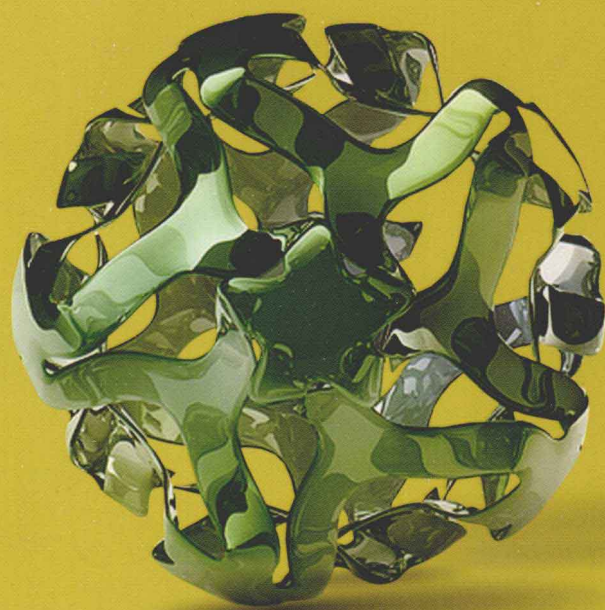


普通高等教育“十二五”规划教材

高等院校重点推荐教材

复杂网络基础理论

郭世泽 陆哲明 编著



科学出版社

内 容 简 介

本书是复杂网络基础理论的入门教材。本书取材广泛,内容全面、新颖,反映了十多年来复杂网络理论的最新研究动态和作者的部分研究成果。全书共分八章。第一章主要介绍与复杂网络有关的基本知识。第二章详细介绍了网络的拓扑结构和静态特征。第三章从机制模型的不同类型入手分别介绍了规则网络、随机网络、小世界网络、无标度网络、层次网络、确定性网络和自相似网络。第四章介绍复杂网络上的流行病传播、舆论传播和知识传播以及数据包传递和拥塞控制机理。第五章介绍复杂网络的混沌同步。第六章介绍复杂网络中的搜索算法与策略、社会网络的分散式搜索以及P2P网络和WWW网络中的搜索。第七章介绍复杂网络中的重要节点挖掘和社团挖掘原理和算法。第八章介绍复杂网络中的演化博弈、抗毁性分析以及抗毁性优化和修复策略等。为便于消化和理解书中内容,每章末附有习题,书末给出大量参考文献。

本书可作为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、社会学等专业本科生和研究生的教材,也可供从事复杂性科学和网络科学等领域工作的教学、科研人员参考。

图书在版编目(CIP)数据

复杂网络基础理论/郭世泽,陆哲明编著. —北京:科学出版社,2012
(普通高等教育“十二五”规划教材·高等院校重点推荐教材)
ISBN 978-7-03-034599-8

I. ①复… II. ①郭…②陆… 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(2012)第115630号

责任编辑:鞠丽娜 / 责任校对:耿 耘
责任印制:吕春珉 / 封面设计:三函设计

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2012年6月第 一 版 开本:787×1092 1/16

2012年6月第一次印刷 印张:22

字数:504 000

定价:40.00元

(如有印装质量问题,我社负责调换<新科>)

销售部电话 010-62134988 编辑部电话 010-62135763-8002

版权所有,侵权必究

举报电话:010-64030229; 010-64034315; 13501151303

前 言

科学根植在理论基础之上,理论是对经验现象或事实的科学解说和系统解释。顾名思义,复杂网络作为一门新兴科学,是对存在的网络现象及其复杂性进行解释的学科。首先,它研究的是网络现象。网络在自然界和人类社会中普遍存在,包括自然界中天然存在的星系、食物链网络、神经网络、蛋白质网络;人类社会中存在的社交网络、传染病传播网络、知识传播网络;人类创造的交通网络、通信网络、计算机网络等。网络科学作为一门交叉学科,主要研究利用网络特性描述物理、生物和社会等现象,进而建立这些现象的预测模型或分析模型,并利用网络的静态特性和动力学特性来解释这些现象。

其次,它代表着对网络自身复杂性的深化认识。从1736年到1958年漫长的400多年里,人们一直用基于图论的规则网络理论来研究与网络有关的问题。1959年,匈牙利数学家Erdős和Rényi首次将随机性引入到网络建模中,创立了随机网络理论,提出了ER随机图模型,使网络科学迈入了第二个重要的发展阶段。1959年到1998年的近40年时间里,随机网络理论曾一度被公认为是正确认识真实网络的基本理论。20世纪80年代,复杂性科学(complexity science)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、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。”1999年,美国“Science”杂志出版了一期以“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。概括起来,复杂系统都有一些共同的特点,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、自相似被认为是复杂系统的共同特征。在这个研究背景下,应用复杂性科学解释网络现象的复杂网络理论应蕴而生。美国康奈尔大学理论和应用力学系博士生Watts及其导师Strogatz于1998年在“Nature”杂志上发表题为“小世界网络的群体动力行为”的论文,提出了一种介于规则网络和随机网络之间的网络模型——小世界网络模型。随之,1999年美国圣母大学物理系Barabási教授及其博士生Albert在“Science”杂志上发表题为“随机网络中标度的涌现”的论文,提出了一个无标度网络模型,引起了全世界的高度重视。小世界网络和无标度网络模型的提出标志着网络科学进入了一个新时代。随后的许多真实网络实证研究表明,真实世界网络既不是规则网络,也不是随机网络,而是兼具小世界和无标度特性,具有与规则网络和随机网络完全不同的统计特性。面对快速发展的Internet和WWW网络,还有其他各种社会、生物、物理网络,科学家们发现已无法用规则网络和随机网络理论来解释它们的结构和演化的许多新问题,因而粗略地称这类网络为复杂网络。

复杂网络至今还没有一个统一公认的定义。人们通常采用复杂网络表现出来的不同于规则网络和随机网络的特性来表征复杂网络,如小世界特性、无标度特性、层次特性、自相似特性、自组织特性等。目前,复杂网络的研究工作集中在以下几个方面:①复杂网络拓扑结构的静态统计分析,包括更广泛的实证研究和更深入的理论刻画;

②复杂网络的演化和机制模型,实证上可以研究实际网络演化的统计规律,如检验 BA 模型的择优连接假设;理论上则可以发展完善的具有形成特定几何性质的网络机制模型;③复杂网络上的动力学研究,包括网络容错性和鲁棒性以及网络上的搜索、传播、演化博弈、同步与共振等各种动力学过程;④有关复杂网络的分析方法与应用研究。总的来说,网络的结构与功能及其相互关系是网络研究的主要内容,结构与功能的相互作用特别是其对网络演化的影响是复杂网络研究需要解决的重要问题。

自 2000 年以来,有关复杂网络的研究论著不断涌现。本书作为基础理论教材,旨在介绍十多年来复杂网络领域公认的广泛提及和深入研究的一些基础理论,尽量以浅显易懂的方式为来自不同学科领域的本科生和研究生学习复杂网络理论提供指导。在学习本书之前,读者需要具备微积分、概率论和图论的基础知识,了解信息论、控制论、博弈论的相关概念,掌握必要的计算机编程语言和仿真验证技能。本书共八章,分别是绪论、网络拓扑结构与静态特征、网络机制模型、复杂网络上的传播动力学、复杂网络的混沌同步、复杂网络上的搜索、复杂网络中的挖掘以及复杂网络中的博弈。为了让读者更好地理解课程内容,各章都配有例题和习题。不管针对何种读者对象,第一章、第二章和第三章是必修课程,每章需要 8 学时课堂授课和 2 学时上机实验。后面的章节针对不同的读者可以有所侧重,例如,对于计算机网络安全人员,第四章、第七章和第八章需要重点学习。针对本科生而言,除了前三章,建议学习第四章、第六章和第七章,每章需要 8 学时课堂授课和 2 学时上机实验。针对研究生而言,建议学习所有章节,一共 64 学时课堂授课和 16 学时上机实验。各章内容简述如下:

第一章:图论是研究网络结构与特性的最有效的理论工具之一,复杂网络中的大多数概念和特性可以从图论中找到依据。其次,数理统计和概率论是进行复杂网络建模和特性分析的一种有力的数学工具。因此,第一章绪论主要介绍与复杂网络有关的基本知识,内容包括:网络科学理论发展的历史、复杂网络的概念与特性、数理统计基础、图论的基本概念、复杂网络的研究内容和意义。

第二章:复杂网络具有很多与规则网络和随机网络不同的统计特征。因此,究竟用哪些统计特性来描述一个给定网络呢?刻画一个网络最基本的三种特征量是度分布、集聚系数和平均距离,其中度分布描述了每个节点的邻居节点个数的分布情况,如幂律度分布是无标度网络的特性;集聚系数描述了同一个节点相连接的两节点之间也相互连接的概率,如对社交网络而言,代表着如下情况出现的程度:你朋友的朋友也是你的朋友或者你的两个朋友彼此也是朋友;平均距离描述了平均意义上网络中两个节点之间的最短路径所经历的边的条数,如小世界网络具有较小的平均距离。本书的第二章详细介绍了网络的拓扑结构和静态特征,首先介绍网络的三个基本静态特征,接着按无向网络、有向网络和加权网络分别讨论网络的各种基本静态特征,然后简要介绍近些年文献中提出的其他一些新的静态特征,最后简要介绍了复杂网络分析软件 Pajek。

第三章:在现实生活中,大家经常会感慨世界很小,尤其在当今发达的通信技术和网络技术背景下,地球上的任何两个人平均经由 6 个朋友的牵线搭桥,就可以连通,这就是一种著名的称为“六度分离”的小世界现象。在 Internet 中,通常是越著名的

网站，链接它的网站就越多，从而形成许多“hub”网站，整个网络呈现无标度特性和“富者愈富”的效应。那么，如何能够构造具有小世界特性的人际关系网络模型和具有无标度特性的网络模型，以便人们进行后续的特性分析和动力学行为分析，这就是网络机制模型问题。本书第三章从机制模型的不同类型入手分别介绍了规则网络、随机网络、小世界网络、无标度网络、层次网络、确定性网络、自相似网络，并在每个小节中比较详细的介绍了与这些模型有关的知识。

第四章：在人类社会，传染病夺去了无数人们的生命，为什么只要将感染率控制在一定的阈值之下，传染病的传播就可以得到有效控制呢？而在计算机网络中，病毒种类越来越多，对 Internet 安全构成极大威胁，为什么没有一种杀毒软件可以查杀所有可能出现的病毒，没有一种防护措施可将所有病毒拒之门外呢？实际上，病毒在计算机网络上的蔓延，传染病在人群中的流行，谣言在社会中的扩散，舆论和知识的传播，网络中的数据包传递等都可以看作是服从某种规律的网络传播行为。本书第四章首先介绍复杂网络上的流行病传播机理，接着介绍复杂网络的免疫策略，然后介绍复杂网络上的舆论传播和知识传播，最后介绍复杂网络上的数据包传递机理和拥塞控制等。

第五章：大家在日常生活中可能都有过如下的发现：夏天的萤火虫会有规律的同时发光，或者同时不发光；窗外青蛙和蟋蟀的叫声一段时间后就会同时叫，或同时不叫；剧场中观众掌声频率逐渐趋于一致等，这就是同步化现象。本书的第五章介绍了复杂网络上的混沌同步，首先简要介绍混沌理论，然后概述混沌同步的概念和方法，接着引出一般意义上的复杂网络完全同步问题及其稳定性分析方法，最后讨论典型复杂动态网络在线性耗散耦合条件下的混沌同步问题。

第六章：在信息爆炸的当今网络时代，如何能够在网络中快速搜索到所需要的信息或文件？在交通网络中，如何快速确定任意两个城市之间的最短路径？在社会网络中，如何快速有效地搜寻罪犯、恐怖分子或失散亲人？这些都是复杂网络中的搜索问题。本书第六章首先介绍三种经典的搜索策略，即广度优先搜索算法、随机行走搜索算法和最大度搜索算法，然后介绍社会网络的快速分散式搜索问题，最后介绍 P2P 网络和 WWW 网络的搜索问题。

第七章：首先，在罪犯关系网络中，发掘网络中的重要节点可以迅速定位犯罪团伙的头目，集中警力进行布控；在电力网络中，对重要的断路器、发电单元等进行保护，可以有效防止由相继故障引起的大范围停电；在谣言传播网络中，通过发掘始作俑者来避免“蝴蝶效应”，这些都是重要节点挖掘问题。其次，在许多现实网络中，社团结构是主要特征之一，如在人际关系网中，社团可能按照人的职业、年龄等因素来划分；在新陈代谢网和神经网络中，社团可能反映了功能单元；在食物链网中，社团可能反映了生态系统中的子系统，这些可以归结为社团挖掘问题。本书第七章从节点挖掘和社团挖掘两个方面入手，详细介绍这两方面的相关知识，并对典型的分析算法进行细致讲解。

第八章：群体合作现象的涌现和稳定维持可以说是自然界中最令人兴趣盎然、也是最令人疑惑的问题。如何理解和解释合作行为的广泛存在和稳定维持是人们面临的

最大的挑战之一，复杂网络的演化博弈论扮演了重要角色，并提供了强有力的理论框架。另外，随着复杂网络理论研究和应用研究的深入，人们开始关注：复杂网络到底有多可靠？在网络遭受损失后如何高效快速的修复网络？例如，生物领域基因网络中的一些核心基因的故障会带来灾难性的后果，军事领域中网络中心战将成为未来战争的主要样式，网络系统的抗毁性将直接关系到整个战争的成败，这些可以归结为复杂网络中的博弈问题。本书第八章首先概述博弈论，接着重点介绍复杂网络的演化博弈理论，然后重点介绍复杂网络中的抗毁性分析，最后简要介绍复杂网络的抗毁性优化和修复策略。

本书可供从事复杂性科学、网络科学、图论、系统工程、计算机网络、统计物理学、生命网络分析、社会网络分析、传播动力学、演化博弈论等研究领域的科技人员与教学人员参考，并可以作为上述专业本科生和研究生的教材。本书的每一章都是由郭世泽研究员提供总体思路和初稿框架，由陆哲明教授完善细化和修改审定。本书广泛参考了国内外复杂网络研究领域的学术论文、学位论文和学术著作，尤其从如下三本专著中得到很多帮助：2006年清华大学出版社出版由汪小帆、李翔和陈关荣合作编著的《复杂网络理论及其应用》、2006年上海科技教育出版社出版由郭雷和许晓鸣编著的《复杂网络》以及2009年高等教育出版社出版由何大韧、刘宗华、汪秉宏等编著的《复杂系统与复杂网络》。本书作者及其研究团队不仅要感谢国际学术大师 Strogatz、Watts、Barabási、Albert、Newman 等为我们开创了复杂网络研究领域，也非常感谢国内学者们在学术界开展的研究工作，包括陈关荣、汪小帆、方锦清、李翔、刘宗华、汪秉宏、章忠志等。本书得到了多个国家自然科学基金项目（项目编号：61171150；项目编号：61003255；项目编号：61070208；项目编号：61071128）和浙江省杰出青年基金项目（项目编号：R1110006）的资助，在此致以深深的谢意。在本书的撰写过程中还得到了总参第五十四研究所、北京邮电大学、浙江大学航天电子工程研究所各位教师、博士生和硕士生的协助，在此表示衷心地感谢。

“放眼何能一叶休，登攀绝顶作身谋。梦中借我春秋笔，挥却丹书始探究。”可以说，我们现在所看到的复杂网络科学只是冰山一角，尚有大量认识不清、理论不足、验证不够的研究领域，亟待我们抱着科学的态度去研究、探索、发现。此书是我们团队关于复杂网络科学的第一本粗习之作，相信随着研究的开展和深入，我们还将为业界奉献更有分量的论著。我们团队的座右铭是：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。

郭世泽

于北京总参第五十四研究所

陆哲明

于杭州浙江大学航空航天学院航天电子工程研究所

2012年5月

目 录

第一章 绪论	1
1.1 引言	1
1.2 网络科学理论发展的三个时期	2
1.2.1 规则网络理论阶段	2
1.2.2 随机网络理论阶段	4
1.2.3 复杂网络理论阶段	5
1.3 复杂网络的概念和特性	7
1.3.1 复杂网络的概念	7
1.3.2 复杂网络的特性	10
1.4 数理统计基础	11
1.4.1 概率论基础	12
1.4.2 数理统计基础	17
1.4.3 统计假设及检验	19
1.4.4 一元线性回归分析	20
1.5 图论的基本概念	23
1.5.1 图的基本概念	23
1.5.2 图的路和连通性	25
1.5.3 图的基本运算	26
1.5.4 树与生成树	27
1.5.5 图的矩阵表示	29
1.6 复杂网络的研究内容和意义	33
1.6.1 复杂网络的研究内容	33
1.6.2 复杂网络的研究意义	37
1.7 本书内容安排	38
习题	39
第二章 网络拓扑结构与静态特征	40
2.1 引言	40
2.2 网络的基本静态几何特征	40
2.2.1 平均距离	40
2.2.2 集聚系数	41
2.2.3 度分布	43
2.2.4 实际网络的统计特征	45
2.3 无向网络的静态特征	45
2.3.1 联合度分布和度-度相关性	45
2.3.2 集聚系数分布和聚-度相关性	48
2.3.3 介数和核度	48
2.3.4 中心性	51

2.3.5	网络密度	53
2.3.6	连通集团(子图)及其规模分布	54
2.4	有向网络的静态特征	55
2.4.1	入度和出度及其分布	56
2.4.2	度-度相关性	58
2.4.3	平均距离和效率	59
2.4.4	入集团和出集团的集聚程度	59
2.4.5	介数和双向比	61
2.4.6	中心性	62
2.5	加权网络的静态特征	63
2.5.1	点权、单位权和权重分布差异性	64
2.5.2	权-度相关性和权-权相关性	65
2.5.3	距离分布和平均距离	66
2.5.4	加权集聚系数	67
2.5.5	介数分布和漏斗效应	68
2.5.6	有向加权网络的最短路径问题	69
2.6	网络的其他静态特征	71
2.6.1	网络结构熵	71
2.6.2	特征谱	72
2.6.3	度秩函数	73
2.6.4	富人俱乐部系数	74
2.7	复杂网络分析软件	75
	习题	78
第三章	网络机制模型	79
3.1	引言	79
3.2	规则网络	79
3.2.1	全局耦合网络	80
3.2.2	最近邻耦合网络	80
3.2.3	星型耦合网络	82
3.3	随机网络	83
3.3.1	随机网络模型	83
3.3.2	随机网络的度分布	85
3.3.3	随机网络的直径和平均距离	86
3.3.4	随机网络的集聚系数	87
3.3.5	随机网络的特征谱	87
3.4	小世界网络	88
3.4.1	小世界网络模型	88
3.4.2	小世界网络的度分布	92
3.4.3	小世界网络的平均距离	92
3.4.4	小世界网络的集聚系数	93
3.4.5	小世界网络的特征谱	93

3.5 无标度网络	94
3.5.1 Price 模型	95
3.5.2 BA 模型	95
3.5.3 BA 无标度网络的度分布和度相关	100
3.5.4 BA 无标度网络的平均距离和集聚系数	102
3.5.5 BA 无标度网络的特征谱	103
3.6 层次网络	104
3.6.1 模块性和模体	104
3.6.2 层次网络概念和特性	105
3.6.3 层次网络构造方法	106
3.7 确定性网络	108
3.7.1 确定性均匀递归树	108
3.7.2 确定性小世界模型	109
3.7.3 确定性无标度网络	112
3.8 自相似网络	113
3.8.1 复杂网络的自相似性	113
3.8.2 自相似复杂网络的构造方法	116
习题	118
第四章 复杂网络上的传播动力学	120
4.1 引言	120
4.2 复杂网络上的流行病传播	121
4.2.1 流行病传播的基本模型	122
4.2.2 均匀网中的流行病传播	125
4.2.3 非均匀网中的流行病传播	128
4.2.4 社团网上的流行病传播	131
4.2.5 有限规模无标度网络和广义无标度网络的传播阈值	133
4.2.6 关联网络的传播阈值	135
4.3 复杂网络上的免疫策略	136
4.3.1 随机免疫	136
4.3.2 目标免疫	137
4.3.3 熟人免疫	138
4.4 复杂网络上的舆论传播和知识传播	139
4.4.1 复杂网络上的舆论演化动力学	139
4.4.2 复杂网络上的舆论传播	143
4.4.3 复杂网络上的知识传播	146
4.5 复杂网络上的数据包传递和拥塞控制	149
4.5.1 复杂网络上的数据包传递模型	149
4.5.2 复杂网络上的数据包传递路由策略	153
4.5.3 复杂网络上的拥塞控制	157
习题	159
第五章 复杂网络的混沌同步	161
5.1 引言	161

5.2	混沌理论	162
5.2.1	混沌	162
5.2.2	混沌模型	165
5.2.3	混沌系统的刻画指标	168
5.3	混沌同步理论	172
5.3.1	混沌同步的定义	172
5.3.2	混沌同步的判定	175
5.3.3	混沌同步的方法	177
5.4	复杂网络的完全同步判据	183
5.4.1	复杂动态网络的完全同步概念	183
5.4.2	复杂动态网络完全同步的稳定性分析	185
5.4.3	连续时间线性耗散耦合网络的完全同步判据	189
5.4.4	连续时间时滞耗散耦合网络的完全同步判据	191
5.4.5	特殊离散时间耦合网络的完全同步判据	192
5.5	复杂网络的混沌同步	193
5.5.1	小世界网络的混沌同步	194
5.5.2	无标度网络的混沌同步	197
5.5.3	提高复杂网络同步能力的方法	199
	习题	201
第六章	复杂网络中的搜索	203
6.1	引言	203
6.2	广度优先搜索	204
6.2.1	复杂网络搜索问题	204
6.2.2	广度优先搜索算法	204
6.2.3	广度优先搜索算法实现	205
6.2.4	广度优先搜索算法的应用和特性	207
6.3	随机行走搜索	208
6.3.1	随机行走搜索算法	209
6.3.2	随机行走的基础理论	209
6.3.3	最近邻耦合网络上的随机行走搜索	211
6.3.4	ER 随机网络上的随机行走搜索	214
6.3.5	WS 小世界网络上的随机行走搜索	216
6.4	最大度搜索	217
6.4.1	最大度搜索算法	217
6.4.2	最大度搜索算法分析	218
6.5	社会网络的分散式搜索	220
6.5.1	引言	220
6.5.2	Kleinberg 网格模型的分散式搜索	220
6.5.3	层次网络模型上的分散式搜索	224
6.5.4	Kleinberg 集合模型上的分散式搜索	227
6.5.5	基于 Kleinberg 网格的动态网络模型的快速分散式搜索	228

6.5.6 复杂网络的可搜索性分析	230
6.6 Internet 中的搜索	232
6.6.1 P2P 网络	233
6.6.2 基于广播方式的 Gnutella 网络搜索	236
6.6.3 基于 K-遍历器随机行走的 Gnutella 网络搜索	239
6.6.4 基于度分布的 Gnutella 网络搜索	240
6.6.5 WWW 网中的搜索	243
习题	245
第七章 复杂网络中的挖掘	246
7.1 引言	246
7.2 重要节点挖掘研究现状及评价指标	246
7.2.1 重要节点挖掘研究现状	247
7.2.2 重要节点指标分析	248
7.2.3 合理评价指标所需条件	249
7.3 常见重要节点挖掘方法	250
7.3.1 基于节点关联性的方法	250
7.3.2 基于最短路径的方法	251
7.3.3 基于模拟流的方法	255
7.3.4 其他分析方法	258
7.4 社团结构挖掘研究现状及评价指标	265
7.4.1 社团结构挖掘研究现状	265
7.4.2 社团结构的定义和模块性函数	266
7.4.3 经典检验网络	268
7.4.4 社团划分结果评价	270
7.5 常见社团挖掘方法	273
7.5.1 Kernighan-Lin 算法	273
7.5.2 谱平分法	274
7.5.3 派系过滤算法	276
7.5.4 分裂算法	279
7.5.5 凝聚算法	282
7.5.6 基于局部信息的算法	284
7.5.7 基于网络动力学的算法	286
习题	289
第八章 复杂网络中的博弈	291
8.1 引言	291
8.2 博弈论概述	291
8.2.1 博弈论基本概念及其发展历史	291
8.2.2 博弈的分类	293
8.2.3 完全信息静态博弈与纳什均衡	295
8.2.4 完全信息动态博弈与子博弈精炼纳什均衡	298
8.2.5 不完全信息静态博弈与贝叶斯纳什均衡	300

8.2.6 不完全信息动态博弈与精炼贝叶斯纳什均衡	301
8.3 复杂网络中的演化博弈	303
8.3.1 演化博弈简介	303
8.3.2 演化网络博弈概述	306
8.3.3 基于囚徒窘境博弈模型的演化网络博弈	308
8.3.4 基于铲雪博弈模型的演化网络博弈	315
8.4 复杂网络的抗毁性分析	319
8.4.1 复杂网络的抗毁性分析背景	319
8.4.2 复杂网络的抗毁性定义	321
8.4.3 复杂网络的抗毁性测度	322
8.4.4 复杂网络的抗毁性分析	328
8.5 复杂网络的抗毁性优化和修复策略	332
8.5.1 复杂网络的抗毁性优化	332
8.5.2 复杂网络的修复策略	333
习题	334
参考文献	336

第一章 绪 论

1.1 引 言

人类从远古走来,很早就构造出山间小路和林中小径,并且把小路连成网络;在农业社会,人又构造出各种水利网络,便于灌溉;通过航海网络,资本主义才走遍全世界;在工业社会,普通的小路被公路、铁路和高速公路所淹没,公路网和铁路网极大方便了人们的交流和贸易;在今天的信息时代,各个国家致力于建设自己的信息高速公路,即新型的信息网路,才会使如今的 Internet/WWW 网络基本覆盖整个世界。有太多的网络与人们生活息息相关,除了上面介绍的网络外,还有通信网络、电力网络、航空网络、银行网络、商业网络等。人类把自己生存的世界变成了网络世界,网络越发达、越有效,世界就越小,人的社会性就越得到强化。现如今,网络显得如此广泛、如此重要,人类已处在网络的重重包围之中。如何开辟出一条研究思路,揭示网络拓扑结构的形成机制,探索网络的演化规律和整体行为,认识网络内部深奥的动力学特性,挖掘网络展现出的广泛、潜在的应用价值等问题,正引起国内外学术界的高度重视^[1]。

21 世纪是复杂性和网络化的世纪。从 20 世纪七八十年代开始,复杂性问题的研究已引起国内外关注,并与非线性科学及其混沌动力学的复杂性研究交错在一起,也因此在国际上形成了非线性科学和复杂性问题的研究热潮^[2]。来自各国的不同学科的科学家,包括物理学家、生物学家、计算机科学家和经济学家等都开始不约而同地讨论和研究各自领域的复杂性问题,例如自组织现象和自组织临界性、自适应问题、计算机与智能问题、生命与生物的演化、全球经济的演化、人类文化和语言的演变等。另一方面,随着人类认识能力的进一步提高,人们发现自然界、人类社会、生物群体中的许多复杂系统都可以通过网络的概念加以描述。尤其是 20 世纪 90 年代以来,随着计算机技术和 Internet 技术的迅猛发展,标志着人类迈入了网络时代。人类已经生活在一个充满各种各样复杂网络的世界中^[3]:从 Internet 到 WWW,从大型电力网络到全球交通网络,从生物体中的大脑到各种新陈代谢网络,从科研合作网络到各种经济、政治、社会关系网络等。甚至像语言和软件等许多在常人眼里并非网络的东西也可以从复杂网络的角度去研究。由于复杂性问题研究与网络的复杂性关系有密切联系,所以近年来它们之间的交叉研究引起了人们的高度重视。在上述背景下,网络科学成为 21 世纪兴起的多学科交叉的研究领域,它关注的是复杂网络的共性和处理它们的有效方法,从而增进人类对各种自然和人工复杂网络的科学理解。从网络观点重新认识事物有可能带来革命性变化的一个典型例子就是 Google 的诞生。与之前的搜索引擎相比,Google 的一个主要突破是它的 PageRank 算法利用了 WWW 的网络结构^[3]。

实际上,随着生命科学的发展、网络时代的到来以及人们交流和经济活动的全球化,人们早就开始观察和思考生命网络、技术网络、交通网络、社会网络等呈现的一些普遍现象或问题。例如,计算机病毒如何在 Internet 网中传播?艾滋病、禽流感、SARS 等疾病如何在人类和动物中传染?为什么美国的次贷危机会引发全球的经济危机?为什么各大城市堵车现象那么严重?生物体的神经系统和新陈代谢系统是如何开展工作的?是否有时你会因为“朋友认识的人刚好也是你的朋友”而感慨这个世界太小?为什么鼓掌时大家的节奏会趋于一致?所有这些问题看上去互不相关,实际上这些都是复杂网络所反映的普遍规律和复杂网络领域学者们所要研究的课题。

近 10 年来,复杂网络的研究正渗透到从物理学到生物学的众多不同的学科,对复杂网络的定性特征与定量规律的深入探索、科学理解以及可能的应用已成为网络时代复杂性科学研究中一个极其重要的挑战性课题。人们对网络的复杂性奥妙的研究只是揭示了冰山一角,大量复杂网络的奥秘有待于深入探索和研究。推进复杂性科学的交叉研究,深入探索和科学理解复杂网络的定性特征与定量规律,使它获得广泛的应用,对全球科学和社会的发展具有十分重大的长远意义^[2]。

1.2 网络科学理论发展的三个时期

1.2.1 规则网络理论阶段

规则网络理论的发展得益于图论和拓扑学等应用数学的发展^[4]。历史上,多位杰出数学家各自独立地建立和研究过图论,其中,欧拉于 1736 年首先开创了图论这门新的数学分支,因此他被誉为“图论之父”。直到两百年后的 1936 年,图论的第一本专著《有限图与无限图的理论》才由匈牙利数学家 König 发表出来。实际上,图论的研究对象就是由一些节点按照一定方式连线组成的图(集合)。用图论的语言和符号可以精确简洁地描述各种网络,为物理学家和数学家提供了共同描述语言和平台,图论的许多研究思想、技巧、成果和结论(如解决网络最短路问题、最大流问题、最小费用最大流问题的算法)能够自然地根植到复杂网络的研究中来,因此图论是一种强有力的研究工具和研究方法。以下是历史上著名的四个图论问题。

1. 哥尼斯堡七桥问题

关于图论的文字记载最早出现在 1736 年瑞士数学家欧拉的论著中,他所考虑的原

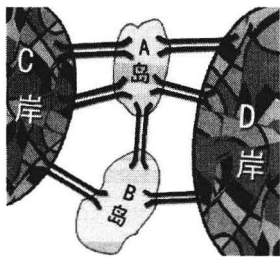


图 1.1 哥尼斯堡七桥问题示意图

始问题具有很强的实际背景,那就是著名的哥尼斯堡七桥问题^[4]。哥尼斯堡是当时东普鲁士的首都,今俄罗斯加里宁格勒市,普莱格尔河横贯其中,这条河上建有七座桥,将河中间的两个岛和河岸联结起来,如图 1.1 所示。有人在闲暇散步时提出:能不能每座桥都只走一遍,最后又回到原来的位置。这个看起来很简单却很有趣的问题吸引了大家,很多人在尝试各种各样的走法,然而无数次的尝试都没有成功。1736 年,有人带着这

个问题找到了当时的大数学家欧拉，欧拉经过一番思考，很快就用一种独特的方法给出了解答。他把两座小岛和河的两岸分别看作四个点，而把七座桥看作这四个点之间的连线，于是这个问题就简化成：能不能用一笔就把这个图形画出来？经过进一步的分析，欧拉得出结论：不可能每座桥都走一遍，最后回到原来的位置，并且给出了所有能够一笔画出来的图形所应具有的条件。这项工作使欧拉成为图论（及拓扑学）的创始人。

2. 哈密顿问题

哈密顿问题也是图论中的著名问题之一。英国数学家哈密顿于 1859 年以游戏的形式提出：把一个正十二面体的二十个节点看成二十个城市，要求找出一条经过每个城市恰好一次而回到出发点的路线，如图 1.2 所示。这条路线就称“哈密顿圈”。换一种说法，对于一个给定的网络，在确定起点和终点后，如果存在一条路径能够穿过该网络，就称该网络存在“哈密顿路径”。一百多年来，对哈密顿问题的研究，促进了图论的发展。哈密顿路径问题在 20 世纪 70 年代初，终于被证明是“NP 完备”的，也就是说具有这样性质的问题，难于找到一个有效的算法。实际上对于某些节点数不到 100 的网络，利用现有最好的算法和计算机可能也需要几百年才能确定是否存在一条这样的路径。

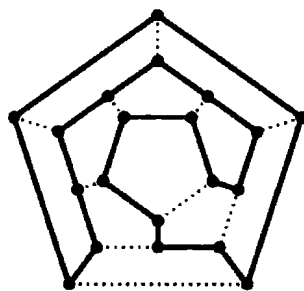


图 1.2 哈密顿问题示意图

3. 四色猜想

四色猜想又称四色问题、四色定理，是世界近代三大数学难题之一^[4]。1852 年，毕业于伦敦大学的格思里来到一家科研单位做地图着色工作时，发现了一个有趣的现象：“看来，每幅地图都可以用四种颜色着色，使得有共同边界的国家着上不同的颜色。”这个结论能不能从数学上加以严格证明呢？他和在大学读书的弟弟格里斯决心试一试，可是研究工作没有进展。他们求教著名数学家摩尔根和哈密顿也均没能解决该问题。1872 年，英国当时最著名的数学家凯利正式向伦敦数学学会提出了该问题，于是四色猜想成了世界数学界普遍关注的问题。1878~1880 年两年间，著名的律师兼数学家肯普和泰勒两人分别提交了证明四色猜想的论文，宣布证明了四色定理。但是到了 1890 年，数学家赫伍德以自己的精确计算指出了肯普的证明存在漏洞，而不久之后，泰勒的证明也被人们否定了。进入 20 世纪以来，随着电子计算机的问世，由于演算速度迅速提高，加之人机对话的出现，大大加快了对四色猜想证明的进程。1976 年，美国伊利诺伊大学哈肯和阿佩尔在大学里的两台不同的电子计算机上，用了 1200 个小时，作了 100 亿判断，终于完成了四色定理的证明。四色猜想的计算机证明，轰动了世界。它不仅解决了一个历时 100 多年的难题，而且成为了数学史上一系列新思维的起点。不过也有不少数学家并不满足于计算机取得的成就，他们还在寻找一种简捷明快的书面证明方法。

4. 旅行商问题

旅行商问题 (traveling salesman problem, TSP) 也叫货担郎问题或中国邮路问题等, 是计算机算法理论历史上的经典问题。在过去几十年中, 它成为许多重要算法思想的测试平台, 同时也促使一些新的理论领域的产生, 比如多面体理论和复杂性理论。该问题描述如下: 给定 N 个节点和任意一对节点 $\{v_i, v_j\}$ 之间的距离为 $\text{dist}(v_i, v_j)$, 要求找出一条闭合的回路, 该回路经过每个节点有且仅有一次, 并且该回路的费用最小 (这里的费用是指每段路径的距离和)。实际上, 旅行商问题就是加权的哈密顿路径问题, 因此求解旅行商问题的精确解是 NP 难的。若将问题限定在欧氏平面上, 就称为欧几里德旅行商问题, 但是它也是 NP 难的。因此, 通常用来解决 TSP 问题的解法都是近似算法。第一个欧几里德旅行商问题的多项式近似算法是由 Arora 于 1998 年使用随机平面分割和动态规划方法给出的 (发表在期刊 “Journal of the ACM” 的 1998 年 45 卷第五期)。

1.2.2 随机网络理论阶段

图论的第一本专著的出版使得图论研究进入了快速发展和突破期。1959 年, 两个匈牙利著名的数学家 Erdős 和 Rényi 又一次对图论作出了第二个里程碑式的贡献。他们建立了著名的随机图理论, 用相对简单的随机图来描述网络, 简称 ER 随机图理论^[4]。ER 随机图理论对图论理论研究的影响长达近 40 年, 以至于在随后的近半个世纪, 随机图一直是科学家研究真实网络最有力的武器。随机网络是指在由 N 个节点构成的图中以概率 p 随机连接任意两个节点而成的网络, 即两个节点之间连边与否不再是确定的事, 而是由概率 p 决定。或简单地说, 在由 N 个节点构成的图中, 可以存在 $N(N-1)/2$ 条边, 从中随机连接 M 条边所构成的网络就叫随机网络。如果选择 $M = pN(N-1)/2$, 则这两种构造随机网络模型的方法就可以联系起来。随机图和经典图之间最大的区别在于引入了随机的方法, 使得图的空间变得更大, 其数学性质也发生了巨大的变化。Erdős 和 Rényi 系统研究了当 $N \rightarrow \infty$ 时随机图性质与概率 p 的关系, 他们发现: 随机网络的许多重要的性质都是随着网络规模的扩大而突然出现的, 也就是说对于给定概率 p , 随着网络规模的扩大, 要么几乎所有的随机图具有某种性质, 要么几乎每一个图都不具有该性质。

Erdős 被称为 20 世纪的欧拉, 于 1984 年获得沃尔夫奖^[4]。他善于与人合作, 打破了数学领域的喜欢个人独立研究的传统, 一生中同 511 位合作者发表过约 1500 篇数学论文, 涉及数学的许多领域, 与那些伟大的物理学家和数学家, 如爱因斯坦、哥德尔、奥本海姆等有都密切学术交往。Erdős 是一个数学“苦行僧”, 他从来没有固定的职位, 从来不定居在一个地方, 也没有结婚, 但更是一个流浪学者, 随时带着手提箱, 穿梭于学术研讨会, 浪迹天涯, 颇富传奇色彩。几乎每一个当代数学家都有一个有限的 Erdős 数, 而且这个数往往非常小, 小得出乎本人的预料。

1.2.3 复杂网络理论阶段

1. 小世界效应的发现

1998年,网络科学又一次取得突破性进展,出现了第三个里程碑,这在很大程度上要归功于计算机技术和Internet技术的迅猛发展^[4]。美国的瓦茨和斯特罗加茨首先突破了随机网络理论的框框,于1998年在“Nature”上发表了题为《“小世界”网络的群体动力行为》的论文^[5],他们推广了“六度分离”的科学假设,提出了小世界网络模型。“六度分离”最早来自于20世纪60年代美国哈佛大学心理学家Milgram对社会调查的推断,是指在大多数人中,任意两个素不相识的人通过朋友的朋友,平均最多通过6个人就能够彼此认识。为了进一步验证六度分离假设,1997年三个美国大学生发明了一个名叫“Kevin Bacon”的游戏(他们认为美国电影演员Kevin Bacon是电影界的中心)。世界上每位电影演员,通过跟他共同演过电影的演员,可以最终联系到Kevin Bacon。如果一个人跟Kevin Bacon演过电影,他的Bacon数就是1,如果一个人跟Kevin Bacon演过电影的人演过电影,他的Bacon数就是2,以此类推。比如成龙演了“Around the World in 80 Days”(2004年),其中有Luke Wilson,而Luke Wilson演了“My Dog Skip”(2000年),其中就有Kevin Bacon。所以成龙的Bacon数是2。人们可以访问网站<http://oracleofbacon.org/>去查任何一个演员的Bacon数,到2010年5月数据库里存有3 585 458个演员信息以及1 604 556部电影电视信息。表1.1是对所有这些近360万个演员所做的统计,左边是Bacon数,右边是拥有这个Bacon数的演员个数,可以看到最大的Bacon数仅仅为8,而近360万个演员的平均Bacon数仅为2.98。

表 1.1 电影演员的 Bacon 数 (截至 2010 年 5 月)

Bacon 数	演员数	Bacon 数	演员数	Bacon 数	演员数
0	1	3	719 767	6	1 040
1	2 251	4	178 784	7	165
2	225 506	5	12 205	8	17

在数学界也有类似的游戏,他们建立关联的方式是看他们是否合作发表过论文。其中当代最伟大的数学家之一Erdős成为了数学界的中心。比如说证明Fermat大定理的Andrew Wiles,他的研究方向与Erdős相去甚远,但他的Erdős数只有3,是通过这个途径实现的:Erdős—Andrew Odlyzko—Chris M. Skinner—Andrew Wiles。借助迅猛发展的Internet,瓦茨领导的研究小组在2003年又发表一个实验报告。他们在全世界范围内检验了上述惊人的“六度分离”假说,有6万多志愿者参与利用电子邮件通信实验,确实不到6步就实现了他们的假设,从而利用Internet初步验证了复杂网络的小世界效应。

2. 社会网络中弱连接优势的发现

在20世纪60年代晚期,哈佛大学Granovetter通过寻访麻省牛顿镇的居民如何找