

INTRODUCTION TO COMPUTATIONAL COMMUNICATION



计算传播学导论

张 伦 王成军 许小可 / 编著



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

INTRODUCTION TO
COMPUTATIONAL COMMUNICATION

计算传播学导论

张 伦 王成军 许小可 / 编著



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

图书在版编目 (CIP) 数据

计算传播学导论 / 张伦, 王成军, 许小可编著. —北京: 北京师范大学出版社, 2018.11
ISBN 978-7-303-24120-0

I. ①计… II. ①张… ②王… ③许… III. ①数据处理-应用-传播学 IV. ①G206-39

中国版本图书馆CIP数据核字 (2018) 第195149号

营 销 中 心 电 话 010-58805072 58807651
北师大出版社高等教育与学术著作分社 <http://xueda.bnup.com>

JISUAN CHUANBOXUE DAOLUN

出版发行: 北京师范大学出版社 www.bnup.com
北京市海淀区新街口外大街19号
邮政编码: 100875

印 刷: 大厂回族自治县正兴印务有限公司
经 销: 全国新华书店
开 本: 787 mm × 1092 mm 1/16
印 张: 15.25
字 数: 315 千字
版 次: 2018 年 11 月第 1 版
印 次: 2018 年 11 月第 1 次印刷
定 价: 49.80 元

策划编辑: 周 粟	责任编辑: 马力敏
美术编辑: 李向昕	装帧设计: 锋尚设计
责任校对: 李云虎	责任印制: 马 洁

版权所有 侵权必究

反盗版、侵权举报电话: 010-58800697

北京读者服务部电话: 010-58808104

外埠邮购电话: 010-58808083

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 010-58805079

本书是张伦、王成军、许小可继《社会网络上的计算传播学》之后的又一著作。张伦和成军是我们香港城市大学互联网挖掘实验室的博士毕业生，而小可和海波则都曾先后来实验室做过访问学者。于公（事关传播学发展之大计）于私，我都很乐意遵嘱写个简序。

也许有些读者会问，本书与前书有何区别？一种简单粗暴的区分是本书为教科书而前书为专题研究。在此我就计算传播学教科书讨论（其实仅能提及）几个问题。首先，有无必要和可能为计算传播学这个新兴领域写教科书？我原本的看法是有其必要但条件尚未成熟。不料本书三位作者居然写出一本，其内容虽然还没达到我心目中的标准（因为这个领域整体上还在引进、模仿和探索的初期），但也是言之有物、自成一体的，对于广大亟需一本入门级教科书的传播学子来说，实为及时雨。因此，我郑重地向读者们推荐此书，也代表读者向三位作者致谢致敬。

也许有些读者（或者也包括本书作者）会问，本书哪些章节需要加强？我觉得最需要的是以下三个方面：研究设计、与传统方法和线下数据的融合、研究伦理道德问题。研究设计指什么？Earl Babbie^[1]在《社会研究方法》中有单独章节详细讲述了实证研究的设计问题，如分析单元（Unit of Analysis）、组间（Between-Subjects）组内（Within-Subjects）比较、时间维度等。我认为这些原理，虽然技术含量不高，但同样适用计算传播学，而又不为来自计算科学（包括数据、网络等的一个统称）的学者所熟悉，所以需要介绍给初学者，以填补理论/研究问题与数据/算法之间的断层地带。

与传统方法和线下数据的融合互补，似乎是个政治正确的流行口号，人人都会喊，但真正实行的并不多。究其原因，客观上有困难，传统方法和线下数据耗时耗财；主观上也有疑虑，觉得旧方法、旧数据过时、老土。我10年前也是这么想的，但随着对计算方法和大数据接触逐渐深入，越来越觉得其中的局限和陷阱。旧方法、旧数据并不能完全解决新方法、新数据的所有问题。但是，如果能在最需要的地方，富有创意地引入一些已被反复验证的旧方法、旧数据（如随机抽样），往往就能收到“四两拨千斤”“化腐朽为神奇”之效。这是一个内涵丰富的话题，值得我们大家反复实验、不断总结。

计算传播学研究中的伦理道德问题，涉及面极其广泛，从个人隐私和网络安全到数据产权和社会责任等。更棘手的是，大多数相关问题迄今并无明文法规可循，学界、

[1] 编辑注：本书是国内第一本关于“计算传播学”的导论教材，书中引用的英文著作、论文等的作者，在国内相关领域还没有统一的中文译名，为了读者查找资料方便，本书中涉及的相关作者直接使用其英文名，不再加注中文名。

业界更无共识。但是，这些问题个个都是躲不过、绕不开的，稍不注意即有可能误踩雷区。当然，入门级的教科书既无可能也无必要长篇大论地论述有关问题，但是确有必要在最易犯规的环节（如数据抓取、在线实验等）上提醒初学者以及提供一些操作清单，列明哪些是严禁的黑色地带和哪些是应避免的灰色地带。

期待在今后的三五年之内，经由海内外学者共同努力，计算传播学对上述问题（以及其他尚未论及的重要问题）展开充分研究，并积累起足够的经验教训和正反案例，以助本书再版更新，也有助其他相关的专著和教科书的涌现。

祝建华

2018年9月20日

自序

1998年, Watts和Strogatz在《自然》(*Nature*)上面发表重要论文《小世界网络集群动力学》(*Collective dynamics of small-world networks*), 标志着网络科学的诞生。社会网络研究和复杂网络研究在研究议题和研究方法层面第一次合流。2008年我进入香港城市大学媒体与传播系, 跟随祝建华老师攻读博士学位, 当时跨学科为主导的网络科学研究方兴未艾。2009年, 我从“校内网”(后来改名为“人人网”)提取了该平台建立前两年的全网数据, 将博士论文的选题定为“在线社交媒体好友关系构建及演化研究”。第一次面对着4万个注册用户、几亿条好友关系记录的“大”数据, 我惶恐不安。只写过SPSS代码的我, 就这样莽莽撞撞地迈进了“计算社会科学”的大门, 从此走上了计算社会科学研究的不归路。

转眼间我进入这个领域已有十年时间, 做过的研究题目也陆续在国内外一些学术期刊发表。只是偶尔回头看时, 当时的那份惶恐犹在。作为一个“过来人”, 对于进入“计算传播学”领域的初学者, 我觉得有几点经验有必要与各位分享。进入这个领域, 需要跨越“三座大山”: 定位理论问题、理解分析方法与掌握数据处理技术。第一, 定位理论问题。计算传播学的核心任务是立足本学科, 回答传播学研究中的重要问题。虽然这是一个需要进行跨学科合作的研究领域, 但交叉学科合作研究的意义在于如何更深入地回答传播学的重要问题, 而非跨入另一个学科, 最终抛弃传播学研究。虽然现在“计算”之风甚猛, 言必称“大数据”, 但纵观顶级期刊发表的论文, 无不是在回答传播学研究中的核心和重要理论问题。第二, 理解分析方法, 以“应用”为最终目的。我经常告诉刚开始学习统计分析的同学, 学习数据分析方法, 要以“会用”为学习目的。统计原理之于统计方法, 就好比汽车制动原理之于汽车驾驶。在学习之初, 你只需要学会如何驾驶汽车, 能够安全地把车开上路即可; 而不需要在汽车动力原理层面花太多时间和精力去理解。随着学习的深入, 对于数据分析方法的反复应用和学习, 学习者自然会逐渐深入理解数据分析方法背后的数学原理。而如果在入门阶段执着于完全理解分析方法原理(如如何迭代、如何估计参数等), 可能方法还没学会, 就已经信心全无准备“跑路”了。当然我并不认为学习者不需要理解数据分析方法原理。只是从“效率”角度出发, 建议初学者对数据分析方法先学会应用, 后学习原理, 这将更有助于尽快入门。第三, 掌握数据处理技术。作为文科生, 编程或许是大家内心深处的终极恐惧。读博士期间, 祝建华老师在实验室里的教导是我迈过这座大山的重要信念。祝老师说, 这个世界上没有学不会的东西, 差别只在于用多长时间学会。祝老师还引用了“一万小时定律”来“安慰”我们。作家格拉德威尔认为, “人们眼中的天才之所以卓越非凡, 并非天资超人一等, 而是付出了持续不断的努力。1万小时的锤炼是任何人从平凡变成世界级大师的必要条件。”的确, 在没有付出必要的努力之前, 我们甚至没有资格怀疑自己的智商。所以简单而言, 学习和练习就是了。

在过去的十年间，每当遇到技术瓶颈需要去翻阅技术论文时，我都在想，什么时候能有一本教材，专门写给初学者？这便是本书的写作初衷之一——站在文科视角，为没有太多数理编程背景的同学写一本通俗易懂的教材。

本书适合致力于“计算社会科学”研究的学生和青年教师自学与教学。本书对读者的基本先修要求是，了解社会科学定量研究的基本范式。对于高年级本科生和研究生同学，建议大家在自学过程中，在读懂每个章节的基础上，学习Python编程技术，并学习如何实现书中所有命令。同时，尽量阅读每个章节所引用的论文原文。本书所引用的大部分研究案例，大多发表在《自然》、《科学》(Science)、*Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 等国际顶尖学术期刊上。建议同学们阅读时仔细想一想，为什么这篇文章能够发表在顶尖学术期刊中？其研究问题的重要性、理论贡献和方法贡献各是什么？对于用于教学目的的青年教师，建议大家利用本书为基本提纲；同时按图索骥，找到自己研究子领域中具有代表性的研究问题，并指导学生，根据每个章节的基本代码，让学生实践相关的数据分析和研究。总而言之，“计算传播学”最有效的学习路径是“实践”，必须要亲手操作，才能体会和明白关键概念和关键步骤。仅仅通过概念的讲解和论文阅读，犹如隔岸观火，无法有效地投入到“计算传播学”研究中。

此外，需要说明的是，“计算传播学”自提出之时起，就是一个跨学科的研究领域。知识和研究问题并无绝对的学科之分，重要的社会科学研究问题往往需要跨越学科合作。例如，信息的传播问题是传播学和物理学共同感兴趣的话题；社会的公平与分配问题是社会学和经济学共同感兴趣的问题。因此，从这个意义上而言，我们很难对某一篇文章进行唯一的学科划分。传播学者认为某一个研究属于传播学研究，而相关领域的学者或许认为该研究同样属于其他学科，这并不奇怪。因此，本书中所例举的经典研究，并未仅限制在传播学领域学术期刊发表的文章，而是首选在上述顶级综合性学术期刊中寻找研究案例。

本书的写作，得益于我们在过去四五年间积累的相关教学实践。在本书的写作中，第一章至第四章由张伦主笔，第五章和第六章由许小可主笔，第七章和第八章由王成军主笔。而向更早追溯，本书的写作得益于我们早年在香港城市大学互联网挖掘实验室的科研经历。我们在相关领域的科研工作中，逐渐习得“计算传播学”的基本研究方法，深深理解作为文科或者社会科学背景的同学学习相关领域知识时的难点和痛点所在。希望这本教材能够作为“踏脚石”，帮助各位同学和老师顺利踏上“计算传播学”科研之路，护送各位披荆斩棘。

最后，我谨代表三位作者，衷心感谢策划编辑周粟和责编编辑马力敏的大力支持。二位老师对本书成稿和顺利出版倾注了大量的心血。马力敏老师对书稿提出了诸多细致的修订建议，使得本书行文更为流畅和规范。周粟老师亲力推进本书出版进度，并对本书的推广给予了大力支持。没有二位老师的帮助和投入，此书不可能如此顺利和迅速地出版。

张 伦

2018年9月25日



第一章 导 论	1
第一节 传统传播学简介	2
第二节 计算传播学之于传统传播学：新的研究范式	4
第三节 基于传统传播学的计算传播学研究	6
本章小结	12



第二章 文本分析简介	13
第一节 文本分析研究现状	14
第二节 文本分析与传播学研究	16
第三节 文本分析的基本步骤：文本的结构化处理	27
第四节 Python语言简介	38
本章小结	40



第三章 情感分析	41
第一节 情感分析简介	42
第二节 情感分析与传播学研究	43
第三节 Python进行情感分析	51
第四节 情感分析的基本算法	52
本章小结	66



第四章 语义建模	69
第一节 语义建模与传播学研究	70
第二节 LDA主题生成模型基本原理	75
第三节 语义模型的Python实现	78
第四节 有监督机器学习分类算法	80
本章小结	82



第五章 网络传播与传播网络	83
第一节 引 言	84
第二节 网络传播中的热点研究问题	86
第三节 社会网络的拓扑结构特征	94



第四节	传播网络的拓扑结构统计量	103
第五节	社会网络和传播网络之间的结构相关性	109
第六节	传播加权网络的多维度测量	113
第七节	传播时效网络的多维度测量	121
	本章小结	129



第六章	网络传播模型与机器学习框架	131
第一节	引 言	132
第二节	信息传播模型	133
第三节	信息传播的机器学习分析框架	147
第四节	影响信息传播的其他因素	149
第五节	特征选择方法	155
第六节	信息传播的机器学习评价指标	158
第七节	基于实证数据的信息流行度预测	161
	本章小结	169



第七章	数据新闻	171
第一节	产生背景	172
第二节	理论源流	174
第三节	实战练习	190
	本章小结	203



第八章	计算广告	205
第一节	引 言	206
第二节	发展历程	208
第三节	优化目标	212
第四节	计算广告市场的博弈、拍卖与匹配	214
第五节	计算广告市场的拍卖机制设计	225
	本章小结	230

后 记	232
-----	-----

第一章

导论

Just as the invention of the telescope revolutionized the study of the heavens, so too by rendering the unmeasurable measurable, the technological revolution in mobile, Web, and Internet communications has the potential to revolutionize our understanding of ourselves and how we interact.

Three hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin.

—Duncan Watts (2011, p. 266)

第一节 传统传播学简介

计算传播学是传播学的新兴分支之一。在介绍何为“计算传播学”之前，我们首先回顾经典传播学的研究范式。传播学是一门研究人类一切传播行为、传播过程及其规律的科学，旨在解释信息传播的机制和本质，描绘传播过程与传播效果，预测未来传播的形势和结构等。传播学按照研究层级，大致可分为人际传播（Interpersonal Communication）、组织传播（Organizational Communication）以及大众传播（Mass Communication）。本节从“研究范式”与“研究领域”两个角度，简要介绍经典传播学。^[1]

从传播学研究范式而言，传统传播学主要分为解释性范式（Explanatory）与阐释性范式（Interpretative）。前者重在解释因果，即给定研究问题的因变量，解释因变量的产生原因。这类研究范式多采用心理实验、社会调查、内容分析等定量研究方法。后者重在阐释因果关系的社会意义（Social Meanings）。这类研究范式多采用观察访谈、档案文献的话语分析和文本分析等定性研究方法进行。例如，利用符号学对于传播现象进行阐释与解读，就是典型的阐释性研究。

如果从研究领域对传播学进行划分，我们或许可以参考传播学最高规格的国际学会——国际传播学会（International Communication Association, ICA）对于研究分论坛的划分。2017年，国际传播学会有研究分支（Division）和研究小组（Interest Group）共计24个，如表1-1所示。

图1-1显示了2017年第67届国际传播学会每个分支领域的论坛（Panel）数。国际传播学会每年年会的论文按照研究分支或研究小组进行征稿。拟参会作者选择其论文应

[1] 王泽华. 新闻学和传播学之比较[J]. 中国广播电视学刊, 1992(2): 90~91.

表 1-1 国际传播学会研究领域划分

Children, Adolescents and Media (青少年与媒介)	Instructional and Developmental Communication (功能与发展传播)
Communication and Technology (传播与技术)	Intercultural Communication (跨文化传播)
Communication History (传播史)	Interpersonal Communication (人际传播)
Communication Law and Policy (传播法律与政策)	Journalism Studies (新闻学研究)
Computational Methods (计算方法)	Language and Social Interaction (语言与社会互动)
Environmental Communication (环境传播)	Mass Communication (大众传播)
Ethnicity and Race in Communication (传播过程中的民族与种族)	Organizational Communication (组织传播)
Feminist Scholarship (女性主义)	Philosophy, Theory and Critique (哲学、理论与批判)
Game Studies (游戏研究)	Political Communication (政治传播)
Global Communication and Social Change (全球传播与社会变迁)	Popular Communication (流行文化传播)
Health Communication (健康传播)	Public Relations (公共关系)
Information Systems (信息系统)	Visual Communication Studies (视觉传播)

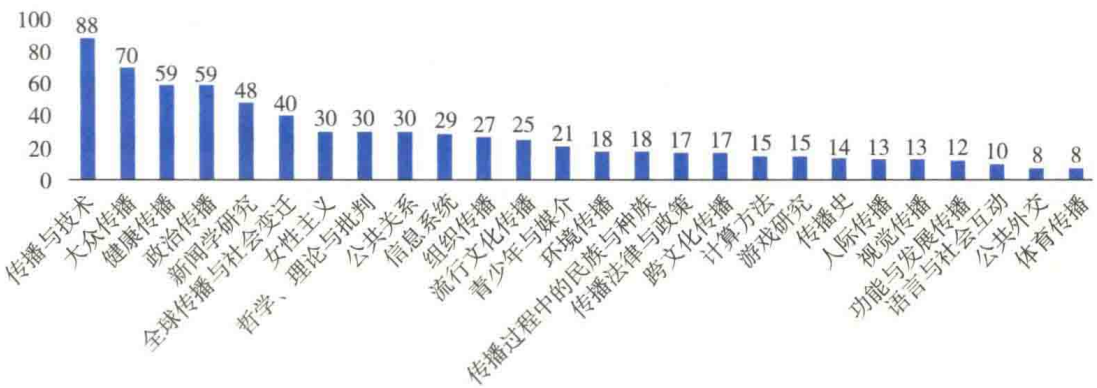


图 1-1 2017 年第 67 届国际传播学会论坛数

属的研究主题进行投稿。国际传播学会基于每个分支（或研究小组）收到的文章总数进行选拔和评审，并按照比较相近的录取率录取论文。因此，每年国际传播年会每个研究分支获得的论坛数基本反映了这个学科分支的论文投稿量，进而反映出该研究分

支或研究小组在当年的热度。

图1-1统计了2017年国际传播学会每个研究分支的论坛数，并按照论文报告论坛数进行排序。由图1-1可以看出，目前传播学研究的主流领域为传播与技术（Communication and Technology）、大众传播（Mass Communication）、健康传播（Health Communication）、政治传播（Political Communication）和新闻学研究（Journalism Studies），分别设立了88、70、59、59和48场论文报告论坛。但是，值得注意的是，2016年刚刚成立的兴趣小组“计算方法（Computational Methods）”于2017年第一次接收论文，就已经获得了15场报告论坛，超过7个传播学早已建立的传统研究领域。这足以说明，以计算方法进行传播学研究已经在一定程度上获得了传播学领域学者的认可。

第二节 计算传播学之于传统传播学：新的研究范式

一、社会科学研究的进路

正如耶鲁大学兼任物理学和社会学教授Christakis博士在《纽约时报》上发表的《让我们动摇社会科学吧》（“Let's Shake Up Social Sciences”）中所言，在过去的100年里，与自然科学相比，社会科学几乎停滞不前；其主要原因在于，社会科学遏制了新知识的创造。这也解释了为何社会科学没有获得如自然科学那般的声望。Golder和Macy^[1]也进一步阐述了社会科学传统研究范式所出现的弊端，即传统的社会科学研究过于强调理论和假设。传统社会科学研究有太多理论，而缺乏数据和恰当的方法对理论进行验证。

互联网和海量数据作为新的数据形式与来源，记录了很多传统社会科学数据获取方法所不能获得的信息。例如，连续的、具有地理和时间戳的时空信息；人们在社交网络中所表达的政治偏见等非结构化文本信息；商业交易信息和健康信息等。这些新型数据使得社会科学研究的已有弊端能够得以改进，其主要原因在于这些海量数据不仅具有数量优势，而且具备传统社会科学数据收集所无法获得的信息。首先，连续的时空信息（Continuous Time-Location Information）提供了个体交往的精确信息，包括交往时间、地点（如IP地址）等，这解决了个体社会生活难以观察的难题。其次，在线信息能够进行“非介入式观察”（Unobtrusively Observation），即不通过直接访问被研究者而对其行为进行研究。这很大程度上有利于研究者探寻无法通过传统的观察或调

[1] GOLDER S A, MACY M W. Digital footprints: Opportunities and challenges for online social research[J]. Annual Review of Sociology, 2014, 40: 129-152.

查方法而获知的人群或人类行为。例如,某些亚文化群体(如同性恋、艾滋病患者等)行为,或人们不愿承认的行为。例如,一项针对在线交友网站(Online Dating Site)的研究发现^[1],白人男性在线交友行为更倾向于规避黑人女性和其他有色人种,而白人女性则更倾向于规避亚裔人种。最后,这些数据能够记录个体的互动行为。社会科学的重要研究问题是人们的社会化行为以及相互影响。而传统社会科学,特别是以随机抽样社会调查为数据收集方法的研究,皆假设个体是独立的(Independent)。在线数据能够解放这个“独立性”假设,从而更好地探究人们具有互动特征的社会行为规律。^[2]

二、什么是“计算社会科学”与“计算传播学”?

2009年,包括Lazer, Pentland等多位著名学者共同署名的文章《计算社会科学》(“Computational Social Science”)^[3]发表,标志着“计算社会科学”这一新兴学科的建立。该文章定义了“计算社会科学”:通过对海量数据的采集和分析,旨在揭示人类个体和群体行为模式的新兴学科。具体而言,“计算社会科学”旨在通过对海量数据的收集、处理、存储,同时利用计算技术(如自动内容分类、语义建模、自然语言处理、模拟和统计模型)分析人类行为模式。

该文章发表4年后,至2013年,Watts^[4]对计算社会科学发展现状做了简要总结。该文章认为,计算社会科学至2013年,已经在“海量数据获取”“计算工具的使用”以及“跨学科合作”等几方面取得了显著进步。但该领域仍有一些重要问题没有解决。例如,对经济危机、传染病和社会运动等重要社会问题的探讨;而这些领域进展缓慢主要缘于社会问题的复杂性。未来,计算社会科学的发展,还需要在以下几个方向着力:对于跨平台数据(Multi-Source Data)的获取,能够帮研究者更全方位地观察个体行为;在线实验的平台搭建以及在线实验的执行,能够更有效地研究因果关系;最重要的是,社会科学家需要更深入地融入计算社会科学研究。社会科学问题的提出和解决,很大程度上依赖于社会科学家,而不是计算机或其他相关学科的研究者。

因此,在“计算社会科学”研究范式下,我们对于计算社会科学分支之一的计算传播学做出如下定义:计算传播学致力于寻找传播学可计算化的基因,以传播网络分析、传播文本挖掘、数据科学等为主要分析工具,大规模地收集并分析人类传播

[1] FELICIANO C, LEE R, ROBNETT B. Racial Boundaries among Latinos: Evidence from Internet daters' racial preferences[J]. Social Problems, 2011, 58(2): 189-212.

[2] KING G. Ensuring the data-rich future of the social sciences[J]. Science, 2011, 331(6018): 719-721.

[3] LAZER D, PENTLAND A, ADAMIC L, et al. Computational Social Science[J]. Science, 2009, 323(5915): 721-723.

[4] WATTS D J. Computational social science: Exciting progress and future directions[J]. The Bridge on Frontiers of Engineering, 2013, 43(4): 5-10.

行为数据，挖掘人类传播行为背后的模式和法则，分析模式背后的生成机制与基本原理。^[1]“新数据”“新方法”以及“重要问题”，是计算传播学不可或缺的三个重要元素。

对于初次接触“计算传播学”这一概念的读者，学习这一新兴领域，大概需要两部分知识：理论与方法。理论层面，包括社会网络、人类行为理论、公共意见形成与演化以及信息传播模型等。方法层面，又分为数据分析方法和编程软件的学习：前者包括社会网络分析（如指数随机图模式），文本挖掘，网络挖掘方法（如数据抓取、机器学习、深度学习），统计分析（如时间序列模型、空间分析），基于个体的模拟建模（Agent-Based Modeling）和可视化分析及技术；而后者则主要包括当前主流开源编程软件，如R、Python、Echart（可视化分析）以及部分商业软件（如Tableau等）。

第三节 基于传统传播学的计算传播学研究

计算传播学是对传统传播学的继承还是颠覆？在笔者看来，计算传播学的发展将深深根植于传统传播学研究的理论土壤，其根本目的是更好地回答传统传播学尚未回答的问题。而从研究问题来讲，计算传播学到底研究什么问题？本节将回到传播学研究的重要领域，从传播结构、传播模式、传播内容与效果以及传播实践四个部分分别举例加以说明（图1-2）。

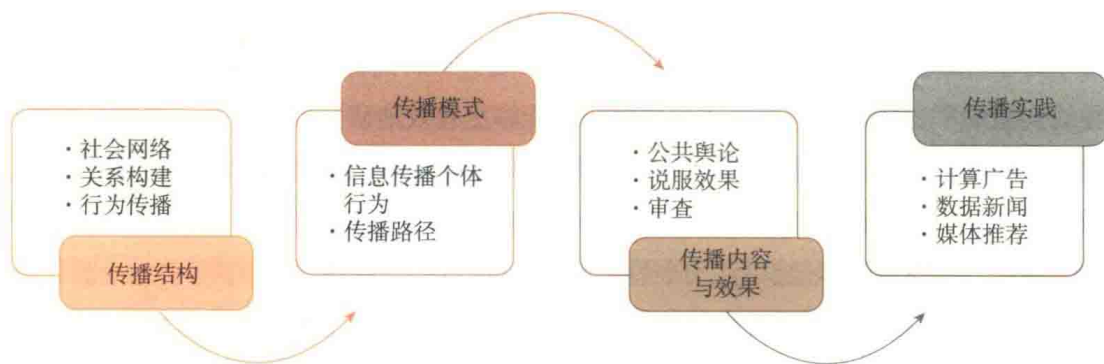


图1-2 计算传播学研究的理论视野

传播学研究自其诞生时起，旨在研究人类信息传播过程。传播过程大致可以分为传播结构、传播模式、传播内容与效果以及传播实践。传播结构指由社会关系构成的信息传播网络结构以及构成社会网络的个体的行为。传播模式从微观层面而言，主要指个体信息传播行为；从宏观层面而言，则指信息传播路径的规律和特点。传播内容

[1] 王成军. 计算传播学：作为计算社会科学的传播学. 中国网络传播研究, 2014(8): 193 ~ 206.

与效果旨在考察传播内容层面的特点以及内容特征所引发的传播效果。传播实践包罗万象,与计算传播学相关的实践领域主要包括数据新闻、计算广告以及媒体推荐算法等。本节将通过列举已有研究案例,依次阐述在上述四个传统传播学所关注的研究领域里,计算传播学如何进行研究,并进一步拓展这四个领域的研究视野和深度。

一、传播结构

对于传播结构的研究,我们需要回溯到社会网络研究。社会网络相关理论阐述了网络结构如何影响传播结构,或者引发信息传播行为发生变化的微观动力。在社会网络研究领域,著名的理论包括三元闭包原则、小世界理论、平衡理论、结构洞特征等。感兴趣的读者可仔细阅读《网络、群体与市场——揭示高度互联世界的行为原理与效应机制》^[1]这本书。该书通过跨学科的视角,深入论述了社会网络与信息传播行为的关系,本书在此不再赘述。

如何通过“计算传播学”的视角,来描述和探讨信息传播结构的规律?我们在此列举一个在线社交关系构建的例子。人们在社交媒体中进行信息传播,其基础结构在于人们的社会关系构建。如果将信息比喻为公路上行走的汽车,那么人们的在线社会关系就好比更为重要的高速公路等基础设施。

探讨在线社交媒体的好友关系构建,一项利用“人人网”早期的全网用户在线行为数据的研究发现,如果将用户在社交网络中的好友添加数量用归一化的累积分布函数(Cumulative Distribution Function)进行表示,则每个个体的自我网络(Ego Network)增长呈现出三种类别:S形曲线(用Logistic Function拟合)、双S形曲线(用Double Logistic Function拟合)以及倒L形曲线^[2](用Power Function拟合)(图1-3)。这三种自我网络的增长模式,反映了个体在社交网络中构建好友关系的不同策略。对于S形曲线类型,个体在构建好友关系时采用“探索”策略,即先小范围添加好友(这些好友大部分是线下已经认识的好友),再通过三元闭包等社会网络演化的微观动力,进行大规模好友关系构建,最终逐渐饱和。对属于双S形曲线类型好友关系构建的用户,其可能出现多个社交团体,即当个体通过“探索”策略完成了其某一个团体(如行政班级)中的好友关系构建之后,又发现了可能存在的新的团体(如兴趣小组、学生社团等),会再完成一遍好友的探索过程。而属于倒L形曲线类型好友关系构建的用户,其好友添加策略则为在进入在线平台后的第一时间迅速添加好友,随即停止添加好友。

该研究得以开展的前提是,首先,研究人员获得了在线社交网站中好友关系构建的时序数据。这类数据很难从线下或者通过调查、观察等其他渠道获得。其次,该研

[1] [美]大卫·伊斯利,乔恩·克莱因伯格. 网络、群体与市场——揭示高度互联世界的行为原理与效应机制[M]. 李晓明,王卫红,杨韞利,译. 北京:清华大学出版社,2011.

[2] ZHANG L, ZHU J J H. Regularity and variability: Growth patterns of online friendships[J]. International Journal of Web Services Research, 2014, 11(4): 19-31.

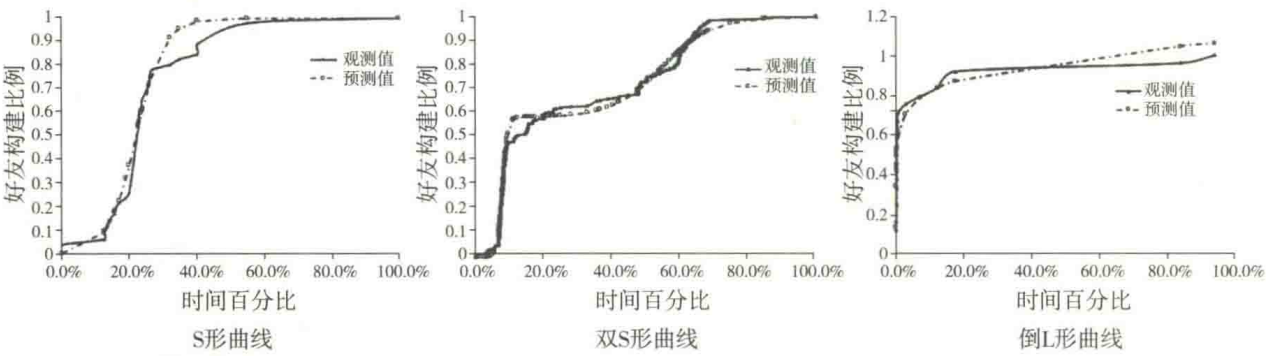


图 1-3 在线社交媒体好友关系构建增长曲线

究通过对在线好友添加行为的数据首先进行“数据导向 (Data-Driven)”的探索，探讨好友增长模式，再通过自动分类的方法，寻找出三种典型模式。这三种典型模式验证并拓展了社会关系增长的微观理论，是对信息传播结构领域研究的一次理论拓展。在该研究的基础上，未来的传播学研究还可以继续探究个体不同的自我网络增长模式对于其信息传播行为所产生的影响。

二、传播模式

对于传播模式的研究，在微观的个体信息传播层面，主要包括对以用户为节点的个体信息传播能力的研究，以及个体信息传播和接受行为研究。对于节点信息传播，在目前基于网络拓扑结构来刻画节点重要性的指标中，基于网络局部属性的主要有度值、局部中心性^[1]等，基于网络全局属性的有特征向量中心性、Katz 指标、紧度中心性、介数中心性、核函数^[2]、可达性 (Accessibility) 指标^[3]等，基于网络位置属性的有 k-核^[4]、混合度分解^[5]，基于随机游走的方法有 PageRank^{[6][7]}、HITS^[8]和

[1] CHEN D B, LÜ L Y, SHANG M S, et al. Identifying influential nodes in complex networks[J]. Physica a: Statistical mechanics and its applications, 2012, 391(4): 1777-1787.
[2] ZHANG J, XU X K, LI P, et al. Node importance for dynamical process on networks: A multiscale characterization[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2011, 21(1): 16107.
[3] TRAVENÇOLO B A N, COSTA L D F. Accessibility in complex networks[J]. Physics Letters A, 2008, 373(1): 89-95.
[4] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888.
[5] ZENG A, ZHANG C. Ranking spreaders by decomposing complex networks[J]. Physics Letters A, 2013, 377(14): 1031-1035.
[6] LANGVILLE A N, MEYER C D. Google's PageRank and beyond: The science of search Engine Rankings[M]. Princeton: Princeton University Press, 2011.
[7] FUSHIMI T, SAITO K, KIMURA M, et al. Finding relation between PageRank and voter model: Pacific Rim Knowledge Acquisition Workshop, 2010[C]. Berlin: Springer, 2010.
[8] KLEINBERG J M. Authoritative sources in a hyperlinked environment[J]. Journal of the ACM , 1999, 46(5): 604-632.