



经典译丛

信息网络技术与网络科学

Networks: An Introduction

网络科学 引论

【美】 M. E. J. Newman 著

郭世泽 陈哲 译

Networks
An Introduction



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

经典译丛·信息技术与网络科学

网络科学引论

Networks: An Introduction

[美] M. E. J. Newman 著

郭世泽 陈 哲 译



电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

作者凭借在计算机、信息论、物理等相关学科的深入研究和丰富经验,系统地分析和论述了网络作为一门科学理论如何应用在现实生活中的方方面面。全书分为5部分,讨论了目前科学研究中的网络类型和用以确定其结构的各种技术,介绍了研究网络的基本数学理论及用以量化网络结构的各类测度与参数,描述了有效分析网络数据的计算机算法,以及有助于预测网络系统行为并理解其生成和演化过程的网络结构数学模型,最后给出了网络上的一些动力学过程,如社会网络中的疾病传染或计算机网络上的搜索过程。

本书不仅是从事网络科学领域的研究人员、专业人员和技术人员必不可少的参考资料,而且还非常适合作为普通高等院校计算机科学与技术、通信工程、应用数学、生物科学、应用物理学、社会学等专业高年级本科生和研究生课程的教材。

Copyright ©2010 by M. E. J. Newman.

Networks: An Introduction was originally published by arrangement with Oxford University Press for sale/distribution in The Mainland (part) of the People's Republic of China (excluding the territories of Hong Kong SAR, Macau SAR and Taiwan) only and not for export therefrom.

本书简体中文版专有出版权由 Oxford University Press, Inc., U. S. A. 授予电子工业出版社出版。未经出版者书面许可,不得以任何方式复制和抄袭本书的任何部分。

本书在中国大陆地区生产,仅限在中国大陆发行。

版权贸易合同登记号 图字:01-2012-3036

图书在版编目(CIP)数据

网络科学引论/(美)纽曼(Newman, M. E. J.)著. 郭世泽,陈哲译. —北京:电子工业出版社,2014.1
(经典译丛·信息网络技术与网络科学)

书名原文:Networks: An Introduction

ISBN 978-7-121-19603-4

I. ①网… II. ①纽… ②郭… ③陈… III. ①计算机网络-高等学校-教材 IV. ①TP393

中国版本图书馆CIP数据核字(2013)第029738号

策划编辑:马 岚

责任编辑:周宏敏

印 刷:涿州市京南印刷厂

装 订:涿州市京南印刷厂

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本:787×1092 1/16 印张:31.25 字数:806千字 彩插:2

印 次:2014年1月第1次印刷

定 价:89.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zltz@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

译者序

科学是“运用范畴、定理、定律等思维形式反映现实世界各种现象的本质规律的知识体系。”每当面临一个新的自然现象或新的现实世界问题时，都会出现一种新的科学理论和方法，对其进行观察、识别、描述、质疑、假设、验证和总结等，正如达尔文所说的：“科学就是整理事实，从中发现规律，得出结论。”

20 世纪 80 年代，人们认识到人类生活与生产活动越来越多地依赖复杂系统——它们普遍存在于自然界和人类社会中，包括自然界天然存在的星系、河流系统、食物链系统、神经系统、新陈代谢系统、基因控制系统等；人类社会中存在的社会系统、疾病传播系统、知识传播系统等；人类创造的经济金融系统、交通系统、通信系统、配送系统、Internet、Web 系统等。为此，复杂性科学兴起，引发了自然科学界方法论的变革，并且日益渗透到哲学和人文社会科学领域。著名物理学家霍金称：“21 世纪是复杂性科学的世纪。”

20 年来，复杂性科学的实践研究迅速发展，分析方法不断更新，技术应用范围日益扩展。对复杂系统的深入研究，不仅产生了新的理论体系，如耗散结构理论、突变论、超循环论、协同论等，而且带来了从还原论到系统论这一研究理念和方法论上的革新。复杂系统的研究成果，极大地拓展了人们理解世界的广度和深度，并在解决若干问题的过程中发挥了巨大的实际作用。1999 年，美国 *Science* 期刊出版了一期以“复杂系统”为主题的专辑，分别就化学、生物学、神经学、动物学、自然地理、气候学、经济学等学科领域中的复杂性研究进行了报道。

总的来说，复杂系统都有一些共同的特点，那就是在变化无常的活动背后，呈现出某种捉摸不定的秩序，其中演化、涌现、自组织、自适应和自相似等，被认为是复杂系统的共同特征。在这个研究背景下，各领域的科学家和研究团队都在尝试从网络的视角探讨复杂系统，建立网络科学理论体系，进而应用网络科学的基本理论和方法研究复杂系统，这已成为一种主流趋势。正如 Alber-László Barabási 在 *Nature Physics* 上所断言的：“还原论作为一种范式已经过时，复杂性作为一个领域也疲惫不堪。而复杂系统的基于数据的数学模型正在提供一种新鲜的视角，并快速发展为一个全新的学科：网络科学。”

本书作者马克·纽曼，于 1991 年获牛津大学物理系理论物理专业博士学位，之后在康奈尔大学做博士后，出站后进入位于新墨西哥州的圣塔菲研究所。从做博士后开始，作者就一直致力于复杂系统的研究工作。2002 年，他离开圣塔菲研究所，任教于密歇根大学物理系，并当选为密歇根大学保罗·狄拉克物理教授 (Paul Dirac Collegiate Professor of Physics)，同时兼任密歇根大学复杂系统研究中心的教授和圣塔菲研究所的特聘研究员。本书完成于 2010 年 2 月，是他在复杂系统和网络科学方面从事了近 20 年的研究工作和近 7 年的教学工作后的总结。截止到中译本付印，英文原版多次重印，合计已销售上万册。

目前，国内外学者已经撰写了许多关于复杂系统和网络科学的著作，但作为一部总结作者在复杂系统方面的理性思考、研究成果和教学实践的专著，本书选材思路开阔、叙述朴实、涉及面广、体系庞大且应用众多，其从网络的视角，利用网络科学的基本理论对复杂系统的物理本

质和数学概念进行深入阐述时的种种考虑，以及追本溯源的研究风格，对读者而言都具有其他书籍难以替代的指导作用。

曾与埃尔温·薛定谔共同获得诺贝尔物理学奖的著名理论物理学家保罗·狄拉克说过：“科学所关注的只是可观察的事物。”本书作者作为密歇根大学物理系的保罗·狄拉克物理教授，也践行着前辈的思想。从书中可以看出，马克·纽曼遵从着“观察——数据获取与分析——统计规律挖掘——建模再现数据规律”的循环，首先从实证研究的角度入手，系统总结了当前研究的网络的主要类型，以及用于确定其拓扑结构的实证方法。之后的章节，紧紧围绕“可观察的事物”的测度展开论述，探讨了与网络科学相关的基本概念和基本分析方法；初步论述了如何基于测度建立网络的数学模型，进行适当的数学分析和求解，并对所得结果给出物理解释、赋予物理意义。本书的基本概念和方法并不复杂，因此读者理解并掌握本学科的关键不在于认识和记忆了多少定义、定理和公式，而要注重体会作者如何运用数学工具和物理概念，分析复杂系统的典型性质，掌握从网络的视角分析和解决问题的一般规律和方法。特别希望读者关注的是，本书仔细分析了大量“测度”，并给出了基于这些“测度”的数学模型和分析方法的精妙之处和不足的地方，同时也涉及了一些学术界尚存的争论，这些深入分析和深刻见解也是本书的精髓所在。

本书涉及的数学内容主要包括线性代数和矩阵计算、图论、概率统计与随机过程、误差理论与实验数据处理、最优化理论、集合论与一般拓扑学、向量算法与场论初步、级数、复变函数等。本书对数学论证的推演严谨但不显繁冗，避免了缺乏数学论证的“通俗”物理解释，同时也基本上避免了脱离特定物理概念的纯数学推演，将物理概念与数学论证完美地结合在一起，注重解决实际问题，强调物理概念的解释。

本书译者之一郭世泽所领导的团队是长期从事网络科学科研和教学的研究群体，正承担着相关领域的“973”、“863”和自然科学基金项目，在 *Physica A* 等期刊上发表了多篇相关领域的论文。郭世泽与陈哲本着打造本学科领域一部经典教材的原则，用了近两年的时间，先在原著首印版本基础上完成初译，得知作者在原著第6印时做了很多细节改进之后，逐句对照第6印的版本对译文进行了更正和完善。特别是，在翻译过程中，为求深刻理解本书的内涵，并且表达得更准确、更专业，译者除延伸阅读了本书中大部分参考文献外，还请教了应用数学、生物化学、应用物理等领域的专业老师和技术人员，并推导了原书中的所有数学公式和证明过程，对于发现的34处笔误和印刷错误，在翻译时进行了更正。在此过程中，译者更加认识到，决定读者对本学科领域基础知识的理解深度、对新知识的领悟能力和对新应用的创新能力的，除了对数学理论的接受能力和推导能力之外，更重要的是理解数学理论背后令人困惑的物理概念。

在本书的翻译过程中，译者与很多同事和朋友进行过交流，他们提出了很多中肯的意见、建议和鼓励，这些都使译者受益匪浅。特别感谢高梁，他组织郭浩、杨子龙、邢富坤、郭婕恬、冀鸿翔、宋晓虎、王宇、贾文波参与了部分章节的初译和数学公式的推导。

“读罢华章叹不休，甘心俯首作耕牛。扬蹄沥血辛勤日，换得芳田万古留。”我们深信，经过译者的精心挑选和深入细致的译校，本书将为读者奉上一部学习精品，为各领域的研究人员开启利用网络科学研究复杂系统的大门！正如译者在编著的《复杂网络基础理论》中所说的，“前面的高山是如此巍峨美丽，让我们一起去攀登吧！”

限于水平，书中难免有错误与不妥之处，恳请读者批评指正。

前 言

与网络如计算机网络、生物网络、社会网络等相关的科学研究，是典型的交叉学科领域，涉及数学、物理学、生物学、计算机科学、社会学及很多其他学科。网络的研究受益于各不同学科已有的大量研究成果，但由于多个学科的不同研究团体及研究者之间缺乏联系沟通，使得已有的研究成果难以被所有研究群体共享。本书的目的就是将相关的网络知识汇集在一起，并使用统一的语言和符号呈献给读者，使其成为一个各部分内容彼此补充、条理清晰的整体，为深入全面地理解网络提供有力支持。

本书分为 5 部分。在给出一个简短的总体介绍后，第 I 部分描述当前研究的网络的主要类型及确定其拓扑结构的实证方法，包括第 2 章到第 5 章。第 II 部分介绍研究网络的主要数学工具，以及分析网络结构的度量方法和统计学方法，包括第 6 章到第 8 章。第 III 部分描述有效分析网络数据的一些常用算法，包括第 9 章到第 11 章。第 IV 部分讨论网络结构的数学模型，借助这些模型能够预测网络化系统的行为特征并理解网络的形成及发展规律，包括第 12 章到第 15 章。最后，第 V 部分描述与网络过程相关的理论，例如与社会网络中的疾病传染过程相关的理论，以及与计算机网络上的搜索过程相关的理论等，包括第 16 章到第 19 章。

本书各个部分对读者已有知识和技术的要求各不相同。第 I 部分基本不需要任何数学知识就可以理解，第 II 部分和第 III 部分需要具有大学本科水平的线性代数和高等数学基础。第 IV 部分和第 V 部分对数学基础要求更高，适合于大学高年级本科生、研究生及本领域的研究者阅读。本书可以作为不同难度等级教学用参考书。第 1 章至第 8 章难度较低，适合于具有中等数学基础的学习者使用；而第 6 章至第 14 章难度较高，可作为选修教材，供那些具有较好数学基础的学生学习。第 II 部分之后的每章都配有习题，供学习者检验学习效果。

完成本书花费了数年时间，期间得到了很多人的帮助。我必须感谢本书的编辑 Sonke Adlung，他的耐心让我非常感动，我和他的合作已经超过 15 年并完成了很多出版项目，在他的不断鼓励与关心下，我与他及牛津大学出版社的合作非常愉快开心。同时还要感谢 Melanie Johnstone、Alison Lees、Emma Lonie 和 April Warman，他们在本书出版的最后阶段都给予了我很大帮助。

在本书写作过程中，我与很多同事和朋友进行过交谈，他们提出了很多意见、建议和鼓励，这些都使我受益匪浅。非常遗憾的是我不能把所有的同事和朋友毫无遗漏地在此提到，但要非常感谢 Steve Borgatti、Duncan Callaway、Aaron Clauset、Betsy Foxman、Linton Freeman、Michelle Girvan、Mark Handcock、Petter Holme、Jon Kleinberg、Alden Klov Dahl、Liza Levina、Lauren Meyers、Cris Moore、Lou Pecora、Mason Porter、Sidney Redner、Cosma Shalizi、Steve Strogatz、Duncan Watts、Doug White、Lenka Zdeborova 和 Bob Ziff 等人，还要感谢我的学生和其他读者们，特别是

Michelle Adan、Alejandro Balbin、Ken Brown、Judson Caskey、Chris Fink、Martin Gould、Ruthi Hortsch、Puck Rombach、Jane Wang、Daniel Wilcox、Yongsoo Yang 和 Dong Zhou 等人，他们提出的反馈意见使得本书更加完善。我还要特别感谢 Brian Karrer，他通读了本书的草稿，提出的中肯和宝贵的意见使我深受启发，并且他还指出了书中很多文字错误。当然，本人对书中仍然存在的错误负有全部责任；同时，也非常欢迎读者对书中的错误提出意见。

最后，我要将我最诚挚的谢意送给我的妻子 Carrie，感谢她在我写作过程中给予的鼓励与支持。虽然没有她，我也要完成这本书，但肯定会失去很多快乐与幸福。

马克·纽曼
密歇根大学安娜堡分校
2010 年 2 月 24 日

目 录

第 1 章 概述	1
1.1 为什么对网络产生兴趣	1
1.2 几个网络示例	2
1.3 网络的性质	5
1.4 本书结构	7

第 I 部分 网络的实证研究

第 2 章 技术网络	10
2.1 Internet	10
2.2 电话网络	16
2.3 电力网络	18
2.4 交通网络	19
2.5 配送网络	19
第 3 章 社会网络	21
3.1 社会网络实证研究	21
3.2 采访与问卷	23
3.3 直接观察	27
3.4 来自于档案或第三方的数据	28
3.5 隶属网络	31
3.6 小世界实验	32
3.7 雪球式抽样、接触者追踪及随机游走	35
第 4 章 信息网络	38
4.1 万维网	38
4.2 引文网络	41
4.3 其他类型的信息网络	44
第 5 章 生物网络	48
5.1 生物化学网络	48
5.2 神经网络	58
5.3 生态网络	61

第 II 部分 网络理论基础

第 6 章 网络的数学基础	66
6.1 网络及其表示方法	66

6.2	邻接矩阵	67
6.3	加权网络	68
6.4	有向网络	69
6.5	超图	75
6.6	二分网络	76
6.7	树	78
6.8	平面网络	80
6.9	度	82
6.10	路径	84
6.11	分支	88
6.12	独立路径、连通度和割集	91
6.13	图拉普拉斯矩阵	95
6.14	随机游走	99
	习题	103
第7章	测度与参数	106
7.1	度中心性	106
7.2	特征向量中心性	106
7.3	Katz 中心性	109
7.4	PageRank	110
7.5	核心顶点与权威顶点	113
7.6	接近度中心性	115
7.7	介数中心性	117
7.8	顶点群组	123
7.9	传递性	126
7.10	相互性	130
7.11	有符号边和结构平衡	131
7.12	相似性	135
7.13	同质性和同配混合	141
	习题	149
第8章	网络的大规模结构	152
8.1	分支	152
8.2	最短路径和小世界效应	155
8.3	度分布	157
8.4	幂律和无标度网络	159
8.5	其他中心性测度的分布	168
8.6	聚类系数	169
8.7	同配混合	172
	习题	173

第Ⅲ部分 计算机算法

第9章 算法基本概念	176
9.1 运行时间和计算复杂度	177
9.2 网络数据的存储	180
9.3 邻接矩阵	180
9.4 邻接表	182
9.5 树	185
9.6 网络的其他表示方法	190
9.7 堆	192
习题	195
第10章 网络基础算法	196
10.1 度和度分布的算法	196
10.2 聚类系数	197
10.3 最短路径和广度优先搜索	200
10.4 加权网络中的最短路径	209
10.5 最大流和最小割	211
习题	217
第11章 矩阵算法与图划分	219
11.1 主特征向量和特征向量中心性	219
11.2 将网络划分成簇	224
11.3 图划分	227
11.4 Kernighan-Lin 算法	228
11.5 谱划分	231
11.6 社团发现	235
11.7 简单模块度最大化	236
11.8 谱模块度最大化	238
11.9 将网络划分为两个以上群组	240
11.10 其他模块度最大化方法	241
11.11 社团发现的其他算法	242
习题	247

第Ⅳ部分 网络模型

第12章 随机图	250
12.1 随机图	250
12.2 边数和度的均值	252
12.3 度分布	252

12.4	聚类系数	253
12.5	巨分支	253
12.6	小分支	257
12.7	路径长度	265
12.8	随机图的问题	267
	习题	268
第 13 章	任意度分布的随机图	270
13.1	生成函数	270
13.2	配置模型	274
13.3	余度分布	281
13.4	聚类系数	283
13.5	度分布的生成函数	284
13.6	一个顶点的两跳邻居顶点数量	285
13.7	小分支的生成函数	288
13.8	巨分支	290
13.9	小分支的规模分布	293
13.10	幂律度分布	297
13.11	有向随机图	299
	习题	305
第 14 章	网络生成模型	307
14.1	优先连接模型	307
14.2	Barabási-Albert 模型	316
14.3	优先连接模型的其他性质	318
14.4	优先连接模型的扩展	325
14.5	顶点复制模型	339
14.6	网络优化模型	343
	习题	348
第 15 章	其他网络模型	351
15.1	小世界模型	351
15.2	指数随机图模型	360
	习题	374

第 V 部分 网络过程

第 16 章	渗流和网络弹性	376
16.1	渗流	376
16.2	顶点的均匀随机删除	378
16.3	顶点的非均匀删除	388
16.4	实际网络中的渗流	392

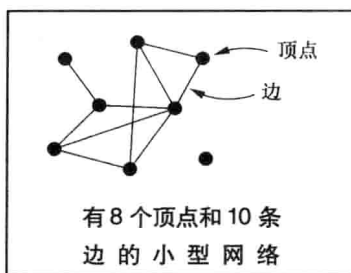
16.5 渗流的计算机算法	393
习题	398
第 17 章 传染病的网络模型	400
17.1 疾病传播模型	400
17.2 SI 模型	400
17.3 SIR 模型	403
17.4 SIS 模型	406
17.5 SIRS 模型	407
17.6 传染病的网络模型	408
17.7 传染病网络模型的晚期特征	409
17.8 SIR 模型的晚期特征	410
17.9 传染病网络模型的时间依赖特性	414
17.10 SI 模型的时间依赖特性	414
17.11 SIR 模型的时间依赖特性	423
17.12 SIS 模型的时间依赖特性	428
习题	431
第 18 章 网络动力系统	433
18.1 动力系统	433
18.2 网络动力学	440
18.3 多变量动力学	446
习题	450
第 19 章 网络搜索	453
19.1 Web 搜索	453
19.2 分布式数据库搜索	455
19.3 消息传递	457
习题	464
参考文献	466
索引	482

第1章 概述

本章简要介绍什么是网络以及为什么要研究网络。

最简单形式的网络，可以表示成点和连接点之间的线的集合。用本领域的术语来描述，点被描述为**顶点**(vertex)^①或**节点**(node)，线被描述为**边**(edge)。在物理学、生物学和社会科学等领域中的很多研究对象都可以抽象成网络，从网络的角度考虑这些研究对象会有更新和更有价值的发现，这也是本书的目的。

本章将重点介绍为什么会对网络有如此浓厚的兴趣，同时还要介绍一些特定类型的网络。本章讨论的内容将在后面章节中给出更加深入详细的讨论。



1.1 为什么对网络产生兴趣

科学家对很多系统都有着浓厚兴趣，这些系统由独立的部件或构件构成，并通过一定的方式联系在一起，例如 Internet 就是由众多计算机及相互间的数据联系构成的系统；再例如人类社会网络，就是由人与人们之间的各种交往关系构成的系统。

这些系统的很多方面非常值得研究。有些研究者专注于研究构成系统的个体的本质规律，例如计算机是如何工作的，人类是如何感知或行动的；而有些研究者则重点研究个体之间的交互及联系的本质规律，例如研究 Internet 的通信协议或人类朋友关系的动力学特征等。但对于这类交互式系统，还有第三个研究领域，那就是研究系统各组成部分之间的**连接模式**(pattern)，该研究领域有时会被忽视，但对研究系统的行为特征具有决定性作用。

对于一个给定系统，其连接模式可以用网络表示，系统的各个构件可抽象成网络中的顶点，构件间的联系抽象成边。显而易见(尽管在某些领域最近才有这种发现)，网络结构和特定的连接模式对系统的整体行为具有重要影响。例如，Internet 上计算机与计算机之间的连接特征，直接影响了网络数据的交换及其传递效率。人类社会中的连接特征影响了人们如何学习、如何思考 and 如何获得信息，还会对其他不太容易观察到的现象产生影响，如疾病的传播等。因此，如果不能够全面地认识网络的结构特征，就不可能完全理解相应系统的运行机制。

网络是对实际系统的一种简化表示，它可将系统简化成一种只保留“连接模式”等基本特征的抽象结构。另外，网络中的顶点与边还可以赋予更多的信息，例如名称或者连接强度等，使其能够更加细致地描述系统特征。尽管如此，当把系统抽象为网络时，仍旧有很多信息会丢失。这样做当然有其不足之处，但也会带来一定的好处，第 6 章将详细讨论最常见的网络类型。

^① 该词单数形式为 vertex，复数形式为 vertices。

很多领域的科学家们多年来已经开发出大量分析、建模和表示网络的工具,包括数学、计算机及统计工具等。这些工具中的大部分都是先将网络简单抽象成顶点和边的集合,然后进行合理计算,从而发现存在于网络中的有价值信息。例如,哪个顶点是连通性最好的顶点,或者从一个顶点到另一个顶点的路径长度是多少。另外一些工具的表现形式为网络模型,利用这些模型能够对网络中发生的行为进行数学预测,例如发现 Internet 上数据流动的方式或者疾病在人类社会传播的规律等。这些工具和模型运行的基础都基于网络的抽象表示,因此从理论上讲能够应用于任何一个可被表示成网络的系统中。也就是说,如果对某个系统感兴趣并且该系统能够被抽象成网络,那么会有成百上千种已开发出的成熟工具可供使用,并迅速应用到对系统的分析中。当然并非所有工具都能够给出有用的分析结果,对于一个特定系统,哪些测度或计算有用不仅取决于这个系统本身的性质和行为特征,也取决于具体的应用问题。不过,如果已经有了具体的应用问题,那么在很多情况下,总可以找到相应的处理工具,来帮助分析和处理感兴趣的网络化系统。

网络是表征系统各部分间联系及交互模式的通用而强大的工具。本书将讨论不同领域中特定类型网络的很多例子,同时也从数学、物理学、计算机科学、信息科学、社会科学、生物学及其他相关学科的角度讨论网络分析方法。为了达到这个目标,我们会将很多学科领域的思想和专业知识汇聚在一起,以便更全面地阐述网络科学。

1.2 几个网络示例

Internet 是网络中最为人熟知并且也是研究得最为深入的例子。Internet 本质上是一个由计算机构成的数据网络,网络的顶点是计算机,边是计算机与计算机之间的物理数据连接,包括光纤或电话线等。图 1.1 展示了 Internet 的基本结构,该图是 2003 年 Internet 的快照,是通过观察其上大量的数据包在不同的数据源与数据目的地之间的传输路径来绘制的。一个特别有意思的现象是,尽管 Internet 是人工创造且精心设计的网络,但人们并不确切知道 Internet 的结构是什么样子,其原因可能是它是由很多不同的团队建设,且各个团队之间缺乏沟通了解和集中控制。目前有关 Internet 结构最有效的认识都来自于实验研究,例如像图 1.1 这样的结果,而不是来自于某个 Internet 中央知识库或负责管理协调的权威部门(2.1 节将更详细地研究 Internet)。

大量的实践需求促使我们对 Internet 结构进行研究。Internet 的基本功能是在分布于不同位置的计算机或其他设备之间传递数据。数据传递时我们首先将数据分解为数据块或数据包(packet),然后将这些数据块或数据包通过网络从一个顶点传递到另一个顶点,依次传递直到数据到达目的地为止。显然,网络的结构会在很大程度上影响数据传输的效率,如果我们了解网络的结构,就能解决很多实际问题,例如如何选择数据的传输路径;为了使传播最快,路径最短是不是总有必要?如果不是,那么什么样的路径传输速度最快,如何找到这样的路径?如何避免数据传输中出现的使数据传输速度降低的瓶颈问题?如果某个顶点或边出现问题,传输一般会受到什么样的影响?如何制定传输方案以应对传输中出现的的问题;如果可以为网络增加一些新的容量,应该添加到哪里?

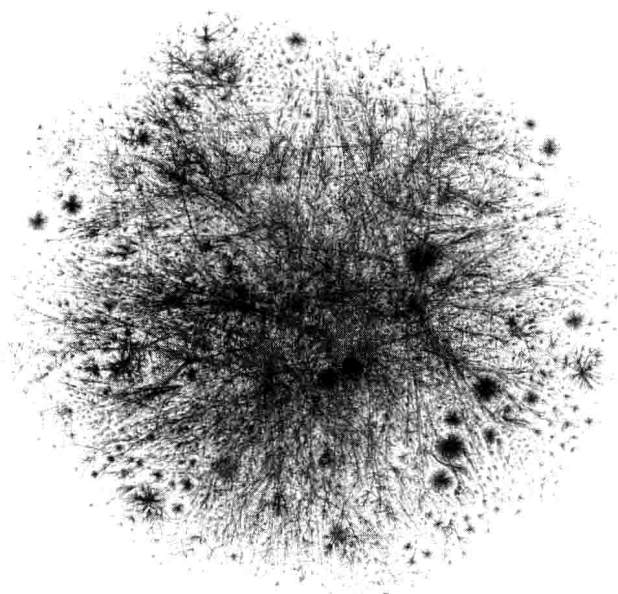


图 1.1 Internet 的网络结构图(又见彩图1)

在图 1.1 中,顶点表示 Internet 中的一个 C 类子网(C 类子网通常是分配给同一个组织机构的 IP 地址集合),它们之间的联系表示 Internet 中的数据在 C 类子网之间的传输路径。图中顶点的几何位置没有特殊含义,完全是出于显示需要而确定的各自位置,与其他因素如顶点的地理位置并不相关。相关内容将在 2.1 节中进行详细介绍。本图由 Opte Project (www.opte.org) 绘制并授权使用。

Internet 结构对于开发新的通信标准至关重要。Internet 通信的新标准及协议不断出台,旧版本不断更新,协议中相关参数的制定和变化都是为了使 Internet 在期望的结构下达到性能最优。网络的初始阶段,在协议的调整过程中,我们只采用了相当简单的网络结构模型;但是,随着可以获得越来越丰富的网络结构数据,使得更好地理解网络结构和提高网络性能成为可能。

网络的另一个更加抽象的例子是万维网。通常情况下,“万维网”与“因特网”(Internet)这两个词经常通用,但是从技术角度看,二者相差甚远(4.1 节将更详细地讨论万维网)。Internet 是一个物理形态的网络,由计算机及相互间的电缆、光纤(或无线信号)连接在一起。而万维网则是由存储在网页上的信息构成的网络。万维网的顶点是网页,边是“超链接”,通过点击强调显示的链接文本或链接按钮,在不同网页之间转换。超链接是一种纯粹的软件结构,借助它能够很轻易地从当前网页跳转到保存在地球另一端机器中的另一个网页上,就像你在宴会厅里能够很容易找到你的朋友一样。新建一个链接不需要建立任何物理结构,例如光缆线路等。新建链接只是告诉计算机一个网络地址,当点击该链接时,机器能够找到对应的网页。

万维网的结构较为抽象,存在其中的海量网页和链接除了能产生巨大的经济效益外,也能对人们带来很大益处,其中网络的链接结构是最本质的也是最重要的原因。由于人们习惯于在内容有联系的网页之间添加超链接,因此链接结构能够反映出网页的内容结构。更重要的是,人们倾向于链接那些自己认为有用的网页,而不去链接那些认为用处很小的网页,因

此网页的被链接数可以作为度量其实用性的指标。这种想法已经被著名的搜索引擎 Google 采纳,只不过其实现得更加复杂,并且很多其他搜索引擎也吸收了类似思想(19.1 节将详细讨论网络搜索的相关机制)。

万维网还表现出了网络理论的另一种概念,即**有向网络**(directed network)。网页上的超链接都具有明确的指向性,即从一个网页发出,指向另外一个网页。给定网页 A 的链接,通过点击该链接可以到达另外一个网页。但是,并不需要网页 B 存在一个指向网页 A 的链接。当然,网页 B 中可以存在一个指向网页 A 的链接,但并没有任何规定要求这样做,并且通常情况下也没有必要这样做。因此,万维网上的边是**有向的**(directed),从链接发出的网页指向被链接的网页。

研究人员不仅对技术领域的网络感兴趣,而且还对社会网络感兴趣。社会网络通常是由人构成的网络,但有时也会是由不同的人群(如公司)构成的网络。个体或人群构成社会网络的顶点,网络的边则代表人或人群之间的关系,例如个体之间的朋友关系,或公司之间的业务关系等。在对现实世界的网络进行实证研究的过程中,社会学也许是历史最悠久并且发展最完善的学科之一,而且研究网络的很多数学和统计工具都直接或间接地借鉴了社会学家的理论成果(第 3 章将更深入地讨论社会网络)。

图 1.2 展示了社会网络研究中的一个非常著名的案例,即 Wayne Zachary 研究的“空手道俱乐部”社会网络。该网络展现了北美一所大学的空手道俱乐部中成员之间的朋友关系。该网络通过直接观察俱乐部成员间的相互关系而建立。这类研究的一个共同特征就是网络规模较小,例如本案例的网络只有 34 个顶点。与此相反的是 Internet 或万维网的网络表征,在这些网络中有成千上万,甚至上百万的顶点。从理论上讲,社会网络同样可以具有很大的规模,全世界的所有人口可以看成一个巨型的社会网络。但在实际研究中,受限于时间精力等因素,构建这么大规模的网络非常困难,因此只能局限于研究较小规模的社会网络。例如构建出图 1.2 所示的网络就需要花费一名研究人员大约两年的时间。近些年来,在某些大规模研究团队的努力下,构建了一些更大规模的社会网络。一些在线社交网络服务,如社交网站 Facebook 或即时聊天工具中的“好友列表”,能够提供以前从未有过的大规模网络数据。不过,有关这些大规模社会网络的研究才刚刚开始。

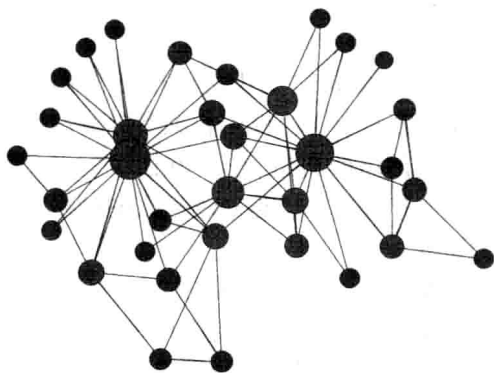


图 1.2 俱乐部成员的朋友关系图

图 1.2 所示的有关社会网络的研究完成于 20 世纪 70 年代,展示出美国一所大学空手道俱乐部成员之间的关系。该数据由 Zachary^[334]收集并公布。

生物学作为网络研究的第三类对象,在近几年逐渐受到关注,网络正在应用于生物学研究的很多方面。有些网络具有非常具体的物理形态,例如神经网络,即在大脑的神经元之间建立联系而形成的网络(5.2 节将详细讨论神经网络),而其他的网络则更加抽象。图 1.3 展示了一张“食物网”,在这个生态网络中,网络顶点是生态系统中的各个物种,网络的边则代表了捕食者与被捕食者之间的关系。也就是说,如果两个物种之间通过边连接在一起,则说明其中一个物种会以另一个物种为食。有关食物网的研究是生态学研究的一个重要分支,能

够帮助人们更好地理解并量化分析生态现象,特别是那些与生态系统中的能量和碳流动相关的现象(5.3节将详细讨论食物网)。食物网也提供了一种与之前讨论的万维网相似的有向网络。食物网中的边是非对称的,通常从被捕食者指向捕食者,表示当被捕食者被吃掉后能量的流动方向。不过,这种方向的选择只是一种约定,完全可以选择另一种表示方法。重要的是表示出捕食者与被捕食者之间的非对称关系。

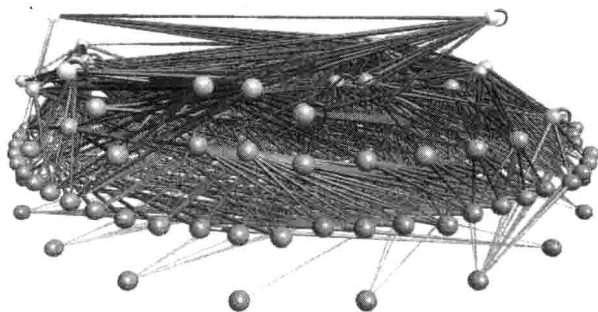


图 1.3 美国威斯康辛州小石湖地区的食物网(又见彩图2)

在图 1.3 中,比较详细地展示出美国北部一个淡水湖地区已知动物种群间的捕食关系。顶点表示物种,边表示捕食者与被捕食者的关系。顶点的上下层次关系可以粗略地认为是表示不同物种间的层级关系。本图由 Richard Williams 和 Neo Martinez^[209] 绘制。

生物网络的另一个分类是生物化学网络,如新陈代谢网络、蛋白质交互网络和基因控制网络等(5.1节将详细讨论生物化学网络)。例如,新陈代谢网络表示了为细胞和组织提供养分的化学反应过程。读者可能在生物化学家的办公室看到过一些作为装饰用的挂图,图上显示了新陈代谢反应的过程。在这些异常精细的挂图中,微小的文字被迷宫一样的箭头连接在一起^①。这些文字实际上就是网络的顶点,代表了代谢物,也就是新陈代谢中的基质和生成物,而带有箭头的线就是有向边,代表了由一种代谢物转变成另一种代谢物的反应过程。将这些反应描述为一个网络,是搞清楚这些错综复杂的生物化学数据的首要步骤之一,而在最新的及正在进行的生物遗传学实验中已经产生了大量这样的数据。

以上只是网络的几个例子,研究这些网络是本书的重点。在本书的后面还要讨论很多其他特定类型的网络,其中有大家熟悉的电话网络、公路网络、铁路网络、航空网络、电力网络、引文网络、推荐网络、P2P 网络、电子邮件网络、合作网络、疾病传播网络、河流网络和单词网络等。

1.3 网络的性质

很多系统都可以表示为网络。如果能够获得关于网络结构的一些数据,那么如何处理这些数据呢?这些数据能够阐述网络所表示的系统在形式和功能方面的哪些信息?能够对该系统的哪些性质进行度量或建模?这些性质如何与关心的实际问题发生联系?这些都是本书要讨论的主要话题。本章无法回答所有问题,不过还是要简单介绍一些具有代表性的概念,以初步接触一下马上要遇到的一些思想和内容。

^① 图 5.2 给出了一个例子。