

复杂网络的结构与演化

郑波尽 著



科学出版社

复杂网络的结构与演化

郑波尽 著

科学出版社

北 京

版权所有，侵权必究

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

内 容 简 介

复杂网络是描述生命、人类社会技术等领域的复杂系统的有力工具。本书从复杂网络的数据分析、复杂网络的层次与演化、复杂网络的抗攻击性及复杂网络的多目标优化建模等方面展开了研究。在数据分析方面，讨论了复杂网络的重要节点和骨干网络的提取方法，将层次结构和抗攻击性关联起来；在层次与演化方面，讨论了层次结构与网络演化之间的关系，并用来解释经济现象；在动力学方面，得出了复杂网络在选择性攻击下具有鲁棒性的结论；在建模方面，利用多目标优化方法来对复杂网络建模，用一个框架得到多种复杂网络，确立了复杂网络之间的关系，修正了复杂网络领域中多个被广泛传播的结论。

本书适用于本领域研究人员和相关专业学生参考。

图书在版编目(CIP)数据

复杂网络的结构与演化/郑波尽著. —北京：科学出版社, 2018.11

ISBN 978-7-03-058164-8

I. ①复… II. ①郑… III. ①计算机网络—研究 IV. ①TP393

中国版本图书馆 CIP 数据核字(2018) 第 139309 号

责任编辑：闫 陶 / 责任校对：肖 婷

责任印制：彭 超 / 封面设计：彬 峰

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>

武汉中科兴业印务有限公司 印刷*

科学出版社发行 各地新华书店经销

*

2018 年 11 月第 一 版 开本：B5 (720 × 1000)

2018 年 11 月第一次印刷 印张：15 插页 6

字数：317 000

定价：75.00 元

(如有印装质量问题，我社负责调换)



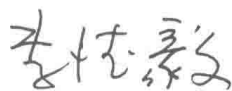
序

复杂网络是一个跨学科的新兴研究领域。在社会科学中，人们可用它来研究科学家的合作关系；在技术方面，人们可用它来研究万维网的性质和计算机软件的诸多性质；在自然科学方面，人们可用它来研究生物的捕食关系以及蛋白质的合作关系。复杂网络的研究得到了人工智能学家、计算机科学家、物理学家、化学家、生物学家、社会学家等的广泛关注。

复杂网络的研究热潮肇始于 1998 年小世界效应以及 1999 年无标度网络的提出。此后，大量研究者进入这个领域，快速地建立起了该领域的基本知识架构，最为人知的几个关键词是：小世界效应、无标度属性、分形网络、社区结构网络等。与这些关键词相关联的还有一些重要的结论，例如，无标度网络呈现既鲁棒又脆弱的特性，无标度网络的偏好连接机制等。

本书以独特的视角来审视现有的知识架构，得出了一些令人意想不到的结论。例如，改变了无标度网络既鲁棒又脆弱的简单结论，发现网络的鲁棒性既与紧致性有关，又与平均度有关，有些无标度网络并不脆弱，而是比较鲁棒的。再例如，当用优化模型来统一描述多种特征网络时，得到了网络特征之间的谱关系，很多曾经难以理解的片段性知识就被关联起来了。还有，发现了具有集散节点聚集行为的新颖分形结构网络，改写了分形结构起源于集散节点排斥行为的结论，还改写了异配性和分形性关联的结论。作者的研究结果，体现了人们对于复杂网络的新理解。

本书围绕复杂网络的结构和演化问题，按作者在复杂网络研究中的不同时期成果作为线索来撰写，体现了作者学术观点的不断发展过程，也体现了人们对复杂网络在不同阶段的不同理解，表现出作者追求科学真理的精神。对于有志从事科研工作的研究生而言，可以提供一個难得的借鉴。这也是本书的另外一个价值所在。



中国工程院院士

前 言

1998 年是复杂网络研究的勃发之年,主要的标志性事件是小世界效应的提出。小世界效应认为在网络直径上存在一个不变量,从现在的研究结果来看,这个不变量的确定性不强,比如超小世界网络和分形结构的网络在网络直径上就不同于小世界网络;小世界效应认为小世界网络是随机网络和规则网络的中间状态,从现在的研究结果来看,小世界效应在刻画网络之间关系的问题上还略显单薄,其他类型的网络并没有涉及。

1999 年,复杂网络的无标度属性被再次发现。研究者从时间维度的累积效应(累积优势、马太效应)上进行了深入的研究,但缺乏空间维度的累积效应方面深入细致的研究。之后,无标度网络被发现具有“随机攻击下鲁棒、选择性攻击下脆弱”的特性。从现有的研究结果来看,这一特性的隐含前提是“删除每个节点付出的代价相等(即与边无关)”。从该前提可以推导出:边攻击下的研究结论和节点攻击下的研究结论不应该一致。为了解决现有文献互相冲突的局面,本书提出了统一的度量方法,得出了一般性的结论:复杂网络的抗攻击性来源于网络的紧致性和平均度,“鲁棒且脆弱”只是特例下的表现。

2002 年,复杂网络的社区结构开始受到关注。但社区结构网络的建模常常采用不优雅的组合法。从现在的研究结果来看,社区结构网络的建模应与网络结构的演化相关联。此外,社区结构的形成原因应归因到节点的类别距离,而非通常认为的拓扑距离上。

2006 年,分形结构的复杂网络被发现。人们普遍认为,分形结构来源于 Hub(集散)节点之间的排斥,并且同/异配性在分形结构网络中具有重要地位。本书给出了一系列分形网络,这些网络中 Hub 节点彼此紧密连接在一起。这些网络构成了原有理论的反例。从而,本书提出了新的关于分形网络的理论。

在目前复杂网络最重要的几个研究结论上,本书都做出了有意义的工作。这是本书在几个点上面的贡献。事实上,本书各个章节之间互相支撑,形成了一个整体。例如,对社会网络的分析结果引导我们质疑无标度网络“鲁棒性与脆弱性并存”;对复杂网络抗攻击性的研究启发我们建立统一的复杂网络建模的多目标优化模型。此外,本书包含了大量的思辨性论述,有助于读者了解具体研究的出发点和研究路径。

总体上来说,作者认为,复杂网络的结构和演化相生相伴,互为因果。在该观点指导下,本书从结构和演化的角度探索了复杂网络的性质,在更一般化的视角上

重新审视了复杂网络中的一些经典结论，得到了一些不一样的结果，并且发展出了一些有用的方法。对于本书读者，建议从经典结论的最基本的逻辑基础这个角度审视经典结论和本书的结论。

对于这本书的出版，最应该感谢的是李德毅院士，在长期的研究工作中，李院士给予了我悉心的指导和巨大的帮助。

感谢吴泓润博士、匡立博士、黄丹女士、苏杨茜女士。吴泓润博士和我一起研究了复杂网络的抗攻击性、复杂网络的优化与演化。匡立博士和我一起研究了分形网络。黄丹女士、苏杨茜女士和我一起研究了复杂网络的抗攻击性。没有他们提供的支持和帮助，很多工作是难以取得成功的。他们也为该书的审校做了大量的工作。

感谢马于涛博士、江健博士、张海粟博士、张书庆先生、韩言妮博士、陈桂生博士、强小利博士、杜文华博士、覃俊博士、周爱民博士以及很多同事、朋友在研究过程中给予的帮助。

感谢王璐、钟凯瑞、齐鑫、李开菊、邓静等同学在稿件校对中给予的帮助。

感谢家人一以贯之的支持。

由于作者学术水平有限，书中难免存在疏漏，恳请读者批评指正。

作 者

2018 年 3 月

目 录

序 前言

第一篇 概论

第 1 章 概述	3
1.1 复杂系统与复杂网络	3
1.2 复杂网络的特征和类型	4
1.2.1 随机网络	4
1.2.2 小世界网络	4
1.2.3 无标度网络	6
1.2.4 超小世界网络	8
1.2.5 网络的社区结构	8
1.2.6 网络的分形结构	9
1.3 本书研究内容	9
1.3.1 网络化数据挖掘	10
1.3.2 无标度网络演化机制	12
1.3.3 网络抗攻击性	15
1.3.4 复杂网络的统一优化建模	16
1.4 讨论	17
1.4.1 复杂网络研究哲学	17
1.4.2 复杂网络研究的弱点	17

第二篇 复杂网络的数据挖掘

第 2 章 复杂网络的节点重要性	21
2.1 相关工作	23
2.1.1 定义节点的重要性	23
2.1.2 冲突消解	24
2.1.3 度量节点重要性排序算法的效能	25
2.2 重要性的基于直觉的规则	25
2.2.1 规则 and 选择规则	25

2.2.2	形式化 5 条规则	26
2.3	用偏序对节点排序	27
2.3.1	基本概念	28
2.3.2	等价类算法	29
2.4	实验结果 I	30
2.4.1	用于测试的网络实例	30
2.4.2	实验网络的等价类	31
2.5	度量排序算法的性能	33
2.5.1	覆盖率	34
2.5.2	覆盖率的算法	35
2.6	实验结果 II	36
2.7	拓扑势对节点按重要性排序	38
2.7.1	拓扑势方法介绍	38
2.7.2	拓扑势方法的实验结果	39
2.8	本章小结	41
第 3 章	复杂网络的骨干	44
3.1	相关研究	44
3.2	骨干网络的度量	45
3.3	基于等价类算法的骨干网络提取算法	47
3.3.1	全骨干网络提取算法	48
3.3.2	全骨干网络提取算法实验结果	49
3.3.3	层次骨干网络提取算法	51
3.3.4	层次骨干网络提取算法实验结果	51
3.4	关于网络分裂的假说	53
3.5	Web 服务网络的骨干网络	55
3.6	Internet 的结构	60
3.7	基于拓扑势方法的骨干网络提取算法	61
3.8	基于拓扑势方法的骨干网络提取算法实验结果	61
3.9	本章小结	63

第三篇 复杂网络的层次与演化

第 4 章	复杂网络的层次演化模型	67
4.1	无标度网络与偏好	67
4.2	对 Web 链接的幂律分布的一个解释	68
4.2.1	Web 链接的隐树模型介绍	69

4.2.2	实验结果	71
4.2.3	Web 链接的幂律分布的讨论	74
4.3	隐树结构重叠带来的效应	75
4.4	隐树模型理论分析	75
4.5	本章小结	77
第 5 章	财富分布建模	79
5.1	马太效应的由来	79
5.2	目前的马太效应的解释	79
5.3	马太效应的新机制和建模	80
5.4	财富分布的实验结果和分析	82
5.5	财富分布的隐树结构模型的讨论	84
5.6	本章小结	85
第四篇 复杂网络的抗攻击性		
第 6 章	复杂网络的抗攻击理论	89
6.1	传统理论及分析	89
6.2	代价攻击理论	90
6.3	相关定义	91
6.3.1	攻击的定义	91
6.3.2	攻击代价的定义	92
6.3.3	攻击策略的定义	93
6.3.4	攻击效果的定义	94
6.4	代价条件下无标度网络的抗攻击性分析	95
6.4.1	紧致性网络的定义	95
6.4.2	理论性讨论	96
6.5	仿真实验	97
6.5.1	CSF 网络	97
6.5.2	Polbook 网络	98
6.5.3	CSF 网络	99
6.5.4	Protein 网络	99
6.5.5	Netscience 网络	100
6.5.6	实验网络相关数据表	101
6.6	实验分析	101
6.6.1	紧致程度对抗攻击性的影响	101
6.6.2	平均度对抗攻击性的影响	104

6.7	理论证明	106
6.8	本章小节	107
第 7 章	复杂网络的抗攻击理论与边攻击	109
7.1	潜在佯谬	109
7.1.1	第一表述	110
7.1.2	第二表述	111
7.1.3	第三表述	112
7.2	定量化方法	114
7.2.1	正则化网络性能	114
7.2.2	基准线的定义	115
7.2.3	抗攻击性指标	116
7.3	定量化指标下小世界网络的抗攻击性分析	117
7.3.1	边攻击定量仿真实验	118
7.3.2	节点攻击定量仿真实验	120
7.4	定量化指标下无标度网络抗攻击性分析	122
7.4.1	实验网络说明	123
7.4.2	鲁棒性指标	123
7.4.3	边攻击策略	124
7.4.4	实验结果与分析	124
7.5	本章小节	127
第五篇 复杂网络的优化与演化		
第 8 章	复杂网络的优化模型	131
8.1	复杂网络的典型类型及传统的演化机制解释	131
8.1.1	无标度网络及演化机制简介	132
8.1.2	小世界网络及演化机制	134
8.1.3	社区结构网络及演化机制	135
8.1.4	分形网络及演化机制	135
8.2	最优化理论与复杂网络演化机制	135
8.3	双优模型	136
8.3.1	方法	137
8.3.2	模型介绍	137
8.4	理论分析结果	139
8.4.1	网络的类型	140
8.4.2	实验仿真	144

8.4.3	网络类型之间的关系	160
8.4.4	改进的快速算法	161
8.5	本章小结	172
第 9 章	复杂网络多目标优化模型的深入研究	173
9.1	社区结构网络的优化建模	173
9.1.1	相关工作	175
9.1.2	基于优化理论建模社区无标度网络的必要性	176
9.1.3	社区无标度网络优化模型	177
9.1.4	仿真实验	179
9.2	分形网络的演化机制研究	187
9.2.1	复杂网络中的分形维度定义及其算法	189
9.2.2	有关分形网络的几个重要观点	192
9.2.3	复杂网络分形的应用研究	195
9.2.4	具有 Hub 吸引行为的分形和无标度的复杂网络模型	195
9.2.5	优化模型所得到网络的性质分析	201
9.2.6	网络的同配性分析	202
9.3	本章小结	204

第六篇 总结与展望

第 10 章	总结与展望	209
10.1	总结	209
10.2	对复杂网络的新理解	212
10.3	展望	215
参考文献		217

第一篇

概 论

第1章 概述

1.1 复杂系统与复杂网络

在现实世界中，我们能轻而易举地找到一些非常复杂的系统。例如，一个基因组、一个细胞、一个人类社会、一个动物种群、一个生态系统等。这些系统中单元类型众多、数量巨大，且单元之间的相互关系复杂多变。例如，人类社会共计有70亿个个体，个体之间职业不同，社会关系纷繁复杂。这样的系统是典型的复杂系统。

复杂系统常常由 10^8 个以上的单元组成，对如此众多的单元进行研究是一件极其困难的事情，尤其是当组成系统的单元具有自主性的时候。在研究较为简单的系统时，人们常常采用还原论的方法。方法大致流程是：对这一系统抽象化处理，为系统建立一个模型，对模型进行研究，探寻模型的性质，然后通过实际数据来验证模型与实际情形的对应程度，从而得出实际系统满足这样性质的可靠程度，进而根据这个模型进行合理的预测。然而，世界是普遍联系的，普遍联系下的整个世界是一个复杂巨系统。这个复杂巨系统在演化的过程中，形成了明显的层次结构。在不同的层次上，单元之间的相互作用往往是大不相同的。也就是说，在不同层级下，系统中涌现出来的行为迥然不同。目前，人类的知识水平还难以准确预测涌现出来的性质和行为。对这样的系统从经典还原论的角度，或者说从细部进行研究常常难以得到全局性的知识。

现在看来，一个有用的办法是：将这些单元抽象成节点，将单元之间的复杂关系抽象成边，从而整个系统就可以抽象成一个图。通过研究图的性质，就可以在大尺度的范围内了解整个系统的性质。这种方法从本质上讲仍然可以称为一种基于还原论的方法，但是研究的对象是系统整体而非局部，因此可以称得上是还原主义上的整体论方法。这种方法既满足了科学方法的基本条件，从而可得到确定性的知识，又可免于笼统和含糊，防止陷入无意义的纯哲学思辨甚至玄学的泥沼，因而得到了快速的发展。

这种从研究复杂系统的关系入手来研究复杂系统的方法是一种新方法。这种方法所构造生成的图称为复杂网络。

目前，复杂网络的方法已经被用于很多科学研究领域。有些研究人员将研究复杂网络的理论和方法统一称为“网络科学”。由此，网络科学 (network science) 成为一个新的专有名词。

本书在复杂系统的背景下研究复杂网络的结构与演化,从而得到了一些有趣的结果。

1.2 复杂网络的特征和类型

通过多年的研究,人们发现复杂网络常常具有一些典型的特征。这些特征可以单独存在,也可以组合存在。人们依据复杂网络的特征将网络进行类型划分。例如,网络常常具有无标度属性^[1],人们就将具有无标度属性的复杂网络称为无标度网络;网络也常常具有小世界效应^[2],人们将具有小世界效应的网络称为小世界网络。但是,存在一些网络既具有小世界效应,又具有无标度属性,因此按照这样的特征对网络进行分类不是一种严格意义上的分类。因为定类需要满足两个原则:①互斥原则,即类与类之间应当互相排斥,也就是说每个研究对象只能归入一类;②无遗漏原则,即每一个研究对象均有归属,不可遗漏。现有的文献介绍复杂网络的分类时,常常不注意定类的原则,这样就导致在讨论复杂网络的类型时,会存在模糊和概念不清楚的地方。在本书中,我们沿用学科领域的历史说法,即通常不深究类型和特征的区分。只在必要时,使用“特征”来代替传统上的“类型”的说法。

人们已经发现了很多复杂网络的特征。主要有随机、小世界、无标度、超小世界、社区结构、分形结构等,沿用历史性说法,分别将具有这些特点的网络称为随机网络、小世界网络、无标度网络、超小世界网络、社区(结构)网络、分形(结构)网络等。下面将分别予以简单介绍。

1.2.1 随机网络

对网络的普适规律和不变量的研究可以追溯到 Erdős 和 Rényi 的研究^[3]。1959年,Erdős 和 Rényi 提出可以通过网络节点间以不变的概率 p 随机地布置连线来有效模拟通信和生命科学中的网络。在 Erdős-Rényi 模型里,节点的度分布遵循泊松分布,这意味着比平均链接度高很多和低很多的节点都很少见,网络非常“民主”,因而也称为均匀网络(homogeneous network)。Erdős-Rényi 模型所生成的随机网络还具有节点之间的聚集程度很低的现象,以及具有比较大的特征路径长度。此外,人们还发现,当概率 p 到达一定值时,网络中最大的连通图的尺寸也呈现出一个相变^[3]。

1.2.2 小世界网络

基于社会网络(social networks)、生物网络以及技术网络的研究,Watts 和 Strogatz 研究了网络的特征路径长度,研究结果显示^[2]:很多实际网络的平均最短路径长度较小,而聚集系数相对较大。

所谓平均最短路径长度,即网络或图中任意两个节点之间的最短路径长度取平均值。假如用 $\text{distance}_{i,j}$ 表示无向图中节点 i 和节点 j 之间的最短距离,则平均最短路径长度 (average shortest path length, ASPL) 可以表示为

$$\text{ASPL} = \frac{2}{N(N-1)} \sum_{i>j}^N \text{distance}_{i,j} \quad (1.1)$$

至于聚集系数,则是用于度量无向图中节点之间聚集程度的指标。聚集程度定义为任意三个节点之间抱团的概率,即节点 i 、节点 j 和节点 k 三者之间均有边的概率。聚集系数 (cluster coefficient, CC) 用公式可以表示为

$$\text{CC} = \frac{1}{N} \sum_{i=1}^N \text{CC}_i = \frac{1}{N} \sum_{i=1}^N \frac{2E_i}{K_i(K_i-1)} \quad (1.2)$$

其中, CC_i 表示节点 i 的局部聚集程度; E_i 表示节点 i 的边数; K_i 表示节点 i 的所有邻点的边数和。

从现实生活经验中很容易理解小世界现象。假如你有在火车上和陌生人聊天的经历,说不定就会发现,陌生人和你是校友或者在同一家公司工作过,甚至你们有共同的朋友。很多人都有这样的经历,他们常常会发出感叹,世界真小啊。这就是小世界效应所描述的现象。

小世界效应很容易理解。假如一个人有一千个熟人,那么他的熟人的熟人就有一百万之多(假如这些人之间不构成小世界,没有交集),当我们继续推导时,数字会指数级地迅猛增加。目前,整个世界只有 70 亿人,这里当然存在大量的重复。因而,我们一方面能够用很短的熟人距离来找到每一个人,又能保证大家聚集成团,常常出现朋友的朋友是朋友的现象。另外,由于人群具有“物以类聚,人以群分”的现象,常常使得小世界效应更加明显。

为了解释小世界效应带来的现象,1998 年, Watts 和 Strogatz 在 *Nature* 上发表论文^[2],提出了“小世界模型”,人们以两人名字的首字母命名为 WS 模型。该模型兼具规则网络较大的聚集系数和随机网络较小的平均距离的特点,能描述完全规则网络到完全随机网络之间的转变。也就是说, WS 模型认为,小世界现象是一个规则网络和完全随机网络的中间形态。

WS 模型使用规则图作为初始网络,令其中的点按一定的概率将最近的某些邻居变更为随机的邻居。以环形网络为例,每个节点都与它左右相邻的节点以一定的概率 p 相连,在合适的参数值下,所得到的网络就是小世界网络。

WS 模型的伪代码如下所示。

1. Initialize a ring network with n nodes and $n * k$ edges
2. Produce a random float p , $0 \leq p \leq 1$

- 3. For each edge of every node
- 4. if random(0, 1) ≤ p then change its destination node randomly
- 5. end for

WS 模型在不同参数值下得到的网络如图 1.1 所示。

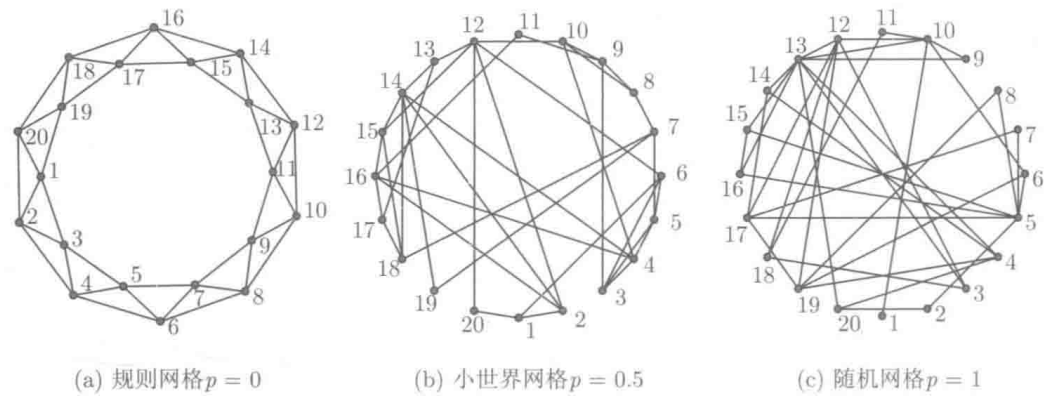


图 1.1 WS 小世界模型下网络的拓扑结构

1.2.3 无标度网络

在 Web 出现以后，人们可以方便地获得这个大规模网络的数据，因此也有机会在这样的实际网络中验证随机网络的理论。

按照小世界理论，大规模网络也具有较小的直径。R.Alert 等的研究表明，整个 World Wide Web 的直径确实非常小 [4]。

另外，由随机网络理论可以推出，假如网页是随机链接，网页的度分布将呈现钟形分布，即泊松分布形态。Barabási、Albert 和 Adamic 发现，Web 网页的链接和随机图论预测的不一样，每个网页的入度非常不均衡，服从幂律分布 [1, 5, 6]。

所谓幂律分布，指变量 x 的分布函数为一个幂函数，即

$$p(x) \approx x^{-r} \tag{1.3}$$

幂函数具有一些有趣的性质，如自相似性。

所谓自相似性，指事物的局部和全局具有相似的形状或者特征。例如，树就是具有自相似性的例子。从树干出发，进行分叉，树的每一个枝又可以作为“树干”，继续分叉，不停重复，就变成了最终的树。对于具有自相似性的事物，不具备特征的尺度或者标度。例如，我们可以把自相似事物的一部分进行放大，这样就可以得到整体的图像，但无法找到一个特征量来定义事物的尺寸大小。这就是“无标度”这一概念的由来。