

格致方法·定量研究系列 吴晓刚 主编



# 指数随机图模型导论

[美] 詹宁·K. 哈瑞斯 (Jenine K. Harris) 著  
杨冠灿 译

- ★ 革新研究理念
- ★ 丰富研究工具
- ★ 最权威、最前沿的定量研究方法指南

51

格致出版社 上海人民出版社

格致方法·定量研究系列 吴晓刚 主编

# 指数随机图模型导论

[美] 詹宁·K. 哈瑞斯 (Jenine K. Harris) 著  
杨冠灿 译

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

指数随机图模型导论/(美)詹宁·K.哈瑞斯著;  
杨冠灿译.—上海:格致出版社:上海人民出版社,  
2016.10

(格致方法·定量研究系列)

ISBN 978-7-5432-2654-8

I. ①指… II. ①詹… ②杨… III. ①社会关系-指  
数模型-研究 IV. ①C912.3②F224.0

中国版本图书馆CIP数据核字(2016)第184147号

责任编辑 张苗凤

格致方法·定量研究系列

指数随机图模型导论

[美]詹宁·K.哈瑞斯 著

杨冠灿 译

出版 世纪出版股份有限公司 格致出版社  
世纪出版集团 上海人民出版社  
(200001 上海福建中路193号 www.ewen.cn)



编辑部热线 021-63914988  
市场部热线 021-63914081  
www.hibooks.cn

发行 上海世纪出版股份有限公司发行中心

印刷 浙江临安曙光印务有限公司  
开本 920×1168 1/32  
印张 6  
字数 120,000  
版次 2016年10月第1版  
印次 2016年10月第1次印刷

ISBN 978-7-5432-2654-8/C·153

定价:30.00元

## 出版说明

---

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书,精选了世界著名的 SAGE 出版社定量社会科学研究丛书,翻译成中文,起初集结成八册,于 2011 年出版。这套丛书自出版以来,受到广大读者特别是年轻一代社会科学工作者的热烈欢迎。为了给广大读者提供更多的方便和选择,该丛书经过修订和校正,于 2012 年以单行本的形式再次出版发行,共 37 本。我们衷心感谢广大读者的支持和建议。

随着与 SAGE 出版社合作的进一步深化,我们又从丛书中精选了三十多个品种,译成中文,以飨读者。丛书新增品种涵盖了更多的定量研究方法。我们希望本丛书单行本的继续出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

# 总序

---

2003年,我赴港工作,在香港科技大学社会科学部教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来,非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课,而第二门课“社会科学中的定量分析”为博士生的必修课(事实上,大部分硕士生修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候,根据社会科学研究生的数理基础比较薄弱的特点,尽量避免复杂的数学公式推导,而用具体的例子,结合语言和图形,帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上,即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”,学完这些课程后,我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章;另一方面,也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容,尽管在线性回归模型的内容上有少

量重复,但各有侧重。“社会科学里的统计学”从介绍最基本的社会研究方法论和统计学原理开始,到多元线性回归模型结束,内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型,以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法,将重点放在因变量为定类数据的分析模型上,包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型,包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生,一直鼓励和支持我将两门课的讲稿结集出版,并帮助我将原来的英文课程讲稿译成了中文。但是,由于种种原因,这两本书拖了多年还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩,每本书都写得通俗易懂,与我的教学理念是相通的。当格致出版社向我提出从这套丛书中精选一批翻译,以飨中文读者时,我非常支持这个想法,因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力,还要有对实质内容有较深的理解能力,而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容,只有语言能力是远远不能胜任的。在短短的一年时间里,我们组织了来自中国内地及香港、台湾地区的二十几位

研究生参与了这项工程,他们当时大部分是香港科技大学的硕士和博士研究生,受过严格的社会科学统计方法的训练,也有来自美国等地对定量研究感兴趣的博士研究生。他们是香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛,硕士研究生贺光辉、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊,应用社会经济研究中心研究员李俊秀;香港大学教育学院博士研究生洪岩璧;北京大学社会学系博士研究生李丁、赵亮员;中国人民大学人口学系讲师巫锡炜;中国台湾“中央”研究院社会学所助理研究员林宗弘;南京师范大学心理学系副教授陈陈;美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛;美国加州大学洛杉矶分校社会学系博士研究生宋曦;哈佛大学社会学系博士研究生郭茂灿和周韵。

参与这项工作的许多译者目前都已经毕业,大多成为中国内地以及香港、台湾等地区高校和研究机构定量社会科学方法教学和研究的骨干。不少译者反映,翻译工作本身也是他们学习相关定量方法的有效途径。鉴于此,当格致出版社和 SAGE 出版社决定在“格致方法·定量研究系列”丛中推出另外一批新品种时,香港科技大学社会科学部的研究生仍然是主要力量。特别值得一提的是,香港科技大学应用社会经济研究中心与上海大学社会学院自 2012 年夏季开始,在上海(夏季)和广州南沙(冬季)联合举办《应用社会科学研究方法研修班》,至今已经成功举办三届。研修课程设计体现“化整为零、循序渐进、中文教学、学以致用”的方针,吸引了一大批有志于从事定量社会科学研究的博士生和青年学者。他们中的不少人也参与了翻译和校对的工作。他们在

繁忙的学习和研究之余,历经近两年的时间,完成了三十多本新书的翻译任务,使得“格致方法·定量研究系列”丛书更加丰富和完善。他们是:东南大学社会学系副教授洪岩璧,香港科技大学社会科学部博士研究生贺光焯、李忠路、王佳、王彦蓉、许多多,硕士研究生范新光、缪佳、武玲蔚、臧晓露、曾东林,原硕士研究生李兰,密歇根大学社会学系博士研究生王骁,纽约大学社会学系博士研究生温芳琪,牛津大学社会学系研究生周穆之,上海大学社会学院博士研究生陈伟等。

陈伟、范新光、贺光焯、洪岩璧、李忠路、缪佳、王佳、武玲蔚、许多多、曾东林、周穆之,以及香港科技大学社会科学部硕士研究生陈佳莹,上海大学社会学院硕士研究生梁海祥还协助主编做了大量的审校工作。格致出版社编辑高璇不遗余力地推动本丛书的继续出版,并且在这个过程中表现出极大的耐心和高度的专业精神。对他们付出的劳动,我在此致以诚挚的谢意。当然,每本书因本身内容和译者的行文风格有所差异,校对未免挂一漏万,术语的标准译法方面还有很大的改进空间。我们欢迎广大读者提出建设性的批评和建议,以便再版时修订。

我们希望本丛书的持续出版,能为进一步提升国内社会科学定量教学和研究水平作出一点贡献。

吴晓刚

于香港九龙清水湾



# 序

---

自 20 世纪初乔治·齐美尔(Georg Simmel)首次论述社会网络相关问题以来(Simmel & Wolff, 1950),社会科学家对于个体之间、组织之间以及其他实体之间相互关联的网络问题一直保持高度的关注(参见例如 Fienberg, 2012)。20 世纪 30 年代,心理医生雅各布·莫雷诺(Jacob Moreno, 1934)的工作为社会网络研究奠定了基础,并将此领域命名为“社会计量学”(sociometry)。在莫雷诺的诸多重要成果中,核心成果便是发明了社群图(sociogram)方法,通过将个体图形化表示为节点,个体之间联系图形化表示为连线的形式,社群图方法就能够用来解释社会结构问题。

在社会网络分析发展的历程中,社群图方法被证明是十分重要的,原因之一是社群图方法将图论的基础理论引入到了社会网络分析中来。图论是一个专门处理由节点(点)以及相连的边(连线)所组成的数学分支,其中,网络图既可以是有序的,即网络中的边通常由从一个节点到另一个节点的箭头所表示,从而展现节点之间潜在的非对称联系;网络图也可以是无序的,直接用线段来表示网络中的边。大多数研

究社会网络的传统方法都是来源于图论的,社会科学中的定量研究方法应用系列丛书(QASS)中,有一本较早的著作,是由诺克和杨(Knoke & Yang, 2008)撰写的《社会网络分析》,该书就主要是采用这种(传统)方法。

传统的网络分析方法主要是描述性的,并不采用具有统计学意义上的随机变量模型构建方法。明确提出以网络结构为中心建立概率模型的思想可以追溯到 20 世纪中叶,即吉尔伯特、艾多斯以及瑞尼(Gilbert, 1959; Erdos & Renyi, 1959)解释了网络结构中最为基础的零模型(null model)。在零模型中,所有的节点对都是以同等的概率建立连线,无论是在有向网络还是无向网络中,简单图模型都是被最广泛采用的模型。

20 年之后,霍兰德和莱因哈特(Holland & Leinhardt, 1981)引入了一种针对有向图的 Gilbert-Erdos-Renyi 零模型的变种。其中,关系形成(tie formation)的概率受到个体的群集性(gregariousness,个体对外与他人建立联系的属性)以及受欢迎程度(popularity,他人与该个体建立联系的属性)的影响。在此之后不久,1981 年,芬博格和沃瑟曼(Fienberg & Wasserman, 1981)将霍兰德和莱因哈特的  $p_1$  模型改造为对数线性模型,对数线性模型是一种为统计学家和社会科学家所熟知的模型,这样一来,学者们就可以方便对模型的参数进行最大似然估计了。此外,芬博格和沃瑟曼还对  $p_1$  模型进行了扩展,将网络的“互惠性”(reciprocity)特征纳入到模型中来,并以“互惠性”特征作为网络连线概率增强的机制——例如,在一个朋友网络中,如果 A 选择 B,那么, B 选择 A 的概率就会提升。

正如詹宁·哈瑞斯(Jenine Harris)在本书中所解释的,吉尔伯特等人的零模型、霍兰德和莱因哈特的  $p_1$  模型,以及芬博格和沃瑟曼(1981)的扩展模型都是指数随机图模型(exponential random graph models, ERGMs)家族的成员。过去30年里,指数随机图模型的研究取得了长足的进展,而且已经成为了目前社会网络分析中最重要的统计工具。在这个进程中,指数随机图模型不断彰显着自己在展现社会网络结构特征分析方面的洞察力,例如对聚类或“聚簇”的分析。

近年来,面对大数据分析所带来的挑战与激励,计算机科学家和统计物理学家,与统计学家、社会科学家并肩作战,对社会网络分析的发展起到了直接推动作用。源于社会生活中的大型网络数据尤为庞大与复杂,如 Facebook 的数据,这也促使研究人员必须不断研究更为复杂的网络模型,不断改进统计软件的计算能力,以确保研究的模型能够适应大数据的环境。哈瑞斯在其书中介绍了由 statnet 团队所研发的最先进的网络分析软件(Handcock et al., 2003),该软件是针对 R 的统计计算环境而开发的(R Core Team, 2013),是一款广泛使用的、免费且开源的统计分析平台。

本书介绍了如何建立指数随机图模型,并解释了如何在实践中使用该模型,詹宁·哈瑞斯的工作对于采用社会网络分析的社会学家而言十分重要。我希望她的这本著作将会有较广泛的读者群,同时,期待该书能够对社会科学中社会网络分析质量的提升产生实质性的影响。

约翰·福克斯(John Fox)

## 参考文献

---

- Erdős, P. and Rényi, A. (1959). On random graphs, I. *Publicationes Mathematicae*, 6:290—297.
- Fienberg, S.E. (2012). A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics*, 21:825—839.
- Fienberg, S.E. and Wasserman, S.S. (1981). Categorical data analysis of single sociometric relations. In Leinhardt, S., editor, *Sociological Methodology 1981*, pages 156—192. Jossey-Bass, San Francisco.
- Gilbert, E.N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30:1141—1144.
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2003). *statnet: Software tools for the Statistical Modeling of Network Data*. Seattle, WA.
- Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76:33—65.
- Knoke, D. and Yang, S. (2008). *Social Network Analysis*. Thousand Oaks CA, second edition.
- Moreno, J. (1934). *Who Shall Survive?* Nervous and Mental Disease Publishing Company, Washington DC.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Simmel, G. and Wolff, K.H. (1950). *The Sociology of Georg Simmel*. The Free Press, New York.

# 目录

---

序	1
第 1 章 网络分析方法的希望与挑战	1
第 1 节 历史与概念	10
第 2 节 网络术语	17
第 2 章 统计网络模型	21
第 1 节 简单随机图	24
第 2 节 ERGM 的发展	30
第 3 节 本章小结	47
第 3 章 建立一个有效的指数随机图模型	49
第 1 节 软件获取与准备	52
第 2 节 数据获取	54
第 3 节 数据探索	59
第 4 节 模型构建	70
第 5 节 曲线指数族模型	112

第 4 章	面向有向网络及二元组属性的应用	131
第 1 节	针对有向网络的研究	132
第 2 节	将二元组和网络协变量作为自变量	146
第 5 章	结论与建议	151
附录		159
参考文献		161
译名对照表		165
译后记		172

第 **1** 章

网络分析方法的希望与挑战

近 30 年来,经验社会研究被抽样调查方法所主导。但正如人们经常指出的那样,通过对个体进行随机抽样的调查方法成为了一台社会学研究的绞肉机,它把个体从其所在的社会情境中剥离开来,使得研究过程中任何个体之间不存在任何关系。这就有点像一个生物学家让他们的实验动物首先经过一台绞肉机的处理,然后,以百计的细胞为观测对象,通过显微镜对(实验动物)进行观察。这样,解剖学和生理学用不着了,结构与功能消失了,唯一剩下的只有细胞生物学……如果我们的目标是去理解人类的行为而不是简单地记录这些行为,那么,我们希望知道诸如主要群体、邻居关系、组织结构、社会圈子、社区关系等问题;还包括交流、沟通、角色期望以及社会控制等问题。

——巴顿(Barton), 1968,

援引自弗里曼(Freeman, 2004:1)

2010 年,南亚的一个内陆小国不丹采用了一种用于测量国家成就的新措施——国民幸福总值(GNH)指数,与之前通用的评价国家生产力的经济指标不同,GNH 指数侧重于观



察国民的福祉，“意在通过改进尚未获得幸福感(not-yet-happy)人群的生存状况，引导人民和国家走向幸福”(Ura, Alkire, Zangmo & Wangdi, 2012)。在科学文献中，GNH 也被称为主观幸福感(subjective well-being, SWB)，通过总结数十年对于幸福问题的研究经验，GNH 指数初步成为一个包含 9 大领域、13 类指标以及 124 项变量的综合指标体系。GNH 指数所涉变量的范围覆盖了整个生态模型，从个体层次人口统计指标，如年龄与教育程度，到中观层次对家庭紧密度、社区融入度的测量，甚至包括对自然和经济环境的认知等。然而，在 GNH 指数所包含的若干测量指标中却遗漏了对特定社会关系的测量。之前的研究发现：相邻个体之间的关系数量、关系属性以及关系之间的同质性(其中，同质性是指人们更愿意与其相似的人群进行联系的行为准则)对于个体知识、态度以及交往活动都起到了重要的作用(McPherson, Smith-Lovin & Cook, 2001)，这种作用也适用于针对幸福感问题的研究(Burt, 1987; Myers & Diener, 1995)。例如，最近对于幸福感问题的研究就发现，至少在我们的在线网络中，感觉幸福的人总是愿意与那些同样感觉到幸福的人交朋友(Bliss, Kloumann, Harris, Danforth & Dodds, 2012; Bollen, Goncalves, Ruan & Mao, 2011)。图 1.1 显示了一个在线交友网络，其中的朋友关系以及他们是否感觉到幸福的状态被标记出来，感觉到幸福的人们被标注为具有较高的主观幸福感，而感觉不幸福的人则具有较低的主观幸福感。