

第 17 讲 复杂网络

司守奎

烟台市, 海军航空大学

Email: sishoukui@163.com

1 引言

随着近年来关于复杂网络理论及其应用研究的不断深入, 人们开始尝试运用这种新的理论工具来研究现实世界中的各种大型复杂系统。其中复杂系统的结构以及系统结构与系统功能之间的关系是人们关注的热点问题。要研究这些复杂系统在结构和功能上的特点, 就需要用一种统一的工具来描述这些复杂系统。在复杂网络理论中, 这种工具就是网络图。而网络图的提出起源于欧拉对七桥问题的研究。

1.1 欧拉与 Königsberg 七桥问题

Königsberg 是东普鲁士一个城镇, 在其城中有条横贯城区的河流, 河中有两个岛, 两岸和两岛之间共架设七座桥。当地居民流传着这样一个有趣的问题: 一个人能否依次走过所有的七座桥且每桥只经过一次, 最后再返回出发点? 这就是著名的“七桥问题”。这个问题看起来似乎非常简单, 但长期以来小镇上没有一个人能够走出这样的一条路径来。

著名数学家欧拉在对这个问题进行分析之后, 巧妙地将被河流分隔开的四块陆地抽象为四个节点, 分别用 A, B, C, D 表示, 而将连接这四块陆地之间的七座桥抽象为这四个节点之间相互连接的七条边。这样就得到了由四个点和七条边构成的一个图 (如图 1 所示)。

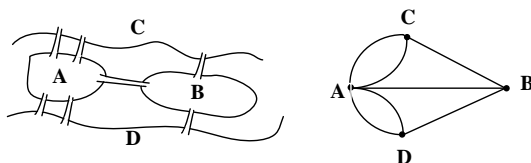


图 1 七桥问题示意图

经过简单抽象, 原来的七桥问题就转化为如下数学问题: 从任一点出发, 恰好经过每条边一次而返回原点的回路是否存在? 欧拉给出了存在回路的充要条件: 如果起点和终点重合 (回到原地), 则每一点只能有偶数条边与之相连。因此我们可以推出上述七桥问题是没有解的。

欧拉对七桥问题的抽象和论证思想, 开创了数学中一个新的分支—图论 (graph theory)。因此, 欧拉被公认为图论之父。事实上, 今天关于复杂网络的研究与欧拉当年关于七桥问题的研究在某种程度上是一脉相承的, 即用抽象的网络图结构来研究复杂系统的性质。

1.2 ER 随机图理论

在欧拉解决七桥问题之后的相当长一段时间里, 图论并未获得足够的发展, 直到 1936 年才出版了图论的第一部学术专著。20 世纪 60 年代匈牙利数学家 Erdős 和 Rényi 建立了随机图理论 (random graph theory)。这一理论被认为在数学上开创了复杂网络理论的系统性研究。在 Erdős 和 Rényi 研究的随机图模型 (称为 ER 随机图) 中, 任意两个节点之间有一条边相连接的概率都为 p 。因此, 一个含 N 个节点的 ER 随机图中边的总数是一个期望值为 $p[N(N-1)/2]$ 的随机变量。由此可以推得, 产生一个有 N 个节点和 M 条边的 ER 随机图的概率为 $p^M (1-p)^{N(N-1)/2-M}$ 。

Erdős 和 Rényi 系统研究了当 $N \rightarrow \infty$ 时 ER 随机图的性质 (如连通性等) 与概率 p 之间的关系。他们最重要的发现在于: ER 随机图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 , 要么几乎每一个图都具有某个性质, 要么几乎每一个图都不具有该性质。

1.3 社会领域的小世界实验

在随后 40 年中, 随机图理论一直是研究复杂网络的基本理论。在此期间, 人们也做了试图解释社会网络特征的一些实验。下面介绍其中比较出名的两个实验。

1. Milgram 的小世界实验

一个社会网络就是一群人或团体按某种关系连接在一起而构成的一个系统。当然, 这里

的关系可以是多种多样的，比如朋友关系、合作关系、姻缘关系和商业关系等等。那么对于地球上任意两个人来说，如果借助第三者、第四者这样的间接关系来建立他们两人的某种联系，平均需要通过多少人呢？社会心理学家 Stanley Milgram 在他的实验中给出的推断是平均中间只要通过 5 个人，即只要通过 6 个联接关系，你与地球上任何角落的任何一个人之间就能建立某种联系，这就是著名的六度分离（six degrees of separation）推断。

Milgram 的小世界实验是这样做的。首先，他选定了两个目标对象：一个是美国马萨诸塞州（Massachusetts）沙朗(Sharon)的一位神学院研究生的妻子，另一位是波士顿（Boston）的一个证券经纪人。然后他在遥远的堪萨斯州（Kansas）和内布拉斯加州（Nebraska）招募到了一批志愿者。Milgram 要求这些志愿者通过自己所认识的人，用自己认为尽可能少的传递次数，设法把一封信最终转交到一个给定的目标对象手中。Milgram 根据最终到达目标者手中的信件的分析发现，从一个志愿者到其目标对象的平均距离只是 6。实验结果在某种程度上反映了人际关系的“小世界”特征。

2. Kevin Bacon 游戏

为检验“六度分离”假设的正确性，人们又做了其他一些小世界实验。其中一个著名的实验就是“Kevin Bacon 游戏（game of Kevin Bacon）”。这个游戏的主角是美国演员 Bacon。游戏的目的是把 Bacon 和另外任意一个演员联系起来。在该游戏里为每个演员定义了一个 Bacon 数：如果一个演员和 Bacon 一起演过电影，那么他（她）的 Bacon 数就为 1；如果一个演员没有和 Bacon 演过电影，但是他（她）和 Bacon 数为 1 的演员一起演过电影，那么该演员的 Bacon 数就为 2，以此类推。这个游戏就是通过是否共同出演一部电影为纽带建立起演员和演员间的关系网。而 Bacon 数就描述了这个网络中任意一个演员到 Bacon 的最短路径。

美国 Virginia 大学的科学家建立了一个电影演员的数据库，放在网上供人们随意查询（<http://www.cs.virginia.edu/oracle/>）。网站的数据库里目前总共存有近 60 万个世界各地的演员的信息以及近 30 万部电影信息。通过简单地输入演员名称就可以知道这个演员的 Bacon 数。表 1 是对近 60 万个演员所做的统计：第一行是 Bacon 数，第二行是具有这个 Bacon 数的演员个数。可以看到最大的 Bacon 数仅仅为 8，而平均 Bacon 数仅为 2.944。

表 1 电影演员的 Bacon 数

Bacon 数	0	1	2	3	4	5	6	7	8
演员数	1	1682	132399	357230	86206	6734	852	103	13

2 复杂网络的特性

复杂网络能很好地描述自然科学、社会科学、管理科学和工程技术等领域的相互关联的复杂模型。它以数学、统计物理学、计算机等科学为分析工具，以复杂系统为研究目标。复杂网络是 21 世纪发展较快的一门交叉学科，然后，到目前为止，在网络科学的研究中，复杂网络没有一个统一的定义。顾名思义，复杂网络就是高度复杂的网络，很难给出一个严格的定义。钱学森给出了一个描述性定义：具有自组织、自相似、吸引子、小世界、无标度中部分或全部性质的网络称为复杂网络。维基百科中将复杂网络定义为由数量巨大的节点和节点之间错综复杂的关系共同构成的网络结构，就是说一个有着足够复杂的拓扑结构特征的图。Bollobás 和 Riordan 将其定义为随机图过程。前者是一个定性定义，过于宽泛；后者涉及数学上太抽象的随机图过程。

复杂网络的优美结构和新奇的规律，越来越吸引着人们去探索更多的奥妙。大多数复杂网络的复杂性表现在如下几个方面：

（1）网络规模庞大。网络节点数可以有成百上千万，甚至更多，但大规模性的网络行为具有统计特性。

（2）连接结构的复杂性。网络连接结构既非完全规则也非完全随机，但却具有其内在的自组织规律，网络结构可呈现多种不同特性。

（3）节点的复杂性。首先表现在节点的动力学复杂性，即各个节点本身可以是各种非线性系统（可以有离散的和连续微分方程描述），具有分岔和混沌等非线性动力学行为。其次表现在节点的多样性，复杂网络中的节点可以代表任何事物，而且一个复杂网络中可能出现各种不同类型的节点。

（4）网络时空演化过程复杂。复杂网络具有空间和时间的演化复杂性，展示出丰富的

复杂行为，特别是网络节点之间的不同类型的同步化运动（包括出现周期、非周期、混沌和阵发行为等运动）。

(5) 网络连接的稀疏性。一个有 N 个节点的具有全局耦合结构的网络的连接数目为 $O(N^2)$ ，而实际大型网络的连接数目通常为 $O(N)$ 。

(6) 多重复杂性融合。若以上多重复杂性相互影响，将导致更为难以预料的结果。例如，设计一个电力供应网络需要考虑此网络的进化过程，其进化过程决定网络的拓扑结构。当两个节点之间频繁进行能量传输时，它们之间的连接权重会随之增加，通过不断的学习与记忆逐步改善网络性能。

除了复杂性，复杂网络一般还具有以下三个特性。

(1) 小世界特性。大多数网络尽管规模很大，但任意两个节点间却有一条相当短的路径。

(2) 无标度特性。人们发现一些复杂网络的节点的度分布具有幂指数函数的规律。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，所以该特性被称为无标度性质。无标度特性反映了网络中度分布的不均匀性，只有很少数的节点与其它节点有很多的连接，成为“中心节点”，而大多数节点度很小。

(3) 超家族特性。2004 年 Sheffer 和 Alon 等在“Science”上发表文章，比较了许多已有网络的局部结构和拓扑特性，观测到有一些不同类型的网络的特性在一定条件下具有相似性。尽管网络不同，只要组成网络的基本单元（最小子图）相同，它们的拓扑性质的重大轮廓外形就可能具有相似性，这种现象被他们称为超家族特性。顾名思义，不同网络之间存在与某个家族的“血缘”相近联系，而出现与该家族相似的特性，究其原因在于它们拥有相同的或相似的网络“基因”，但问题是网络“基因”是不是找准了？是否存在网络“基因”排序等更深层次的问题。目前，对于超家族特性在研究理论方法和技术上都有待进一步改进和发展，需要更多的不同网络的实证研究和严格的理论证明。

3 矩母函数、特征函数和概率母函数

利用分布函数的变换有时可以较容易地确定独立随机变量和的分布。矩母函数就是这种变换之一。对一个非负随机变量 X ，其矩母函数定义为

$$m_X(t) = E[e^{tX}], \quad -\infty < t < h,$$

其中 h 为某个常数。因为我们特别要用到矩母函数在 0 点附近的小区间里的取值，所以要求 $h > 0$ 。随机变量的矩母函数与分布函数一一对应。如果 X 和 Y 相互独立，则

$$m_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}]E[e^{tY}] = m_X(t)m_Y(t).$$

对于某些具有重尾的分布，如柯西分布，其矩母函数不存在。但是特征函数总是存在的。特征函数定义为

$$\phi_X(t) = E[e^{itX}], \quad -\infty < t < +\infty.$$

利用展开式可以得到

$$m_X(t) = \sum_{k=0}^{\infty} \frac{E[X^k]t^k}{k!},$$

所以 X 的 k 阶矩等于

$$E[X^k] = \frac{d^k}{dt^k} m_X(t) \Big|_{t=0}.$$

随机变量的特征函数与分布函数一一对应。

概率母函数 (pgf) 用于离散型随机变量，定义为

$$g_X(t) = E[t^X] = \sum_{k=0}^{\infty} t^k P\{X = k\}.$$

若 $|t| \leq 1$ ，则上式中的级数总是收敛的。

概率母函数、特征函数和矩母函数之间有如下的关系：

$$g_X(t) = m_X(\log t); \quad \phi_X(t) = m_X(it).$$

例 1 如果参数为 (μ, σ^2) 的正态分布随机变量 X 的概率密度函数为

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty,$$

证明正态分布的矩母函数为 $M_x(t) = e^{\mu t + \frac{\sigma^2}{2} t^2}$ 。

证明 矩母函数

$$\begin{aligned} M_x(t) &= E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{[x-(\sigma^2 t + \mu)]^2}{2\sigma^2}} e^{\mu t + \frac{\sigma^2}{2} t^2} dx \\ &\text{令 } \frac{x - (\sigma^2 t + \mu)}{\sigma} = y, \text{ 得} \\ M_x(t) &= e^{\mu t + \frac{\sigma^2}{2} t^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = e^{\mu t + \frac{\sigma^2}{2} t^2}. \end{aligned}$$

例 2 如果 Poisson 随机变量 N 的分布律为

$$P\{N = k\} = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

计算 Poisson 随机变量的概率母函数和矩母函数。

解 Poisson 随机变量 N 的概率母函数为

$$P(z) = E(z^N) = \sum_{k=0}^{\infty} z^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(z\lambda)^k}{k!} = e^{\lambda(z-1)}.$$

于是它的矩母函数为

$$M(t) = P(e^t) = e^{\lambda(e^t - 1)}.$$

4 网络的基本静态几何特征

实际网络都兼有确定和随机两大特征，确定性的法则或特征通常隐藏在统计性质中，因此，对复杂网络各种性质的统计描述十分重要。虽然目前提出的大多数网络统计性质仅描述网络的拓扑性质，但是由于网络节点的连边表示它们形形色色的相互作用，所以这种统计描述也包含动力学的成分，具有重要的意义。

人们在刻画复杂网络结构的统计特性上提出了许多概念和方法，其中包含三个基本的概念：度分布、平均路径长度和聚类系数。下面的网络如果没有特别指明，都是指简单无向网络。

4.1 度与度分布

1. 节点的度

节点 v_i 的度 k_i 定义为与该节点连接的边数。直观上看，一个节点的度越大就意味着这个节点在某种意义上越“重要”。网络中所有节点 v_i 的度 k_i 的平均值称为网络的平均度，记为 $\langle k \rangle$ ，即

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i.$$

无向无权图的邻接矩阵 A 与节点 v_i 的度 k_i 的函数关系很简单：邻接矩阵二次幂 A^2 的对角元素 $a_{ii}^{(2)}$ 就等于节点 v_i 的度，即

$$k_i = a_{ii}^{(2)}.$$

实际上，无向无权图的邻接矩阵 A 的第 i 行或第 i 列元素之和也是度，从而无向无权网络的

平均度就是 A^2 对角线元素之和除以节点数，即

$$\langle k \rangle = \text{tr}(A^2) / N,$$

式中， $\text{tr}(A^2)$ 表示矩阵 A^2 的迹（trace），即对角线元素之和。

2. 度分布

网络中节点的度分布情况可用分布函数 $P(k)$ 来描述， $P(k)$ 表示的是网络中度为 k 的节点在整个网络中所占的比率，也就是说，在网络中随机抽取到度为 k 的节点的概率为 $P(k)$ 。一般地，可以用一个直方图来描述网络的度分布（degree distribution）性质。

对于规则的网络来说，由于所有的节点具有相同的度，所以其度分布集中在一个单一尖峰上，是一种 Delta 分布。网络中的任何随机化倾向都将使这个尖峰的形状变宽。完全随机网络的度分布近似为 Poisson 分布，其形状在远离峰值 $\langle k \rangle$ 处呈指数下降。这意味着当 $k > \langle k \rangle$ 时，度为 k 的节点实际上是不存在的。因此，这类网络也称为均匀网络（homogeneous network）。

近几年的大量研究表明，许多实际网络的度分布明显地不同于 Poisson 分布。特别地，许多网络的度分布可以用幂律形式 $P(k) \propto k^{-\gamma}$ 来更好地描述。图 2 是两种分布在同一坐标系下的对比，可以看出幂律分布（ $P(k) = \frac{6}{\pi^2} k^{-2}$ ）曲线比 Poisson 分布（ $P(k) = \frac{30^k}{k!} e^{-30}$ ）

曲线下降要缓慢得多。

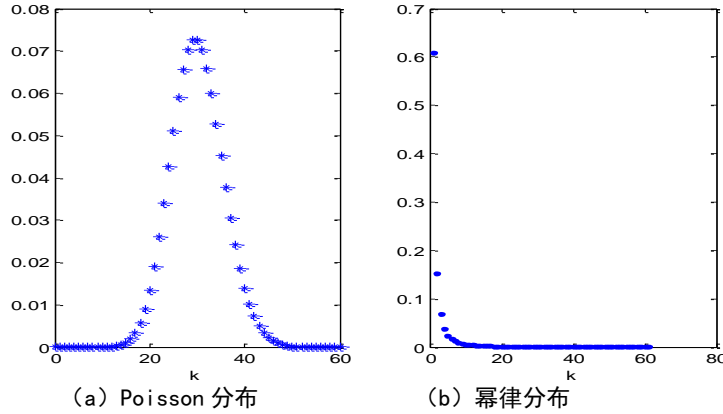


图 2 两种网络度分布比较

幂律分布也称为无标度（scale-free）分布，具有幂律度分布的网络也称为无标度网络。这是因为幂律分布函数具有如下无标度性质。

引理 1 考虑一个概率分布函数 $F(x)$ ，如果对任意给定常数 a ，存在常数 b 使得函数 $F(x)$ 满足“无标度条件” $F(ax) = bF(x)$ ，那么必有（假定 $F(1)F'(x) \neq 0$ ）

$$F(x) = F(1)x^{-\gamma}, \quad \gamma = -F'(1)/F(1). \quad (1)$$

也就是说，幂律分布函数是唯一满足“无标度条件”的概率分布函数。

证明 取 $x=1$ ，我们有 $F(a) = bF(1)$ ，从而 $b = F(a)/F(1)$ ，则

$$F(ax) = \frac{F(a)F(x)}{F(1)}, \quad (2)$$

由于上述方程对任意的 a 都成立，方程两边对 a 求导可得

$$x \frac{dF(ax)}{d(ax)} = \frac{F(x)}{F(1)} \frac{dF(a)}{da}, \quad (3)$$

若取 $a=1$ ，则有

$$x \frac{dF(x)}{d(x)} = \frac{F(x)}{F(1)} F'(1), \quad (4)$$

微分方程（4）的解为

$$\ln F(x) = \frac{F'(1)}{F(1)} \ln x + \ln F(1), \quad (5)$$

两边取指数，即得公式（1）。

在一个度分布为具有适当幂指数（通常为 $2 \leq \gamma \leq 3$ ）的幂律形式的大规模无标度网络中，绝大部分的节点的度相对很低，但存在少量的度相对很高的节点。因此，这类网络也称为非均匀网络，而那些度相对很高的节点称为网络的“集线器”。例如，美国高速公路网就可近似看作是一个均匀网络，因为不可能有上百条高速公路都经过同一个城市；而美国航空网则可看作是一个无标度网络，大部分机场都是小机场，但却存在少量连接众多小机场的非常大的机场。

3. 累积度分布

除了度分布，人们还可以用累积度分布函数（cumulative degree distribution function）来描述度的分布情况，它与度分布的关系为

$$P_k = \sum_{i=k}^{\infty} P(i), \quad (6)$$

它表示的是度不小于 k 的节点的概率分布。

如果度分布为幂律分布，即 $P(k) \propto k^{-\gamma}$ ，那么累积度分布函数符合幂指数为 $\gamma - 1$ 的幂律

$$P_k \propto \sum_{i=k}^{\infty} i^{-\gamma} \propto k^{-(\gamma-1)}. \quad (7)$$

如果度分布为指数分布，即 $P(k) \propto e^{-\frac{k}{\lambda}}$ ，其中 $\lambda > 0$ 为一常数，那么累积度分布函数也是指数型的，且具有相同的指数

$$P_k \propto \sum_{i=k}^{\infty} e^{-\frac{k}{\lambda}} \propto e^{-\frac{k}{\lambda}}. \quad (8)$$

4.2 平均路径长度

网络中两个节点 v_i 和 v_j 之间的距离 d_{ij} 定义为连接这两个节点的最短路径上的边数，它的倒数 $1/d_{ij}$ 称为节点 v_i 和 v_j 之间的效率，记为 ε_{ij} 。通常效率用来度量节点间的信息传递速度。当 v_i 和 v_j 之间没有路径连通时， $d_{ij} = \infty$ ，而 $\varepsilon_{ij} = 0$ 。

网络中任意两个节点之间的距离的最大值称为网络的直径，记为 D ，即

$$D = \max_{1 \leq i < j \leq N} d_{ij},$$

其中， N 为网络节点数。网络的平均路径长度 L 则定义为任意两个节点之间的距离的平均值，即

$$L = \frac{1}{C_N^2} \sum_{1 \leq i < j \leq N} d_{ij}.$$

一个含有 N 个节点和 M 条边的网络的平均路径长度可以用时间量级 $O(MN)$ 的广度优先搜索算法来确定。

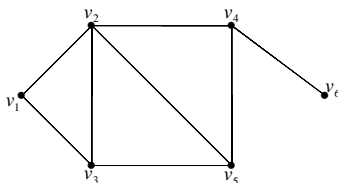


图3 一个简单网络的直径和平均路径长度

例 3 对于图 3 所示的一个包含 6 个节点和 8 条边的网络，求网络直径 D 和平均路径长度 L 。

解 首先构造图 1.1 对应的图 $G = (V, E)$ 的邻接矩阵

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

然后应用 Floyd 算法求出任意节点之间的最短距离见表 2。其中的最大距离为网络直径，网络直径 $D = d_{16} = d_{36} = 3$ ，把所有节点对之间的距离求和（只需求对应矩阵上三角元素的和），再除以对应完全图的边数 C_6^2 ，求得网络的平均路径长度 $L = 1.6$ 。

表 2 6 个节点对之间的最短距离

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	1	1	2	2	3
v_2	1	0	1	1	1	2
v_3	1	1	0	2	1	3
v_4	2	1	2	0	1	1
v_5	2	1	1	1	0	2
v_6	3	2	3	1	2	0

计算的 Matlab 程序如下

```
clc, clear
a=zeros(6); %邻接矩阵初始化
a(1,[2 3])=1; a(2,[3:5])=1; %输入邻接矩阵的上三角元素
a(3,5)=1; a(4,[5 6])=1;
a=a+a'; %构造完整的邻接矩阵
[D,L,dist]=myAPL(a) %调用我们自己编写的函数，计算网络直径 D，平均路径长度 L，
和所有节点对之间的最短距离 dist
```

我们编写的计算网络的最大直径及平均路径长度的 Matlab 函数如下

```
function [D,L,dist]=myAPL(a);
%计算网络直径 D，平均路径长度 L，和所有节点对之间的最短距离 dist；输入参数 a
为网络的邻接矩阵
a=tril(a); %截取邻接矩阵的下三角部分，满足 Matlab 工具箱的要求
a=sparse(a); %转换成稀疏矩阵,Matlab 工具箱的要求
dist=graphallshortestpaths(a,'directed',false);
D=max(max(dist)); %计算网络直径
Ldist=tril(dist); %提取最短距离矩阵的下三角部分
he=sum(nonzeros(Ldist)); %求所有边的和
n=length(a); %求节点个数
L=he/nchoosek(n,2); %求平均路径长度，这里使用 Matlab 求组合数命令
```

注：要调用上面的函数，必须输入无向图完整的邻接矩阵，即一个实对称矩阵。在 Matlab 工具箱中求无向图的最短路径，只使用邻接矩阵的下三角元素。

在朋友关系网络中， L 是连接网络内两个人之间最短关系链中的朋友的平均个数。近期的研究发现，尽管许多实际的复杂网络的节点数巨大，网络的平均路径长度却惊人的小。具体地说，对于一个网络而言，如果对于固定的网络节点平均度 $\langle k \rangle$ ，平均路径长度 L 的增加速度至多与网络规模 N 的对数成正比，则称此网络是具有小世界效应的。

4.3 聚类系数

在你的朋友关系网络中，你的两个朋友很可能彼此也是朋友，这种属性在复杂网络理论里称为网络的聚类特性。一般地，假设网络中的一个节点 v_i 有 k_i 条边将它和其它节点相连，这 k_i 个节点就称为节点 v_i 的邻居。显然，在这 k_i 个节点之间最多可能有 $C_{k_i}^2$ 条边。而这 k_i 个节点之间实际存在的边数 E_i 和总的可能的边数 $C_{k_i}^2$ 之比就定义为节点 v_i 的聚类系数 C_i ，即

$$C_i = \frac{E_i}{C_{k_i}^2}. \quad (9)$$

从几何特点看，上式的一个等价定义为

$$C_i = \frac{\text{与点 } i \text{ 相连的三角形的数量}}{\text{与点 } i \text{ 相连的三元组的数量}} = \frac{n_1}{n_2}, \quad (10)$$

其中，与节点 i 相连的三元组是指包括节点 i 的三个节点，并且至少存在从节点 i 到其它两个节点的两条边（图4）。

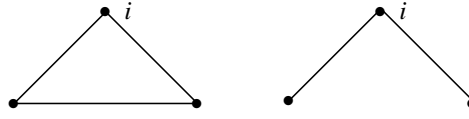


图4 以节点 i 为顶点之一的三元组的两种可能形式

下面讨论如何根据非赋权无向图的邻接矩阵 A 来求节点 v_i 的聚类系数 C_i 。显然，邻接矩阵二次幂 A^2 的对角元素 $a_{ii}^{(2)}$ 表示的是与节点 v_i 相连的边数，也就是节点 v_i 的度 k_i 。而邻接矩阵三次幂 A^3 的对角元素 $a_{ii}^{(3)}$ 表示的是从节点 v_i 出发经过三条边回到节点 v_i 的路径数，也就是与节点 v_i 相连的三角形数的两倍（正向走和反向走）。因此，由聚类系数的表达式（10）可知

$$C_i = \frac{n_1}{n_2} = \frac{2n_1}{2C_{k_i}^2} = \frac{2n_1}{k_i(k_i-1)} = \frac{a_{ii}^{(3)}}{a_{ii}^{(2)}(a_{ii}^{(2)}-1)}, \quad (11)$$

整个网络的聚类系数 C 就是所有节点 v_i 的聚类系数 C_i 的平均值，即

$$C = \frac{1}{N} \sum_{i=1}^N C_i.$$

显然， $0 \leq C \leq 1$ 。当所有的节点均为孤立节点，即没有任何连接边时， $C = 0$ ； $C = 1$ 当且仅当网络是全局耦合的，即网络中任意两个节点都直接相连。

对于一个含有 N 个节点的完全随机的网络，当 N 很大时， $C = O(N^{-1})$ 。而许多大规模的实际网络都具有明显的聚类效应，它们的聚类系数尽管远小于 1 但却比 $O(N^{-1})$ 大得多。事实上，在很多类型的网络中，随着网络规模的增加，聚类系数趋向于某个非零常数，即当 $N \rightarrow \infty$ 时， $C = O(1)$ 。这意味着这些实际的复杂网络并不是完全随机的，而是在某种程度上具有类似于社会关系网络中“物以类聚，人以群分”的特性。

例4 计算例3简单网络的聚类系数。

解 下面以节点 v_2 的聚类系数计算为例，节点 v_2 与 4 个节点相邻，而这 4 个节点之间可能存在的最大边数为 $C_4^2 = 6$ ，而这 4 个节点间实际存在的边数为 3，由定义可得

$$C_2 = 3 / C_4^2 = 1/2,$$

同理可求得其它节点的聚类系数为

$$C_1 = 1, C_3 = 2/3, C_4 = 1/3, C_5 = 2/3, C_6 = 0.$$

整个网络的聚类系数 $C = 19/36$ 。

计算的 Matlab 程序如下

```
clc, clear
```



```

format rat %有理分数的数据格式
a=zeros(6); %邻接矩阵初始化
a(1,[2 3])=1; a(2,[3:5])=1; %输入邻接矩阵的上三角元素
a(3,5)=1; a(4,[5 6])=1;
a=a+a'; %构造完整的邻接矩阵
[TC,c]=mycluster(a) %调用我们自己编写的 Matlab 函数，TC 是整个网络的聚类系数，
c 为各个节点的聚类系数

```

我们利用聚类函数的定义，编写的计算整个网络的聚类系数及各个节点的聚类系数的 Matlab 函数如下

```

function [TC,c]=mycluster(a);
%输出参数 TC 是整个网络的聚类系数，c 为各节点的聚类系数，输入参数是邻接矩阵
n=length(a);
for i=1:n
    m=find(a(i,:)); %找第 i 行非零元的地址
    ta=a(m,m); %提取节点 vi 的所有邻居节点所构成的邻接矩阵
    Lta=tril(ta); %提取邻居节点所构成邻接矩阵的下三角元素
    if length(m)==0 | length(m)==1
        c(i)=0; %孤立节点或只有 1 个邻居的节点的聚类系数定义为 0
    else
        c(i)=sum(sum(Lta))/nchoosek(length(m),2); %求节点 vi 的聚类系数
    end
end
TC=sum(c)/n; %求整个网络的聚类系数

```

我们也可以利用式 (11) 计算聚类系数，计算的 Matlab 程序如下

```

clc, clear
format rat %有理分数的数据格式
a=zeros(6); %邻接矩阵初始化
a(1,[2 3])=1; a(2,[3:5])=1; %输入邻接矩阵的上三角元素
a(3,5)=1; a(4,[5 6])=1;
a=a+a'; %构造完整的邻接矩阵
n=length(a);
b3=a^3; b33=b3(1:n+1:end); %提出对角线元素
b2=a^2; b22=b2(1:n+1:end); %提出对角线元素
c=b33./(b22.*(b22-1));
c(isnan(c))=0 %把分母为 0 的不确定值替换成 0
TC=mean(c) %求整个网络的聚类系数

```

4.4 实际网络的统计性质

在 Watts 和 Strogatz 关于复杂网络的小世界现象的研究，以及 Barabasi 和 Albert 关于复杂网络的无标度特征的工作之后，人们对来自不同领域的大量实际网络的拓扑特征进行了广泛的实证性研究，表 3 列出了部分结果。测量的性质包括：有向或无向、节点总数 N 、边的总数 M 、平均度数 $\langle k \rangle$ 、平均路径长度 L 、聚类系数 C 。表中的空格表示没有可靠的数据。

表 3 各种实际网络的特征参数统计

网络		类型	N	M	$\langle k \rangle$	L	C
社会 领域	电影演员	无向	449913	25516482	113	3.48	0.78
	公司董事	无向	7673	55392	14.4	4.6	0.88
	数学家合作	无向	253339	496489	3.92	7.57	0.34
	电子邮件	有向	59912	86300	1.44	4.95	0.16

	学生关系	无向	573	477	1.66	16	0
信息领域	WWW(nd.edu)	有向	269504	1497135	5.55	11.3	0.29
	引用网络	有向	783339	6716198	8.57		
	罗氏词典	有向	1022	5103	4.99	4.87	0.15
	单词搭配网络	无向	460902	1.7E+07	70.1		0.44
技术领域	电力网	无向	4941	6594	2.67	19	0.08
	铁路网	无向	587	19603	66.8	2.16	0.69
	软件包	有向	1439	1723	1.2	2.42	0.08
	对等网络	无向	880	1296	1.47	4.28	0.01
生物领域	代谢网络	无向	765	3686	9.64	2.56	0.67
	蛋白质网络	无向	2115	2240	2.12	6.8	0.07
	海洋食物网	有向	135	598	4.43	2.05	0.23
	神经网络	有向	307	2359	7.68	3.97	0.28

5 网络机制模型

5.1 随机网络

随机网络就具有上面提到的“小世界”特点，即顶点之间的平均最短路径比较短。经典的随机网络是这样构造的：给定 N 个顶点，任意两点之间都尝试以某个概率 p 连接，所构成的网络就叫随机网络。数学家在上个世纪50年代对随机网络进行了深入细致的研究，得到了网络性质突现的临界连接概率等一系列结果，发现对于同样规模的网络，随机网络具有短的平均最短路径，这一点和许多实证研究结果相符。但另一方面，许多实际网络都具有很高的平均聚类系数，而随机网络却是一个局部集团化很差的网络，平均聚类系数很低。所以，随机网络不是实际网络的一个很好的刻画。

和随机网络相对应的是完全规则网络，规则网络的性质恰与随机网络相反，一般情况下具有大的聚类系数和长的平均最短路径。所以，无论是规则网络还是随机网络都无法作为描述真实网络的合适的研究基础。或者说，把规则网络和随机网络作为复杂系统的抽象都过于简单化了，离真实的复杂性太远。

5.2 小世界网络

Watts 和 Strogatz 结合规则网络和随机网络的特点，给出了小世界网络的生成机制模型。WS 小世界网络是在规则网络的基础上加入随机性产生的，即对规则网络的每一个顶点的所有边，以概率 p 断开一个端点，并重新连接，连接的新端点从网络中的其它顶点里随机选择。WS 小世界模型的构造算法如下

(1) 从规则图开始：考虑一个含有 N 个点的最近邻耦合网络，它们围成一个环，其中每个节点都与它左右相邻的各 $K/2$ 节点相连， K 是偶数。

(2) 随机化重连：将上面规则图中的每条边以概率 p 随机地重新连接，即将边的一个端点保持不变，而另一个端点以概率 p 变为网络中其余 $N - K - 1$ 个节点中随机选择的一个节点。其中规定，任意两个不同的节点之间至多只能有一条边，即若重连的两个节点之间有边，则该边就不进行重连。

例 5 WS 小世界网络仿真。

仿真的 Matlab 程序如下

```

clc, clear, hold on
N=20; K=4; p=0.2; %N 为网络节点总数，K 为邻域节点个数，p 为重连概率
t=0:2*pi/N:2*pi-2*pi/N; %生成最近邻耦合网络各节点坐标的参数方程的角度
x=100*sin(t); y=100*cos(t);
plot(x,y,'ro','MarkerEdgeColor','g','MarkerFaceColor','r','markersize',6);
A=zeros(N); %邻接矩阵初始化
for i=1:N %该层循环构造最近邻 K 耦合网络的邻接矩阵
    for j=i+1:i+K/2
        jj=(j<=N)*j+(j>N)*mod(j,N); %如果 j 超过了 N，要取除以 N 的余数
        A(i,jj)=1; A(jj,i)=1;
    end
end
end

```

```

for i= 1:N    %该层循环进行随机重连
    for j=i+1:i+K/2
        jj=(j<=N)*j+(j>N)*mod(j,N);
        ChangeV=randint(1,1,[1,N]); %产生随机整数，为可能重连的另外一个节点
        if rand<=p & A(i,ChangeV)==0 & i~=ChangeV %重连的条件
            A(i,jj) = 0; A(jj,i) = 0; %删除原边
            A(i,ChangeV)=1; A(ChangeV,i)=1; %重连新边
        end
    end
end
end
for i=1:N-1
    for j=i+1:N
        if A(i,j)~=0
            plot([x(i),x(j)], [y(i),y(j)], 'linewidth', 1.2);
        end
    end
end
end
Matlab_to_Pajek(A) %把邻接矩阵 A 转换为 Pajek 格式的数据，这里函数
Matlab_to_Pajek 是我们自己编写的函数，见下面的复杂网络的可视化。

```

在上述模型中， $p = 0$ 对应于完全规则网络， $p = 1$ 则对应于完全随机网络，通过调节 p 的值就可以控制从完全规则网络到完全随机网络的过渡。 p 值存在一个很大的区域，网络同时拥有大聚类系数和小最短路径，在复杂网络研究中，网络同时具备这两个性质时我们就称其具有小世界特性。

WS 小世界模型构造算法中的随机化过程有可能破坏网络的连通性。另一个研究较多的小世界模型是由 Newman 和 Watts 稍后提出的，称为 NW 小世界模型。该模型是通过用“随机化加边”取代 WS 小世界模型构造中的“随机化重连”而得到的。NW 小世界模型构造算法如下

(1) 从规则图开始：考虑一个含有 N 个点的最近邻耦合网络，它们围成一个环，其中每个节点都与它左右相邻的各 $K/2$ 节点相连， K 是偶数。

(2) 随机化加边：以概率 p 在随机选取的一对节点之间加上一条边。其中，任意两个不同的节点之间至多只能有一条边，并且每一个节点都不能有边与自身相连。

在 NW 小世界模型中， $p = 0$ 对应于原来的最近邻耦合网络， $p = 1$ 则对应于全局耦合网络。在理论分析上，NW 小世界模型要比 WS 小世界模型简单一些。当 p 足够小和 N 足够大时，NW 小世界模型本质上等同于 WS 小世界模型。

例 6 NW 小世界网络仿真。

仿真的 Matlab 程序如下

```

clc, clear, hold on
N=100; K=4; p=0.15; %N 为网络节点总数,K 为邻域节点个数,p 为随机化加边概率
t=0:2*pi/N:2*pi-2*pi/N; %生成最近邻耦合网络各节点坐标的参数方程的角度
x=100*sin(t); y=100*cos(t);
plot(x,y,'ko','MarkerEdgeColor','k','MarkerFaceColor','r','markersize',6);
A=zeros(N); %邻接矩阵初始化
for i=1:N    %该层循环构造最近邻 K 耦合网络的邻接矩阵
    for j=i+1:i+K/2
        jj=(j<=N)*j+(j>N)*mod(j,N); %如果 j 超过了 N，要取除以 N 的余数
        A(i,jj)=1; A(jj,i)=1;
    end
end
end
A(1:N+1:N^2)=1; %邻接矩阵对角线元素置 1，下面找无边的节点对时，要排除 A(i,i)=0
[u,v]=find(A==0); %找无边的节点对
m=length(u); %计算无边节点对的个数

```

```

z=rand(1,m); %生成 m 个随机数
ind=find(z<=p); %找 z 中小于等于 p 的随机数的地址，对应的地址将来加边
for k=1:length(ind);
    A(u(ind(k)),v(ind(k)))=1; A(v(ind(k)),u(ind(k)))=1;
end
for i=1:N-1
    for j=i+1:N
        if A(i,j)~=0
            plot([x(i),x(j)], [y(i),y(j)], 'linewidth', 1.2);
        end
    end
end
end
Matlab_to_Pajek(A) %生成 Pajek 数据

```

5.3 无标度网络

大量实际网络还存在着另一个突出的结构特征—幂律度分布，我们称其为无标度网络。网络中的大部分节点度值都很低，但存在着度数非常高的中枢节点。幂律度分布使网络在小世界特征的基础上又具有了许多新的性质。如不存在传染病传播的临界阈值等。对网络攻击的研究结果表明，随机攻击基本上不会破坏无标度网络的连通性，但在有目的的最大度攻击下，很小比例的顶点移除就会对网络的连通性造成根本性的破坏。鲁棒性和脆弱性奇妙地结合在了一起，人们戏称其为“阿基里斯的足踝”。这与现实世界中许多复杂系统的表现完全类似。

为了解释幂率分布的产生机理，Barabási 和 Albert 提出了一个无标度网络模型，现被称为 BA 模型。他们认为以前的许多网络模型都没有考虑到实际网络的如下两个重要特性：

(1) 增长(growth)特性：即网络的规模是不断扩大的。例如每个月都会有大量的新的科研文章发表，而 WWW 上则每天都有大量新的网页产生。

(2) 优先连接(preferential attachment)特性：即新的节点更倾向于那些具有较高连接度的“大”节点相连接。这种现象也称为“富者更富(rich get richer)”或“马太效应(Matthew effect)”。例如，新发表的文章更倾向于引用一些已被广泛应用的重要文献。

基于网络的增长和优先连接特性，BA 无标度网络模型的构造算法如下

(1) 增长：从一个具有 m_0 个节点的网络开始，每次引入一个新的节点，并且连到 m 个已存在的节点上，这里 $m \leq m_0$ 。

(2) 优先连接：一个新节点与一个已经存在的节点 v_i 相联接的概率 Π_i 与节点 v_i 的度 k_i 满足如下关系

$$\Pi_i = \frac{k_i + 1}{\sum_j (k_j + 1)}.$$

例 7 BA 无标度网络仿真。

仿真 BA 无标度网络的 Matlab 程序为

```

clc, clear
m0=input('请输入未增长前的网络节点个数 m0: ');
m=input('请输入每次引入新节点时新生成的边数 m: ');
N=input('请输入增长后的网络节点总数 N: ');
disp('初始网络时 m0 个节点的连接情况：1 表示都是孤立点；2 表示构成完全图；3 表示随机连接一些边');
se=input('请选择初始网络情况 1, 2 或 3: ');
if m>m0
    disp('输入参数 m 不合法'); return;
end
x=100*rand(1,m0); y=100*rand(1,m0); %构造初始 m0 个节点用于画图的坐标
if se==1

```

```

    A=zeros(m0);
elseif se==2
    A=ones(m0); A(1:m0+1:m0^2)=0; %对角线元素置 0
else
    A=zeros(m0); B=rand(m0); B=tril(B); %截取下三角元素
    A(B<=0.1)=1; %按照概率 0.1 进行连边
    A=A+A'; %构造完整的邻接矩阵
end
for k=m0+1:N
    x(k)=100*rand; y(k)=100*rand; %生成当前节点用于画图的坐标
    p=(sum(A)+1)/sum(sum(A)+1); %计算所有节点的连接概率
    pp=cumsum(p); %求累积分布
    A(k,k)=0; %加入新的连边之前, 邻接矩阵扩充维数
    for i=1:m %从已有的节点中随机选择 m 个节点与新加入的节点相连
        while 1 %若新选择的节点之间有边, 重新选择
            ind=find(pp>=rand);jj=ind(1); %节点 jj 为用赌轮法选择的节点
            if A(k,jj)==0
                A(k,jj)=1; A(jj,k)=1; break %选择成功, 退出该层循环
            end
        end
    end
end
end
plot(x,y,'ro','MarkerEdgeColor','g','MarkerFaceColor','r','markersize',8);
hold on, A2=tril(A); [i,j]=find(A2); %找邻接矩阵下三角元素的非零元素
for k=1:length(i)
    plot([x(i(k)),x(j(k))],[y(i(k)),y(j(k))],'linewidth',1.2)
end
deg=sum(A) %计算邻接矩阵的列和,即各节点的度
ave_degree=sum(deg)/N %计算平均度
figure, bar([1:N],deg); %画各节点度的直方图
title('网络图各节点度大小');
xlabel('$v_{i}$','Interpreter','Latex'), ylabel('$k$','Interpreter','Latex')
degrange=minmax(deg); %求度的取值范围
pinshu=hist(deg,[degrange(1):degrange(2)]);%求度取值的频数
df=pinshu/N; %求度的频率分布
figure, bar([degrange(1):degrange(2)],df,'r') %画度分布图
title('网络图的度分布');
xlabel('$k$','Interpreter','Latex'), ylabel('$P$','Interpreter','Latex')
Matlab_to_Pajek(A) %生成 Pajek 数据, 文件名为 Pajek_data.net

```

为了以后方便使用, 我们编写的求度的频数分布函数如下

```

function dp=mydegree(a);
%返回值 dp 是度的频数分布表, 即它的第一行是度的取值, 第二行是对应度的频数
%输入参数 a 为邻接矩阵
N=length(a); %计算节点的个数
deg=sum(a); %求各节点的度
degrange=minmax(deg); %求度的取值范围
pinshu=hist(deg,[degrange(1):degrange(2)]);%求度取值的频数
ind=find(pinshu==0); %找频数为 0 的地址
dp=[degrange(1):degrange(2); pinshu]; %构造频数分布表
dp(:,ind)=[]; %删除频数为 0 的列
df=dp(2,:)/N; %求度的频率分布
figure, bar(dp(1,:),df,'r') %画度的频率分布图

```

```
title('网络图的度分布');
xlabel('$k$', 'Interpreter', 'Latex'), ylabel('$P_k$', 'Interpreter', 'Latex')
```

5.4 复杂网络的可视化

复杂网络中的顶点数和边数一般较多，使用通常的画图软件如 Matlab 软件等，很难看清网络之间的拓扑结构。需要一些专用的可视化软件来画复杂网络。

Pajek软件是网络分析软件工具的一种，它有出色的大型网络处理能力和强大的可视化功能。Pajek可以处理拥有多达几百万节点的大型网络，突破了很多网络分析软件只能处理较小规模数据的瓶颈。它可以从大规模网络中提取出若干小网络，以便于使用经典算法实现更加细致的研究，并通过强大的可视化功能将网络及分析结果展示出来。

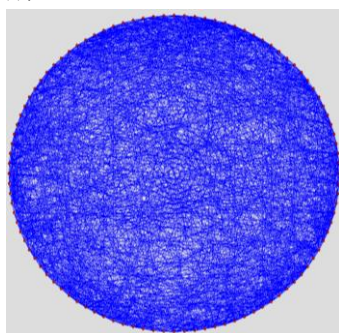
Pajek的输入方式比较灵活，可以直接定义一个小网络，也可以从外面导入数据生成网络，除了本身的数据格式之外，它还支持很多其他软件数据格式的导入。软件的结构是建立在网络、分类、向量、排序、群和层级6种数据结构之上的。

用 Pajek 软件画网络图，必须先构造适合 Pajek 的数据，构造 Pajek 数据的 Matlab 函数如下

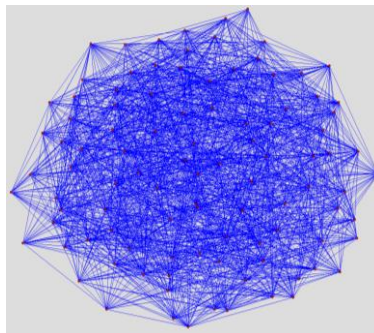
```
function Matlab_to_Pajek(A,k)
%Matlab 邻接矩阵 A 转换成 Pajek 数据的函数
%k 是第 k 次转换，生成的文件命名成 Pajek_datak.net,
%如果不输入第 2 个参数 k,默认文件名为 Pajek_data1.net
if nargin==1
    str='Pajek_data1.net';
else
    str=['Pajek_data',int2str(k),'.net'];
end
n=length(A); v=1:n;
fid=fopen(str,'w'); %创建纯文本文件 Pajek_data.net
fprintf(fid,'%s%d\n','*Vertices ',n); %写入字符串并换行
for i=1:n
    fprintf(fid,'%d ',v(i)); %写入节点编号
    fprintf(fid,'" %d" \n',v(i)); %写入双引号节点字符串并换行
end
fprintf(fid,'%s\n%s\n','*Arcs','*Edges'); %写入两个字符串并各自换行
A=tril(A); %截取邻接矩阵的下三角元素
[u,v]=find(A); n=length(u); %找非零元素，并计算个数
for i=1:n
    fprintf(fid,' %d %d 1\n',u(i),v(i)); %逐条边写入信息并换行
end
fclose(fid);
```

使用 Pajek 软件，调入上述函数生成的数据文件 Pajek_datak.net，就可以画出网络图。

注：必须使用较高版本的 Pajek 软件，否则的话，Matlab 构造的纯文本文件 Pajek 软件不识别。



(a) 未调整布局的网络



(b) 调整布局的网络

图5 N=100; K=4; p=0.15 时的 NW 小世界网络

按 NW 小世界模型的构造算法, 得到原始的小世界网络见图 5 (a), 由于节点众多且节点间距离太短, 不便于认识网络中各节点的联系, 采用 Pajek 软件提供的 Kamada-Kawai 算法, 对网络空间进行重新布局, 得到图 5 (b)。对网络重新布局的操作步骤为, 在 Pajek 图形窗口, 依次选择 Layout 菜单中的 Energy→Kamada-Kawai→Separate Components。

6 各种类型复杂网络的构造算法

6.1 Waxman模型及仿真

20 世纪 80 年代 Internet 的网络规模比较小, 处于发展的初级阶段。1988 年, Waxman 提出了随机图产生器, 它较好地再现了 ARPANET 网络。其建模步骤如下

- (1) 在 $m \times n$ 的平面上均匀放置 N_1 个节点。
- (2) 两节点 u 和 v 之间建立边的概率为

$$\Pi(u, v) = \alpha e^{-\frac{d(u, v)}{\beta L}}, \quad (12)$$

其中 $0 < \alpha, \beta \leq 1$, $d(u, v)$ 是节点 u 和 v 之间的欧氏距离, α 为平均连接度, L 是图中相距最远的两节点间的欧氏距离, β 决定了边的平均长度。如果 α 增大, 则边的生成概率 $\Pi(u, v)$ 增大, 使得最后生成的图中边的数目增加; 如果 β 增加, 则使得最后生成的图中较长的边相对于较短的边出现的概率增加。

Waxman 图中相距较远的两节点间存在边的可能性较小, 从而获得连通图的可能性也比较小, 通常用来对网络中的最大连通子图进行研究。仿真的 Waxman 图的直观表示如图 6 所示, 用频数 f_d 表示网络中度为 d 的节点的个数, 其对数分布如图 7 所示。Waxman 产生器是随机图产生器的典型代表, 后来的很多随机图模型都受其启发。

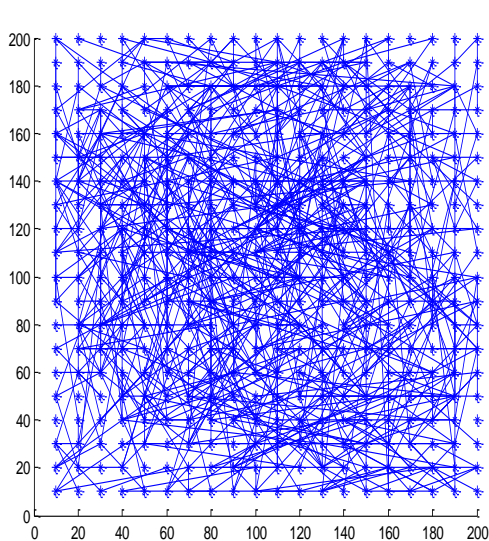


图6 400 个节点的 Waxman 随机图

其中 $\alpha = 0.15$, $\beta = 0.1$

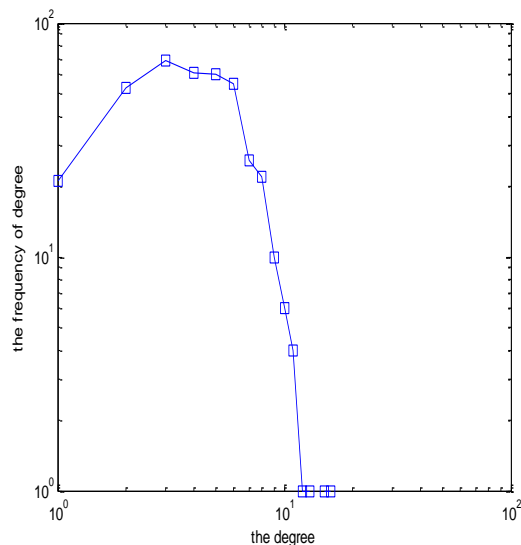


图7 频数 f_d 关于 d 的对数分布图

由于节点众多且节点间距离太短, 使用 Matlab 软件画图不便于认识网络中各节点的联系, 我们使用 Pajek 软件对上述 Waxman 模型构造的网络进行可视化, 并采用 Pajek 软件提供的 Kamada-Kawai 算法, 对网络空间进行重新布局, 得到图 8。从图 8 可以看出网络中存在一些孤立点。

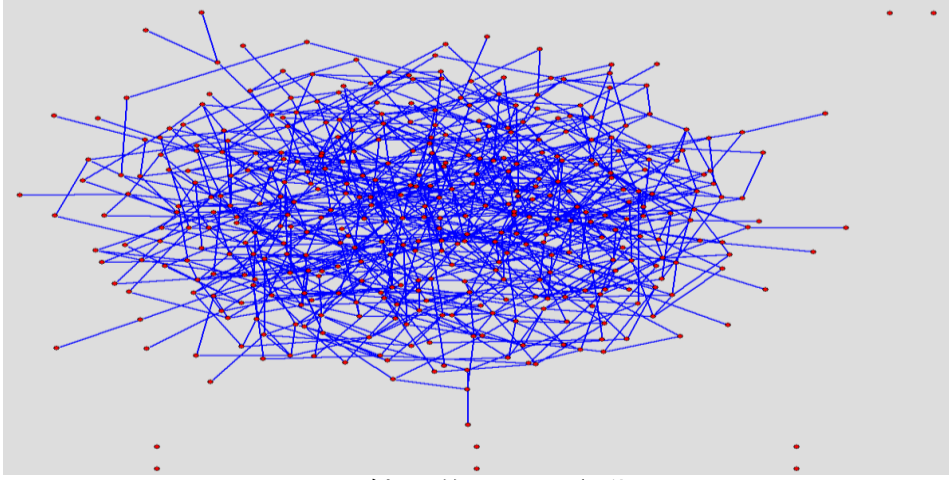


图 8 Waxman 随机图的 Pajek 可视化

Waxman 随机图仿真的 Matlab 程序如下

```
clc, clear
alpha=0.15; beta=0.1;
m=20; n=20; %水平方向和垂直方向的点数
h=10; %水平方向和垂直方向的间隔步长
u=h:h:m*h; v=h:h:n*h; %水平方向
[u,v]=meshgrid(u,v); %生成网格节点的横坐标和纵坐标
u=u(:); v=v(:); %转换成列向量
hold on, plot(u,v,'*') %画 mn 个点
uv=[u';v']; %构造 mn 个点的坐标
d=dist(uv); %求 mn 个点之间的两两距离矩阵
L=max(max(d)); %求最大距离
gailv=alpha*exp(-d/(beta*L)); %计算概率矩阵
gailv=tril(gailv); %提取概率矩阵的下三角元素
gailv(1:m*n+1:end)=0; %概率矩阵的对角线元素置 0
randnum=rand(m*n); %生成服从[0,1]上均匀分布随机数的 mn×mn 矩阵
[i,j]=find((randnum<=gailv)); %节点 i 和 j 之间连边
a=zeros(m*n); %邻接矩阵初始化
for k=1:length(i)
    a(i(k),j(k))=1; a(j(k),i(k))=1; %构造邻接矩阵
    plot([u(i(k)),u(j(k))],[v(i(k)),v(j(k))]); %节点 i 和 j 之间画线
end
deg=sum(a); %计算所有节点的度
dminmax=minmax(deg) %求度的最小值和最大值
pinshu=hist(deg,[dminmax(1):dminmax(2)]); %计算度取值的频数
hold off,figure,loglog([dminmax(1):dminmax(2)],pinshu,'-s') %对数坐标系画图
xlabel('the degree'), ylabel('the frequency of degree')
```

6.2 连通的局域网构造算法

为了构造连通的网络,下面我们利用最小生成树算法构造一个连通的局域网。连通局域网的构造步骤如下

(1) 水平方向取 m 个点,垂直方向取 n 个点,在平面上设置 mn 个网格节点。随机地取一个网格节点,如果所取的网格节点与已经选定的网格节点相距非常近,就将其舍去,这里所取的距离非欧氏距离,而是绝对值距离,即节点 (x_i, y_i) 和节点 (x_j, y_j) 的距离为

$$d_{ij} = |x_i - x_j| + |y_i - y_j|.$$

(2) 当选定的网络节点的个数达到所需的目标个数 N_2 时, 停止选择节点。我们先构造一个赋权完全图 $G=(V,E,W)$, 其中 $V=\{v_1,\dots,v_{N_2}\}$ 为节点集合, $W=(d_{ij})_{N_2\times N_2}$ 为距离权重矩阵。然后在图 G 中利用 Prim 算法构造最小生成树 $T=(T,E_2,W_2)$ 。

(3) 最小生成树 T 就是一个连通的最简单网络。为了保证网络的可靠性, 对于一些重点的关键节点, 要使它的度大于等于某个阈值 k_0 , 若度小于该阈值 k_0 , 则需要添加它与网络中最近节点的边 (节点之间没有重复边)。仿真时, 我们随机地检查 γN_2 ($\gamma=0.15$) 个节点, 若它们不满足阈值条件 k_0 , 则需要添加额外的边。

仿真时参数的取值如下, $m=n=100$, $N_2=100$, $k_0=8$; 仿真图见图 9 (b)。

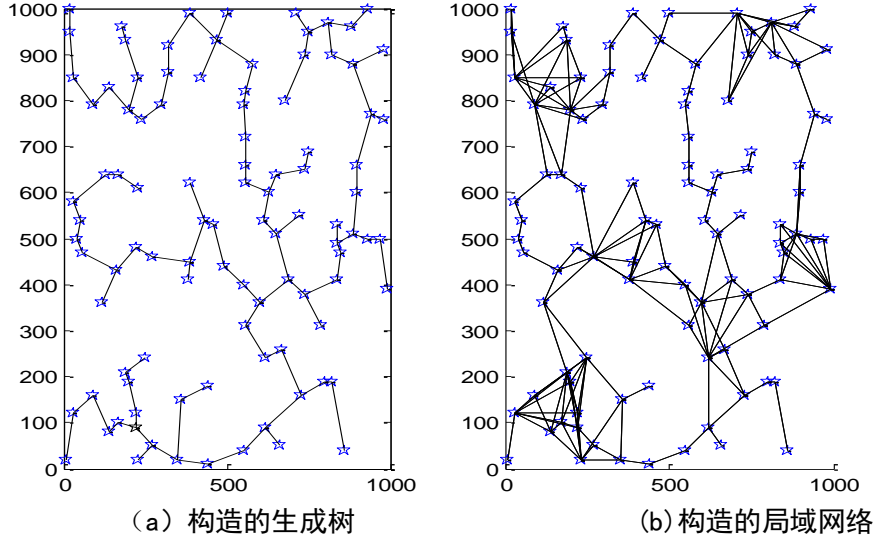


图 9 连通的局域网示意图

6.3 AB 模型

AB 模型的构建过程如下

初始有 m_0 个孤立节点, 每一步执行下面三个步骤中的一个。

(1) 以概率 p 增加 m ($m \leq m_0$) 条新的内部连接, 即在已存在的节点间添加新的边, 随机选取一个节点作为新的边的起始点, 边的另一个端点由以下概率决定

$$\Pi_{ab}(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)}, \quad (13)$$

重复此过程 m 次。

(2) 以概率 q 重新配置 m 条边。随机选取节点 i 和连接到 i 的一条边 l_{ij} , 然后移走此边, 以连接节点 i 和节点 j' 的新边 $l_{ij'}$ 取代。每次根据式 (13) 所示的概率选取 j' 来配置一条边, 并重复此过程 m 次。

(3) 以概率 $1-p-q$ 增加一个新节点。根据 (13) 所示概率分别与网络中已存在的 m 个节点相连接。

其中, $0 < p < 1$, $0 < q < 1-p$, 式 (2) 中采用 $(k_i + 1)$, 保证了孤立节点建立新连接的概率为非零。

AB 模型的度分布是服从广义的幂律分布。

仿真参数 $m_0=m=2$, $p=0.6$, $q=0.1$, 总共执行 $t=2000$ 次。仿真得到的网络图见图 10, 度分布见图 11。

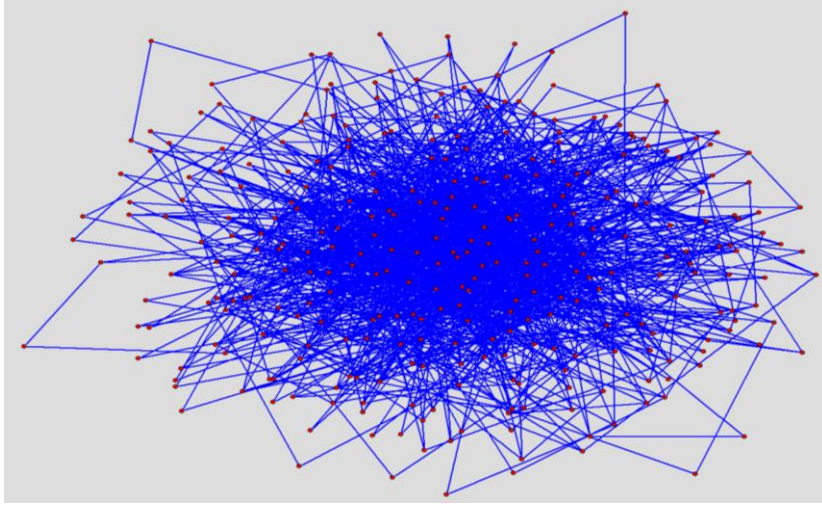


图10 AB模型的仿真图

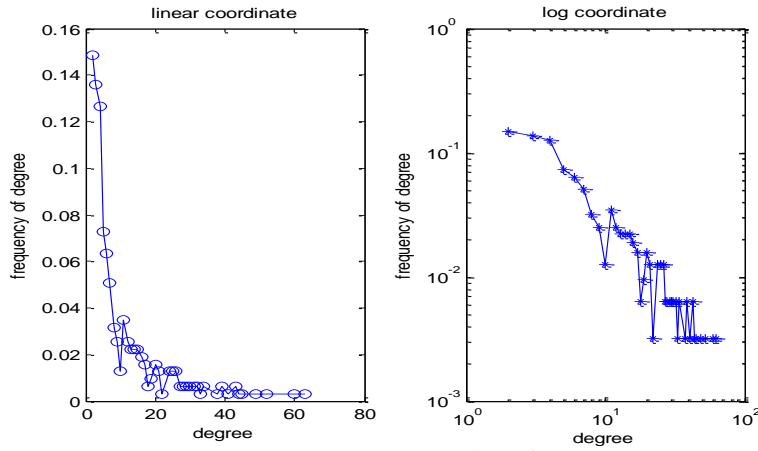


图11 AB模型的度分布

6.4 距离优先和度优先连接的演化网络模型

由于有距离优先连接，我们所构造的网络节点必须有位置关系，为此在 x 轴上取 m 个点，在 y 轴上取 n 个点，然后画网格线，在 xoy 平面上就形成 mn 个网格节点。我们构造的网络节点在网格节点上。网格节点 (x_i, y_i) 和 (x_j, y_j) 之间的距离定义为

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

我们的基于距离优先和度优先连接的演化网络模型的构建过程如下

初始时随机地在 mn 个网格节点中选取 m_0 个节点，接着构造连接这 m_0 个节点的最小生成树。每一步执行下面四个步骤中的一个。

(1) 以概率 p 增加一个新节点。新节点与网络中离它最近的 m_1 ($m \leq m_0$) 个节点相连接。

(2) 以概率 q 增加 m_1 条新的内部连接，即在已存在的节点间添加新的边；随机选取一个节点作为新边的起始点，边的另一个端点由以下概率决定

$$\Pi(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)}, \quad (14)$$

重复此过程 m_1 次。

(3) 以概率 r 重新配置 m_1 条边。随机选取节点 i 和连接到 i 的一条边 l_{ij} ，然后移走此边，以连接节点 i 和节点 j' 的新边 $l_{ij'}$ 取代。每次根据式 (14) 所示的概率选取 j' 来配置一

条边，并重复此过程 m_1 次。

(4) 以概率 s 去掉 m_2 ($m_2 < m_1$) 条边，这是刻画边的消亡。首先随机地选取边的一端，另一端由如下概率选定

$$\Pi'(k_i) = \frac{1}{N-1}(1-\Pi(k_i)), \quad (15)$$

其中 N 为网络内的节点总数。重复此过程 m_2 次。为了保证网络的连通性，删除过程中不删除网络的割边。

上述参数满足 $0 < p < 1$, $0 \leq q, r, s < 1$, $p + q + r + s = 1$ 。

仿真时参数 $m_0 = 3$, $p = 0.4$, $m_1 = 2$, $m_2 = 1$, $q = 0.3$, $r = 0.25$, $s = 0.05$, 总共执行 $t = 2000$ 次，仿真得到的网络图见图 12。网络的度分布见图 13。

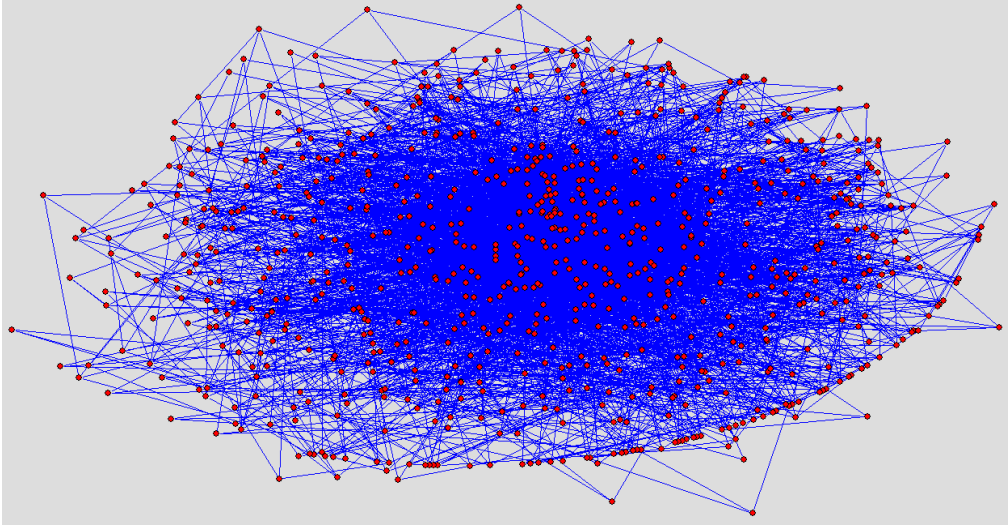


图 12 仿真网络图

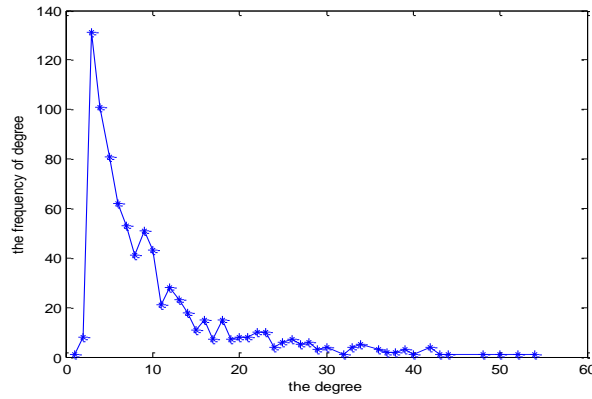


图 13 仿真网络的度分布图

6.5 多局域世界演化模型

6.5.1 多局域世界演化模型的基本算法

多局域世界演化模型的基本思想是：改进 BA 模型中的增加机制，细分为局域世界个数的增长，局域世界内节点的增长、边的增长、边的消亡，局域世界之间远程边的增长；利用局域世界优先连接机制代替 BA 模型中的全局优先连接机制；增加边的消亡机制。多局域世界演化模型的构造算法如下：

起始为 m 个孤立的局域世界，在每个局域世界内部均有 m_0 个节点、 e_0 条边。每一步随机地进行如下五个操作中的一个。

(1) 以概率 p 增加一个拥有 m_0 个节点、 e_0 条边的局域世界。

(2) 以概率 q 将一个新节点加入到一个已存在的局域世界中，它与同一个局域世界中的节点建立 m_1 条边。

首先随机地选取一个局域世界 Ω ，此局域世界中，新节点将要连接的节点由如下概率选取

$$\Pi(k_i) = \frac{k_i + \alpha}{\sum_{j \in \Omega} (k_j + \alpha)}, \quad (16)$$

其中参数 $\alpha > 0$ 表示节点 i 的吸引力，是为了保证孤立点能够以一定的概率得到新的连边。重复此过程 m_1 次。

(3) 以概率 r 增加 m_2 条边到一个选定的局域世界。

首先随机地选取一个局域世界，边的一端随机选取，另一端根据概率式 (16) 选定，重复此过程 m_2 次。

(4) 以概率 s 在一个选定的局域世界内去掉 m_3 条边，这是刻画边的消亡。

首先随机地选取一个局域世界 Ω ，边的一端随机选择，另一端由如下概率选定

$$\Pi'(k_i) = \frac{1}{N_\Omega - 1} (1 - \Pi(k_i)), \quad (17)$$

其中 N_Ω 为局域世界 Ω 内的节点个数。重复此过程 m_3 次。

(5) 以概率 u 在一个选定的局域世界与其它已存在的局域世界间建立 m_4 条长程边。

首先随机选取一个局域世界，在其内部根据概率式 (16) 选定一个节点，作为边的一端，边的另一端位于另一个随机选取的局域世界内，在其中以概率式 (16) 选取节点，重复此过程 m_4 次。

上述参数满足 $0 < q < 1$ ， $0 \leq p, r, s, u < 1$ ， $p + q + r + s + u = 1$ 。

仿真参数取值为 $m = 5$ ， $m_0 = 8$ ， $e_0 = 15$ ， $p = 0.3$ ， $q = 0.2$ ， $r = 0.25$ ， $s = 0.05$ ， $u = 0.2$ ， $m_1 = 3$ ， $m_2 = 5$ ， $m_3 = 4$ ， $m_4 = 6$ ， $\alpha = 0.5$ 时，得到的仿真网络图见图 14，网络中有很多孤立的局域世界。仿真网络图的度分布见图 15，从图 15 观察，它的度分布不服从幂律分布或广义幂律分布，度分布似乎服从 Poisson 分布，通过非参数 χ^2 拟合检验法，推断出它的度也不服从 Poisson 分布，即基本算法构造的网络不是一个随机网络。

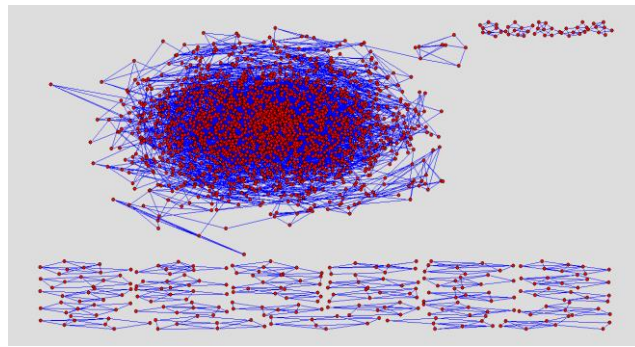


图 14 基本算法的仿真网络图

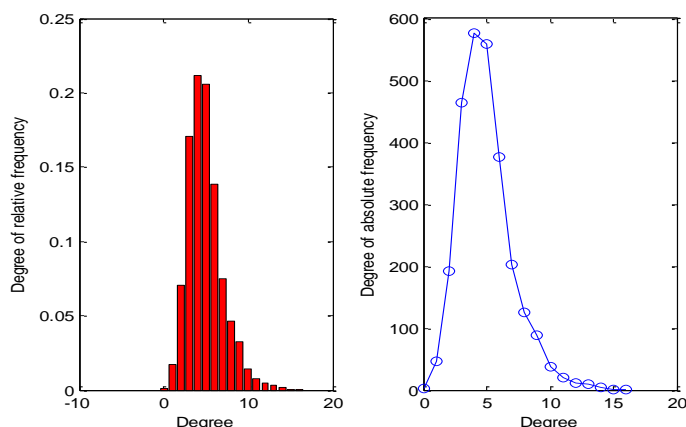


图 15 基本算法的仿真网络图的度分布

6.5.2 改进的多局域世界模型

多局域世界演化模型引进边的消亡机制和局域世界之间远程边的增长机制,使网络结构的演化更加符合现实复杂网络,但可能产生以下两个问题

- (1) 边的消亡机制可能导致某一局域世界内的节点成为孤立节点。
- (2) 局域世界之间远程边的连接具有随机性,可能会导致一些局域世界与其它局域世界无连接,成为孤立局域世界。

为了提高网络连接的冗余度,我们试着把度小于 2 的节点按照概率式(4)和整个网络中的某个节点之间加一条边,仿真得到的网络仍然不连通。为了提高构造连通网络的速度,本文对多局域世界模型的基本算法做出以下改进。

在上面基本算法构造的不连通网络图基础上。

(1) 对于网络中的孤立节点和叶子节点,按照概率式(4)在整个网络中随机地选择一点,作为加边的另外一个端点。直到整个网络中不存在孤立节点和叶子节点为止。

(2) 对所有的孤立局域世界,逐个相连并首尾相连,即各个连通分支构成一个环形连通网络。所加连边的两个端点在各自局域世界内按照概率式(16)随机选取。

改进的多局域世界模型的仿真图见图 16, 它的度分布见图 17。

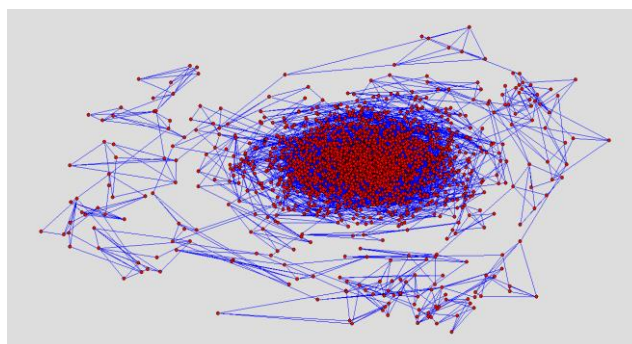


图 16 改进算法的仿真网络图

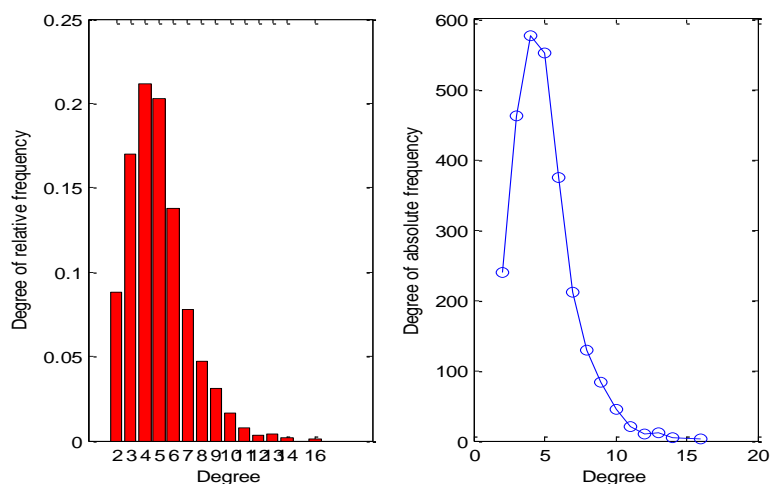


图 17 改进算法的仿真网络图的度分布

7 结语与展望

关于网络的研究，数学家早在两百多年前就开始了，他们已经发展出了成体系的理论与技术。数学家经典的网络理论，要么是分析包含几十数百个顶点，可以画在一张纸上从而形成直观印象的网络；要么是讨论不含有限尺度效应，可以精确求解的网络性质。“随机移走一个顶点会对网络的性能产生什么样的影响？”这个问题对于研究有限规则网络的数学家是有意义的，但对于拥有几千万个节点，连接方式复杂多样的真实网络而言，或许“随机移走3%的顶点会对网络性能产生什么样的影响？”这个问题更有意义。

复杂网络理论的研究目前已经发展成物理学、控制科学、生物学、管理学等几乎所有学科的前沿热点问题，复杂网络抗毁性研究也日益引起更多的注意力，其研究成果（尤其是在无标度网络上）已经在很大程度上改变和拓展了我们对网络抗毁性的认识。由于复杂网络不仅仅是一种常见的复杂系统的结构形态，它还可以作为复杂系统结构拓扑特性的模型，因而复杂网络抗毁性研究必将对复杂系统的抗毁性研究产生重大的推动作用，成为系统科学的一个重要研究内容。