

数据分析与定量研究丛书

H 型指数和 H 型测度研究

H-type Indices and
H-type Measures

赵星 李盛庆 叶鹰 等著



科学出版社

数据分析与定量研究丛书

H 型指数和 H 型测度研究

H-type Indices and H-type Measures

赵 星 李盛庆 叶 鹰 等著

科 学 出 版 社

北 京

内 容 简 介

H 指数是 21 世纪以来科学计量和信息计量领域最具创新的发现之一,学术界不仅已经研究了各种 H 型指数,而且发展出多种 H 型测度和独具特色的抽取核心结构的量化测度方法。本书是研究 H 型指数和 H 型测度的一部学术专著,在阐明 H 型指数和 H 型测度来龙去脉的基础上,对 H 型指数和 H 型测度的理论机理、实证研究和应用拓展等进行系统探索,可供读者参考。

本书可作为科研管理与评价、科技政策、科学计量学、信息计量学、文献计量学、图书情报与档案管理等相关领域研究者和工作者的业务参考用书,也可作为信息资源管理、信息管理与信息系统、图书馆学、情报学与文献学等相关专业的本科生和研究生的专题教材。

图书在版编目(CIP)数据

H 型指数和 H 型测度研究 / 赵星等著. —北京:科学出版社, 2018.9
(数据分析与定量研究丛书)

ISBN 978-7-03-058638-4

I. ①H… II. ①赵… III. ①文献计量学—研究 IV. ①G250.252

中国版本图书馆 CIP 数据核字 (2018) 第 199545 号

责任编辑:惠 雪 / 责任校对:彭 涛

责任印制:张克忠 / 封面设计:许 瑞

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

文林印务有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018 年 9 月第 一 版 开本:720mm×1000mm

2018 年 9 月第一次印刷 印张:19 3/4

字数:375 000

定价:139.00 元

(如有印装质量问题,我社负责调换)



“数据分析与定量研究丛书”编委会

主 编：叶 鹰

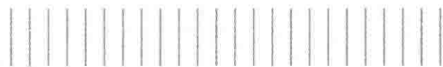
编 委（姓氏拼音序）：

曹 冲	黄 萃	李 江	刘晓钟	潘云涛	彭国超
唐 莉	王贤文	萧 璐	王 建	武夷山	叶 鹰
岳卫平	张 琳	赵 星	赵宇翔		

支持单位（单位拼音序）：

大连理工大学科学学与科技管理研究所暨 WISE 实验室
华东师范大学学术评价与促进研究中心
江苏省数据工程与知识服务重点实验室
科睿唯安
清华大学科教政策研究中心

丛书序



自从西学东渐，科学思想及应用遍及各种学问，而量化作为科学的重要特征却未见得深入人心，尤其是属于“文科”的社会科学中有些学者以为不必定量、甚或以当代西方社会科学对过度量化的反思为由，用中国学术长于思辨、容忍模糊的文化传统拒定量研究于千里之外。实则西方定性是建立在定量传统基础上的定性，中国量化远未及格，切不可尚未浸染量化即论过度定量，而宜取强化定量且定性定量并重之方略，既倡导质性研究需要量化支持（从 Nvivo 软件看质性研究），也鼓励量化研究需有质性判断（从统计相关到因果关联）。

不仅科学技术需要量化，2009 年以来，伴随计算社会科学的兴起，数据驱动的研究模式方兴未艾，在互联网和大数据蓬勃发展的推动下社会科学研究范式正在发生变革，乃至人文研究也出现了数字人文浪潮，数据分析和定量研究如同促进学术腾飞的双翼，正在学术研究中发挥引擎般作用，特别是在面向问题展开的跨学科研究中尤其如此。

在此学术发展新时期，科学出版社和本编委会以成熟一本推出一本的方式出版“数据分析与定量研究丛书”，旨在把数据分析和定量研究的研究成果系统化，既深化社会网络分析、社会仿真模型、复杂系统建模、自动信息抽取等方法，也强化复杂网络、信息分析、知识服务等应用，内容总体上以量化为方法特征，以数据、信息、知识、社会为主要解析对象，既处理大千世界的万千变化，也分析人类社会的复杂机理，期望能为理论界带来新思想，能为学术界迎来新知识。

收入本丛书的专著有两大特点：一是研究方法以定量主导，研究基础以数据支撑；二是跨学科性明显，既可能是计算社会科学的新领域，也可能是科技新方法新技术在人文社会科学中的新应用，覆盖科学计量、政策分析、用户测评等研究领域，体现知识的交叉融合。

丛书的作者大多是年富力强、充满活力的青年学者，他们都学有专长、术有



专攻，对自己的研究具有独特的思考，取得了令人瞩目的成绩。丛书荟萃他们的研究成果，可以彰显当代数据分析和定量研究之主流，期望能丰富学术研究、推动学术进步，为学术发展提供参考。

是为序。

叶 鹰

欧洲文理科学院院士，南京大学教授

2018年8月1日

前言



何为最好的测度？在信息计量学发展与成熟的数十年过程中，这一问题被频繁提起，其答案被孜孜以求，却依旧没有定论。

自 20 世纪 60 年代普赖斯 (D. J. Price) 提出引文分析、加菲尔德 (Eugene Garfield) 引入影响因子之后数十年，科学计量学界再没有产出具有同等开创意义和全局影响的测度或方法。以至于在 21 世纪之初，即使是图书情报学内部，也对这一领域持悲观态度，认为这一号称图书情报学最具有学科知识输出的领域在理论与方法创新上已走到尽头，剩下的仅是应用而已。当时，一部分信息计量学者也判定：测度、指标和方法，已不再是本领域能突破的重点研究方向。

然而，科学探索的本质是解决问题。学术数据测评这一问题客观存在，无可回避。学界要求我们面对这一问题，无关难度。而历史的进程也呼唤天才，无关学科。2005 年，似乎从来不曾关心科学计量学的物理学家乔治·赫希 (Jorge E. Hirsch) 横空出世，给出 21 世纪以来最有全局影响的信息计量学发现——H 指数。与普赖斯和加菲尔德一样，乔治·赫希并非图情学科出身。与引文分析和影响因子类似，H 指数也没有发表在图情学刊物上，而是发表于知名的综合性刊物《美国国家科学院院刊》(PNAS)，其中之巧合与必然或许值得我们深思。

然而，H 指数带来的启示远不只这些。在理论上，后来的研究显示它能与包括洛特卡定律等情报学基本模型无缝衔接。在方法上，它的计算方法具有独特性，而并非像其他所有信息计量学指标那样仅是移植于统计学或数学的算法。在应用上，它仿若从“石头中蹦出”，却能在短短数年内被很多主流学科学者所自发地接受和使用，在这个抵制定量评价和影响因子大行其道的年代，H 指数起到了逆流而上的作用。

本书即在这样的背景下产生。我们课题组于近十年中，对 H 指数、H 型指数和 H 型测度进行坚持不懈的探索，先后发表了约 30 篇国际论文和 10 余篇中文论

文,完成和正在完成三个国家自然科学基金项目[“h-指数和类 h-指数的机理分析与实证研究”(批准号:70773101)、“专利 h 指数与专利信息网络测度研究”(批准号:71173187)和“h 型信息网络测度的机理与实证研究”(批准号:71503083)]。其中精华,皆在本著作中。

全书分为五篇,从各类代表性的 H 型指数说起,随后分别在理论研究、应用研究和拓展研究三个方面进行系统性探索,最后导出有特色的 H 型网络测度。全书第一篇主要由赵星和叶鹰等主笔,第二篇由叶鹰和赵星撰写,李盛庆、叶鹰和赵星等合作完成了第三篇,第四篇主要由赵星和叶鹰等完成,最后一篇由赵星主笔。先后还有多位学者和师生参与了本书或之前 2011 年出版的《h 指数与 h 型指数研究》的相关研究工作,他们是 Ronald Rousseau、高小强、潘有能、唐建辉、张力、黄娟、彭晓东、谭旻、王舒月、张家榕、威尔鹏、陈路遥、蔡前黎、苏林伟、罗瑞脩、张品慧、张瑜婷等,特致谢意。十载光阴,若有遗漏,还盼见谅。

本书的出版获得国家自然科学基金项目(h 型信息网络测度的机理与实证研究,批准号:71503083)的经费资助。

纵观全书,H 指数虽可谓影响深远,但仍难以称为“最好的测度”。欣慰的是,它让我们看到最好测度应有之特点:简洁、巧妙、富含信息、独一无二。较完整地呈现 H 指数及相关拓展是本书的目的,但不是我们全部的目标,通过 H 型测度研究追寻最好测度的足迹才是本书的立著之意。限于水平,不足难免,只盼对读者有所启示,引发更多的批判、合作与创新。

作 者

2018 年夏

目 录



丛书序

前言

第一篇 H 指数和 H 型指数的方法发展

第 1 章 从 H 指数到 H 型指数	3
1.1 Hirsch H 指数	3
1.2 针对高被引论文权重的 H 型指数	6
1.2.1 g 指数及 hg 指数	6
1.2.2 h(2)指数及广义 h(2)指数	10
1.2.3 w 指数及广义 w 指数	11
1.2.4 h_w 指数	12
1.3 区分论文作者贡献的 H 型指数	12
1.3.1 h_l 指数与 h_m 指数	12
1.3.2 $\bar{h}$ (hbar) 指数	14
1.3.3 h_c 指数	15
1.4 考虑时间维度的 H 型指数	17
1.4.1 现时 H 指数与趋势 H 指数	17
1.4.2 H 序列与 H 矩阵	18
1.4.3 动态 H 指数	18
1.5 H 核基础上更细粒度的 H 型指数	20



1.5.1	A 指数、AR 指数与 R 指数	20
1.5.2	e 指数	21
1.5.3	m 指数与 q^2 指数	22
1.5.4	J 指数	22
1.5.5	核心区域与核心区间指数	23
1.5.6	双边 H 指数	24
1.5.7	k 指数和 w 指数	26
1.6	测度 H 指数增量的 H 型指数	28
1.6.1	锥形 H 指数	28
1.6.2	随机 H 指数	30
1.6.3	实 H 指数与有理 H 指数	31
1.7	区分施引来源的 H 型指数	32
1.7.1	H_1 指数	32
1.7.2	WL 指数	34
1.8	跨学科或跨层面的 H 型指数	35
1.8.1	h_r 指数	35
1.8.2	连续 H 指数	36
1.9	其他指数	37
1.9.1	社会化 H 指数	37
1.9.2	Zynergy 指数	38
	参考文献	39

第二篇 H 指数和 H 型测度的理论机理

第 2 章	H 指数和 H 型指数的数学模型	47
2.1	Hirsch 模型	47
2.2	Egghe-Rousseau 模型	49
2.3	Glänzel-Schubert 模型	50

2.4 Burrell 假说	50
2.5 各类模型的检验与贯通	51
参考文献	65
第3章 H型指数与信息计量理论研究	67
3.1 H核-H尾分布及其测度	67
3.2 H指数框架下的多变量测度集	73
3.3 广义H指数	76
参考文献	80
第三篇 H型指数和H型测度的应用研究	
第4章 学术主体(作者、大学、专利权人)H型指数测评	85
4.1 作者H型指数测评	85
4.2 大学H型指数测评	90
4.3 专利权人H型指数测评	95
参考文献	99
第5章 学术客体(期刊、论著、专利)H型指数测评	101
5.1 期刊H型指数测评	101
5.2 论著H型指数测评	110
5.3 专利H型指数测评	114
参考文献	119
第6章 特殊对象H型指数测评	120
6.1 科学基金H型指数测评	120
6.1.1 科学基金H指数	120
6.1.2 数据处理	121
6.1.3 科学基金H指数与其他指标的相关关系	122
6.1.4 科学基金H指数对论文数量和影响力的代表性分析	124
6.1.5 科学基金H指数的不足	126

6.2 图书馆 H 型指数测评	126
6.2.1 图书馆 H 指数的含义与计算	127
6.2.2 省级公共图书馆馆员论文 H 指数的总体情况	128
6.2.3 省级公共图书馆馆员论文 H 指数与其他指标的联系	131
6.2.4 省级公共图书馆与高校图书馆的馆员论文 H 指数比较	133
6.2.5 小结	135
参考文献	136
第 7 章 国家 H 测度及其与 GDP 和 R&D 投入关系研究	137
7.1 数据与方法	137
7.2 国家 H 测度的实证拟合	139
7.3 国家 H 测度与 GDP 的相关性	143
7.4 国家 H 测度与 R&D 投入的关系	146
7.5 小结	149
参考文献	149
第四篇 H 型指数和 H 型测度的扩展研究	
第 8 章 H 指数与王冠指数和 MNCS 比较研究	153
8.1 H 指数与王冠指数相关分析	153
8.2 H 指数与 MNCS 相关分析	167
参考文献	169
第 9 章 H 指数与特色期刊计量参数比较研究	170
9.1 H 指数与 SJR	170
9.2 H 指数与特征因子	171
9.3 总量型与平均型——期刊计量参数的两类因子	172
参考文献	175
第 10 章 构建于 g 指数的研究热点分析方法	177
10.1 g 指数与主题词 g 指数的计算	177

10.2 数据	178
10.3 主题词 g 指数与词频的关系	179
10.4 主题词 g 指数与该主题论文被引之间的关系	180
10.5 小结	181
参考文献	182
第 11 章 多维 H 分布与多维 H 测度	183
11.1 多维 H 分布	183
11.2 广义 H 核心集	185
11.3 数据处理	186
11.4 <i>Nature</i> 与 <i>Science</i> 的多维 H 分布比较	187
11.5 社会网络和复杂网络研究的多维 H 分布比较	190
11.6 小结	195
参考文献	196
第 12 章 H 指数研究的知识图谱	197
12.1 数据输入	197
12.2 图谱示意	198
12.2.1 总体态势	198
12.2.2 关键词与研究主题	202
12.2.3 核心作者	207
参考文献	211

第五篇 H 型信息网络测度的探索研究

第 13 章 信息网络	215
13.1 背景	215
13.2 社会网络、复杂网络与信息网络	217
13.2.1 社会网络分析的探索之路	217
13.2.2 复杂网络的方兴未艾	220

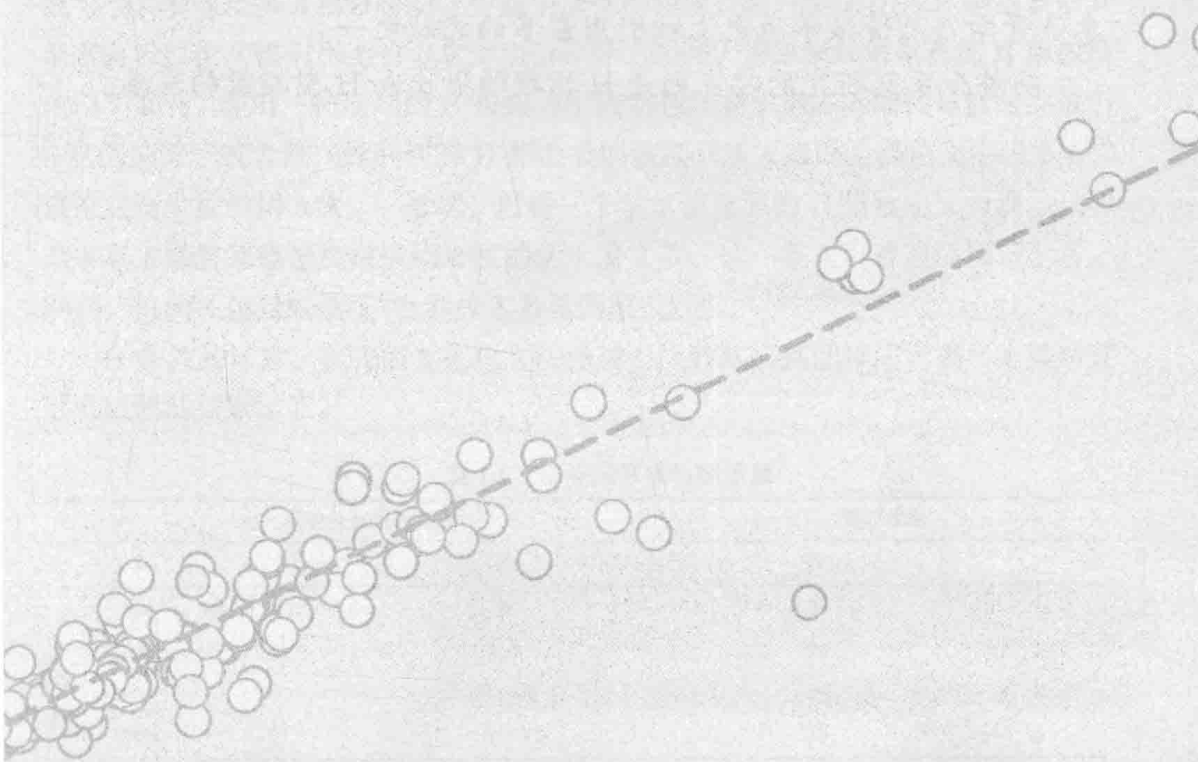
13.2.3 信息网络: 矛盾的涌现	222
参考文献	228
第 14 章 节点测度: H 度及其延拓	231
14.1 H 度及其特质	231
14.1.1 H 度的定义	231
14.1.2 H 度的理论性质	233
14.2 H 度的方法扩展	235
14.2.1 H 中心度	235
14.2.2 H 中心势	237
14.2.3 H 比率	238
14.3 实证研究: 核心文献的 H 度	238
14.3.1 数据采集	238
14.3.2 结果与讨论	240
14.4 有向带权信息网络中的 H 度	242
14.4.1 有向 H 度测度集	243
14.4.2 LIS 学科期刊与中国人文社会学科的网络分析	247
参考文献	254
第 15 章 联系测度: 幂律、H 强度与核心联系	256
15.1 联系权重的幂律分布	256
15.1.1 方法和数据处理	257
15.1.2 实证结果	259
15.1.3 理论讨论: 优先重连机制	261
15.2 H 强度方法	263
15.2.1 H 强度的含义与特点	263
15.2.2 H 强度的扩展: H 子网与 I_h	265
15.2.3 实证研究: 论文共引网络与国家合作网络	266
15.2.4 理论讨论: 理论性质、Glänzel-Schubert 与 Hirsch 模型	270



参考文献	273
第 16 章 全网测度：带权信息网络的骨干抽取	275
16.1 核心节点、核心联系与桥重构	275
16.1.1 步骤 1：确定核心节点	276
16.1.2 步骤 2：确定核心联系集	276
16.1.3 步骤 3：重构网络“桥”	278
16.2 相关理论性质分析	279
16.2.1 H 骨干的核心节点性质	279
16.2.2 H 骨干的核心联系性质	280
16.2.3 H 骨干中的桥及整体性质	281
16.3 实证研究：LIS 学科共引网络	282
16.3.1 数据描述	282
16.3.2 结果与讨论	283
参考文献	285
第 17 章 H 型信息网络测度的整体讨论	286
17.1 从局部到整体的带权信息网络测度框架	286
17.2 从计量属性到结构揭示的信息学研究	288
参考文献	290
附录	292
后记	297

第一篇

H 指数和 H 型指数的方法发展



2005 年 11 月 15 日,一篇名为“An Index to Quantify an Individual's Scientific Research Output”的论文发表在美国科学院院报(PNAS)第 102 卷第 46 期上(Hirsch, 2005),署名 J. E. Hirsch,虽然该文章于 2005 年 8 月已在 *arXiv* 上公开发表,但在 *PNAS* 的正式发表强化了学术界对此关注。*Nature* 评论员 Philip Ball 在当年 *Nature* 第 436 卷 900 页上的评论正面肯定了 H 指数的效果(Ball, 2005)。尽管后来也有不同意见(Lehmann, 2006),但 H 指数仍以其简单新颖的特点引起学术界的兴趣。很快, H 指数研究的学术帷幕开启,并成为学术热点。

现今, H 指数已全面进入学术评价、信息测度等领域的应用, Hirsch 的原始论文已在 Web of Science 平台中被引用近 3500 次,成为信息计量学乃至情报学历史上被引用最多的文献之一。

而整个系列研究的起点则是 H 指数的提出和 H 型指数的发展。