



国家科学技术学术著作出版基金资助

网络科学与工程丛书

N S E

11

# 非对称信息共享网络 理论与技术

Asymmetric Information  
Sharing Network  
Theory and Technology

■ 任 勇 徐 蕾 姜春晓 王景璟 杜 军 编著

高等教育出版社



国家科学技术学术著作出版基金资助

N S E

11

网络科学与工程丛书

# 非对称信息共享网络 理论与技术

Asymmetric Information  
Sharing Network  
Theory and Technology

■ 任勇 徐蕾 姜春晓 王景璟 杜军 编著

高等教育出版社·北京

## 图书在版编目 (CIP) 数据

非对称信息共享网络理论与技术 / 任勇等编著. --  
北京: 高等教育出版社, 2019.6  
ISBN 978-7-04-051855-9

I. ①非… II. ①任… III. ①网络信息资源-研究  
IV. ①G255.76

中国版本图书馆 CIP 数据核字 (2019) 第 077348 号

策划编辑 刘 英      责任编辑 刘 英      封面设计 李卫青      版式设计 马敬茹  
插图绘制 于 博      责任校对 高 歌      责任印制 赵义民

---

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 三河市潮河印业有限公司  
开 本 787mm × 1092mm 1/16  
印 张 21.25  
字 数 380 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.hepmall.com.cn>  
<http://www.hepmall.com>  
<http://www.hepmall.cn>  
版 次 2019 年 6 月第 1 版  
印 次 2019 年 6 月第 1 次印刷  
定 价 69.00 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 51855-00

## 作者简介

任勇，清华大学教授，博士生导师。1999年创建清华大学电子系复杂工程系统实验室并担任主任，从事复杂系统理论与方法及其在通信网络与电子系统中的应用研究。近年来承担了多项国家自然科学基金研究项目，发表学术论文百余篇，申请发明专利多项。获省部级科技进步二等奖2项、三等奖1项，清华大学优秀教学成果一等奖1项、二等奖1项，清华大学科技成果转化效益显著奖2项。



徐蕾，北京理工大学计算机学院特别副研究员。2008年于清华大学电子工程系获工学学士学位，2015年获清华大学信息与通信工程专业博士学位。主要研究方向为网络信息共享、大数据中的隐私保护等。近年来参与多项国家自然科学基金项目，在国内外刊物、会议上发表论文10余篇。



姜春晓，清华大学信息科学技术学院助理研究员。2008年于北京航空航天大学获工学学士学位，2013年于清华大学电子工程系获博士学位。现从事空间信息网络通信理论研究，发表IEEE期刊论文64篇，谷歌学术引用1300余次。曾获2013年IEEE Globecom最佳论文奖，2014年中国人工智能学会优秀博士论文，2015年IEEE GlobalSIP最佳论文奖，2016年入选中国科协“青年人才托举工程”。IEEE高级会员。



王景璟，清华大学电子工程系博士研究生。2014年于大连理工大学获工学学士学位。近年来参与多项国家自然科学基金项目，先后在国外期刊、会议发表论文10余篇。主要研究方向为复杂网络建模、信息传播理论、网络化信息处理与决策等。



杜军，清华大学电子工程系博士后。2009年于北京理工大学获工学学士学位，2014年、2018年于清华大学分别获得工学硕士及博士学位。近年来参与多项国家自然科学基金项目，先后在国外期刊、会议发表论文30余篇。主要研究方向为空间信息网络资源协同与优化理论、信息传播理论、网络经济学理论及应用等。



## “网络科学与工程丛书” 编审委员会

名誉主编：郭 雷院士    金芳蓉院士    李德毅院士

主    编：陈关荣

副 主 编：史定华    汪小帆

委    员：（按汉语拼音字母排序）

曹进德    陈增强    狄增如    段志生

方锦清    傅新楚    胡晓峰    贾  韬

来颖诚    李  翔    刘宗华    陆君安

吕金虎    吕琳媛    任  勇    汪秉宏

王青云    王文旭    谢智刚    周  涛

# 序

随着以互联网为代表的网络信息技术的迅速发展，人类社会已经迈入了复杂网络时代。人类的生活与生产活动越来越多地依赖于各种复杂网络系统安全可靠和有效的运行。作为一个跨学科的新兴领域，“网络科学与工程”已经逐步形成并获得了迅猛发展。现在，许多发达国家的科学界和工程界都将这个新兴领域提上了国家科技发展规划的议事日程。在中国，复杂系统包括复杂网络作为基础研究也已列入《国家中长期科学和技术发展规划纲要（2006—2020年）》。

网络科学与工程重点研究自然科学技术和社会政治经济中各种复杂系统微观性态与宏观现象之间的密切联系，特别是其网络结构的形成机理与演化方式、结构模式与动态行为、运动规律与调控策略，以及多关联复杂系统在不同尺度下行为之间的相关性等。网络科学与工程融合了数学、统计物理、计算机科学及各类工程技术科学，探索采用复杂系统自组织演化发展的思想去建立全新的理论和方法，其中的网络拓扑学拓展了人们对复杂系统的认识，而网络动力学则更深入地刻画了复杂系统的本质。网络科学既是数学中经典图论和随机图论的自然延伸，也是系统科学和复杂性科学的创新发展。

为了适应这一高速发展的跨学科领域的迫切需求，中国工业与应用数学学会复杂系统与复杂网络专业委员会偕同高等教育出版社出版了这套“网络科学与工程丛书”。这

套丛书将为中国广大的科研教学人员提供一个交流最新研究成果、介绍重要学科进展和指导年轻学者的平台，以共同推动国内网络科学与工程研究的进一步发展。丛书在内容上将涵盖网络科学的各个方面，特别是网络数学与图论的基础理论，网络拓扑与建模，网络信息检索、搜索算法与数据挖掘，网络动力学（如人类行为、网络传播、同步、控制与博弈），实际网络应用（如社会网络、生物网络、战争与高科技网络、无线传感器网络、通信网络与互联网），以及时序网络分析（如脑科学、心电图、音乐和语言）等。

“网络科学与工程丛书”旨在出版一系列高水准的研究专著和教材，使其成为引领复杂网络基础与应用研究的信息和学术资源。我们殷切希望通过这套丛书的出版，进一步活跃网络科学与工程的研究气氛，推动该学科领域知识的普及，并为其深入发展做出贡献。

金芳蓉 (Fan Chung) 院士  
美国加州大学圣迭戈分校  
二〇一一年元月

# 前言

经过几十年的发展，当前互联网的规模与应用方式已与其诞生之初大不相同。特别地，以 WWW、P2P 为代表的信息共享业务已成为目前互联网的主流业务。信息共享业务的非对称传输特征是导致网络结构、服务模式产生矛盾而造成带宽浪费的原因。如何解决互联网在承载信息共享业务方面存在的缺陷，这一问题引起了作者和国内外相关领域研究者的广泛关注。

作者在“十五”和“十一五”期间先后承担了多项与信息共享理论及其关键技术相关的国家级项目。通过国家自然科学基金项目“共享信息网络新结构及关键技术研究”（60672142）研究了信息网络结构和信息共享模式的改进方法，提出了一种信息汇聚模型。通过国家自然科学基金重点项目“非对称广域覆盖信息共享网络理论与技术”（60932005）研究了信息共享业务的非线性整体动力学特征，提出了以信息资源汇聚、用户广泛共享及偏远地区覆盖为特征的信息共享服务模式。通过国家自然科学基金面上项目“内容共享网络的非对称性及流量优化理论与策略”（61271267）研究了共享业务的非对称性和动态演化规律，探索了拓扑失配问题的解决方案和分布式缓存机制的优化方法。

全书分为五篇。第一篇，信息共享网络理论与体系结构，介绍了信息共享的基本概念和网络中的非对称现象，分析了常见的信息共享网络体系结构与服务模式，讨论了信息共享网络的新体系。第二篇，信息共享网络的拓扑与动力学

特性，介绍了典型的网络演化模型，提出了基于效益的网络演化模型，分析了 P2P 网络的整体统计特征，用博弈论的方法对 P2P 信息共享行为进行了分析。第三篇，信息共享网络的拓扑动态性，介绍了网络容量的基本模型，提出了基于网络流量分布的网络容量优化方法，探讨了拓扑失配问题的解决方案。第四篇，非对称信息共享网络中的信息处理，介绍了网络信息的获取方法和 Web 文本挖掘方法，提出了面向特定领域的网络信息聚合方法。第五篇，网络信息共享中的隐私安全，对数据挖掘、在线社交网络中的隐私等安全问题进行了分析，提出了“反挖掘”研究框架，探讨了博弈论、契约理论在个人信息共享决策中的应用。

本书的撰写汇集了很多人的辛勤劳动。任勇教授负责全书的内容规划和体系架构搭建，以及撰写过程中的组织和协调工作，孙若姿、李一鹏、叶王毅、彭兆盟、孙权、桂琨智等人参加了相关章节的初稿编写，徐蕾、杨超群、邵旦参加了部分章节的初稿和终稿编写；此外，本书引述了山秀明、袁坚、姜春晓、王景璟、许晨敏、何能强等人的前期工作。全书由徐蕾统稿，姜春晓、王景璟对全书进行了校对。在此一并表示衷心的感谢。

同时也感谢在本书撰写前后，很多前辈、同行和学者对于信息共享理论的热情参与、研讨，以及给予我们的鼓励。

由于作者水平有限，书中难免存在不妥之处，恳请读者批评指正。

作者

2018年5月于清华大学

# 目录

## 第一篇 信息共享网络理论与体系结构

第 1 章 非对称现象及互联网的非对称性 .....	3
1.1 对称与非对称 .....	4
1.1.1 对称的定义 .....	4
1.1.2 对称与守恒的关系 .....	7
1.1.3 非对称的定义 .....	8
1.2 非对称现象 .....	9
1.2.1 宇宙中的非对称 .....	9
1.2.2 生命中的非对称 .....	12
1.2.3 经济学中的非对称 .....	14
1.2.4 多样性与非对称 .....	16
1.3 网络中的非对称 .....	17
1.3.1 互联网业务的非对称 .....	17
1.3.2 社交网络中的非对称 .....	17
参考文献 .....	19
第 2 章 网络信息共享的内涵与外延 .....	23
2.1 网络信息的基本概念与特征 .....	24
2.1.1 信息的定义 .....	24
2.1.2 信息的分类 .....	26
2.1.3 信息的基本性质 .....	28

2.2 网络信息共享的概念 .....	29
2.2.1 网络信息的涵义和特征 .....	29
2.2.2 信息不守恒原理 .....	32
2.2.3 信息的共享增益 .....	33
2.3 信息共享系统模型 .....	34
2.3.1 信息共享系统的基本特征 .....	34
2.3.2 信息共享系统的一般模型 .....	35
2.3.3 信息共享系统的栅格模型 .....	36
参考文献 .....	38
<b>第3章 信息共享网络的体系结构与服务模式 .....</b>	<b>41</b>
3.1 网络体系结构 .....	42
3.1.1 OSI 模型 .....	42
3.1.2 TCP/IP .....	45
3.2 网络服务模式 .....	47
3.2.1 C/S 模式 .....	48
3.2.2 P2P 模式 .....	49
3.2.3 广播模式 .....	52
3.3 网络共享技术 .....	52
3.3.1 网格计算技术 .....	53
3.3.2 云计算技术 .....	54
3.3.3 CDN 技术 .....	56
参考文献 .....	58
<b>第4章 信息共享的新体系与新模式 .....</b>	<b>61</b>
4.1 以内容共享为中心的未来网络 .....	62
4.1.1 典型技术 .....	62
4.1.2 典型方案 .....	64
4.1.3 未来研究方向 .....	67
4.2 非对称广域覆盖的信息共享网络 .....	68
4.3 社会化媒体 .....	70
4.3.1 社会化媒体的定义 .....	70
4.3.2 社会化媒体的分类 .....	71
4.3.3 微信 .....	72

参考文献 .....	74
------------	----

## 第二篇 信息共享网络的拓扑与动力学特性

<b>第 5 章 基于效用的网络演化模型 .....</b>	<b>79</b>
5.1 网络演化模型的基本概念 .....	80
5.1.1 复杂网络拓扑性质 .....	80
5.1.2 常见的网络演化模型 .....	81
5.2 基于兴趣效用的网络演化模型 .....	82
5.2.1 基于效用的网络演化基本算法 .....	82
5.2.2 基于兴趣的效用度量 .....	83
5.2.3 仿真与理论分析 .....	84
5.3 实证研究 .....	87
5.3.1 社交网络分析 .....	87
5.3.2 仿真和实际数据对比分析 .....	89
参考文献 .....	95
<b>第 6 章 P2P 网络相变的非线性成因及整体对策 .....</b>	<b>97</b>
6.1 应用层网络与物理网络双重耦合模型简介 .....	98
6.1.1 网络性能的相变与分布式代理对策 .....	99
6.1.2 实验结果及分析 .....	102
6.2 P2P 网络流量局部化整体对策 .....	105
6.2.1 改进的二维元胞自动机模型 .....	106
6.2.2 P2P 网络流量局部化对策 .....	108
6.2.3 实验结果及分析 .....	109
参考文献 .....	113
<b>第 7 章 P2P 信息共享行为分析 .....</b>	<b>115</b>
7.1 P2P 网络的不稳定性 .....	116
7.2 博弈策略对 P2P 网络稳定性的影响 .....	116
7.2.1 囚徒困境 .....	117
7.2.2 雪堆博弈 .....	118
7.3 合作率的影响因素 .....	119
7.4 基于累积收益的雪堆博弈模型 .....	120

7.4.1 基本假设 .....	120
7.4.2 模型的建立 .....	120
7.5 典型网络模型中的仿真 .....	122
7.5.1 数据来源和数据格式 .....	122
7.5.2 APBSG 在随机网络下的仿真结果 .....	123
7.5.3 APBSG 在 WS 小世界网络下的仿真 结果 .....	125
7.5.4 APBSG 在 BA 无标度网络下的仿真 结果 .....	127
参考文献 .....	128

### 第三篇 信息共享网络的拓扑动态性

<b>第 8 章 网络容量与流量分布</b> .....	133
8.1 网络容量基本模型 .....	134
8.2 网络容量优化方法 .....	136
8.3 网络流量分布对网络容量的影响 .....	137
8.3.1 路由距离 .....	137
8.3.2 缓存节点的分布 .....	139
8.3.3 服务器节点的分布 .....	141
参考文献 .....	142
<b>第 9 章 信息共享网络的拓扑失配和控制问题</b> .....	145
9.1 拓扑失配的概念及影响 .....	146
9.2 拓扑失配的解决方案 .....	148
9.2.1 基于网络延迟的方法 .....	148
9.2.2 基于位置信息的 Overlay 匹配方法 .....	149
9.2.3 基于用户兴趣的 Overlay 匹配方法 .....	149
9.3 拓扑控制 .....	150
9.3.1 基本概念 .....	151
9.3.2 均一功率拓扑控制算法 .....	152
9.3.3 非均一功率拓扑控制算法 .....	152
参考文献 .....	154

## 第四篇 非对称信息共享网络中的信息处理

<b>第 10 章 网络信息获取</b> .....	161
10.1 网络信息资源获取 .....	162
10.1.1 信息获取方式概述 .....	162
10.1.2 搜索引擎 .....	163
10.1.3 信息推送 .....	165
10.2 网络爬虫 .....	167
10.2.1 网络爬虫基本原理 .....	167
10.2.2 面向主题的网络爬虫 .....	170
10.2.3 AJAX 网络爬虫 .....	173
10.3 网页内容提取 .....	175
10.3.1 网页净化 .....	176
10.3.2 Web 信息抽取 .....	177
10.3.3 基于 Xpath 网页内容提取 .....	178
参考文献 .....	179
<b>第 11 章 Web 文本挖掘</b> .....	183
11.1 文本挖掘基础 .....	184
11.1.1 文本表示 .....	184
11.1.2 文本分类 .....	186
11.1.3 文本聚类 .....	187
11.1.4 文本摘要 .....	189
11.2 微博信息挖掘 .....	191
11.2.1 微博信息挖掘概述 .....	191
11.2.2 基于深度学习的微博文本降维 .....	193
参考文献 .....	204
<b>第 12 章 面向特定领域的网络信息聚合</b> .....	207
12.1 语义网 .....	208
12.1.1 语义网体系结构 .....	208
12.1.2 本体研究 .....	210
12.1.3 语义网的应用 .....	212
12.2 信息集成 .....	213

12.2.1	信息集成概述 .....	213
12.2.2	基于本体的信息集成 .....	215
12.2.3	示例：面向农产品价格领域的信息集成框架 .....	217
	参考文献 .....	231

## 第五篇 网络信息共享中的隐私安全

<b>第 13 章</b>	<b>个人信息共享与隐私安全 .....</b>	<b>235</b>
13.1	在线社交网络中的隐私问题 .....	236
13.1.1	隐私悖论与隐私计算 .....	236
13.1.2	社交网络中的隐私度量 .....	237
13.2	推荐系统中的隐私问题与用户博弈策略 .....	239
13.2.1	协同过滤推荐系统 .....	240
13.2.2	系统模型 .....	241
13.2.3	满足博弈 .....	242
13.2.4	均衡学习 .....	244
13.2.5	收敛性分析 .....	246
13.2.6	仿真分析 .....	253
	参考文献 .....	258
<b>第 14 章</b>	<b>数据挖掘中的隐私保护 .....</b>	<b>261</b>
14.1	反挖掘研究框架 .....	262
14.1.1	关键问题 .....	262
14.1.2	反挖掘的基本构成 .....	264
14.1.3	研究现状 .....	265
14.1.4	未来研究方向 .....	274
14.2	博弈论在隐私保护数据发布中的应用 .....	278
14.2.1	博弈模型描述 .....	280
14.2.2	子博弈精炼纳什均衡 .....	282
14.2.3	分析示例： $k$ 匿名 .....	283
14.2.4	仿真分析 .....	287
	参考文献 .....	289

第 15 章 个人信息共享中的理性决策 .....	293
15.1 隐私拍卖与隐私定价 .....	294
15.1.1 隐私拍卖 .....	294
15.1.2 隐私定价 .....	295
15.2 基于契约的隐私数据交易 .....	296
15.2.1 问题建模 .....	297
15.2.2 契约设计 .....	300
15.2.3 仿真分析 .....	309
参考文献 .....	312
结论与展望 .....	315