

网络科学与工程丛书

N S E

1

网络度分布理论

Theory of Network Degree
Distributions

■ 史定华 著

 高等教育出版社
HIGHER EDUCATION PRESS

网 络 科 学 与 工 程 丛 书

N S E

1

WANGLUO DU FENBU LILUN

网络度分布理论

Theory of Network Degree
Distributions

■ 史定华 著



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

图书在版编目(CIP)数据

网络度分布理论/史定华著. —北京: 高等教育出版社, 2011.3

(网络科学与工程丛书/陈关荣主编)

ISBN 978-7-04-031513-4

I. ①网… II. ①史… III. ①系统复杂性-研究
IV. ①N94

中国版本图书馆 CIP 数据核字(2011)第 011979 号

策划编辑	刘 英	责任编辑	廖肇源	封面设计	李卫青
责任绘图	尹 莉	版式设计	李卫青	责任校对	杨雪莲
责任印制	刘思涵				

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

经 销 蓝色畅想图书发行有限公司
印 刷 北京中科印刷有限公司

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.landracom.com>
<http://www.landracom.com.cn>
畅想教育 <http://www.widedu.com>

开 本 787×1092 1/16
印 张 14
字 数 190 000
插 页 2

版 次 2011 年 3 月第 1 版
印 次 2011 年 3 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 31513-00

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的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被列入《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、统计力学、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广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），以及时间序列网络分析（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，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。我们热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，推动该学科领域的普及，并为其深入发展作出贡献。

金芳蓉（Fan Chung）院士
美国加州大学圣地亚哥分校
2011 年元月

前言

20 世纪的一项重大科研成果是认识到简单系统会展现复杂行为。这些简单系统通常采用差分、微分或迭代过程来描述，由于非线性而展现出复杂行为，这是复杂性的一种重要范式。世纪之交，在《自然》和《科学》期刊上相继发表了三篇重要的网络文章：小世界网络模型说明了少量的随机捷径（长程连线）会改变网络的拓扑结构，从而涌现出小世界效应；无标度网络模型揭示了增长和择优机制在复杂网络自组织演化过程中的普遍性和幂律的重要性；可导航网络模型则解释了如何利用局部信息来找到这条最短路径。上述模型都是基于几条简单规则反复迭代而得到复杂的网络结构，这说明大量复杂系统也可由某些简单规则自组织演化而形成。这项新的重大科研发现是复杂性的另一种重要范式。

三篇重要的网络文章的发表立即引起了世界科学界和工程界对网络科学和工程应用研究的巨大热情。其中十年前 Barabási 和 Albert 在《科学》上发表的论文《随机网络中的标度涌现》尤其引人注目。论文的主要思想有：从许多实际复杂网络度分布的统计结果发现具有幂律尾部是其普遍的特征。他们提出一个择优增长的动态模型来解释产生幂律的机制，并且认为正确的模型其网络度分布应该独立于时间，从此具有幂律度分布的网络被称为无标度（scale-free）

网络。为了更好地理解无标度网络的拓扑结构和特性，Barabási 等人又在《自然》和《科学》上发表了针对细胞代谢网络的确定层次网络模型，探讨复杂网络实际应用和建模分析，还在《自然》上阐述了无标度网络具有“稳健而又脆弱”的特性。这些思想新颖动人，触发了许多后继研究工作，也留下了广阔的发展空间。

首先，Bollobás 和 Riordan 指出 BA 模型有两点不够明确：一是初始网络没有设定；二是有多条连线时如何进行择优连接？模型不明确定义将难以进行理论研究。

其次，加州理工学院 Li 等人认为仅以度分布是幂律作为无标度网络的定义不尽合理，因为幂律是方差跨度极大的高可变分布，由于存在极大的波动性，所以幂律相同的度分布网络完全可能具有截然不同的拓扑结构和特性。因特网和代谢网就是两个典型的案例，它们面对蓄意攻击仍然是“稳健而非脆弱”，并给出了相应的解释。

最近，我们发现 BA 模型网络度分布独立于时间只是理想化的个案，这个要求对许多实际网络统计和模型网络模拟都无法满足，必须考虑网络有限规模的影响。最近许多研究工作报道了网络动力学行为明显依赖于网络规模的大小，从而引发了对有限模型网络最大度重要性和度-度相关性的讨论，这里有许多概念似乎需要厘清。

另外，实际网络的统计和模型网络的模拟得到的都是确定的数据，如何判断它们的度分布为无标度？又如何估计它们的度指数？从确定的层次网络模型引入起，层次网络、伪分形图及阿波罗网，中间都有许多度（几何增长）的概率为零。这类几何增长网络的度分布和度指数，从提出起就一直是颇有争议的论题。在权威期刊上对这个问题的认识迄今也没有令人信服的说法，而且实际网络度分布的统计也会遇到类似问题。

基于上述背景,对于复杂网络的重要指标网络度分布,人们不能只满足于简单的统计、模拟和启发式推导,而需要从理论高度给予严谨的探索。为此,首先我们对度分布常用统计方法给出评述,阐述正确的统计方法对网络科学的重要性。接着我们引入两个图值马氏过程——结点 i 度数非齐次马氏链和度 k 结点数向量马氏链,考虑三个重要理论问题——网络度分布的稳定性,网络度 k 结点数的极限定理,有限网络最大度的扰动和发散规律。然后我们对网络最大度和度-度相关性展开了讨论,我们同意 Barabási 的观点:“除非(深入)探讨网络拓扑,否则无法理解复杂系统。共性是存在的,(只是)还没有发现能够解释它们普遍性的框架。”

本书就是围绕上述框架来探讨网络度分布理论,由六章和两个附录组成。第一章从图论到网络,介绍某些有代表性的复杂网络模型并提炼出网络度分布需要研究的理论问题。第二章度分布的统计,介绍实际网络和确定的模型网络度分布正确统计方法。这些确定的模型网络每步的结点数按几何级数增长,它们在文献上分别称为层次网络、伪分形图及阿波罗网等,其度指数的确定至今仍存在争议。第三章度分布的计算,介绍四种模型网络度分布的计算方法,它们是平均场方法、率方程方法、主方程方法和马氏链方法。第四章度分布的稳定性,通过引入两个网络马氏链(图值马氏过程),首先严格证明了几个类似 BA 模型网络度分布的稳定性;其次导出一类增长网络模型的稳定性条件和成为无标度网络的条件;最后给出该类网络瞬时度分布的迭代计算公式,进一步,若收敛还得到了收敛速率和误差上界。第五章最大度的重要性,讨论网络有限规模的影响,包括网络有限标度的问题,最大度的发散和扰动规律,部分结点随机失效稳健和蓄意攻击中枢点集脆弱的特性,以及物理学家与数学家对这个问题所作的分析。第

六章度-度的相关性，从无标度网络是否“稳健而又脆弱”出发，讨论网络同配和异配，网络的无标度程度，提出度-度相关性的正确测量问题，以及网络联合度分布的计算，说明问题将涉及网络结构与功能关系的研究。附录 A 网络马氏链的程序，列出了本书用到的与网络马氏链有关的计算程序。附录 B 度分布的极限定理，介绍了随机图论学者在增长随机树方面利用鞅论等工具所取得的初步成果。

现在全世界对复杂网络研究都寄予很高的期望，发达国家如英、法、美等相继提出了“复杂系统科学研究路线图”，美国国家自然科学基金委员会也将“网络科学与工程”提上了议事日程。复杂网络要成为一门新科学，加强基础理论研究势在必行且有重要意义。特别是建立复杂性两翼“混沌和网络”相互作用和协调统一的理论框架，仍然是我们需要攻克的下一个科学难题。在此寄希望于年轻有为的中国学者。

本书的写作过程中得到许多专家学者的支持和鼓励，特别要致谢的有中国科学院郭雷院士，香港城市大学陈关荣讲座教授，北京师范大学狄增如教授和中南大学侯振挺教授，其中与侯教授在度分布稳定性方面的讨论使我获益匪浅。同时本书的研究阶段还得到国家自然科学基金项目（编号为 60874083 和 10872119）的资金支持。在书中我们引用了国内外同行的许多重要成果，我的学生周晖杰帮助编写了相关程序，在此一并表示由衷的感谢。

由于时间匆促，加之水平有限，错误在所难免，敬请读者指正。希望这本小书能够起到抛砖引玉的作用，繁荣我国网络科学与工程的研究。

史定华
上海大学数学系
2010 年 9 月

符号一览表

初始网络结点数: m_0

初始网络总度数: k_0

结点的初始权重: $\omega_0 > -1$

网络演化时间步: $t = 1, 2, \dots$

第 i 步加入的新结点: i

新结点 i 带入的度数分布: $\alpha_i(h)$, $h = 0, \dots, i-1+m_0$

新结点带入的最小度数: $m = \min_i \{h_i\} \geq 0$, 其中 $h_i = \min \{h \mid \alpha_i(h) > 0\}$

结点 i 在 t 步结束时的度: $k_i(t)$, $i \leq t$

在 $t+1$ 步择优选择度为 $k_i(t)$ 的旧结点 i 的概率:

$\Pi[k_i(t)]$

在 t 步度为 k 的旧结点 i 在 $t+1$ 步获得度加 1 的概率:

$f_i(k, t)$

在 $t+1$ 步第 h 条新连线所连接的旧结点的度: $\xi_h(t)$

结点 i 在 t 步结束时的邻域: $O_i(t) = \{j \mid i \leftrightarrow j\}$

结点 i 在 t 步结束时的权: $x_i(t) = k_i(t) + \omega_0$, $i \leq t$

结点 i 在 t 步结束时度为 k 的概率: $P(k, i, t)$

网络在 t 步结束时的最大度: $k_m(t)$

网络在 t 步结束时度为 k 的结点数: $N_k(t)$

网络在 t 步结束时度为 k 的结点平均数: $S(k, t) =$

$$\sum_i P(k, i, t) = E[N_k(t)]$$

网络在 t 步结束时瞬时度分布: $P(k, t)$

网络在热力学极限下的稳态度分布: $P(k)$

网络平均度: $\langle k \rangle$ 或 $\bar{k}$

网络平均度函数: $\bar{k}(s, t)$

网络稳态度的补分布: $F_c(k)$

网络稳态度的概率密度函数: $p(k)$

网络稳态剩余度分布: q_k

度 k 结点的群集系数: $C(k)$

网络平均群集系数: $\bar{C}$

有限网络结点数: $N=t+m_0$

有限网络连线数: E

网络规模为 N 时的结点度序列: $\{d_1, d_2, \dots, d_N\}$

网络动力学指数: β

网络度分布指数: $\gamma=1+(1/\beta)$

网络中心化变量: $C_k(t) = N_k(t) - (t+m_0)P(k)$

网络中心化均值: $\bar{C}_k(t) = S(k, t) - (t+m_0)P(k)$

度 k 结点的邻结点平均度: $\langle k_{nn} \rangle$

网络 g 的相关系数: $r(g)$

网络 g 的无标度程度: $S(g)$

结点 i 的混合系数: $I(i)$

网络 g 的混合系数: $I(g)$

网络 (在 t 步结束时度为 k 和 l) 相邻结点对 ($i \leftrightarrow j$) 的数目: $N_{k,l}(t)$

网络 (在 t 步结束时度为 k 和 l) 相邻结点对的瞬时联合度分布: $P(k, l, t)$

网络在热力学极限下 (度为 k 和 l) 相邻结点对的稳态

度分布: $P(k, l)$

Kronecker 符号: $\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

δ -函数: $\delta(x-x_0) = \begin{cases} \infty, & x=x_0, \\ 0, & x \neq x_0, \end{cases} \int_0^\infty \delta(x-x_0) dx = 1$

增量: $\Delta_k(t) = Y(t+1) - Y(t)$

目 录

第一章 从图论到网络	1
§ 1.1 网络研究的三个里程碑	2
§ 1.2 复杂网络分类和度分布	9
§ 1.3 复杂网络的建模和模拟	15
§ 1.4 无标度网络概念的讨论	24
参考文献	29
第二章 度分布的统计	33
§ 2.1 实际网络的统计方法	34
§ 2.2 确定的层次网络统计	42
§ 2.3 确定的伪分形图统计	49
§ 2.4 确定的阿波罗网络统计	53
参考文献	57
第三章 度分布的计算	59
§ 3.1 平均场方法	60
§ 3.2 率方程方法	65
§ 3.3 主方程方法	72
§ 3.4 马氏链方法	77
参考文献	84
第四章 度分布的稳定性	87
§ 4.1 模型网络的稳定性问题	88

目 录

§ 4.2 网络结点数向量马氏链	93
§ 4.3 网络结点度非齐次马氏链	99
§ 4.4 一类增长网络的稳定性	103
§ 4.5 网络瞬时度分布计算公式	108
参考文献	113
第五章 最大度的重要性	117
§ 5.1 网络有限标度的确定	118
§ 5.2 最大度的扰动和发散	123
§ 5.3 稳健而又脆弱的特性	128
§ 5.4 剩余度与巨集团涌现	133
参考文献	139
第六章 度-度的相关性	141
§ 6.1 网络同配抑或异配	142
§ 6.2 无标度程度的测量	146
§ 6.3 相关性测度的讨论	156
§ 6.4 网络的联合度分布	163
参考文献	168
附录 A 网络马氏链的程序	171
§ A.1 马氏链矩形迭代程序	172
§ A.2 马氏链公式计算程序	181
附录 B 度分布的极限定理	189
§ B.1 随机树与 Pólya 罐	190
§ B.2 大数定律	193
§ B.3 中心极限定理	197
§ B.4 大偏差原理	198
参考文献	201
名词索引	203

第一章 从图论到网络

本章首先介绍从图论到网络的三个发展阶段和代表性的工作。针对几个实际的复杂网络，如演员网、引文网、万维网、因特网、代谢网的数据统计，说明这些不同网络度分布都具有幂律尾部。由于实际的复杂网络还在不断发展，如何抓住它们生成的主要机制来建立恰当的网络模型，是本章论述的重点内容。我们将介绍某些有代表性的复杂网络模型，然后从中提炼出极其重要的且逻辑上正确的无标度网络概念，以及需要研究的理论问题。

§ 1.1 网络研究的三个里程碑^[1]

几何图形在 Euclid (欧几里得) 时代就是数学研究的重要对象之一. 但网络作为一门科学, 目前公认的看法是应从 Euler (欧拉) 开创图论学科算起. 第二个重要发展阶段始于 20 世纪中叶, 由 Solomonoff 和 Rapoport^[2] 以及 Erdős 和 Rényi^[3] 引入随机图论 (网络). 然而, 影响深远的进展则是近十年来在统计物理中出现的小世界网络^[4] (small-world networks)、无标度网络^[5] (scale-free networks) 和可导航网络^[6] (navigable networks) 的研究热潮.

粗略地说, 网络是结点与连线的集合. 如果结点按确定的规则连线, 所得到的网络就称为规则网络. 例如, 将结点排成一条直线, 假定每个结点与它最邻近的 $K=2k=4$ 个结点连线, 我们就得到一维无限规则网, 见图 1.1.1 (a). 若将结点排成一个圆圈, 让它们首尾相连, 仍然假定每个结点与它最邻近的 4 个结点连线, 我们就得到一维有限规则网, 见图 1.1.1 (b). 如不允许重复连线, 则实际上每个结点只能向一边连线, 即有 $k=2$ 条连线. 类似地, 将结点排在二维平面整数坐标格点上, 假定每个结点只与它最邻近的结点连线, 我们就得到二维无限规则网格, 见图 1.1.2. 若要得到二维有限规则网格, 就要将结点排在球面经纬线的交点上. 当然, 还可以类似地定义高维规则网格.

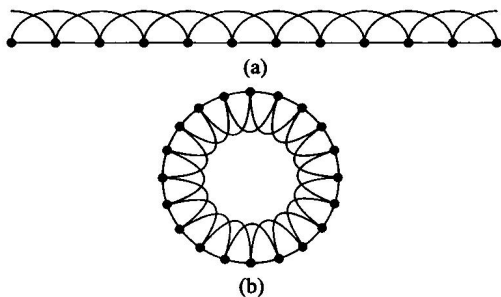


图 1.1.1 (a) 一维无限规则网; (b) 一维有限规则网

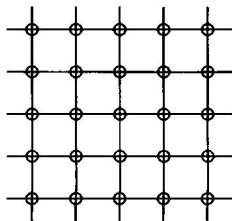


图 1.1.2 二维无限规则网格

如果结点不是按确定的规则连线,譬如按纯粹的随机方式连线,所得到的网络就称为随机网络.如果结点按照某种(自)组织原则方式连线,将演化成各种不同的网络,称为复杂网络.注意:复杂网络与图论有点不同,一般不考虑一个结点自身到自身的连线,也不考虑两个结点的重复连线.但有向网络中两个结点方向不同的两条连线不算重复连线.

随机网络的第一个模型^[2]:给定网络结点总数 N ,网络中任意两个结点以概率 p 连线,生成的网络全体记为 $G(N, p)$,构成一个概率空间.由于网络中连线数目是一个随机变量 X ,取值可以从 0 到 $N(N-1)/2$,有 n 条连线的网络数目为 $\binom{N(N-1)/2}{n}$,其中一个特定网络出现的概率 $P(G_n) = p^n (1-p)^{[N(N-1)/2]-n}$.因此,该模型可生成的不同网络的总数为 $2^{N(N-1)/2}$,它们服从二项分布.网络中平均连线数目为 $pN(N-1)/2$.

随机网络的第二个模型^[3]:给定网络结点总数 N 和连线总数 n ,而这些连线随机放置在总共 $N(N-1)/2$ 个可能连线位置上,生成的网络全体记为 $G(N, n)$,构成一个概率空间.这样,可生成的不同网络的总数为 $\binom{N(N-1)/2}{n}$,它们出现的概率相同,服从均匀分布.网络中两个结点连线的概率为 $p = 2n/[N(N-1)]$.

随机网络只是一个理想的数学模型,实际网络往往并不都是纯粹随机连线.

在社会系统中,朋友或熟人关系网络是最基本的,也是研究比较深入的一种网络.社会心理学家 Milgram 等人^[7]曾经做过一个著名的实验,实验要求参与者把一封信通过熟人传送给指定的某个人,借此探明熟人关系网络中路径长度(指给定两个结点,使它们首尾相连并且点边都相异的连线数)的分布.虽然实验中大多数信被丢弃了,但仍有四分之一的信送达目标人.统计显示平均路径长度为 5.2,即依次经过 6 个人就可传达,这就是著名的“六度分隔”(six degree of separation)概念的起源.这一实验的重要发现是,尽管人际关系网络极其庞大而又复杂,然而推测该网络的平均路径则是相对短的.

另外,人际关系网络中往往表现出很高的集群性.例如,住在同一小区的邻居都很熟悉,在同一单位工作的人也彼此认识,等等,即相互皆为朋友.社