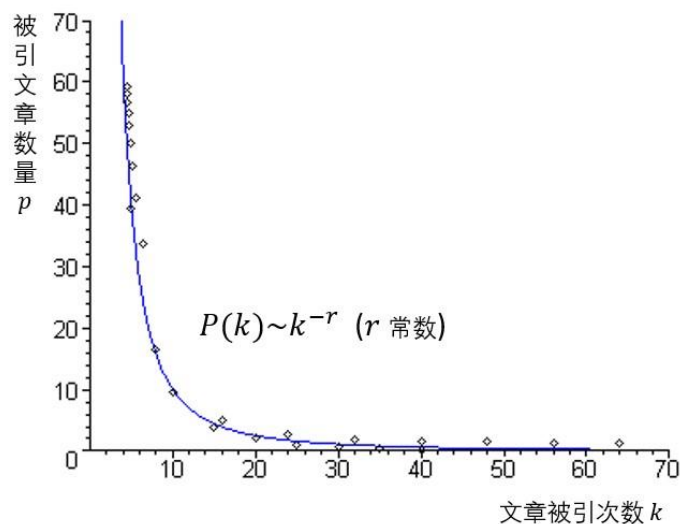


图5 幂律分布示例

普莱斯对科学论文之间的引文网络进行过许多定量研究。他1965年发表在《科学》杂志的文章“科学论文网络”第一次指出引文网络的入度和出度均符合幂律分布。这篇论文为今天周知的“无标度网络”（Scale-free Network）提供了一个早期发现的具体例子。

Networks of Scientific Papers

The pattern of bibliographic references indicates the nature of the scientific research front.

Derek J. de Solla Price

This article is an attempt to describe in the broadest outline the nature of the total world network of scientific papers. We shall try to picture the network which is obtained by linking each published paper to the other papers directly associated with it. To do this, let us consider that special relationship which is given by the citation of one paper by another in its footnotes or bibliography. I should make it clear, however, that this broad picture tells us something about the papers themselves as well as something about the practice of citation. It seems likely that many of the conclusions we shall reach about the network of papers would still be essentially true even if citation became much more or much less frequent, and even if we considered links obtained by subject indexing rather than by citation. It happens, however, that we now have available ma-

chine-handled citation studies, of large and representative portions of literature, which are much more tractable for such analysis than any topical indexing known to me. It is from such studies, by Garfield (1, 2), Kessler (3), Tukey (4), Osgood (5), and others, that I have taken the source data of this study.

Incidence of References

First, let me say something of the incidence of references in papers in serial publications. On the average, there are about 15 references per paper and, of these, about 12 are to other serial publications rather than to books, theses, reports, and unpublished work. The average, of course, gives us only part of the picture. The distribution (see Fig. 1) is such that about 10

percent of the papers contain no references at all; this notwithstanding, 50 percent of the references come from the 85 percent of the papers that are of the "normal" research type and contain 25 or fewer references apiece. The distribution here is fairly flat; indeed about 5 percent of the papers fall in each of the categories of 3, 4, 5, 6, 7, 8, 9, and 10 references each. At the other end of the scale, there are review-type papers with many references each. About 25 percent of all references come from the 5 percent (of all papers) that contain 45 or more references each and average 75 to a paper, while 12 percent of the references come from the "fattest" category—the 1 percent (of all papers) that have 84 or more references each and average about 170 to a paper. It is interesting to note that the number of papers with n references falls off in this "fattest" category as $1/n^2$, up to many hundreds per paper.

These references, of course, cover the entire previous body of literature. We can calculate roughly that, since the body of world literature has been growing exponentially for a few centuries (6), and probably will continue at its present rate of growth of about 7 percent per annum, there will be about 7 new papers each year for every 100 previously published papers in a given

The author is Avalon Professor of the History of Science, Yale University, New Haven, Connecticut. This article is based on a paper presented 17 March 1964 at the National Bureau of Standards, Washington, D.C., in a Symposium on Statistical Methods for Mechanized Documentation. Part of this research was supported by grant GN-299 from the National Science Foundation.

图6 普莱斯的《科学》杂志论文（1965）

说到无标度网络，就不能不提及“普莱斯模型”（Price Model）。1976年，普莱斯在《美国信息学会杂志》上发表论文“文献计量学和其它累积优势过程的一般理论”，获得了该杂志该年度最佳论文奖。受“通才人物”司马贺（赫伯特·西蒙，Herbert A. Simon, 1916-2001）幂律分布随机模型的启发，普莱斯在这篇论文中引进了一个具体的网络数学模型，用来描述引文网络增长的过程和度分布生成的规律。该网络模型基于一个文献数目的“增长”（Growth）过程以及一个具有累计优势即今天称为“偏好性连接”（Preferential Attachment）的文献引用机制。普莱斯从数学上证明了他新模型的节点度累积分布在某种统一形式下反映了上面提到的布拉德福德定律、洛特卡定律、帕累托法则和齐夫幂律分布。

也许是由于在普莱斯那个时代人们尚未有高速计算机和大规模数据库，特别是还没有互联网可以使用，计量科学领域的专家们对普莱斯模型不太重视，其它领域绝大多数科学家甚至不知道它的存在，因而该模型并不为大众知晓。直到1999年，阿伯特-拉斯洛·巴拉巴西（Albert-László Barabási, 1967-）和他的博士生雷卡·阿伯特（Réka Albert, 1972-）在《科学》杂志上发表了影响极为广泛的论文“随机网络中标度的涌现”，才激发了研究者们对这个后来被称为“BA无标度网络模型”的兴趣和热情。现在知道，BA网络本质上是“重新发现”了普莱斯模型。不过，普莱斯模型描述有向网络而BA模型刻画无向网络，两者并非全同。回顾二十多年来的历史，BA模型最重要的贡献是在新千禧之年带起了一波研究网络科学的热潮，在科学发展进程中功不可没。

【五】

1956年，普莱斯在国际科学联盟理事会（International Council of Science Unions）旗下创建了国际科学史与科学哲学联盟（International Union of History and Philosophy of Science and Technology）以及国际科学政策研究理事会（International Council for Science Policy Studies）并任第一届主席。

1960年代，普莱斯被美国国家科学基金会（NSF）任命为科学信息委员会成员。这项工作让他有机会深入研究了大量的科学文献，启动了他后来的科学计量学奠基工作，并让他有条件分析各种科学政策和进行“科学学”研究。普莱斯后来还担任过联合国教科文组织的科学政策顾问，应邀到过阿根廷、澳大利亚、巴西、丹麦、埃及、德国、印度、以色列、巴基斯坦、瑞士、苏联等国家进行调查咨询和政策研究。

1976年，普莱斯获得国际技术历史学会（Society for the History of Technology）的最高奖达芬奇奖章（Leonardo da Vinci Medal）。1981年，他又获得国际社会科学研究学会（Society for Social Studies of Science）的伯纳尔奖（Bernal Award）。1983年初，他被选为瑞典皇家学院的外籍院士，同年受美国科学促进会（AIAA）邀请作萨顿讲座（Sarton Lecture）的演讲。

1978年，普莱斯创办了《科学计量学》（Scientometrics）杂志并担任主编。

普莱斯1983年去世后，国际科学计量学和信息计量学学会（International Society for Scientometrics and Infometrics）自1984年起设立了德里克·德·索拉·普莱斯纪念奖章（Derek de Solla Price Memorial Medal），每年颁发一次，授予在科学计量研究领域做出杰出贡献的科学家。

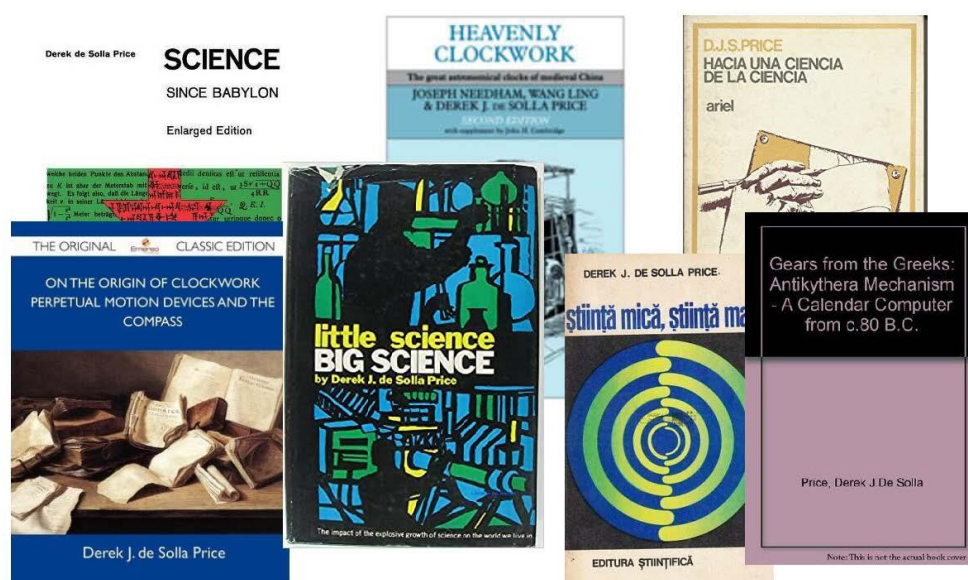


图7 普莱斯书籍选示

延伸阅读

汪小帆：[从无标度网络研究历史看想法传播](#)

陈关荣：[司马贺之问：学习还是创造？](#)