

雷 敏 孔祥星 著 侯振挺 主审

复杂网络的度分布 及其仿真算法

FUZA
WANGLUODE
DUFENBU
JIQI
FANGZHEN
SUANFA



同濟大學出版社
TONGJI UNIVERSITY PRESS

复杂网络的度分布 及其仿真算法

雷 敏 孔祥星 著
侯振挺 主审

内 容 提 要

本书应用马氏链理论及分析技巧,对 BA 模型、增长网络模型、演化网络模型和合作网络模型度分布给出了存在的证明及表达式,并用马尔科夫骨架过程对 PERT 网络进行了分析,最后简单介绍了基于 Matlab 平台的复杂网络及其拓统计量的仿真方法。

本书共分为 7 章。第 1 章结论,讲述了复杂网络的发展过程和研究现状。第 2 章复杂网络的理论基础,叙述了复杂网络的基本概念和经典模型。第 3 章增长网络模型,应用马氏链理论及分析技巧对 BA 模型和一般的增长网络模型择优概率、度分布的存在性和表达式问题进行了严格分析。第 4 章演化网络模型,重点应用马氏链理论及分析方法对演化网络的度分布进行了严格分析。第 5 章广义合作网络模型,对连边数为随机变量的广义合作网络的度分布进行了研究。第 6 章 PERT 网络研究,应用马尔科夫骨架过程对 PERT 网络的项目完工时间进行了分析,得到了项目完工时间分布的解析结果。第 7 章复杂网络仿真程序简介,基于 Matlab 平台对复杂网络的生成、画图及拓统计量的仿真方法进行了介绍。

本书可作为概率论、统计学专业复杂网络方向研究人员的参考用书。

图书在版编目(CIP)数据

复杂网络的度分布及其仿真算法 / 雷敏,孔祥星著。

-- 上海: 同济大学出版社, 2016. 4

ISBN 978-7-5608-6274-3

I. ①复… II. ①雷… ②孔… III. ①概率分布函数—

仿真算法 IV. ①0211. 3

中国版本图书馆 CIP 数据核字(2016)第 062119 号

复杂网络的度分布及其仿真算法

雷 敏 孔祥星 著 侯振挺 主审

责任编辑 张智中 助理编辑 亢福军 责任校对 徐逢乔 封面设计 陆冬霞

出版发行 同济大学出版社 www.tongjipress.com.cn

(地址:上海市四平路 1239 号 邮编:200092 电话:021-65985622)

经 销 全国各地新华书店

印 刷 江苏句容排印厂

开 本 787 mm×960 mm 1/16

印 张 9.25

字 数 185 000

版 次 2016 年 10 月第 1 版 2016 年 10 月第 1 次印刷

书 号 ISBN 978-7-5608-6274-3

定 价 38.00 元

序

在新世纪之交,随着计算机技术的高度发展、计算机运算能力的不断提高,人们拥有各种网络的数据库,因此对大规模的网络进行实证研究有了可能性,这使得对网络科学的研究取得了突破性的进展。1998年Watts和Strogatz提出了小世界网络模型,说明了少量的随机捷径会改变网络的拓扑结构,从而涌现出小世界效应。1999年Barabási和Albert在研究万维网时发现,万维网的度分布不像随机网络和小世界网络那样具有对称的泊松分布,而是偏倚的幂率分布,为此他们提出了无标度网络模型(简称“BA模型”)。小世界网络和无标度网络的提出开启了复杂网络研究的热潮,在众多知名杂志中发表了大量的有关复杂网络研究论文,有的是对BA模型进行实证研究,有的对BA模型进行推广,有的论文试图对BA模型度分布问题给出严格的处理。

应此,我们团队(雷敏、孔祥星、童金英、谭利、赵清贵、张玄、康宁)在侯振挺教授的悉心指导和带领下,于2008年开始研究复杂网络,得到一系列成果,现把这些研究成果整理成册出版。我们的研究主要集中在对BA模型、增长网络模型、演化网络模型和合作网络模型度分布存在性的证明和表达式的计算上,应用马氏链理论(仅在第6章用到比马氏链更复杂的马尔科夫骨架过程)及分析技巧,给出了上述提到诸模型度分布的一个统一处理方法(这从对BA模型度分布的研究过程中,可窥见一斑)。从文章中可以看出择优概率是可实现的,推导过程是严格的,得到的结果是精确的,这是我们方法的独到之处。

由于笔者水平有限,本书中难免有一些缺点和错误,恳请批评指正。

雷 敏

2016年3月

目 录

序

第 1 章 绪论	1
1.1 复杂网络的研究概述	1
1.1.1 复杂网络的发展	1
1.1.2 复杂网络的研究现状	6
1.2 复杂网络的研究内容	8
1.2.1 复杂网络的生成机制及拓扑建模	8
1.2.2 复杂网络上的传播机理与动力学特性	12
1.2.3 复杂网络上的信息搜索	14
第 2 章 复杂网络理论基础	16
2.1 复杂网络的基本概念	16
2.1.1 复杂网络的定义	16
2.1.2 复杂网络的统计量	18
2.2 复杂网络的基本模型	25
2.2.1 ER 随机网络模型	25
2.2.2 WS 小世界网络模型	28
2.2.3 BA 无标度网络模型	31
2.2.4 LCD 网络模型	33
2.2.5 具有随机和择优连线规则的网络模型	34

2.2.6 HK 网络模型	35
2.3 复杂网络度分布的求解方法	36
2.3.1 平均场方法	36
2.3.2 率方程方法	37
2.3.3 主方程方法	38
2.3.4 鞍方法	38
2.3.5 马氏链方法	39
第 3 章 增长网络模型	41
3.1 BA 模型的严格数学基础	41
3.2 增长网络马氏链理论及应用	50
3.2.1 不允许重复连线的增长网络马氏链理论	50
3.2.2 允许重复连线的增长网络马氏链理论	54
3.2.3 增长网络马氏链的应用	57
第 4 章 演化网络模型	61
4.1 简单的演化网络模型	61
4.2 演化网络马氏链理论及应用	69
4.2.1 演化网络马氏链理论	69
4.2.2 演化网络马氏链的应用	80
第 5 章 广义合作网络模型	82
5.1 随机连接广义合作网络模型	84
5.2 择优连接广义合作网络模型	89
5.3 混合连接广义合作网络模型	95
第 6 章 PERT 网络研究	98
6.1 PERT 网络基础知识介绍	99

6.2 PERT 网络的马尔科夫骨架过程建模	100
6.3 PERT 网络实例	105
6.4 本章小结	106
第 7 章 复杂网络仿真程序简介.....	107
7.1 复杂网络仿真程序	107
7.1.1 全局耦合网络	108
7.1.2 最近邻耦合网络	110
7.1.3 随机网络	112
7.1.4 小世界网络	115
7.1.5 BA 模型	120
7.2 复杂网络统计量仿真程序	123
7.2.1 度分布	123
7.2.2 聚类系数	125
7.2.3 最短路径长度	127
7.2.4 介数	128
7.2.5 社团	131
7.2.6 完全子图	133
参考文献.....	137

第1章 绪论

1.1 复杂网络的研究概述

20世纪90年代以来,以Internet为代表的信息技术迅猛发展使人类社会大步迈入了网络时代。从Internet到WWW,从大型电力网络到全球交通运输网络,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,以及手机联系人网络和手机短信网络等,可以说,人们已经生活在一个充满着各种各样的复杂网络的世界中。人类社会的日益网络化需要人类对各种人工和自然的复杂网络的行为有更好的认识,复杂网络理论所要研究的就是各种看上去互不相同的复杂网络之间的共性和处理它们的普适方法。因此,人们致力于揭示复杂网络的拓扑结构和功能的形成机制、演化规律和动力学特性。国内外众多学者非常重视复杂网络理论的基础研究,并取得了丰硕的成果,世界上发达国家(如美国、法国等)纷纷提出复杂网络研究的发展路线图。

1.1.1 复杂网络的发展

随着人们对各种实际复杂系统的深入研究,发现复杂系统的很多性质不能仅由其各个组成部分简单组合而得到,这就需要我们把一个系统看成一个由很多组成部分以及它们间的相互作用共同构成的整体来研究。近年来复杂网络研究的兴起,使得人们开始广泛关注网络结构复杂性及其与网络行为之间的关系。要研究各种不同的复杂网络在结构上的共性,首先需要有一种描述网络的统一工具,这种工具在数学上称为图(Graph)。任何一个系统都可以看做是由一些节点按某种方式连接在一起而构成的一个网络。具体网络的抽象图,就是用点表示具体网络中的节点(系统的组成部分),并用节点之间的连线来表示具体网络中节点之间的连接关系(组成部分间的相互作用)。网络的发展大致经历了规则网络、随机网络和复杂网

络三个阶段.

第一阶段规则网络理论:图论的发展始于 18 世纪伟大的数学家 Euler 对著名的柯尼斯堡(Konigsberg)七桥问题的研究. 柯尼斯堡是东普鲁士的一个城镇,城中有一条河流穿过,河上有两个小岛,有七座桥把这两个小岛和河岸连接起来(图 1-1). 有人提出过一个问题:一个步行者怎样才能不重复、不遗漏地一次走完七座桥,最后回到出发点. 1735 年 Euler 仔细研究了这个问题,他将被河流分隔开的四块陆地抽象为四个点,将连接这四块陆地的七座桥抽象为七条边,这样就得到了一个由四个点和七条边构成的图,如图 1-2 所示. 于是 Euler 将七桥问题转化为图 1-2 能否用一笔画成的问题,并给出了能一笔画成的图的充分必要条件. Euler 对七桥问题的抽象和论证思想,开创了数学的一个分支——图论的研究.

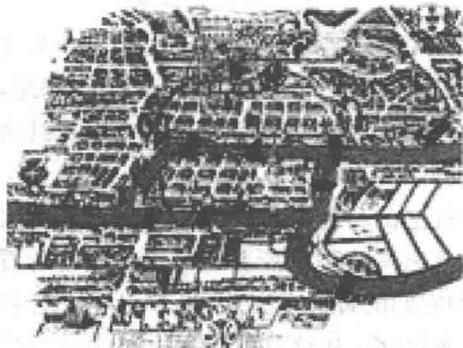


图 1-1 柯尼斯堡七桥问题

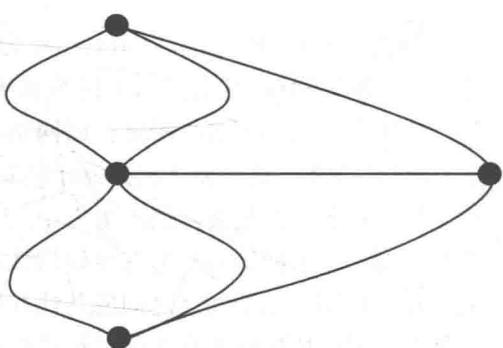


图 1-2 柯尼斯堡七桥抽象图

如果节点之间按某种确定的规则连线,所得到的网络称为规则网络. 例如,将节点排成一条直线,假定每个节点与它最近邻的 4 个节点连接,就得到一维无限规则网络,如图 1-3 所示. 将节点排在二维平面整数坐标格点上,假定每个节点只与它最近邻的节点连接,就得到二维无限规则网络,如图 1-4 所示.

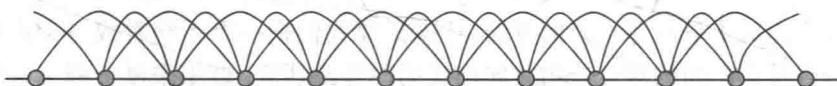


图 1-3 一维无限规则网络

研究表明规则网络具有较大的聚类系数和较大的平均最短路径长度,其所有的节点都有相同的度,节点度分布为 $\delta(d - d_0)$.

第二阶段随机网络理论:如果节点之间不按确定的规则连接,而是以一定的概率随机连接,所得到的网络就是随机网络. 20 世纪 60 年代,两个著名的匈牙利数

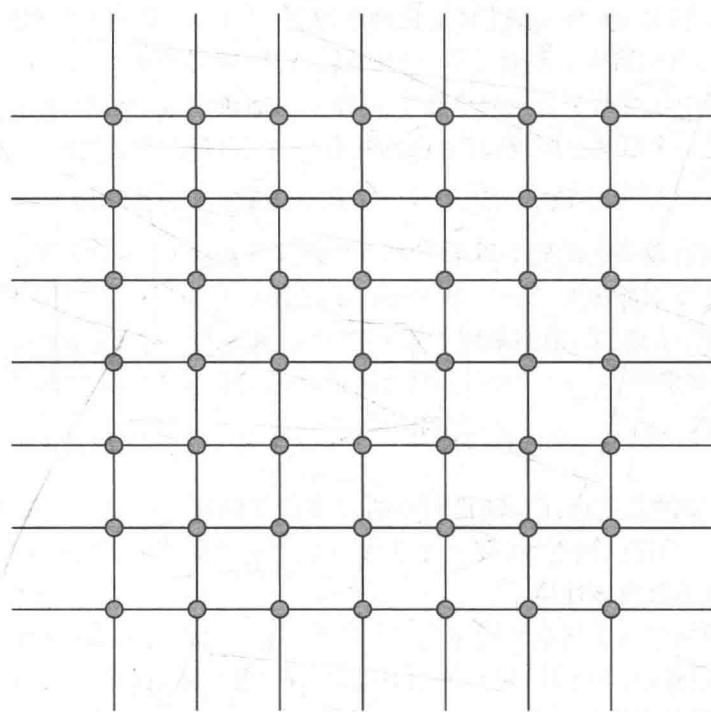


图 1-4 二维无限规则网络

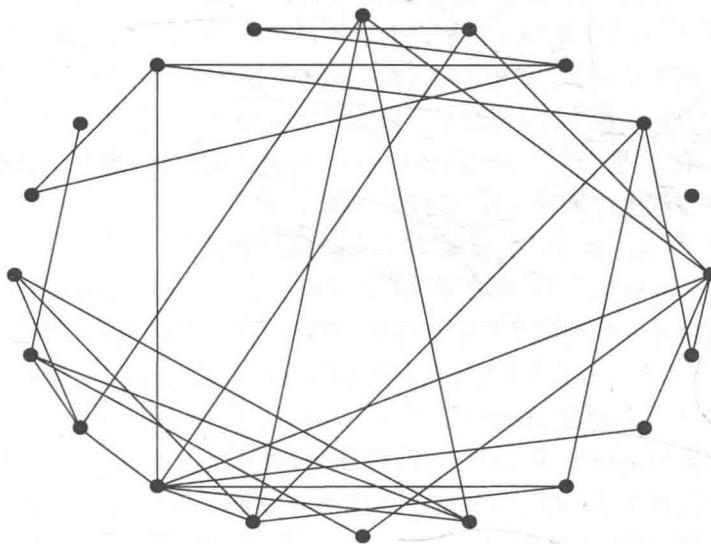


图 1-5 $N = 20, p = 0.2$ 的 ER 随机图

学家 Erdős 和 Rényi 提出的 ER 随机网络模型(图 1-5),开创了随机网络研究的新局面. ER 随机网络模型是通过节点间的随机连线而得到的网络. 给定网络节点总数 N , 网络中任意两个节点间以概率 p 独立的随机连线, 生成的网络的全体记为 $G(N, p)$ 是一个概率空间. 所有生成的具有 N 个节点和 M 条边的网络的概率都为 $p^M (1-p)^{(N \choose 2)-M}$, 参数 p 可以看成一个权重函数, 当 p 从 0 增加到 1 时, 模型生成的网络含有较多边的概率越来越大. 特别地, 当 $p = 0.5$ 时, $G(N, p)$ 中 $2^{(N \choose 2)}$ 个网络生成的概率都相同. Erdős 和 Rényi 系统研究了当 $N \rightarrow \infty$ 时 ER 随机网络的性质与概率 p 的关系, 他们采用了如下的定义: 几乎每一个 ER 随机网络都具有某种性质 Q, 如果当 $N \rightarrow \infty$ 时产生具有这种性质 Q 的 ER 随机网络的概率为 1. 如当 $p = \frac{\ln N}{N}$, $N \rightarrow \infty$ 时, $G\left(N, \frac{\ln N}{N}\right)$ 中几乎每一个网络都是连通的. Erdős 和 Rényi 最重要的发现是: ER 随机网络的许多重要的性质都是突然涌现的, 也就是说, 对于任一给定的概率 p , 要么几乎每一个网络都具有某个性质 Q, 要么几乎每一个网络都不具有该性质.

随机网络包含了从空图到完全图的所有可能, 为此研究其几何性质需要对每一种可能进行统计和平均. 随机网络具有较小的聚类系数和较小的平均最短路径长度, 节点度分布为泊松分布.

第三阶段复杂网络理论: 在世纪之交, 随着计算机技术的高度发展、计算机运算能力的不断提高, 人们拥有各种网络的数据库, 因此对大规模的网络进行实证研究有了可能, 网络科学取得了突破性的进展. 《自然》和《科学》相继发表了三篇重要的网络研究的文章, 国际顶级杂志如此青睐关于网络的研究论文, 这预示着一个新的学科研究方向的诞生. 问题的起因可追溯到 20 世纪 60 年代, 美国心理学家曾做过的一个所谓“六度分隔”(six degrees of separation)实验. 当时已有的规则网络和随机网络模型都无法解释这种“小世界现象”.

1998 年 Watts 和 Strogatz 提出了小世界网络模型(图 1-6)说明了少量的随机捷径(random shortcuts)会改变网络的拓扑结构, 从而涌现出“小世界效应”. 小世界网络模型介于规则网络和随机网络之间, 具有大的聚类系数和短平均路径长度, 节点度分布为泊松分布, 比规则网络和随机网络更能模拟真实网络的特性.

1999 年 Barabási 和 Albert 在研究万维网时发现, 万维网的度分布不像随机网络和小世界网络那样具有对称的泊松分布, 而是偏倚的幂率分布. 在该网络中, 大多数节点仅有少量连线, 而少数节点拥有大量的连线. 为此他们提出了无标度(scale-free)网络模型(图 1-7), 该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. 无标度网络中的节点具有很强的异质性, 不

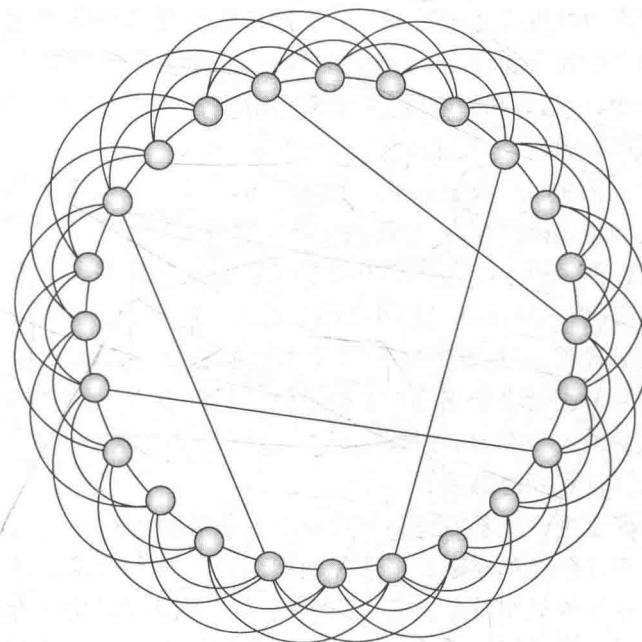


图 1-6 小世界网络

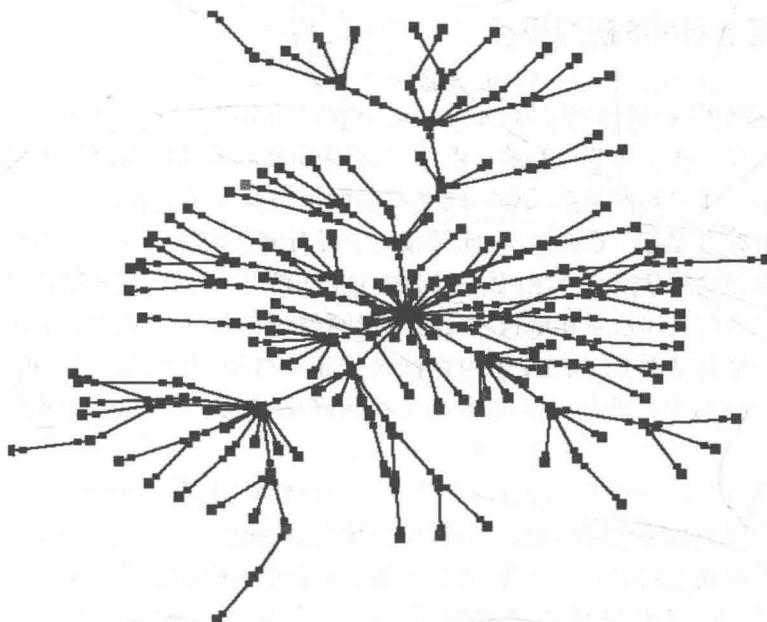


图 1-7 无标度网络

同度的节点在网络中的重要性不同。例如，在城市航空交通网络中，度分布表示城市之间的航线的多少和重要程度，度越大的城市，其重要性也就越大；在社会网络中，度可表示个体的作用力和影响度，一个节点的度越大，一般表示在整个网络系统组织中的作用和影响越大，反之亦然。

因为幂指数函数在双对数坐标中是一条直线，这个分布与系统特征长度无关，所以这个特性被称为无标度性质。它反映网络中度分布的不均匀性，只有很少数的节点与其他节点有很多的连接，成为“中心节点”，而大多数节点度很小。这个无标度特性是一把双刃剑，一是网络对意外故障具有惊人的抗攻击能力；另一方面网络对协同式攻击则很脆弱，一旦击中少数“中心节点”，就会导致整个网络崩溃。因此，人们为了避免网络因遭受攻击或意外事故导致的崩溃发生，最有效的办法就是保护好网络中节点度最大和次大的少数“中心节点”。了解到这一点对于我们设计和维护网络的安全性有重要的意义。

可导航网络模型则解释了如何利用局部信息能够找到最短（邮件传送）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。复杂网络的研究工作中最基本的、也是最重要的是对真实网络系统的仔细观察和实证研究，从中能够促使人们去进一步理解真实网络的形成机制，并捕获它们的拓扑特征，共性和个性。

1.1.2 复杂网络的研究现状

复杂网络可以描述自然界和人类社会中大量的复杂系统，一个典型的网络是由许多节点和一些连接节点间的边组成的，其中节点代表真实系统中不同的个体，而边则表示个体之间的关系。例如，航空交通网络中每个节点代表一个城市，每条边表示两个城市间有一条航线。不同领域的学者对复杂网络研究的重点不同，一般而言数学家和物理学家在考虑网络的时候，往往只关心节点之间有没有边相连，而忽略节点和边的具体位置和形状。在这里，把网络不依赖于节点的具体位置和边的具体形态就能表现出来的性质叫做网络的拓扑性质，相应的结构叫做网络的拓扑结构。目前，复杂网络的研究主要为社会网络、信息网络、技术网络、生物网络等。

复杂网络在社会网络中的应用十分广泛，社会网络是指社会个体成员之间因为互动而形成的相对稳定的关系体系，社会网络关注的是人们之间的互动和联系，社会互动影响到人们的社会行为。社会网络研究如个人间的朋友关系，公司商业伙伴关系，电影演员合作网络，性接触网络，科学家合作网络（图 1-8）等。

在信息网络类别中，目前研究较充分的有学术论文之间的引文网络。大部分学

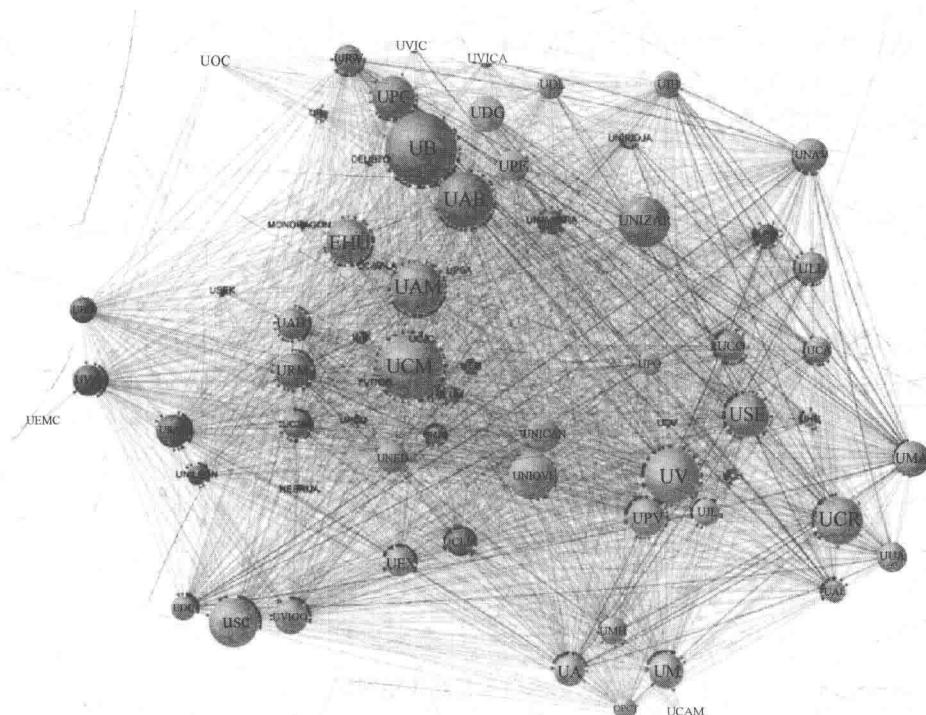


图 1-8 科学家合作网络

术论文都经由相关主题的其他文章来引用以前做的工作,这些引用就形成了一个网络,其中节点代表论文,从论文 A 到论文 B 的有向边代表论文 A 引用了论文 B. 其他研究有邮件网络、语言网络等.

在现实生活中人们为了有效地分配或分享一些商品和资源设计了很多技术网络,如电力网络、交通网络等.主要有 Watts 和 Strogatz, Amaral 等人关于电力的研究.还有航空路线网络、道路网络、铁路网络以及人行交通网络.互联网作为当今人类社会信息化的标志,其规模的增长非常迅速,但现在对于互联网宏观拓扑结构的研究还处于初级阶段,对互联网宏观拓扑结构演化及其内在机制的研究,是在较高层次上开发利用互联网的基础.互联网是一个复杂网络,构建互联网拓扑结构模型,刻画互联网拓扑结构在宏观上的特征,反映一种总体趋势,可以帮助人们用其来对互联网进行分析、预报、决策和控制.

很多生物系统也可以表示成网络。生物网络的典型例子是新陈代谢路径网络，一些学者对其统计学属性进行了研究，如 Jeong 等人，Fell 和 Wagner，以及 Stelling 等人。另一个不同的网络是蛋白质之间的力学物理相互作用网络，它经常被指为蛋白质相互作用网络（图 1-9）。很多学者对其进行了研究。

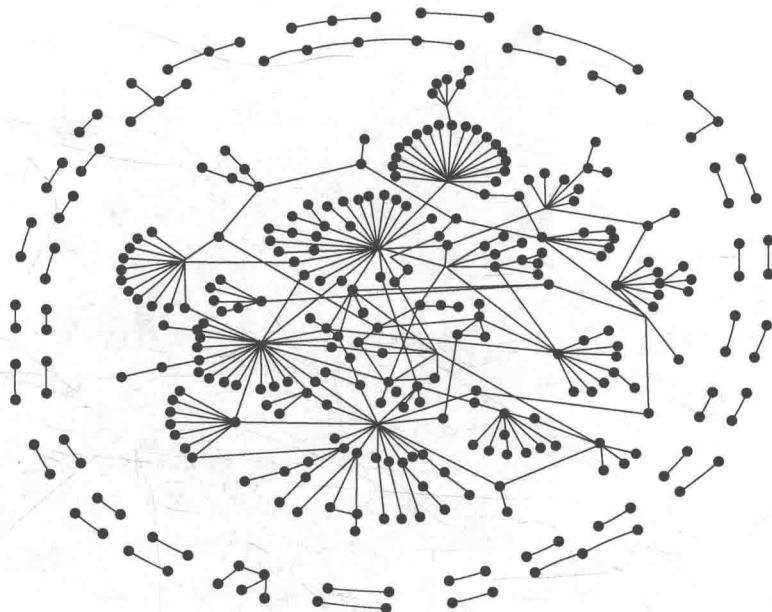


图 1-9 蛋白质相互作用网络

1.2 复杂网络的研究内容

在世纪之交,随着计算机技术的高度发展、计算机运算能力的不断提高,人们拥有各种网络的数据库,因此对大规模的网络进行实证研究有了可能,复杂网络的研究取得了突破性的进展。目前,复杂网络研究的主题主要包括:复杂网络的拓扑性质,复杂网络的生成机制,复杂网络演化的统计规律,复杂网络的结构稳定性,复杂网络上的传播机理与动力学特性,复杂网络中的搜索,复杂网络中的同步与控制以及把网络的拓扑结构与具体系统结合起来。

1.2.1 复杂网络的生成机制及拓扑建模

20世纪末,Watts 和 Strogatz 在通过对大量的生物网络、技术网络和社会网络研究后,发现很多网络具有完全规则网络和完全随机网络都不具有的特殊的拓扑特性。通过向一个规则网络中增加一些随机连线从而增加规则网络的无序性,得到的新网络具有同规则网络一样较大的聚类系数和随机网络一样较小的最短路径

长度,具有这种较小的最短路径长度和较大的聚类系数现象称为小世界现象.根据向规则网络增加随机连线的机制,1998年他们提出了一个小世界网络模型.美国西部电力网络和电影演员合作网络都是小世界网络.小世界网络提出有利于研究社会关系网络和传染病在网络中的传播等问题,具有重大的科学意义.

Barabási 和 Albert 对大量实际网络(电影演员合作网络,WWW,美国西部电力网络)研究后发现这些大规模网络节点度分布是幂率衰减的,即网络节点具有度 k 的概率 $P(k)$ 是幂率衰减的,具有 $P(k) \sim k^{-\gamma}$ 的形式.这说明这些网络会自组织演化为一个无标度状态.他们认为产生这种现象源于增长和择优两个内在的机制,为此提出了一个具有增长和择优机制的网络模型——BA 模型. BA 模型的算法描述如下:(1)给定一个具有 N_0 个节点和若干条边的初始网络;(2)在每时间步增加一个新节点和 $m(m \leq N_0)$ 条新连线;(3)新节点以择优概率选择旧节点并与之相连.如图 1-10 所示是一个具有 100 个节点和 99 条边的 BA 网络.

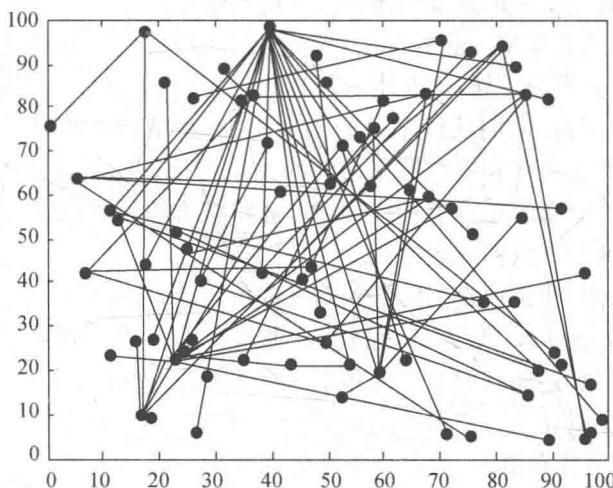


图 1-10 含有 100 个节点和 99 条边的 BA 网路

侯振挺等人发现 BA 模型的网络度瞬时分布具有马尔科夫性,他们用马尔科夫链理论的方法和技巧,给出了求解 BA 模型网络稳态度分布的一个普适的方法.给出了 BA 无标度网络模型稳态度分布存在的严格证明,并且从数学上推导了度分布的精确解析表达式.

Barabási 等人通过平均场方法得出 BA 模型的标度指数为 3,然而 Dorogovtsev 发现很多具有幂率分布的网络的标度指数并不全为 3,为此他们提出了一个具有原始吸引度的模型,其中择优概率与 $d_v + A$ 成正比.原始吸引度模型反映了真实网络中孤立节点也能得到连线的事实. Buckley 等人推广了 Dorogovtsev 的吸引

度模型,择优概率与 $k_i + ra$ 成正比,其中 r 是新节点所带的边数, a 是一个常数. 当 $a, r \in \mathbb{N}$ 时得到了度分布的渐近表达式.

在 BA 模型中择优连线的概率与节点的度成正比即与节点度呈线性关系,Krapivsky 等人提出了一个新的模型,将择优概率与节点度呈线性关系改为与节点度的函数呈线性关系,即择优连线的概率与 A_k 成正比,其中 $A_k \sim k^\gamma$, $\gamma \geq 0$. 对于不同 γ 生成的网络具有不同的性质,只有当 $\gamma = 1$ 时网络度分布服从幂率分布. 研究说明,非线性择优会破坏网络的无标度特性.

然而,Bollaás 等人指出 BA 模型有两点不足. 一是初始网络没有确定,只说明开始给定 N_0 个节点,它们之间如何连线没有阐明. 如果初始是孤立节点,那么择优连线将无法进行,所以初始网络不能有孤立节点. 对于不同的初始网络,其后演化的结果网络会有所不同. 二是当每次加入的连线数大于 1 时, m 条连线如何择优加入到网络中没有说清楚. 为此,他们提出了 LCD 模型,该模型修正了 BA 模型的两点不足,并给出了网络度分布存在性的数学证明.

侯振挺等人认为运用一般的方法只能得到 LCD 模型度分布的近似表达式,虽然可以利用鞅方法严格求解,但是求解过程比较复杂,而且鞅方法只适用于带有重连和自连的网络模型. 他们从概率论的角度出发,将节点的度数看成一系列马氏链. 根据马氏链首达概率和度数概率之间的关系严格证明了这两类模型的度分布存在性,得到了度分布的精确解析式,此外,还将该方法简化,同样严格证明度分布存在性,并得到度分布的精确表达式.

在引文网络中,我们通常引用那些新发表的文章,很少引用那些很长时间以前发表的论文. Dorogovtsev 等人认为新节点在选择旧节点相连时的概率不仅要依赖于旧节点的度数,还要和旧节点的年龄 τ 有关系,为此提出了一个带有年龄的网络模型,择优概率正比于节点的度与 τ^α 之积,其中 α 是一个常数. 并且分析了对于不同的 α 值网络度分布情况. 通过分析和仿真发现,当 $\alpha < 1$ 时,度分布服从幂率分布. 当 α 从 $-\infty$ 增大到 1 时,度指数 γ 从 2 开始增大. 当 $\alpha = 1$ 时,即为 BA 模型度指数 $\gamma = 3$. 当 $\alpha > 1$ 时,度分布服从指数分布.

无标度网络具有高度的自组织特性,其节点具有高度的异质性,且存在一小部分节点具有很高的度数,网络度分布服从严格的幂率分布,而随机网络则缺乏这些特性,网络度分布 $P(k)$ 服从指数分布. 现实中的很多复杂网络应该介于两者之间,即现实中有很多网络的度分布是幂率分布和指数分布,也有很多网络其度分布既不是幂率也不是指数的. 为了揭示上述现象,刘宗华等人提出了带有随机与择优连接的混合网络模型,新连线在选择节点时的概率与 $(1-p)k_i + p$ 成正比,其中 $0 \leq p \leq 1$ 是一个权重参数,调节随机和择优的比重. 当 $p = 1$ 时是完全随机的网络,而 $p = 0$ 时即为 BA 模型,当 $p < 1$ 时度分布服从幂率分布. 我们在参考文献