丁连红　时　鹏　编著

网络社区发现

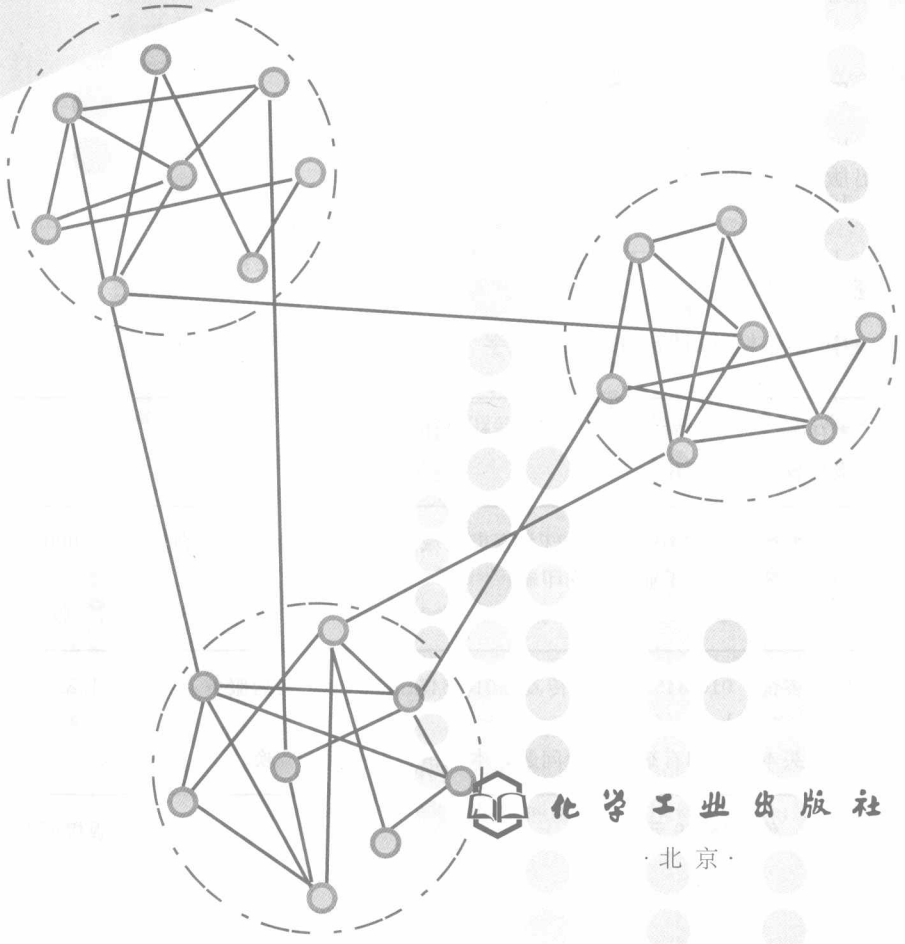
Network Community Detection

化学工业出版社

丁连红　时　鹏　编著

网络社区发现

Network Community Detection



化学工业出版社
·北京·

本书综合了国内、外社区发现技术的研究成果，阐述了社区现象的背景和原理、社区发现的技术以及社区发现的典型应用。

社区现象是复杂网络中的一种普遍现象，表达了多个个体具有的共同体特性。社区的发现技术，从最初的图分割方法、W-H 算法、层次聚类法、GN 算法等基本算法，逐渐发展和改进，形成了包括改进 GN 算法、派系过滤算法、局部社区算法和 Web 社区发现方法在内的更具可操作性的方法。网络的社区发现可为个性化服务、信息推送等提供基本数据，尤其是在信息时代，社区的存在更加普遍，发现技术应用更加方便，其商业价值和服务价值更大。

本书提供的基本原理和实现技术为国内学者与技术人员深入理解社区发现技术提供了有益参考，是一本不可多得的好书。

图书在版编目（CIP）数据

网络社区发现/丁连红，时鹏编著. —北京：化学工业出版社，2008.9
ISBN 978-7-122-03524-0

Ⅰ.网… Ⅱ.①丁…②时… Ⅲ.计算机网络-文化-研究 Ⅳ.TP393-05

中国版本图书馆 CIP 数据核字（2008）第 22920 号

责任编辑：刘亚军　　装帧设计：张　辉
责任校对：王素芹

出版发行：化学工业出版社（北京市东城区青年湖南街 13 号　邮政编码 100011）
印　　装：化学工业出版社印刷厂
720mm×1000mm　1/16　印张 9　字数 162 千字　　2008 年 8 月北京第 1 版第 1 次印刷

购书咨询：010-64518888（传真：010-64519686）　售后服务：010-64518899
网　　址：http://www.cip.com.cn
凡购买本书，如有缺损质量问题，本社销售中心负责调换。

定　　价：28.00 元

前　言

现实世界中的很多系统都可以用复杂网络的形式来描述，复杂网络已逐渐成为研究复杂系统的一种重要方法和跨学科的研究热点。社会网络是一种复杂网络，反映了社会成员及其相互关系。通过对社会网络的理论研究，尝试挖掘隐藏在表面关系之下的隐性关系，可进行电子商务、信息推荐等有益的应用。

随着对网络性质的物理意义和数学特性的深入研究，人们发现许多实际网络都具有一个共同性质——社区结构。也就是说，网络是由若干个“群”或“团”构成的。每个群内的节点之间的连接非常紧密，而群之间的连接相对比较稀疏。揭示网络中的社区结构，对于了解网络结构与分析网络特性具有极为重要的意义。社区结构分析在生物学、物理学、互联网、商业活动和社会学中都有广泛的应用。本书第一章和第二章介绍了复杂网络和社会网络的基本特征以及社区现象的基本原理。

为了研究网络中社区结构的特性，研究人员对寻找网络中社区结构的方法进行了探索和研究。其目的主要是通过有效的算法，利用尽量少的信息得到尽量准确的网络社区结构。目前已经存在若干社区发现方法，其中最具有代表性的是计算机科学中的图分割（Graph Partitioning）方法，社会学领域中的层次聚类（Hierarchical Clustering）方法、W-H 算法和 GN 算法。社区发现技术还处在不断的发展和更新过程中，如以降低算法复杂度为主要目的的改进 GN 算法、以解决社区重叠和嵌套的派系过滤算法、不以掌握网络全部拓扑结构为前提进行社区发现的局部社区算法，主要针对万维网和 Internet的 Web 社区发现方法。这些算法分别在本书的第三章和第四章进行了阐述。

社区发现技术对科学研究和商业应用都具有很高的价值。在科学研究方面，可应用于生命科学、社会学以及信息科学等许多领域。在商业应用方面，最具代表性的是个性化服务和互联网应用。基于社区发现的个性化服务

系统可以克服传统系统的很多缺陷，如缺乏建立用户模型的信息、缺乏用户评价信息等。在网络飞速发展的今天，互联网上的社区应用具有广泛适用性。将社区发现技术应用于电子商务，不但可以帮助商家通过服务水平的提高创造更大的商业价值，而且可以通过人性化的服务增强用户的忠诚度。基于社区的网络文化安全评估和预警，既能最大限度保证网民利益，又能够维护网络文化安全。本书第五章和第六章详细介绍了社区发现技术在个性化服务和互联网上的应用，包括基本流程和所涉及的技术细节。

目前国内科研人员已经开始进行复杂网络和社区结构的学术研究，但仅有少量关于复杂网络和社区结构的学术论文发表，虽然已经有阐述复杂网络的原理和特征的专著面世，但关于社区结构及其发现技术，尚未有专门的中文作品。为了加速国内对社区发现技术的研究和应用进程，笔者查阅了大量资料，在科学研究和应用实践的基础之上，撰写了本书。

本书受北京市属市管高等学校人才强教计划资助项目、北京物资学院科技创新平台资助项目和国家“十一五”科技支撑计划“网络文化安全预警技术研究”（No. 2006BAK11B03）项目的资助。本书共分六章，北京物资学院丁连红撰写了第一、三、四、五章；北京科技大学时鹏完成了其余部分的撰写，并且对全书做了文字修订和润色加工。

由于笔者水平有限，书中难免有疏漏之处，敬请广大读者批评指正。

编著者

2008 年 6 月

目　　录

第一章　复杂网络与社会网

从最初的规则网络，之后的随机网络，到近几年的复杂网络，越来越多的关于网络的研究成果被发掘并应用，为人们更深刻认识现实中的复杂系统，并对之进行控制或应用提供了有效帮助。现实世界中的很多系统都可以用复杂网络的形式来描述，这些复杂网络具有网络平均路径长度较小、聚类系数较大、节点度分度服从幂律分布等相同特性。近年来，复杂网络已逐渐成为研究复杂系统的一种重要方法，对复杂网络的研究正受到来自不同领域的越来越多的研究人员的关注，复杂网络已经成为一个跨学科的研究热点。

社会网是一种复杂网络，反映了社会成员及其相互关系。通过对社会网的理论研究，尝试挖掘隐藏在表面关系之下的隐性关系，可进行电子商务、信息推荐等有益的应用。

第一节　复杂网络及其特点

一、复杂网络的定义及来源

现实世界中的许多系统都可以采用网络的形式来加以描述，可以将网络看作由节点和连接节点的边组成的集合。通常用节点来表示现实系统中的个体，用边表示个体间的某种关联，有边相连的两个节点被称作相邻节点，有点相连的两条边被称作相邻边。若网络中的边具有方向性，称为有向网络；反之，称为无向网络。本书中未特别指明的网络为无向网络。图论中的图与本书中的网络类似，图是抽象化的网络，图论中的方法可以用于解决复杂网络中的问题。

现实世界中的许多系统都可以利用网络图进行描述。例如，如果用一个节点表示一个人，一条边表示它所连接的两个节点（即所表示的两个人）之间的交往，就能构成反映人际关系的社会网络；如果用节点表示城市，用边

表示城市之间的铁路，就能构建反应交通路线状况的铁路网；如果用节点表示物种，用边表示从被捕食者指向捕食者的能量传递关系，就构成了食物链网；如果用节点表示协同团队中的成员，边表示知识在成员之间的传播，就构成了知识流网。这样的例子随处可见，如 Internet、World Wide Web、神经网络、代谢网络、分布式的血管网络等。研究网络的结构，并发现其内在共同特性，以便多个领域相互参考借鉴，是科学家们一直所关注的问题。

网络研究的初次尝试可以追溯到 1736 年，瑞士数学家欧拉（Euler）在他的一篇论文中讨论了哥尼斯堡七桥问题。在二百多年的发展过程中，网络理论的研究先后经历了规则网络、随机网络和复杂网络三个阶段。在最初的一百多年里，研究人员普遍认为真实系统各因素之间的关系可以用一些规则的结构表示，例如二维平面上的欧几里得格子，它看起来像是格子衬衫上的花纹；又或者最近邻环网，它容易让人想到一群手牵着手围着篝火跳圆圈舞的人们。1960 年，数学家 Erdös 和 Rényi 提出了随机图理论，为构造网络提供了一种新的方法。在这种方法中，两个节点之间是否有边连接不再是确定的事情，而是根据一个概率决定，这样生成的网络称作随机网络。随机图的思想主宰复杂网络研究长达四十年之久，直到近几年，科学家们对大量的现实网络的实际数据进行计算研究后得到的许多结果，既不是规则网络，也不是随机网络，而是具有与前两者皆不同的统计特征的网络。这样的一些网络称为复杂网络，对于复杂网络的研究标志着网络研究的第三阶段的到来。由 Watts 和 Strogatz 于 1998 年提出的 WS 小世界网络模型，刻画了现实世界中的网络所具有的大的凝聚系数和短的平均路径长度的小世界特性。1999 年，Barabási 和 Albert 提出的无尺度网络模型，刻画了实际网络中普遍存在的“富者更富”的现象。小世界网络和无尺度网络的发现掀起了复杂网络的研究热潮。

二、复杂网络的特征及度量

（一）平均路径长度与小世界现象

在网络研究中，如果网络中的两个节点可以通过一些首尾相连的边连接起来，则称这两个节点是可达的，并把连接两者的路径中边数最少的路径称为最短路径，最短路径的边数称为两个节点之间的距离。显然两个点之间的距离总是比网络拥有的节点总数要小。网络的直径定义为网络中任意两个节点间的最大距离。把所有节点对的距离进行平均，就得到了网络的平均距

离，它描述了网络中节点间的分离程度，即网络的大小或尺寸。

“小世界现象”源于社会心理学家 Stanley Milgram 在 20 世纪 60 年代所做的试验。他要求从奥马哈市（Omaha）随机选取的 300 人尝试寄一封信给波士顿市（Boston）的一位证券业务员，寄信的规则是每个参与者只能转发给一个他们认识的人。直觉告诉我们，从茫茫人海中找到一条相续认识的链，把最初的寄信人跟目标业务员连接起来，应该会费尽周折。然而，实验结果表明：完整的链的平均长度为 6 个人。

小世界特性容易使人联想起疾病、谣言、或数据在网络中的传播或传输问题，这些问题很多时候恰恰是很关键的问题。除了具有平均最短距离较小以外，小世界网络还要具有高聚集性，同时具有这两个方面特性的网络才可以被称为是小世界的。实验结果说明，在以细胞中的化学物质为节点、化学反应关系为边构成的网络中，节点之间的典型间隔为 3；在以好莱坞演员作为节点、同在一部电影中出演作为边的网络中，演员之间的平均间隔为 3；在具有 153127 个节点的万维网（World Wide Web）中，节点之间的平均路径长度为 3.1。另外，Erdös 和 Rényi 已经证明，经典的随机网络中，任何两个节点间的典型距离为网络节点数的对数数量级，所以也具有小世界的特点。

（二）聚类系数与聚集性

在一个社会网络中，一个人的朋友的朋友可能也是他的朋友，或者他的两个朋友可能彼此也是朋友。聚集性用于描述这类可能性的程度，即，网络有多紧密。聚集性表达了网络连接的聚集程度。

通常用聚类系数（Cluster Coefficient）来描述网络中节点的聚集情况，其定义为：假设节点 i 与其他 k_i 个节点相连接，如果这 k_i 个节点都相互连接，它们之间应该存在 $k_i(k_i-1)/2$ 条边，而这 k_i 个节点之间实际存在的边数只有 E_i 的话，则它与 $k_i(k_i-1)/2$ 之比就是节点 i 的聚类系数。相应的计算公式为：

$$C_i=\frac{2E_i}{k_i(k_i-1)} \tag{1-1}$$

显然聚类系数表达了节点的紧邻之间也是紧邻的程度。所有节点的聚类系数的平均值称为平均聚类系数 C 或整个网络的聚类系数。公式表示为(1-2)，其中 N 为节点总数。

$$C=\frac{1}{N}\sum_i C_i \tag{1-2}$$

平均聚类系数也是复杂网络中的一个重要的全局几何量，在全连通网络（每个节点都与其余所有的节点相连接）中，聚类系数才能等于1，其他情况均小于1。对于随机网络，则有 $C=p$，p 为节点间的连接概率。Watts 和 Strogatz 首先指出，许多实际网络的聚集系数远大于具有相同节点数和边数的随机网络。也就是说，许多实际网络趋于具有集团的特性，就像人的社会关系网络一样。这个定义被广泛使用，在社会学领域常称为网络密度。

（三）度和度分布

节点的度（Degree）是网络研究中的一个重要概念，是描述网络局部特性的基本参数。在 N 个节点的网络中，任意一个节点 i 的度 k_i 等于与该节点相连的其他节点的数目（连接数）。若网络的邻接矩阵为 $A=[a_{ij}]_{N\times N}$，则节点 i 的度为：

$$k_i = \sum_{j\in N} a_{ij} \tag{1-3}$$

在有向网络中，节点的度分为出度（Out-degree）和入度（In-degree）。节点的出度，是指从该节点指向其他节点的边的数目；节点的入度，是指从其他节点指向该节点的边的数目。度用于描述网络节点连接数目的分布情况。直观上看，一个节点的度越大，表明其在网络拓扑中的地位越重要。事实上度在不同的网络中含义不同。如，社会网络中，度可以表示个体的影响力和重要程度，度越大的个体，其影响力就越大，在整个组织中的作用也就越大；反之亦然。

节点的平均度是指所有节点的度的平均值，用符号$<k>$表示。

$$<k>=\frac{1}{N}\sum_{i=1}^{N} k_i \tag{1-4}$$

度分布（Degree distributions）是对节点的度的规律的一种描述，通常用度分布函数 $P(k)$ 表示任意选择一个网络节点，其度恰好为 k 的概率。其值等于网络中度为 k 的节点的个数占网络节点总个数的比值。由于连接的随机性，随机网络的所有节点的度应该接近网络的平均度$<k>$。随机网络的度分布为二项分布（Binomial）或大规模极限下的泊松分布（Poisson Distribution），其峰值为$<k>$，在远离峰值处呈指数下降。在无尺度网络中，如论文引用网络、WWW、Internet、代谢网络，电话呼叫网络和人之性关系网络等，其度分布都呈一种幂律分布（Power-law Distribution），也就是分布函数的形式为 $P(k)\sim k^{-\gamma}$，其中 γ 一般介于2～3之间。

同时研究者也发现，在非泊松度分布的真实网络中，除了幂律分布外，还存在其他形式的度分布。如电力网络的度分布服从指数分布，在单对数坐标系下是一条下降的直线；也存在幂律加指数截断（Cutoff）的度分布的网络，如电影演员合作网络以及蛋白质相互作用网络。

（四）度和聚类系数之间的相关性/选型连接性（Assortativeness）

网络中度和聚类系数之间的相关性被用来描述不同网络结构之间的差异，包括两方面内容：节点的度相关性和节点度分布与其聚类系数之间的相关性。前者也称为网络选型连接性（或选型相关性），指的是网络中与高度数（或低度数）节点相连接的节点的度数偏向于高还是低。若连接度大的节点趋向于和其他连接度大的节点连接，则认为网络呈现协调混合；若连接度大的节点趋向于和其他连接度小的节点连接，则认为网络呈现非协调混合。研究中常用相关系数来描述网络的选型连接性。

相关系数的定义为：

$$\Gamma=\frac{c\sum_i j_i k_i-\left[c\sum_i \frac{1}{2}(j_i+k_i)\right]^2}{c\sum_i \frac{1}{2}(j_i^2+k_i^2)-\left[c\sum_i \frac{1}{2}(j_i+k_i)\right]^2} \tag{1-5}$$

式中，j_i，k_i 为与第 i 条边关联的两个节点的度；$c=1/m$，m 是网络中边的条数。实际的网络的选型连接性有一些呈现协调混合（$\Gamma>0$），一些呈现非协调混合（$\Gamma<0$）。如，社会网络（演员合作网络、公司董事网络、电子邮箱网络）中节点具有正的度的相关性，而节点度分布与其聚类系数之间却具有负的相关性。其他类型的网络（信息网络、技术网络、生物网络）则相反。因此，这两种相关性也被认为是社会网络区别于其他类型网络的重要特征，在社会网络的研究中引起了人们的高度重视。

（五）网络健壮性（Robustness）/网络弹性

许多实际复杂系统表现出惊人的容错能力，这引起研究者的广泛关注。举例来说，复杂的通信网络呈现高度的健壮性，常规的局部失效及关键部件的故障很少会导致网络的整体信息承载传送能力的丧失，这种网络的稳定性常被人们归因于网络的冗余连接。但是除了冗余之外，网络的拓扑是否对其稳定与健壮性有一定作用呢？网络对部件失效或者连接失败的抗拒能力称为

网络的健壮性或者恢复力（Resilience）。

网络的功能依赖其节点的连通性，即，依赖于节点间存在的路径。网络节点的删除对网络连通性的影响称为网络弹性，其分析方式有两种：随机删除和有选择的删除，分别称为网络的健壮性分析和网络的脆弱性分析。Albert和 Barabási 对度分布服从指数分布的随机网络模型和度分布服从幂律分布的无尺度网络进行了研究，结果显示：随机删除节点基本上不影响无尺度网络的平均路径长度，即对随机节点的删除具有高度弹性；相反，有选择的删除度数最大的节点时，无尺度网络的平均路径长度较随机网络的增长快得多。这表明，无尺度网络相对随机网络具有较强的鲁棒性和易受攻击性。出现上述现象的原因在于：幂律分布网络中存在的少数具有很大度数的节点，它们在网络连通中扮演着关键角色，一般也称它们为 Hub 节点。

（六）介数/居间中心性（Betweenness Centrality，BC）

介数分为边介数和节点介数，节点的介数为网络中所有的最短路径中经过该节点的数量比例，节点 k 的介数定义为：

$$g_k = \sum_{i \neq j} g_k(i,j) = \sum_{i \neq j} \frac{C_k(i,j)}{C(i,j)} \tag{1-6}$$

式中，$C_k(i,j)$ 表示节点 i 和 j 之间最短路径中经过节点 k 的次数；$C(i,j)$ 则表示节点 i 和 j 之间最短路径的总数目。介数反映了相应的节点或者边在整个网络中的作用和影响力，具有很强的现实意义。社会学中常用这个指标描述指定的人在社会中的影响力，介数在社会关系网络或技术网络中的分布特征反映了不同人员、资源和技术在相应社会关系或生成关系中的地位，这对于在网络中发现和保护关键资源和技术具有重要意义。

边的介数与节点介数的含义类似，是指网络中所有的最短路径中经过该边的数量比例，多应用于网络中的社区结构的识别，这方面的内容将在第三章给出详细介绍。

第二节　复杂网络模型

真实网络所表现出来的小世界特性、无尺度幂律分布或高聚集度等现象促使人们从理论上构造出多样的网络模型，以解释这些统计特性，探索形成这些网络的演化机制。本节介绍了几个经典网络模型的原理和构造方法，包

括 ER 随机网络模型、BA 无尺度网络模型和小世界模型。

一、ER 随机网络模型

Erdös-Rényi 随机网络模型（简称 ER 随机网络模型）是匈牙利数学家 Erdös 和 Rényi 提出的一种网络模型。1959 年，为了描述通信和生命科学中的网络，Erdös 和 Rényi 提出，通过在网络节点间随机地布置连接，就可以有效地模拟出这类系统。这种方法及相关定理的简明扼要，导致了图论研究的复兴，数学界也因此出现了研究随机网络的新领域。ER 随机网络模型在计算机科学、统计物理、生命科学、通信工程等领域都得到了广泛应用。

ER 随机网络模型是个机会均等的网络模型。在该网络模型中，给定一定数目的个体（节点），它和其他任意一个个体（节点）之间有相互关系（连接）的概率相同，记为 p。因为一个节点连接 k 个其他节点的概率，会随着 k 值的增大而呈指数递减。这样，如果定义 k 为每个个体所连接的其他个体的数目，可以知道连接概率 $P(k)$ 服从钟形的泊松（Poisson）分布，有时随机网络也称作指数网络。

随机网络理论有一项重要预测：尽管连接是随机安置的，但由此形成的网络却是高度民主的，也就是说，绝大部分节点的连接数目会大致相同。实际上，随机网络中连接数目比平均数高许多或低许多的节点，都十分罕见。

在过去 40 多年里，科学家习惯于将所有复杂网络都看作是随机网络。在 1998 年研究描绘万维网（以网页为节点、以超级链接为边）的项目时，学者们原以为会发现一个随机网络：人们会根据自己的兴趣，来决定将网络文件链接到哪些网站，而个人兴趣是多种多样的，可选择的网页数量也极其庞大，因而最终的链接模式将呈现出相当随机的结果。

然而，事实并非如此。因为在万维网上，并非所有的节点都是平等的。在选择将网页链接到何处时，人们可以从数十亿个网站中进行选择。然而，我们中的大部分人只熟悉整个万维网的一小部分，这一小部分中往往包含那些拥有较多链接的站点，因为这样的站点更容易为人所知。只要链接到这些站点，就等于造就或加强了对它们的偏好。这种“择优连接（Preferential Attachment）”的过程，也发生在其他网络中。在 Internet 上，那些具有较多连接的路由器通常也拥有更大的带宽，因而新用户就更倾向于连接到这些

路由器上。在美国的生物技术产业内，某些知名公司更容易吸引到同盟者，而这又进一步加强了它在未来合作中的吸引力。类似地，在论文引用网络（论文为节点，引用关系为边）中，被引用次数较多的科学文献，会吸引更多的研究者去阅读并引用它。针对这些网络的“择优连接”的新特性，学者提出了 BA 无尺度网络模型。

二、BA 无尺度网络模型

无尺度网络的发现，使人类对于复杂网络的认识进入了一个新的天地。无尺度网络的最主要特征是节点的度分布服从幂次定律。BA 模型是无尺度网络（Scale-free Network）的第一个抽象模型。由于考虑了系统的成长性（Growth）和择优连接性，BA 模型给我们带来了很多启发，并且可以应用于多种实际网络。但是 BA 模型的两个基本假定，对于解释许多现实中的现象来说过于简单，与现实的网络还有较大的距离。有学者试图对 BA 模型进行扩展，即根据现实中的网络，增添某些假定，以便进一步探索复杂网络系统的规律。对 BA 模型的扩充可以考虑三个因素：择优选择的成本、边的重新连接、网络的初始状态。扩充的 BA 模型可以更好地模拟现实世界中的网络现象。

（一）无尺度网络

1999 年，A. Barabási 和 R. Albert 在对互联网的研究中发现了无尺度网络，使人类对于复杂网络系统有了全新的认识。过去，人们习惯于将所有复杂网络看作是随机网络，但 Barabási 和 Albert 发现互联网实际上是由少数高连接性的页面组织起来的，80％以上页面的链接数不到 4 个。只占节点总数不到万分之一的极少数节点，却有 1000 个以上的链接。这种网页的链接分布遵循所谓的“幂次定律”：任何一个节点拥有 k 条连接的概率，与 $1/k$ 成正比。它不像钟形曲线那样具有一个集中度很高的峰值，而是一条连续递减的曲线。如果取双对数坐标系来描述幂次定律，得到的是一条直线。Scale-free 网络指的是节点的度分布符合幂律分布的网络，由于其缺乏一个描述问题的特征尺度而被称为无尺度网络。其后的几年中，研究者们在许多不同的领域中都发现了无尺度网络。从生态系统到人际关系，从食物链到代谢系统，处处可以看到无尺度网络。

图 1-1 描述了一个随机网络和无尺度网络的例子：美国公路系统为典型

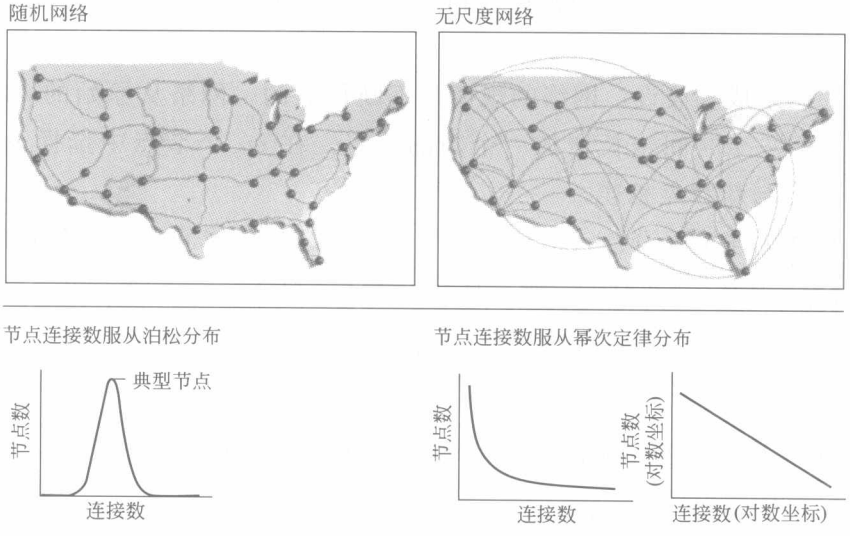


图 1-1　随机网络与无尺度网络的例子

的随机网络（左上图），其节点间的连接数服从钟形的泊松分布（左下图）；美国航空网则是典型的无尺度网络（右上图），存在少数拥有大量连接的集散节点，而大多数节点拥有较少连接，其节点连接数服从幂次定律分布（右下图）。

（二）BA 模型及其机制

为什么随机模型与实际不相符合呢？Barabási 和 Albert 在深入分析了 ER 模型之后，发现问题在于 ER 模型讨论的网络是一个既定规模的，不会继续扩展的网络。正是由于现实当中的网络往往具有不断成长的特性，早进入的节点（老节点）获得连接的概率就更大。当网络扩张到一定规模以后，这些老节点很容易成为拥有大量连接的集散节点。这就是网络的“成长性”。其次，ER 模型中每个节点与其他节点连接时，建立连接的概率是相同的。也就是说，网络当中所有的节点都是平等的。这一情况与实际也不相符。例如，新成立的网站选择与其他网站链接时，自然是在人们所熟知的网站中选择一个进行链接，新的个人主页上的超文本链接更有可能指向新浪、雅虎等著名的站点。由此，那些熟知的网站将获得更多的链接，这种特性称为“择优连接”。这种现象也称为“马太效应（Matthew Effect）”或“富者更富（Rich Get Richer）”。

“成长性”和“择优连接”这两种机制解释了网络当中集散节点的存在。

Barabási 和 Albert 根据这两种特性和假设提出了 BA 模型，从理论上解释了无尺度网络的现象。

（1）网络成长假设：网络的规模是不断扩大的。网络从原始的 m_0 个节点开始，每一个时间步长增加一个新的节点，在 m_0 个节点中选择 $m(m<m_0)$ 个节点与新节点相连。

（2）择优连接假设：一个新节点与一个已经存在的节点 i 相连接的概率 Π_i 与节点 i 的度 k_i 成正比，即：$\Pi(k_i)=\alpha k_i$，其中 $\alpha=1/\sum_j k_j$ 。

经过 t 步后，这种算法产生一个有 $N=t+m_0$ 个节点、$m\times t$ 条边的网络。

假设 k_i 是一个连续随机变量，k_i 变化的速率与 $\Pi(k_i)$ 成正比，因而 k_i 满足动力学方程：$\frac{\partial k_i}{\partial t}=m\,\Pi(k_i)=m\frac{k_i}{\sum_j k_j}$，每一步加入 m 条边，即增加了 $2m$ 个度值，于是分母求和项为 $\sum_j k_j=2mt$ 。则有：$\frac{\partial k_i}{\partial t}=\frac{k_i}{2t}$。

因为初始条件为 $k_i(t_i)=m$，故此方程的解为：$k_i(t)=m\left(\frac{t}{t_i}\right)^{\beta}$，其中$\beta=1/2$。由上式可以写出度少于 k 的节点的概率：$P[k_i(t)<k]=P\left[t_i>\frac{m^{1/\beta}t}{k^{1/\beta}}\right]$，若以等时间隔地向网络中增加节点，则 t_i 值就是一个常数概率密度 $P(t_i)=1/(m_0+t)$。因此，网络中度大于 k 的节点的概率为：

$$P[k_i(t)>k]=P(t_i)\frac{m^{1/\beta}t}{k^{1/\beta}}=\frac{m^{1/\beta}t}{k^{1/\beta}(t+m_0)} \tag{1-7}$$

网络中所有节点的概率之和为 1，所以度小于 k 的节点的概率为：

$$P[k_i(t)<k]=1-\frac{m^{1/\beta}t}{k^{1/\beta}(t+m_0)} \tag{1-8}$$

于是得到的度分布函数为：

$$P(k)=\frac{\partial P[k_i(t)<k]}{\partial k}=\frac{2m^{1/\beta}t}{(m_0+t)k^{1/\beta+1}} \tag{1-9}$$

当 $t\to\infty$时，有 $P(k)\sim 2m^{1/\beta}k^{-\gamma}$，其中 $\gamma=1/\beta+1=3$，由此可以看出 γ 与 m 无关。

由以上推导过程可知，BA 模型中的度分布 $P(k)$ 具有幂律特征，度的分布曲线是一条随着 k 增加、$P(k)$ 不断下降的递减曲线。