

# 第五章 网络模型

## 5.1 随机图

## 5.2 任意度分布的随机图

### 5.2.1 生成函数

### 5.2.2 配置模型

## 5.3 小世界模型

## 5.4 网络生成模型

### 5.4.1 Price模型

### 5.4.2 Barabási-Albert模型

## 5.5 网络渗流过程



## 5.1 随机图

随机图是某些参数确定，其他参数随机取值的网络模型。模型表示为 $G(n, p)$ ，即点数 $n$ 是确定的，连边的概率 $p$ 也是确定的，但与谁相连以及连边的个数都是不确定的。故 $G(n, p)$ 模型是一系列图的集合，恰好有 $m$ 条边的某个图的出现概率为 $P(G) = p^m(1 - p)^{\binom{n}{2} - m}$ ，非简单图出现的概率为零。

### 边数和度的均值

有 $m$ 条边的图的总概率 $P(m) = \binom{\binom{n}{2}}{m} p^m(1 - p)^{\binom{n}{2} - m}$ ，故可得：

$$\text{边的均值 } \bar{m} = \sum_{m=0}^{\binom{n}{2}} m P(m) = \binom{n}{2} p;$$

$$\text{度的均值 } \bar{k} = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n} \binom{n}{2} p = (n - 1)p = c。$$

## 度分布

某点恰好与 $k$ 个点相连的总概率  $p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$ ,

当 $n \rightarrow \infty$ 时, 有  $p_k = \frac{(n-1)^k}{k!} p^k e^{-c} = \frac{(n-1)^k}{k!} \left(\frac{c}{n-1}\right)^k e^{-c} = e^{-c} \frac{c^k}{k!}$ 。

## 聚类系数

任意点的邻居点互连(也是邻居)的概率就是聚类系数  $C = p$ 。

这也是随机图模型与许多实际网络有较大不同之处。

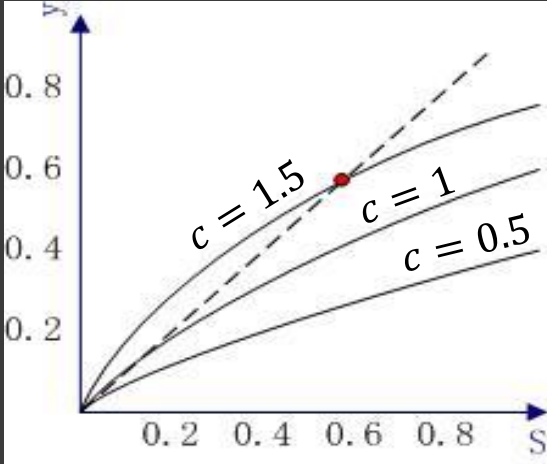
## 巨分支

$p = 0$ , 图为 $n$ 个孤立点;  $p = 1$ , 图为由 $n$ 个点组成的最大分支。

随着 $p$ 由 $0 \rightarrow 1$ , 图也会在某个 $p$ 值处发生突变或称变相。

若网络分支的规模与 $n$ 成比例增长, 则称该网络分支为**巨分支**。

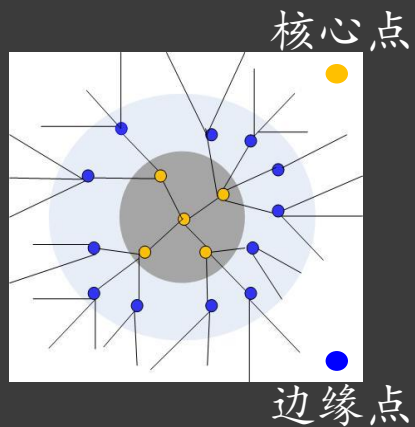
巨分支节点所占比例为  $S = 1 - e^{-cS}$ , 可推导得  $c = -\frac{\ln(1-S)}{S}$ 。



$\frac{d}{dS}(1 - e^{-cS}) = 1$ , 若  $c = 1$ , 则  $S = 0$ 。

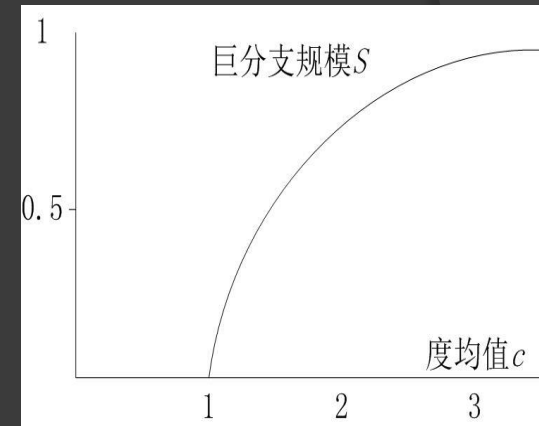
而  $c \leq 1$  时, 随机图中不存在巨分支, 图由若干规模固定的分支和孤立的点组成;

$c > 1$  时, 随机图有且仅有一个巨分支。



设一个集合有  $s$  个点, 则其一个边缘点与集合外部点的连接个数平均值为:

$$p(n-s) = (n-s) \frac{c}{n-1} \approx c。$$



设以  $p = c/(n-1)$  的概率生成一个随机图, 以  $p' = c/(n-1)^{3/2}$  继续在没连边的点间添加一些额外的边。当  $n$  充分大时, 网络中连边的概率为  $p + p'$ , 度均值  $c' = (n-1)(p + p') = c \left(1 + \frac{1}{\sqrt{n-1}}\right) \approx c$ 。

可见, 用  $p'$  随机添加的额外边对网络特性无本质影响。

假设网络存在节点比例为 $S_1$ 和 $S_2$ 的两个巨分支，巨分支间可构成的节点对数量为 $S_1 S_2 n^2$ 。以 $p'$ 添加额外的边，且额外边仍使两巨片不连通的概率为 $q = (1 - p')^{S_1 S_2 n^2} \approx e^{-c S_1 S_2 \sqrt{n}}$ ，当 $n \rightarrow \infty$ 时， $q \rightarrow 0$ 。所示，网络中不可能存在两个巨片。

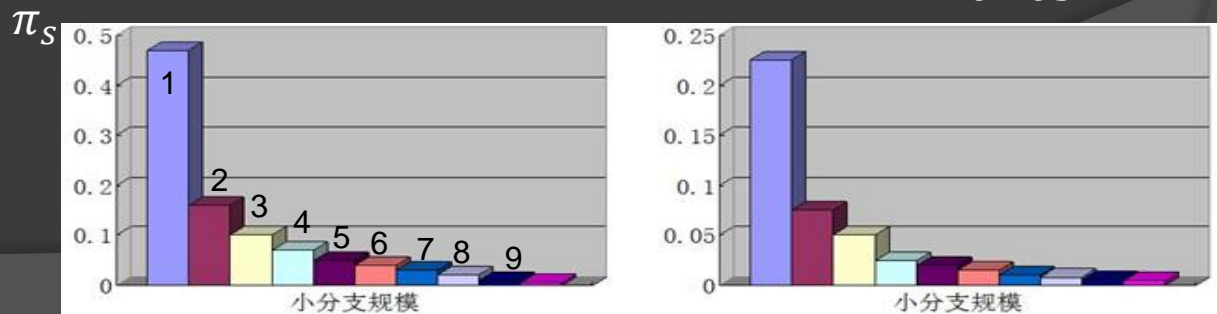
## 小分支

分支规模不与网络规模 $n$ 对应成比例增长的一类分支。设 $\pi_s$ 为一个点属于规模为 $s$ 的小分支的概率，则有 $\sum_{s=0}^{\infty} \pi_s = 1 - S$ 。

- 随机选取一个点，可能所属的小分支的平均规模 $\bar{s} = \frac{1}{1-c+CS}$ 。
- 随机图模型中所有小分支规模的平均值为 $\langle s \rangle = \frac{2}{2-c+CS}$ 。

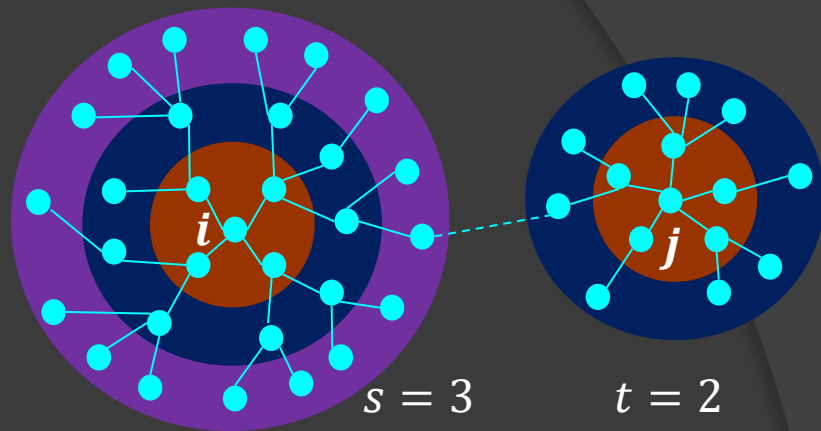
小分支规模分布：

$$\pi_s = \frac{e^{-cS} (cS)^{s-1}}{s!}。$$



## 路径长度

随机图中任选一点，其距离 $d$ 内的平均点的个数为 $c^d$ 。右图共有 $c^{s+t}$ 对点，设 $l = s + t + 1$ ，有



$P(d_{ij} > l) = (1 - p)^{c^{l-1}} = \left(1 - \frac{c}{n}\right)^{c^{l-1}}$ 。因 $\ln \left(1 - \frac{c}{n}\right)^{c^{l-1}} \approx -\frac{c^l}{n}$ ，  
 $P(d_{ij} > l) = \exp\left(-\frac{c^l}{n}\right)$ 。故网络直径为使得 $e^{-c^l/n} = 0$ 的 $l_{min}$ 。

$l_{min}$ 应满足 $c^{l_{min}} = an^{1+\epsilon}$ ，其中 $a$ 为常数， $\epsilon$ 从正方向趋于0，

故有 $l_{min} = \frac{\ln a}{\ln c} + \lim_{\epsilon \rightarrow 0} \frac{(1+\epsilon) \ln n}{\ln c} = A + \frac{\ln n}{\ln c}$ ，随 $n$ 缓慢增加。

## 随机图模型与实际网络的偏差：

- 1) 度的泊松分布无法解释很多实际现象；
- 2) 随网络规模增加，聚类系数趋于零；
- 3) 随机图模型中不存在“社团”结构。

## 5.2 任意度分布的随机图

### 1、概率生成函数(z变换)

点度数= $k$ 的概率为 $p_k$ ,  $g(z) = p_0 + p_1z + p_2z^2 + \cdots = \sum_{k=0}^{\infty} p_k z^k$

称为概率分布 $p_k$ 的生成函数, 求微分 $\frac{1}{k!} \frac{d^k g}{dz^k} \big|_{z=0}$ 可获得 $p_k$ 的值。

### 幂律分布的生成函数

一般幂律分布 $p_k = Ck^{-\alpha}$ , 因 $\sum_k p_k = 1$ , 归一化 $C = 1 / \sum_{k=1}^{\infty} k^{-\alpha}$ ,

$$p_k = \begin{cases} 0, & k = 0 \\ k^{-\alpha} / \sum_{k=1}^{\infty} k^{-\alpha}, & k \geq 1 \end{cases}, \text{ 生成函数 } g(z) = \frac{1}{\sum_{k=1}^{\infty} k^{-\alpha}} \sum_{k=1}^{\infty} k^{-\alpha} z^k$$

- 因为 $Li_{\alpha}(z) = \sum_{k=1}^{\infty} k^{-\alpha} z^k$ 和 $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ 的许多性质已知, 故可进行许多生成函数运算。生成函数的导数非常重要, 其中

$$\frac{\partial Li_{\alpha}(z)}{\partial z} = \frac{\partial}{\partial z} \sum_{k=1}^{\infty} k^{-\alpha} z^k = \sum_{k=1}^{\infty} k^{-(\alpha-1)} z^{k-1} = \frac{\partial Li_{\alpha}(z-1)}{z}。$$

注意: 现实网络的度分布通常在 $k > k_{min}$ 值时服从幂律分布。



## 归一化和矩

- 生成函数中令 $z = 1$ ，则有生成函数 $g(1) = 1 = \sum_{k=0}^{\infty} p_k$ 。
- $g(z)$ 的导数 $g'(z) = \sum_{k=0}^{\infty} k p_k z^{k-1}$ ，有 $g'(1) = \sum_{k=0}^{\infty} k p_k = \bar{k}$ 。  
类似地，度的 $m$ 阶矩， $\overline{k^m} = \left[ \left( z \frac{d}{dz} \right)^m g(z) \right]_{z=1} = \frac{d^m g}{d(\ln z)^m} \Big|_{z=1}$ 。

## 生成函数的幂

- 已知概率分布 $p_k$ 的生成函数为 $g(z)$ ，设整数 $k_i (i = 1, \dots, m)$ 是从分布中随机取出的 $m$ 个相互独立的随机数，那么，这 $m$ 个整数之和 $\sum_{i=1}^m k_i$ 的概率分布的生成函数为 $[g(z)]^m$ 。

□ 因 $m$ 个 $p_k$ 独立选取，整数组集合 $\{k_i\}$ 出现的概率为 $\prod_i p_{k_i}$ ； $\{k_i\}$ 所有元素之和等于 $s$ 的概率分布 $\pi_s = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \delta(s, \sum_i k_i) \prod_{i=1}^m p_{k_i}$ ， $\delta$ 为克罗内克函数；故 $\pi_s$ 的生成函数 $h(z) = \sum_{s=0}^{\infty} \pi_s z^s = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} z^{\sum_i k_i} \prod_{i=1}^m p_{k_i} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \prod_{i=1}^m p_{k_i} z^{k_i} = [\sum_{k=0}^{\infty} p_k z^k]^m = [g(z)]^m$ 。



## 2、配置模型

配置模型是一种给定度序列和边数的广义随机图  $G(n, m)$  模型。

- 度数为  $k_i$  和  $k_j$  的任意两点之间的**连边概率**:  $p_{ij} = \frac{k_i k_j}{2m}$ 。
- 当  $n \rightarrow \infty$ , 重边和自边的密度以  $1/n$  的速率趋于0。
- 点  $v_i$  和  $v_j$  共有同一个邻居的概率:  $n_{ij} = p_{ij} \frac{\overline{k^2} - \bar{k}}{\bar{k}}$ 。

**余度分布**: 随机沿一条边到达某点, 除到达边之外的其他边的个数为该点的余度, 其概率分布  $q_k = \frac{(k+1)p_{k+1}}{\bar{k}}$ , 且  $\sum_{k=0}^{\infty} q_k = 1$ 。

**聚类系数**: 某点的邻居  $v_i$  和  $v_j$  存在边  $e_{ij}$  的概率为  $k_i k_j / 2m$ , 配置模型的聚类系数为  $C = \sum_{k_i, k_j=0}^{\infty} q_{k_i} q_{k_j} \frac{k_i k_j}{2m} = \frac{1}{n} \frac{[\overline{k^2} - \bar{k}]^2}{\bar{k}^3}$ 。

**度和余度的生成函数**:  $g_0(z) = \sum_{k=0}^{\infty} p_k z^k$ ;  $g_1(z) = \sum_{k=0}^{\infty} q_k z^k$ 。

幂律分布有,  $g_0(z) = \frac{\sum_{k=1}^{\infty} k^{-\alpha} z^k}{\sum_{k=1}^{\infty} k^{-\alpha}}$ ;  $g_1(z) = \frac{\sum_{k=1}^{\infty} k^{-(\alpha-1)} z^{k-1}}{z \sum_{k=1}^{\infty} k^{-(\alpha-1)}}$ 。

**一个点的 $d$ 跳“邻居”数量：**一个点拥有 $k$ 个两跳邻居的概率

$p_k^{(2)} = \sum_{m=0}^{\infty} p_m p^{(2)}(k|m)$ ,  $m$ 为1跳邻居数,  $p^{(2)}$ 的生成函数

$g^{(2)}(z) = g_0(g_1(z))$ , 且可推导出 $g^{(d)}(z) = g^{(d-1)}(g_1(z))$ 。

两跳邻居的均值 $c_2 = \overline{k^2} - \bar{k}$ ;  $d$ 跳邻居的均值 $c_d = c_1 \left(\frac{c_2}{c_1}\right)^{d-1}$ 。

配置模型有巨分支的充要条件:  $c_2 > c_1$  即  $\overline{k^2} - 2\bar{k} > 0$ 。

**度幂律分布：**设度为“纯”幂律分布，

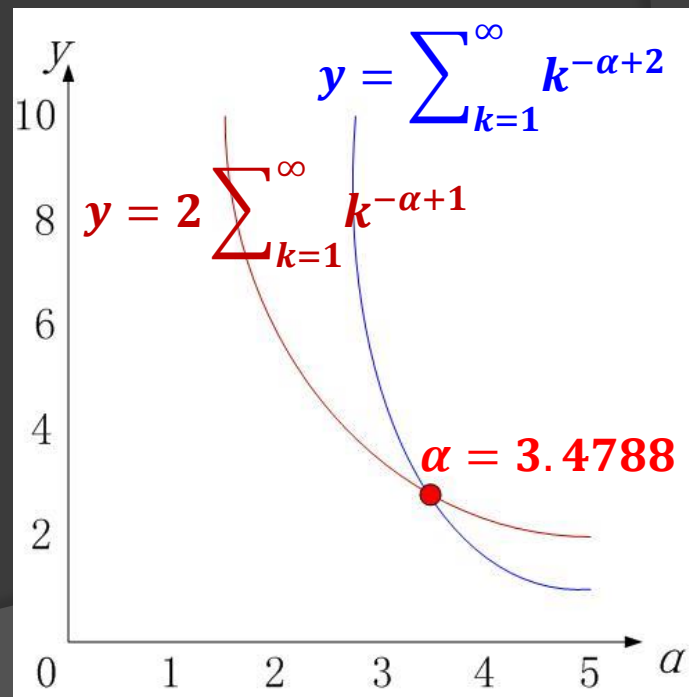
有 $\bar{k} = \frac{\sum_{k=1}^{\infty} k^{-\alpha+1}}{\sum_{k=1}^{\infty} k^{-\alpha}}$ ,  $\overline{k^2} = \frac{\sum_{k=1}^{\infty} k^{-\alpha+2}}{\sum_{k=1}^{\infty} k^{-\alpha}}$ 。

$\alpha > 3.4788$ , 网络只由小分支组成;

$2 \leq \alpha \leq 3.4788$ , 存在一个巨分支和

若干小分支;  $\alpha < 2$ , 存在巨分支且

可认为是整个网络。



## 5.3 小世界模型

社会网络的“小世界”实验产生了“六度分离”问题。

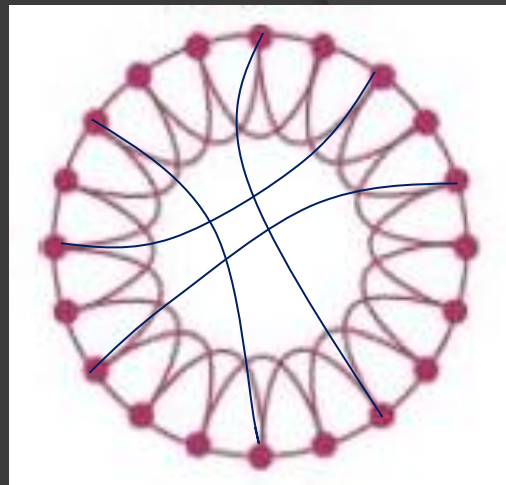
### 传递性与短路径

规则网络可具有高传递(聚类)性，如右图示例，

聚类系数  $C = \frac{3(\bar{k}-2)}{4(\bar{k}-1)}$ ，但平均路径较长  $\bar{L} = \frac{n}{2\bar{k}}$ 。

随机网络具有较短路径，平均长度  $\bar{L} \sim \ln n / \ln \bar{k}$ ，

缓慢增长，但聚类系数很低， $C = p = \frac{\bar{k}}{n-1}$ 。



**小世界模型：**现实网络既不是严格的规则网络，也不是纯粹的随机网络，而是两者特性结合的网络系统。

**WS构造方法：**基于最近邻耦合图，以概率 $p$ 重新连接原有的边。

**NW构造方法：**基于最近邻耦合图，以概率 $p$ 随机添加新的边。

**度分布：**在度数为 $c$ 的最近邻耦合网络中，每个点以概率 $p$ 增加一条连接到随机位置的捷径，则任意一点的度分布在 $k > c$ 时为

$$p_k = e^{-cp} \frac{(cp)^{k-c}}{(k-c)!}; \quad k < c \text{ 时}, p_k = 0。$$

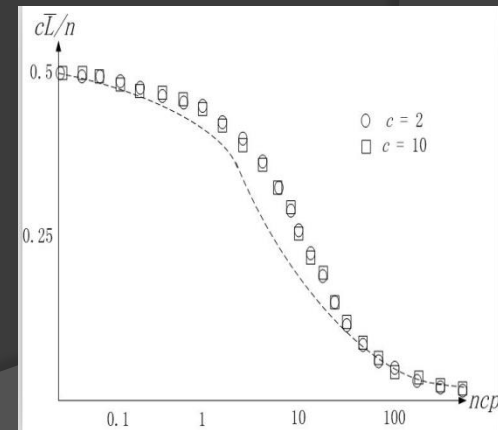
与实际网络不符

**聚类系数：**当 $n \rightarrow \infty$ 时，可忽略随机远程捷径所构成的三角形。

可能由捷径与原有的边构成的三元组共有 $nc^2p$ ，可能由2条捷径构成的三元组共有 $\frac{1}{2}nc^2p^2$ 。故聚类系数 $C = \frac{3(c-2)}{4(c-1)+8cp+4cp^2}$ 。

**平均路径长度：**设 $f(x)$ 为普适函数，则平均路长 $\bar{L} = \frac{n}{c} f(ncp)$ 。令 $f(x) = \frac{2}{\sqrt{x^2+4x}} \operatorname{artanh} \sqrt{\frac{x}{x+4}}$ ，则在 $ncp \gg 1$ 时，有 $\bar{L} = \frac{\ln(ncp)}{c^2p}$ 。

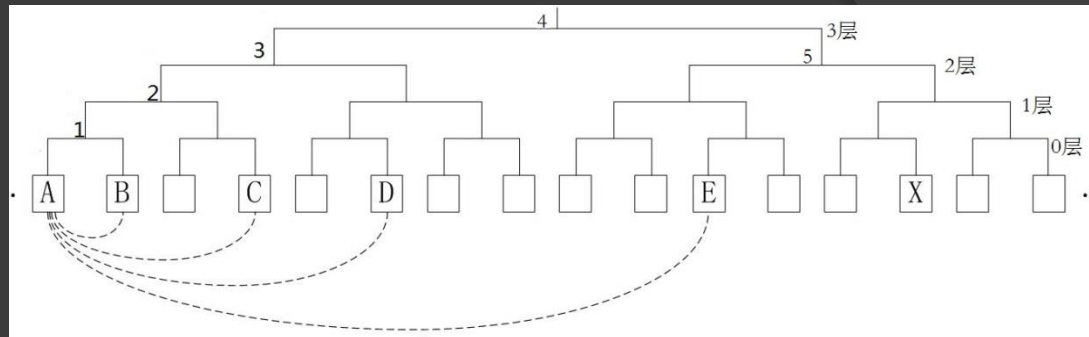
- 小世界模型同时具有传递性和短路径效应，且二者并不相互排斥。



标度函数 $f(x)$



**层次模型：**将社会结构中影响相互关联的关系用树形图表示，一直划



分到每个单元中所有元素都直接关联(相互认识)为止。设群组规模为 $g$ ，群组数量为 $n/g$ ，二叉树状图层次为 $\log_2(n/g)$ 。

设在 $m$ 层同属一个祖先的两点连边的概率为 $p_m = C2^{-\beta m}$ ，则任意一点与其 $m$ 层祖先分枝下的所有其它点的连边数量的期望值为：

$2^m g p_m = C g 2^{(1-\beta)m}$ 。在 $m$ 层祖先分枝下搜索到最佳捷径所需的最大步长为 $1/C g 2^{(1-\beta)m}$ ；搜索到去往目标的路径所需的最大步长为：

$$l_{max} = \begin{cases} D(n/g)^{1-\beta}, & \beta < 1 \\ E(\log n/g)^2, & \beta = 1 \\ F(n/g)^{\beta-1}, & \beta > 1 \end{cases}。$$

$D, E, F$ 为常数。

- 与Kleinberg模型有相似的最大消息传递路径计算模式。



## 5.4 网络生成模型

网络的生成机制模型是通过研究假设的生成机制来观察所生成的网络的结构特性。

### 1、Price模型

核心思想：新发表论文随机引用已发表论文的概率正比于先前已发表论文的被引用次数，且赋予新论文相同的非零初始值。

网络构造算法：

- 增长：从 $n$ 个孤立点的网络开始，每次引入一个新节点并连接 $k(\bar{k} = c)$ 条有向边指向已有的 $k$ 个旧节点； $n = n + 1$ 。
- 累积优势：新节点 $v_j$ 以概率 $p_i$ 创建指向旧节点 $v_i$ 的边 $e_{ji}$ 。设点 $v_i$ 的入度为 $q_i$ ，选择因子 $a$ 为正常数，有 $p_i = \frac{q_i + a}{n(c + a)}$ 。



## 度分布演化方程

令 $p_q(n)$ 为网络中入度为 $q$ 的点所占比例，则点总数为 $np_q(n)$ 。

新增的一点(以均值 $c$ )与入度为 $q$ 的点集新增连边的总概率为

$np_q(n)c \frac{q+a}{n(c+a)} = \frac{c(q+a)}{c+a} p_q(n)$ 。度分布随时间变化的主方程为：

$$(n+1)p_q(n+1) = np_q(n) + \frac{c(q-1+a)}{c+a} p_{q-1}(n) - \frac{c(q+a)}{c+a} p_q(n);$$

$$\text{对 } q=0, (n+1)p_0(n+1) = np_0(n) + 1 - \frac{ca}{c+a} p_0(n)。$$

$$\text{令 } n \rightarrow \infty, \text{ 有 } \begin{cases} p_q = \frac{c}{c+a} [(q-1+a)p_{q-1} - (q+a)p_q], & q \geq 1 \\ p_0 = 1 - \frac{ca}{c+a} p_0, & q = 0 \end{cases}$$

重新整理有 $p_0 = \frac{1+a/c}{a+1+a/c}$ ,  $p_q = \frac{q+a-1}{q+a+1+a/c} p_{q-1}$ 。迭代计算，代

入 $p_0$ 求 $p_1$ 、 $p_2$ 、 $\dots$ ，有 $p_q = \frac{(q+a-1)(q+a-2)\cdots a}{(q+a+1+a/c)\cdots(a+2+a/c)} p_0$ 。

已知  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ , 对  $x > 0$  有  $\Gamma(x+1) = x\Gamma(x)$ ,

迭代上式有  $\frac{\Gamma(x+n)}{\Gamma(x)} = (x+n-1)(x+n-2)\cdots x$ ; 代入  $p_q$  算式,

得到新的算式为  $p_q = (1 + \frac{a}{c}) \frac{\Gamma(q+a)\Gamma(a+1+a/c)}{\Gamma(a)\Gamma(q+a+2+a/c)}$ , 分子分母同乘以

$\Gamma(2 + a/c)$ , 则有  $p_q = \frac{\Gamma(q+a)\Gamma(2+a/c)}{\Gamma(q+a+2+a/c)} \times \frac{\Gamma(a+1+a/c)}{\Gamma(a)\Gamma(1+a/c)}$ 。

已知欧拉  $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ , 利用与上式有  $p_q = \frac{B(q+a, 2+a/c)}{B(a, 1+a/c)}$ 。

**Stirling** 公式  $\Gamma(x) \approx \sqrt{2\pi} e^{-x} x^{x-1/2}$ ,  $B(x, y) \approx \frac{e^{-x} x^{x-1/2}}{e^{-(x+y)} (x+y)^{x+y-1/2}} \Gamma(y)$ 。

因  $(x+y)^{x+y-1/2} \approx x^{x+y-1/2} e^y$ ; 当  $x \rightarrow \infty$  时,  $B(x, y) \approx x^{-y} \Gamma(y)$ 。

将  $B(x, y)$  代入  $p_q$  算式,  $q$  足够大时,  $p_q \sim (q+a)^{-(2+\frac{a}{c})}$ 。

当  $q \gg a$  时, 可得  $p_q \sim q^{-\gamma}$ ,  $\gamma = 2 + \frac{a}{c} > 2 (a > 0, c > 0)$ 。

实际的引文网络中  $\gamma \approx 3$ , 只需令  $a \approx c$  即可。

## 时间函数的度分布

令  $p_q(t, n)$  为  $t$  时刻  $k^{in} = q$  的度分布 (节点比例), 则主方程为:

$$(n+1)p_q(t, n+1) = np_q(t, n) + \frac{q+a}{c+a} [(q-1+a)p_{q-1}(t, n) - (q+a)p_q(t, n)].$$

令  $\tau = \frac{t}{n}$ ; 设  $\pi_q(\tau, n)$ , 使得  $\pi_q(\tau, n)d\tau$  为  $\Delta\tau$  内入度  $= q$  的点的概率;

$$\pi_q\left(\frac{n}{n+1}\tau, n+1\right) = \pi_q(\tau, n) + \frac{c}{c+a} \left[ (q-1+a) \frac{\pi_{q-1}(\tau)}{n} - (q+a) \frac{\pi_q(\tau, n)}{n} \right].$$

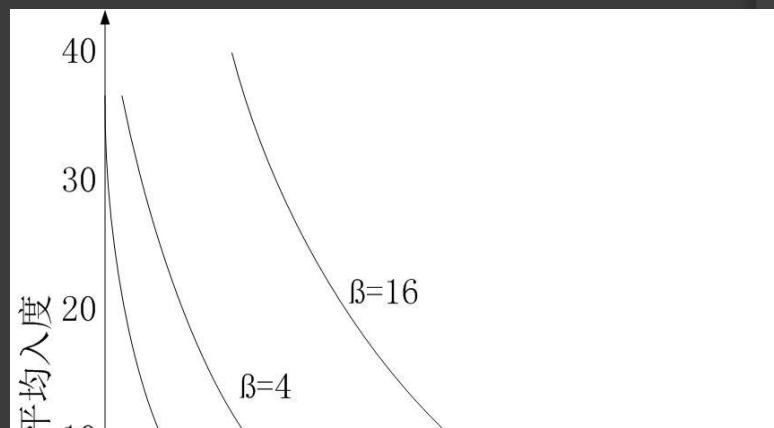
当  $n \rightarrow \infty$ , 可推导出  $\pi_q(\tau) \sim q^{a-1} (1 - \tau^{c/c+a})^q$ 。时间  $\tau$  创建的新点的

的平均入度:  $\bar{q}(\tau) = \sum_{q=0}^{\infty} q \pi_q(\tau)$ ;

$$\bar{q}(\tau) = \beta (\tau^{-c/c+a} - 1) \sim \beta \tau^{-\frac{1}{2}}.$$

在  $t_j > t_i$  两个时间分别添加一点,

$$\text{则有 } \frac{\bar{q}_i\left(\frac{t_i}{n}\right)}{\bar{q}_j\left(\frac{t_j}{n}\right)} = \sqrt{\frac{t_j}{t_i}} > 1.$$



越早加入的点平均度数越高。

归一化时间

设点的创建时间为 $t$ ，随后经过时间为 $s$ ，那么 $n = t + s$ 。有

$$\bar{q}_t(s) = \beta \left[ \left( 1 + \frac{s}{t} \right)^{c/c+a} - 1 \right], \text{ 可得 } \begin{cases} \bar{q}_t(s) \approx \frac{ca}{c+a} \left( \frac{s}{t} \right), & s \ll t; \\ \bar{q}_t(s) \approx \beta \left( \frac{s}{t} \right)^{c/c+a}, & s \gg t. \end{cases}$$

- 再次证明较早加入网络的节点将具有很大的连边优势。

## 内向分支规模

规模= $n$ 的网络中，设 $p_s(n)$ 为内向分支规模= $s$ 的概率。一新点加

入后与规模= $s$ 分支连接的概率为： $\frac{c}{c+a} [(1+a)s - 1] p_s(n)$ 。可得

$$(n+1)p_s(n+1) = np_s(n) + \frac{c+ca}{c+a} \left[ \left( s - 1 - \frac{1}{1+a} \right) p_{s-1}(n) - \left( s - \frac{1}{1+a} \right) p_s(n) \right].$$

当 $n \rightarrow \infty$ ，可以推导有 $p_s \sim s^{-\beta}$ ， $\beta = 1 + \frac{1+a/c}{1+a}$ ， $c \geq 1$ 时， $1 < \beta \leq 2$ 。

- 大多数分支很小且为树形，可能存在个别大型网状分支。

## 2、Barabási-Albert模型

### 网络构造算法:

- 增长: 从有 $m_0$ 个点 $M_0$ 条边的连通网络开始, 每次引入一个新的点并连接到 $m$ 个已存在的点, 这里 $m \leq m_0$

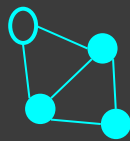
- 优先连接: 一个新点与一个已存在点 $i$ 的连边概率 $= \frac{k_i}{\sum_j k_j}$ 。
- 经过 $t$ 步, 算法产生一个包含 $t + m_0$ 个点和 $mt + M_0$ 条边的网络,

$0 < M_0 \leq \frac{1}{2}m_0(m_0 - 1)$ 。示例中 $m_0 = M_0 = 3$ ,  $m = 2$ 。

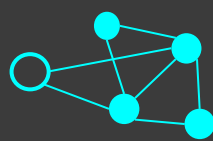
$t = 1$



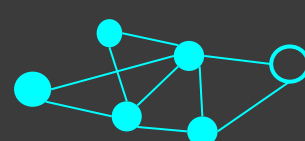
$t = 2$



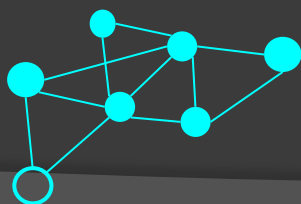
$t = 3$



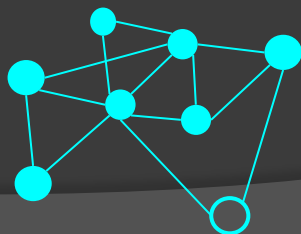
$t = 4$



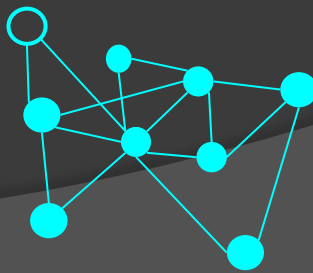
$t = 5$



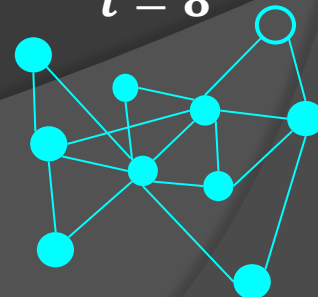
$t = 6$



$t = 7$



$t = 8$



## 度分布

$$a = c = m, \quad p_q = \frac{B(q+m,3)}{B(m,2)}; \quad k = q + m, \quad p_k = \begin{cases} \frac{B(k,3)}{B(m,2)}, & k \geq m \\ 0, & k < m \end{cases}.$$

因有  $B(k,3) = \frac{\Gamma(k)\Gamma(3)}{\Gamma(k+3)} = \frac{\Gamma(3)}{k(k+1)(k+2)}$ ,  $B(m,2) = \frac{\Gamma(2)}{m(m+1)}$ , 对  $k \gg m$ , 有

$$p_k = \frac{\Gamma(3)}{\Gamma(2)} \frac{(m+1)m}{(k+2)(k+1)k} = \frac{2m(m+1)}{k(k+1)(k+2)} \sim 2m^2 k^{-3}, \quad \text{可得 } p_k \sim k^{-3}.$$

这与基于平均场理论的推导结果一致。相比Price模型，BA模型形式简单，因此，关注较多。但是，简化的代价是 $\gamma$ 被限定为3。

另外，模型还需要改进的问题：

- Internet或Web网络中连接关系并不固定，会频繁地出现或者消失，并且，连接建立时间并不一定与节点建立时间同步。
- 没有理由要优先连接过程一定与度数呈线性关系。
- 现实中并不是所有节点建立时都是完全平等的。

**添加额外的边：**每添加一个点和 $m$ 条边的同时，额外地添加 $w$ 条边；每条额外边随机连接2个点，点的连边概率都为 $k/\sum k$ 。

令 $p_k(n)$ 为度= $k$ 的点的比例；新增一点有 $m+2w$ 个点度数发生改变；类似地得概率： $np_k(n)(m+2w)\frac{k}{2n(m+w)} = \frac{m+2w}{2(m+w)}kp_k(n)$ ；

主方程 $p_k = \frac{m+2w}{2(m+w)}[(k-1)p_{k-1} - kp_k]$ ,  $p_m = 1 - \frac{m+2w}{2(m+w)}mp_m$ ；

$p_k = \frac{B(k,\alpha)}{B(m,\alpha-1)}$ ,  $\alpha = 2 + \frac{m}{m+2w}$ ；可推导得  $p_k \sim k^{-\alpha}$ ,  $2 < \alpha \leq 3$ 。

**删除已有的边：**每添加一点的同时，平均随机删除 $v$ 条已有的边 ( $v < m+w$ )，采用反向优先选择(度数越高删除边概率越大)。

添加与删除同时考虑，网络中总边数为 $n(m+w-v)$ ；度为 $k$ 的

点集中： $np_k(n)(m+2w)\frac{k}{2n(m+w-v)} = \frac{m+2w}{2(m+w-v)}kp_k(n)$ 为获得边

的概率； $np_k(n)2v\frac{k}{2n(m+w-v)} = \frac{v}{(m+w-v)}kp_k(n)$ 为失去边的概率。



主方程:  $(n+1)p_k(n+1) = np_k(n) + \delta_{k=m} + \frac{m+2w}{2(m+w-v)}(k-1)p_{k-1}(n)$   
 $+ \frac{v}{m+w-v}(k+1)p_{k+1}(n) - \frac{m+2w+2v}{2(m+w-v)}kp_k(n)$ 。 ( $k=m$ ;  $k \neq m$ 时的合并式)

$p_k$  的生成函数  $g(z) = \sum_{k=0}^{\infty} \delta_{k=m} z^k + \frac{m+2w}{2(m+w-v)} \sum_{k=0}^{\infty} (k-1)p_{k-1} z^k$   
 $+ \frac{2v}{2(m+w-v)} \sum_{k=0}^{\infty} (k+1)p_{k+1} z^k - \frac{m+2w+2v}{2(m+w-v)} \sum_{k=0}^{\infty} kp_k z^k$ 。

• 第一项 =  $z^m$ ; 令  $k = k-1$ , 第二项和式 =  $\sum_{k=0}^{\infty} kp_k z^{k+1} = z^2 \sum_{k=0}^{\infty} kp_k z^{k-1}$ ,

微分形式为  $z^2 \frac{dg}{dz}$ ; 类似地, 第三项和式 =  $\frac{dg}{dz}$ ; 第四项和式 =  $z \frac{dg}{dz}$ 。

重新整理得到  $z^c = g(z) + \frac{(c+2w)z-2v}{2(c+w-v)}(1-z) \frac{dg}{dz}$ 。解微分方程,

$$p_k = Ak^{-\alpha} \int_0^k \frac{(1-x/k)^k}{(1-\gamma x/k)^k} x^{\alpha-2} dx, \quad \alpha = 2 + \frac{v-w}{m+2w-2v}, \quad \gamma = \frac{2v}{m+2w}。$$

$$\left[1 - \frac{\gamma x}{k}\right]^k \rightarrow e^{\gamma x}, \quad p_k \sim k^{-\alpha} \int_0^{\infty} e^{-(1-\gamma)x} x^{\alpha-2} dx = \frac{\Gamma(\alpha-1)}{(1-\gamma)^{\alpha-1}} k^{-\alpha}。$$

控制好删除边的数量 ( $v < \frac{1}{2}m + w$ ), 度分布仍满足幂律规则。

**适应度**：每个点在创建时赋予一个适应度 $\varphi$ ，其分布为 $\rho(\varphi)$ 。

新添加点与旧点之间的连边概率正比于旧点的连接核 $a_k(\varphi)$ ，

而 $a_k(\varphi)$ 由旧点的度数和适应度两个因素共同决定。

在规模 $n$ 的网络中，定义 $p_k(\varphi, n)d\varphi$ ：度为 $k$ 且适应度为 $[\varphi, \varphi + d\varphi]$

的点所占的比例。参照前面的推导可得其主方程： $n \rightarrow \infty$ 时，有

$$(1) \quad k > m, \quad p_k(\varphi) = \frac{m}{\mu} [a_{k-1}(\varphi)p_{k-1}(\varphi) - a_k(\varphi)p_k(\varphi)];$$

$$(2) \quad k = m, \quad p_c(\varphi) = \rho(\varphi) - \frac{ma_c(\varphi)}{\mu} p_c(\varphi), \quad \mu = \frac{1}{n} \sum_{i=1}^n a_{k_i}(\varphi_i)。$$

上两式反复迭代，得  $p_k(\varphi) = \rho(\varphi) \frac{\mu}{ma_k(\varphi)} \prod_{r=m}^k \left[ 1 + \frac{\mu}{ma_r(\varphi)} \right]^{-1}。$

$$\text{令 } a_k = \varphi k, \text{ 有 } p_k(\varphi) = \rho(\varphi) \frac{B(k, 1 + \mu/m\varphi)}{B(m, \mu/m\varphi)} \sim \rho(\varphi) k^{-\alpha}, \quad \alpha = 1 + \frac{\mu}{m\varphi}。$$

**度分布特性取决于 $\varphi$ 的分布，可能符合也可能不符合幂律。**

因每个添加的点都添加 $m$ 条边，网络的平均度数应为 $\bar{k} = 2m$ 。

又 $\bar{k} = \sum_{k=m}^{\infty} k p_k(\varphi) d\varphi = m \int_0^{\varphi_0} \frac{\rho(\varphi) d\varphi}{1 - m\varphi/\mu}$ ， $\varphi_0 < \frac{\mu}{m}$ 才有意义。

积分关于 $\mu$ 单调递减， $\mu = [m\varphi_0, \infty]$ 。该积分无法达到最大，即

$\bar{k} < 2m$ 。这说明网络存在个别度数很高的“凝聚物”点。

设所有凝聚物点度数之和为 $K$ ，则 $\bar{k} = \frac{K}{n} + m \int_0^{\varphi_0} \frac{\rho(\varphi) d\varphi}{1 - m\varphi/\mu} = 2m$ ，

那么 $K = nm \left[ 2 - \int_0^{\varphi_0} \frac{\rho(\varphi) d\varphi}{1 - m\varphi/\mu} \right]$ ，凝聚物规模与网络规模成正比。

**非线性优先连接：**设 $a_k = k^\gamma$ ，则 $p_k = \frac{\mu}{mk^\gamma} \prod_{r=m}^k \left[ 1 + \frac{\mu}{mr^\gamma} \right]^{-1}$ ，

无确定幂律尾部。度分布的幂律形式对优先连接策略敏感。

$\frac{1}{2} < \gamma < 1$ ， $p_k \sim k^{-\gamma} e^{-\frac{\mu k^{1-\gamma}}{m(1-\gamma)}}$ ； $\gamma > 1$ ，出现极个别“领袖”点。

**优先连接生成模型：先到者优、适应者胜、赢者通吃。**

## 平均路径长度

基于时间的度  $k_i(t) = m \sqrt{\frac{t}{t_i}}$ ,  $m \leq k \leq m\sqrt{n}$ ;  $p_k \approx 2m^2 k^{-3}$ ;

$$\bar{k} = \int_m^{m\sqrt{n}} k p_k dk \approx 2m; \quad \overline{k^2} = \int_m^{m\sqrt{n}} k^2 p_k dk = m^2 \ln n [\ln n - 2]。$$

设某个点为第0跳，令常数  $A = m^2 \ln m$ ;  $B = 2A + 2m$ 。

其1跳邻居个数的期望值:  $c_1 = \bar{k} = 2m$ ;

其2跳邻居个数的期望值:  $c_2 = \overline{k^2} - \bar{k} = A \ln n - B$ ;

其 $L$ 跳邻居个数的期望值:  $c_L = \left(\frac{c_2}{c_1}\right)^{L-1} c_1 = 2m \left(\frac{A \ln n - B}{2m}\right)^{L-1}$ ;

有:  $1 + c_1 + \dots + c_L = 1 + 2m + A \ln n - B + \dots + 2m \left(\frac{A \ln n - B}{2m}\right)^{L-1} = n$ 。

解方程，有  $l \sim \frac{\ln(n)}{\ln \ln(n)}$ 。BA网络平均路径长度很短，小于随机

$2 < \gamma < 3, l \sim \ln \ln(n)$ ;

网络和小世界网络的平均路径长度。

$\gamma > 3, l \sim \ln(n)$ 。

## 聚类系数

$t \rightarrow \infty$ 时, 有 $\sum_i k_i \approx 2mt$ ,  $k_i = m \sqrt{\frac{t}{t_i}}$ , 则点 $j$ 与点 $i$ 连接的概率为 $p_{ij} = k_j \frac{k_i}{\sum_l k_l} = \frac{m}{2\sqrt{t_i t_j}}$ 。局部聚类系数 $C_i$ , 网络聚类系数 $C$ , 新点

同时与点 $i$ 及其 $x$ 个邻居连接的微分方程为:  $\frac{dC_i}{dt} = \int_{x=0}^{m-1} \tilde{p}_{ix} \Delta C_{ix}$ ,

其中,  $x \leq m-1$ ;  $C_i(\text{新}) - C_i(\text{旧}) = \Delta C_{ix} = -\frac{2C_i}{k_i+1} + \frac{2x}{k_i(k_i+1)}$ ;

这一事件发生的概率  $\tilde{p}_{ix} = \frac{k_i}{2t} \binom{m-1}{x} \left(\frac{k_i \ln t}{4t}\right)^x \left(1 - \frac{k_i \ln t}{4t}\right)^{m-1-x}$ 。

解方程得:  $C_i = \frac{(m-1)}{8(\sqrt{t} + \sqrt{t_i}/m)^2} \left( (\ln t)^2 + \frac{4m}{(m-1)^2} (\ln t_i)^2 \right)$ ; 因 $C = \sum_i C_i / n$ , 有

$$C = \int_1^t C_i dt_i / t \approx \frac{6m^2[(m+1)^2(\ln t)^2 - 8m \ln t + 8m]}{8(m-1)(6m^2 + 8m + 3)t} \sim \frac{(m-1)(\ln t)^2}{8t}。$$

在 $t$ 很大、 $m \geq 1$ 时, 因 $(\ln t)^2 \ll t$ , 故BA模型的聚类系数按负指数性衰减。这可能是因为存在“结构洞”导致的。

## 5.5 网络渗流过程

把一个网络中的部分点及其连边去除的过程称为“**渗流**”。

概率 $\varphi$ 为占有(未被删除)点比例, 当 $\varphi$ 很大时, 网络存在巨分支。

巨分支的形成或解体称为**渗流转变**, 其临界值为**渗流阈值**。

### 1、均匀随机删除

**配置模型的均匀删除**: 设网络中一个点的一条边不与巨分支相连

的平均概率为 $u$ , 则该点不属于巨分支的概率为 $1 - \varphi + \varphi u^k$ 。因

$$q_k = \frac{(k+1)p_{k+1}}{\bar{k}}, \text{ 有 } u = \sum_{k=0}^{\infty} q_k (1 - \varphi + \varphi u^k) = 1 - \varphi + \varphi g_1(u)。$$

一个点属于巨分支的概率 $S = \varphi [1 - \sum_{k=0}^{\infty} p_k u^k] = \varphi [1 - g_0(u)]。$

渗流阈值出现的条件为 $\left[ \frac{d}{du} (1 - \varphi + \varphi g_1(u)) \right]_{u=1} = 1$ , 求导可得

临界值 $\varphi_c = \frac{1}{g'_1} = \frac{\bar{k}}{k^2 - \bar{k}}。$  **幂律分布中 $\varphi_c \rightarrow 0$ , 鲁棒性很好。**

实际网络规模是有限的，故用“最大分支”代替巨分支概念。

**幂律网络中的最大分支规模：**幂律网络余度分布生成函数如图。

设 $g_1(u)$ 满足 $1 - g_1(u) = c(1 - u)^\beta$ ， $u$ 算式为

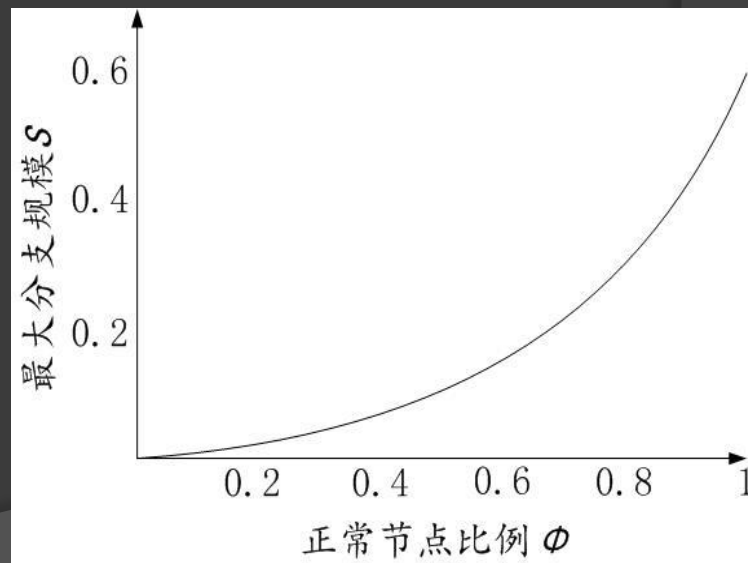
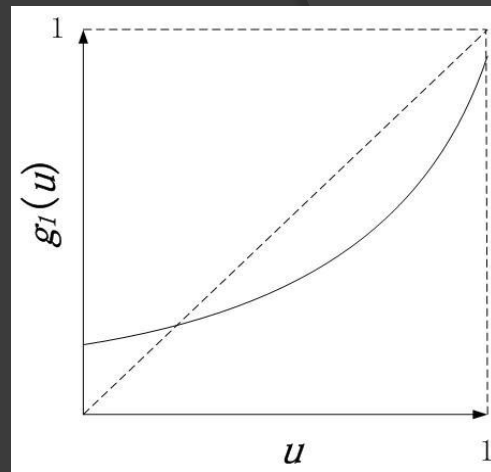
$1 - u = (c\varphi)^{1/(1-\beta)}$ ，有 $g_0(u) \approx 1 + \bar{k}(u - 1)$ 。

在 $u = 1$ 附近，只要 $\alpha > 2$ ，则 $\bar{k}$ 是有限的，故

最大分支规模 $S = \varphi[1 - g_0(u)] \sim \varphi^{(2-\beta)/(1-\beta)}$ 。

当 $\varphi \rightarrow 0$ 时， $S$ 比线性更快地趋于0。

虽然在 $\varphi_c \rightarrow 0$ 时最大分支仍存在，但其规模已经非常小，对整个网络来说意义不大。所以，**鲁棒性应结合渗流阈值和最大分支规模两项指标评估。**





## 2、非均匀随机删除

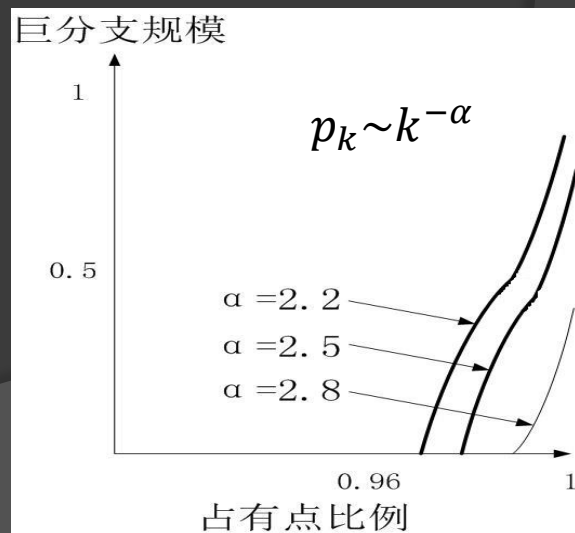
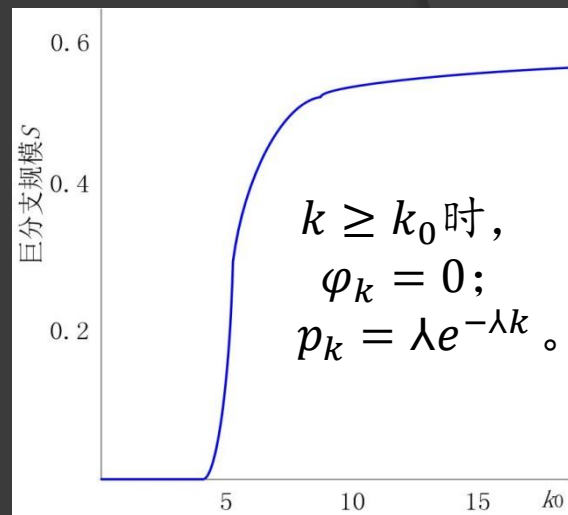
$\varphi_k$ 表示度为 $k$ 的点出现的概率。某点连接巨分支的概率 $1 - u^k$ ，其属于巨分支的概率 $\varphi_k(1 - u^k)$ ，该点属于巨分支的平均概率：

$$S = \sum_{k=0}^{\infty} p_k \varphi_k (1 - u^k) = \bar{\varphi} - \sum_{k=0}^{\infty} p_k \varphi_k u^k;$$

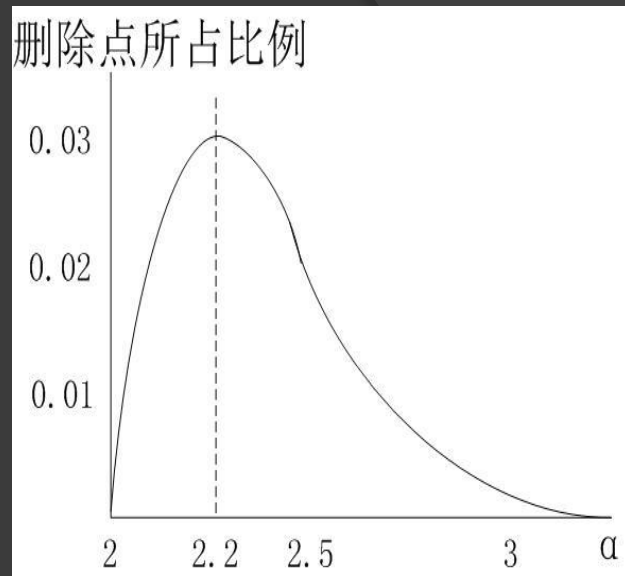
$$u = 1 - \frac{1}{\bar{k}} \sum_{k=1}^{\infty} k p_k \varphi_k + \frac{1}{\bar{k}} \sum_{k=1}^{\infty} k p_k \varphi_k u^{k-1}。$$

在幂律分布 ( $2 \leq \alpha \leq 3$ ) 的网络中，不能求出巨分支规模 $S$ 的闭合形式解，但可以容易地得到其数值解，如右图所示。

可见，只要删除很少一部分高度数节点，巨分支就迅速消失。这与均匀随机删除效果正相反，也称为幂律网络的**脆弱性**。

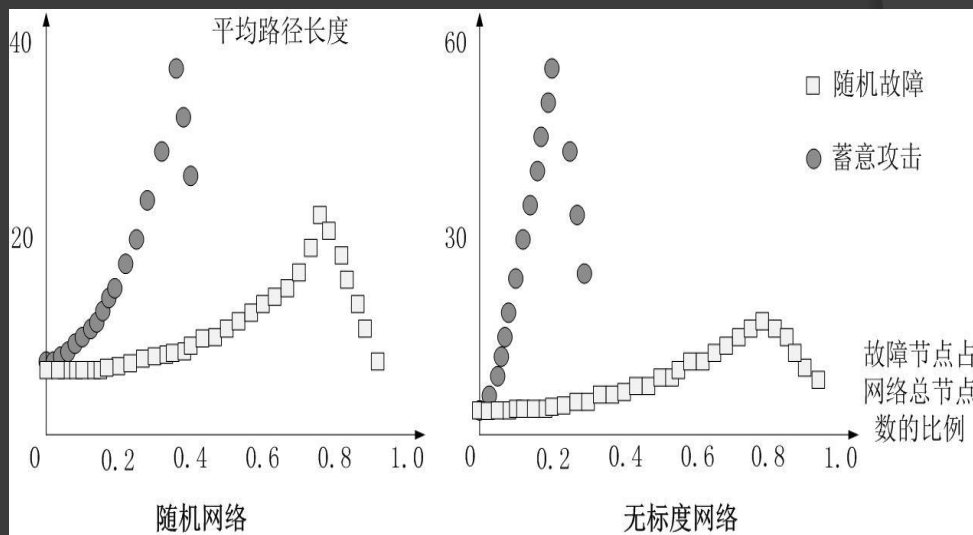


$u$ 算式非平凡解的出现或消失点就是渗流阈值。设  $f_1(u) = \frac{1}{k} \sum_{k=1}^{\infty} k p_k \varphi_k u^{k-1}$  , 则临界点的一般判断为  $f'_1(1) = 1$ 。幂律网络为破坏巨分支所需删除高度数点所占的比例如右图所示。



**互联网鲁棒性与脆弱性并存：**  
对随机故障的容忍度高、抗毁能力强；但病毒扩散迅速、范围广。

核心点故障会导致整个系统失效；但核心点的免疫也可以有效控制病毒传播。



幂律网络的低聚类特性导致高度数点失效后路径长度快速增加。

作业：

45、图8是两种模型产生的网络拓扑和度分布情况，A和B分别对应哪种类型的网络模型？

46、用小世界模型生成的网络拓扑会与图8. A和图8. B存在哪些差异？

47、怎样在一个规则网络上构造“捷径”，才能使网络系统具备“小世界”特性？分析该方式的优缺点。

48、度分布为 $p_k = (1 - e^{-\lambda})e^{-\lambda k}$ 的配置网络， $\lambda > 0$ 。分别写出度分布和余度分布的生成函数 $g_0(z)$ 和 $g_1(z)$ 。设网络规模为 $n$ ，写出某一节点的两跳邻居个数的计算表达式。

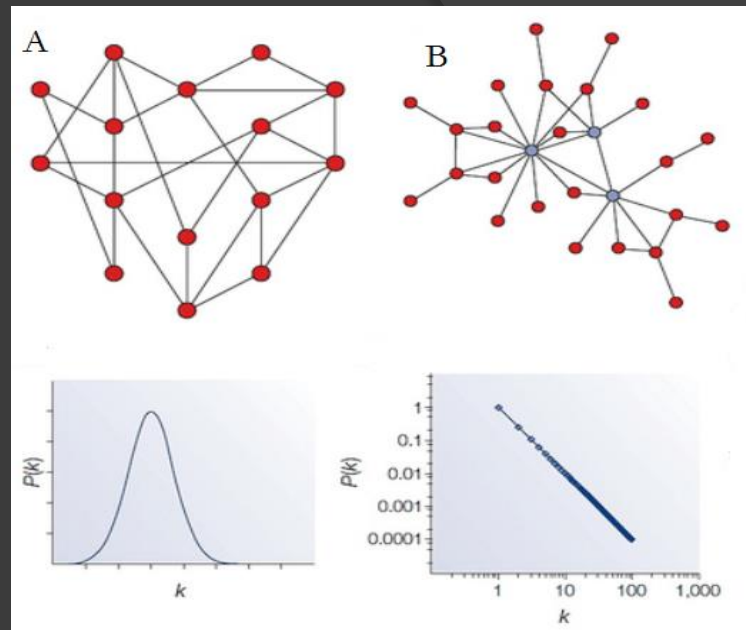


图8

49、你认为BA模型会产生图9这种比较奇异拓扑图形的原因是什么？

50、在BA模型的基础上，增加新点时，适当添加额外的边，并删除部分与有的边。

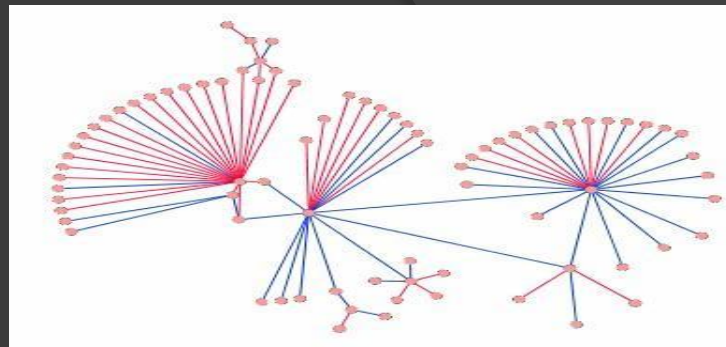


图9

1) 写出该条件下度分布的演化方程；

2) 简单说明方程中每一项的含义；

3) 如果删除的边过多( $v > \frac{m}{2} + w$ )会出现什么现象？

51、1) 在BA模型的基础上，加入适应度的好处是什么？

2) 试写出其度分布的演化方程。

3) 假设适应度 $a_k(\varphi) = \varphi \cdot k$ ，如果所有节点都具有相同的适应度会发生什么现象？

52、考虑一种配置模型网络，该网络中节点的度只有1、2和3，且比例分别为 $p_1$ 、 $p_2$ 和 $p_3$ ，求网络中点渗流发生时节点占有率 $\varphi_c$ 的临界值。

53、为什么说互联网既具有鲁棒性，又具有脆弱性。