

# 社交网络上的 计算传播学

许小可 胡海波 张伦 王成军 编著

高等教育出版社

# 社交网络上的 计算传播学

SHEJIAO WANGLUO SHANG DE JISUAN CHUANBOXUE

许小可 胡海波 张伦 王成军 编著

高等教育出版社·北京

## 内容简介

本书着重分析了社交网络中影响传播行为的各种可计算因素,以网络科学为理论基础,以计算新闻学、计算广告学等为实践框架,全书共11章。第1~5章分析了社交网络中信息传播的测量工具和影响传播的可计算因素;第6~8章介绍社交网络中计算传播学的相关应用;第9~11章介绍与社交网络上计算传播学相关的一些研究方法和数据处理手段。

本书着力于融合多个学科的力量,系统性地为社交网络上的信息传播建立一个完整的可计算框架。本书可作为研究生以及高年级本科生的社交网络分析和计算传播学教材,也可供自然科学、工程技术和社会科学领域的研究人员和学者参考。

关键词:计算社会学,计算传播学,社交网络,信息传播,复杂网络,信息扩散

## 图书在版编目(CIP)数据

社交网络上的计算传播学/许小可等编著. -- 北京:高等教育出版社,2015.7

ISBN 978-7-04-043069-1

I. ①社… II. ①许… III. ①互联网络-传播学-研究 IV. ①G206

中国版本图书馆CIP数据核字(2015)第129510号

策划编辑 刘英 责任编辑 刘英 封面设计 李卫青 版式设计 于婕  
插图绘制 杜晓丹 责任校对 高歌 责任印制 田甜

出版发行 高等教育出版社  
社 址 北京市西城区德外大街4号  
邮政编码 100120  
印 刷 三河市吉祥印务有限公司  
开 本 787mm×1092mm 1/16  
印 张 20.25  
字 数 300千字  
购书热线 010-58581118

咨询电话 400-810-0598  
网 址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.landaco.com>  
<http://www.landaco.com.cn>  
版 次 2015年7月第1版  
印 次 2015年7月第1次印刷  
定 价 59.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换  
版权所有 侵权必究  
物料号 43069-00

# 前言

社交网络和社交新媒体已成为大众日常生活中不可缺少的组成部分，这些新的媒介改变了信息交流和信息传播的方式。本书最初命名为《社交网络上的信息传播》，考虑到信息传播这个名词不同领域的人对它的内涵与外延理解差异非常大，而本书侧重于信息传播可计算性方面的研究，就改成现在的书名。计算传播学是一个初露锋芒的研究领域，我们希望本书能够成为该领域的入门教科书。

我们四位作者来自三个不同的学科专业，信息科学、管理科学与传播学专业，这是一次真正意义上的跨学科合作。我们也来自不同的高等院校，能一起合作撰写书稿是因为我们都曾受教于香港城市大学祝建华教授。祝老师作为华人传播学领域的权威学者，他对我们在新媒体和社交网络方面的研究产生了深刻的影响。这种影响不仅改变了我们的知识结构，更改变了我们对自然科学和社会科学之间关系的看法。虽然自然科学与社会科学之间的边界早已变得模糊，但如何处理好二者之间的关系并使之相互借鉴相互补充却并不像表面上看起来那么容易，尤其是在计算传播学这一交叉学科领域。

自然科学与社会科学的碰撞与融合不是近来才有的事情。18世纪以来牛顿的物理学取得了巨大的成功，而它建

立在数学分析基础之上的定量方法成为一门成熟的学科必备的研究方法，由此导致 19 世纪成为社会科学中各学科模仿物理学的时代。例如，被誉为“社会学之父”的法国社会学家孔德明确将社会学称之为社会物理学，而英国经济学家李嘉图也被物理学吸引，倡导采用定量分析的方法使经济学获得更高的地位和认可度。

随着移动互联网技术的蓬勃发展，当自然科学与社会科学随着社交网络的兴起而再度重逢，两者都在以复杂的心态审视着对方。自然科学领域的学者开始关注社会系统，而社会科学的专家也已学会了自然科学所使用的各种复杂精致的分析工具。作为多年从事跨学科研究的人员，我们深刻体会到这种跨学科融合的强大威力。然而我们也深知，社会科学中并非所有的研究对象都可以量化。此外，社会科学的研究具有某种反身性，研究结果会对研究对象产生影响，比如对网络社区中用户行为的研究结果可能会对用户本身产生影响，对股票市场的研究也可能影响股民行为，这意味着在很大程度上社会科学的研究比自然科学要更为艰难。

近 50 年来，社会科学领域开始大量使用数学模型来进行研究，最明显的现象是经济学系开始大量招收来自数学系的研究生。但是对数学模型的滥用也导致了负面结果，凯根在《三种文化》中指出，这种滥用允许研究者沉浸在种种自负的迷官和无用的符号中，而对真实世界的种种复杂性和相互依赖视而不见。要求模型清晰性的代价是对复杂系统进行尽可能的简化，最终缺失了严格忠于实际存在现象的精神。因此在这本书中，我们虽然给出了信息传播的一些数学模型，但更多关注的是计算传播学的实证研究。从这个角度讲，我们自诩是实证研究的开拓者孔多塞、圣西门和孔德的继承人。

历史上不乏不同学科间合作经典的先例。比如物理学家为生物学家发明了显微镜；而作为回报，生物学家也参与了能量守恒定律的发现过程。同样，当新媒体与社交网络兴起，以此为研究对象的自然科学家和工程技术人员，也需要从社会科学关于人的思想、行为及人类社会的研究成果那里汲取营养。我们深知，自然科学与社会科学的研究对象和目标存在显著差异，它们之间在短时间内不可能完全融合，自然科学和社会科学的研究者应该采取一种更谦卑的姿态来审视自己，因为虽然每个团体在自己的领地内是强大的，但在他人的领地内却往往是无能为力的。在这里作为一个起点，同时作为我们为自己设定的目标，我们希望超越自然科学与社会科学传统上的对立与不成熟的融合，以一种发展的观点来审视不同学科间的关系。

在写作这本书的过程中，我们遇到了各种各样的困难。因为信息传播的文献浩如烟海，各个学科均有所涉及。在资料搜集过程中，我们不能采用仅在单一学科搜集材料的纵向搜索，而只能采用在多个学科中搜集的横向搜索策略。这些文献来自于社会学、传播学、计算机科学、物理学、管理学和社会心理学，在阅读这些文献的过程中，我们遇到了各种陌生的专业术语和概念，即使同一术语在不同学科中的含义也存在很大区别，比如 Attachment 在网络科学中我们马上想到了偏好连接，而在心理学中则译为依恋，如婴儿与照看者之间的关系；对于 Attribution 也很少有人知道在社会心理学中它译为归因(理论)。不同学科的研究假定和研究方法也存在显著差异，对于这些困难，没有跨学科的合作几乎是不可能克服的。

虽然写作初期，我们每位作者都有明确的分工，确保每人负责互不重叠的几个主题。但当完成初稿之后，发现仍然存在文稿重叠现象，这是因为计算传播学的每个分支

都有很强的外延性，跟其他分支都有千丝万缕的联系。当然读者看到的终稿中这种重叠已经被修正，每一章都自成一体，各章之间虽然有联系但即使随机挑选一章阅读也应该不太会有理解上的障碍。计算传播学是一个很庞大的体系，一本书或许不必每一页都仔仔细细读完，读者在阅读本书时可以直接跳过某些章节直接选择自己感兴趣的研究主题阅读。

本书的具体分工如下：许小可负责第1章、第9章和第10章，对第8章也有贡献，最后对全书进行了统稿；胡海波负责第4、5章和第8章，对第3章也有贡献；张伦负责第6、7章，对第4章也有贡献；王成军负责第2、3章和第11章，对第4章也有贡献。此外，尚可可对第10章有贡献；彭泰权、梁海、王成军、秦洁和陈鹤鑫对第6章有贡献。

本书的写作过程中也得到了许多专家学者的支持和帮助，其中汪小帆教授审阅了本书的重要章节，陈关荣、史定华和祝建华教授的建议使本书在各方面均有提高，李澄清教授对本书部分章节进行了文字润色。同时本书研究阶段得到了国家自然科学基金（61104143、61104139、61473119、61004104和61374170）以及辽宁省高等学校杰出青年学者成长计划的支持（LJQ2013125）。作者们特别感谢高等教育出版社刘英编辑对本书出版给予的大力支持，她的严格要求和鞭策让本书增色不少。最后，也非常感谢家人们对我们为这本书持续忙碌的理解和大力支持。

我们在这本书上付出了大量的热情和努力，这种努力远远超过我们撰写研究论文所付出的精力。即便如此，作为一个正在发展中的领域，很多的研究分支都尚不成熟，参考文献所拼凑出的领域图景仍支离破碎。一方面是作者们才学疏浅，另一方面也是该领域目前的真实写照。因此

我们希望这本书能抛砖引玉，引领更多的青年学者进入到这个正在成长中的充满挑战性的领域中。

10年前，当我们这些作者刚迈入科学研究大门，有幸阅读了沃尔德罗普的《复杂》、考夫曼的《科学新领域的探索》、巴克的《大自然如何工作》等一些学术大师的作品，由此对复杂系统产生了浓厚的兴趣。正是这些深入浅出的科普书籍引领我们进入了复杂性科学的领域，并一步步引导我们走进网络科学的宏大领域。我们满怀憧憬地希望这本书也能发挥这样的作用，至少能带动这一领域涌现出更多、更好的学术作品。同时，我们也希望通过本书结交更多的朋友，共筑我们的精神家园。“人以群分”既是社交网络重要的基础理论，也是中国传统文化中高山流水遇知音的一种准确描述。我们今后和读者们交流的重要渠道是计算传播学网站 <http://computational-communication.com/>。

由于成书时间仓促，加之我们几位年轻作者水平有限，错误在所难免，敬请各位读者批评指正。我们非常希望收到读者的反馈，我们也会适时更新本书的内容，在条件允许的情况下，我们会开发与本书配套的网站，不仅会提供每一章的讲义和课件，也会提供其他相关的补充材料。作为一个正在成长中的领域，相信随着研究的深入，以后陆续会出版计算传播学相关的图书，从而构建出从理论研究到实践应用的完整体系。

许小可 胡海波 张伦 王成军

2015年5月



## 作者简介

许小可，大连海事大学通信与信息系统专业博士，香港理工大学博士后，现为大连民族大学信息与通信工程学院教授。主要从事在线社交分析和非线性时间序列方面的研究工作，2013年获CCF-腾讯犀牛鸟科研基金，阿里巴巴数据平台创新大赛一等奖。电子邮箱：xiaokeeie@gmail.com。

胡海波，上海交通大学控制理论与控制工程专业博士，现为华东理工大学管理科学与工程系副教授。主要研究领域为复杂系统、社交网络，曾在中科院自动化所、香港城市大学从事网络科学相关的研究工作，近年来致力于社会网络在社会经济系统中的应用研究。电子邮箱：sdhuzi@gmail.com。

张伦，香港城市大学媒体与传播学博士，现为中国科学院大学新闻传播学系讲师。研究方向为社会化媒体中用户行为分析，研究成果发表于 *Computers in Human Behavior*, *Internet Research* 以及 *New Media & Society* 等 SSCI 及 SCI 期刊。电子邮箱：zhanglun1525@gmail.com。

王成军，香港城市大学媒体与传播学博士，现为南京大学新闻传播学院助理研究员。主要研究兴趣为采用计算社会科学视角分析人类传播行为。电子邮箱：wangchj04@gmail.com。

## 郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010) 58581897 58582371 58581879

反盗版举报传真 (010) 82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

# 目录

第1章 引论 .....	1
1.1 社交网络分析：从大众传播到人际传播的 必然选择 .....	2
1.2 网络科学领域的结构主义倾向 .....	2
1.3 计算传播学简介 .....	3
1.4 基于计算传播学的信息传播过程可计算框架 .....	4
1.5 本书内容简介 .....	5
第2章 单条信息传播的多维度测量 .....	7
2.1 引言 .....	8
2.2 扩散规模及扩散率 .....	9
2.3 扩散网络 .....	9
2.3.1 高度和宽度 .....	10
2.3.2 级联率 .....	11
2.4 扩散阈值 .....	13
2.5 扩散时间 .....	17
2.5.1 速度 .....	19
2.5.2 爆发与持续性 .....	20
2.6 扩散空间 .....	24
2.7 本章小结和展望 .....	26

参考文献 .....	26
<b>第 3 章 影响信息传播的可计算因素分析 .....</b>	<b>29</b>
3.1 引言 .....	30
3.2 社交网络结构特征对传播的影响 .....	30
3.2.1 网络连通性 .....	31
3.2.2 度相关性 .....	32
3.2.3 空间结构 .....	33
3.2.4 Hub 节点 .....	34
3.2.5 小世界特性 .....	35
3.2.6 信息传播对于社交网络结构演化的反作用 .....	36
3.3 信息属性的影响 .....	39
3.4 用户属性以及属性相似性的影响 .....	41
3.4.1 用户自身属性 .....	41
3.4.2 好友属性的趋同性 .....	44
3.4.3 好友间的同伴作用 .....	45
3.5 用户阵发活动模式的影响 .....	47
3.6 本章小结与展望 .....	49
参考文献 .....	50
<b>第 4 章 节点传播能力的测量 .....</b>	<b>55</b>
4.1 引言 .....	56
4.2 社交网络中的意见领袖 .....	56
4.2.1 意见领袖与二级传播理论 .....	56
4.2.2 意见领袖成因分析 .....	57
4.3 基于网络结构的节点传播能力测量 .....	59
4.3.1 节点连接度 .....	60
4.3.2 结构洞特征 .....	60
4.3.3 PageRank 算法 .....	61

4.3.4	LeaderRank 算法	62
4.3.5	$k$ -核分解算法	63
4.4	基于信息传播的节点传播能力测量	65
4.4.1	有效用户数量	65
4.4.2	被转发次数与被提及次数	66
4.4.3	TunkRank 排序算法	67
4.4.4	明星指数	68
4.5	影响最大化问题	68
4.6	本章小结与展望	71
	参考文献	72
<b>第 5 章</b>	<b>信息扩散模型</b>	<b>79</b>
5.1	引言	80
5.2	经典的信息扩散模型	81
5.2.1	创新扩散模型	81
5.2.2	局部相互作用博弈模型	84
5.2.3	阈值模型	85
5.2.4	级联模型	86
5.2.5	Markov 随机场图模型	87
5.2.6	传染病模型	87
5.3	数据驱动的信息扩散模型	88
5.3.1	分支过程模型	88
5.3.2	数据驱动的级联模型	92
5.3.3	基于信息亲和性的模型	93
5.3.4	基于信息扩散升降模式的模型	95
5.3.5	基于外部影响的模型	96
5.3.6	基于权威节点影响的模型	97
5.3.7	其他数据驱动模型	98
5.4	信息扩散模型的最新进展	99

5.4.1 谣言传播模型 .....	99
5.4.2 竞争性的信息扩散模型 .....	101
5.4.3 其他模型 .....	104
5.5 本章小结与展望 .....	105
参考文献 .....	106
<b>第6章 计算传播学在新闻和公共舆论领域的应用 .....</b>	<b>111</b>
6.1 引言 .....	112
6.2 传播者研究 .....	114
6.2.1 新闻传播中传播者研究的现状 .....	114
6.2.2 公众参与公共舆论传播的不平等性 .....	115
6.3 受众研究 .....	116
6.4 内容研究 .....	118
6.5 渠道研究 .....	121
6.6 效果研究 .....	122
6.6.1 公共舆论与受众的认知、态度及行为改变 .....	122
6.6.2 公共舆论的宏观效果 .....	123
6.6.3 在线公共舆论与社会运动 .....	124
6.6.4 社会化计算与在线公共舆论效果研究 .....	127
6.7 本章小结与展望 .....	128
参考文献 .....	130
<b>第7章 计算传播学在电子商务中的应用 .....</b>	<b>135</b>
7.1 引言 .....	136
7.2 消费者参与口碑传播的心理动机 .....	137
7.3 消费者参与口碑传播的行为分析 .....	138
7.3.1 用户评分的统计特征 .....	138
7.3.2 用户评分的时间特征 .....	139
7.4 用户评论内容的文本分析 .....	140

7.4.1	用户评论质量的判定	141
7.4.2	用户评论的语义分析	141
7.4.3	用户评论的情感倾向分析	143
7.5	口碑营销的效果测量	144
7.5.1	口碑营销的信息传播效果	144
7.5.2	口碑营销中信息传播的影响因素	145
7.5.3	口碑传播对品牌认知的影响	147
7.5.4	口碑传播对购买行为的影响	148
7.6	广告受众的精准定向	149
7.6.1	基于用户行为的受众定向	150
7.6.2	基于社会关系的受众定向	151
7.6.3	基于用户自创内容的受众定向	152
7.7	本章小结与展望	154
	参考文献	156
<b>第8章</b>	<b>行为、情绪和感受的计算传播学</b>	<b>163</b>
8.1	引言	164
8.2	行为的计算传播学研究	165
8.2.1	恶性行为的传播	165
8.2.2	合作行为的传播	167
8.2.3	导致肥胖的行为传播	168
8.2.4	吸烟行为的传播	171
8.2.5	饮酒行为的传播	174
8.2.6	睡眠行为的传播	177
8.3	情绪和感受的计算传播学研究	179
8.3.1	幸福感的传播	179
8.3.2	孤独感的传播	183
8.3.3	抑郁的传播	185
8.3.4	不同极性情绪的传播特性差异	188

8.3.5 情绪传播研究中的学术道德 .....	189
8.4 影响行为、情绪和感受传播的因素分析 .....	191
8.4.1 网络结构的影响 .....	191
8.4.2 趋同性的影响 .....	193
8.5 社会影响、趋同性与共因原理 .....	196
8.6 本章小结与展望 .....	198
参考文献 .....	199

## 第9章 社交网络的抽样策略及在计算传播学中的

应用 .....	209
9.1 在线社交网络的常用抽样策略 .....	210
9.1.1 随机抽样 .....	210
9.1.2 广度优先搜索抽样 .....	212
9.1.3 滚雪球抽样 .....	212
9.1.4 森林火灾抽样 .....	213
9.1.5 随机行走抽样 .....	214
9.1.6 Metropolis-Hastings 随机行走抽样 .....	215
9.1.7 有偏随机行走抽样 .....	216
9.1.8 融合随机抽样和依存式抽样的混合抽样 .....	216
9.2 评估和比较各种抽样策略的性能 .....	217
9.2.1 常用统计量 .....	217
9.2.2 常用统计方法 .....	218
9.2.3 性能比较 .....	219
9.3 其他类型社交网络中的抽样策略 .....	221
9.3.1 有向社交网络的抽样 .....	221
9.3.2 具有强社团结构网络的抽样 .....	222
9.4 局部大数据的分析与应用 .....	222
9.4.1 局部大数据与随机样本数据的区别 .....	222
9.4.2 局部数据中边界效应的影响 .....	223



9.5 信息传播研究中的抽样策略 .....	226
9.5.1 衡量信息传播抽样效果的统计量 .....	226
9.5.2 基于网络拓扑结构的抽样 .....	227
9.5.3 基于信息传播特性的抽样策略 .....	227
9.5.4 融合信息传播特性和拓扑结构的抽样 .....	228
9.5.5 信息源确定与信息传播路径重构 .....	229
9.6 本章小结与展望 .....	231
参考文献 .....	232

## 第 10 章 基于置乱算法的社交网络零模型构造及

应用 .....	237
10.1 引言 .....	238
10.2 静态无权网络零模型的构造 .....	239
10.2.1 基于网络模型构造零模型的局限性 .....	240
10.2.2 基于随机断边重连的 1 阶零模型 .....	240
10.2.3 强同配特性或异配特性的断边重连零模型 .....	242
10.2.4 基于随机断边重连的 2 阶零模型 .....	243
10.2.5 局部断边重连算法 .....	244
10.3 静态加权网络零模型的构造 .....	246
10.3.1 权重置乱算法 .....	246
10.3.2 等权重置乱算法 .....	247
10.3.3 局部权重置乱算法 .....	248
10.4 时变网络零模型的构造 .....	249
10.4.1 时变网络简介 .....	250
10.4.2 连边置乱算法 .....	251
10.4.3 时间置乱算法 .....	251
10.4.4 时间随机化算法 .....	253
10.4.5 时权置乱算法 .....	254
10.4.6 等权置乱算法 .....	255