

# Chapter 6 随机网络模型

张庆凯

复旦大学管理学院

2020 年 11 月 28 日

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 常见规则网络

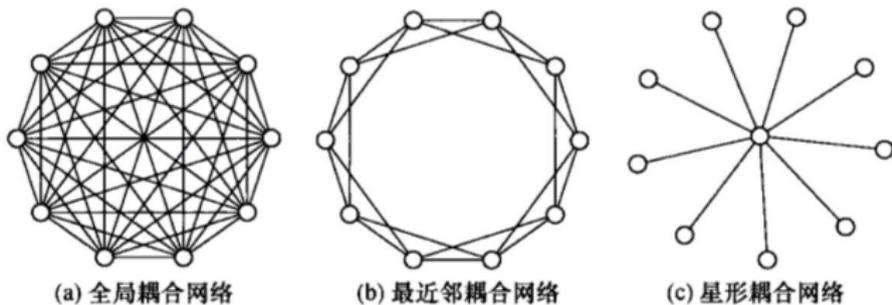


图 6-1 三种规则网络

# 基本拓扑性质

- 聚类系数

$$C = \frac{3 \times \text{网络中三角形的数目}}{\text{网络中连通三元组的数目}}$$

- 平均路径长度

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i>j} d_{ij}$$

其中  $d_{ij}$  为连接这两个节点的最短路径上的边的数目

# 基本拓扑性质

- 全耦合网络

- $C_{gc} = 1$
- $L_{gc} = 1$

- 最近邻耦合网络

$$C_{nc} = \frac{3 \times N \times \binom{K/2}{2}}{N \times \binom{K}{2}}$$
$$= \frac{3(K-2)}{4(K-1)}$$

# 基本拓扑性质

- 最近邻耦合网络

$$L_{nc} \approx \frac{1}{(N/2)} \sum_{m=1}^{N/2} [2m/K] \approx \frac{N}{2K}$$

对固定的  $K$  值, 当  $N \rightarrow \infty$  时,  $L_{nc} \rightarrow \infty$

- 星形网络

- $C_{star} = 0$

- $L_{star} = 2 - \frac{2(N-1)}{N(N-1)} \rightarrow 2 \quad (N \rightarrow \infty)$

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 模型描述

ER 随机图具有两种形式的定义

- 具有固定边数的 ER 随机图  $G(N, M)$
- 具有固定连边概率的 ER 随机图  $G(N, p)$

# 具有固定边数的 ER 随机图 $G(N, M)$

## 算法 6-1 ER 随机图 $G(N, M)$ 构造算法

- ① 初始化：给定  $N$  个节点和待添加的边数  $M$
- ② 随机连边：
  - step 1 随机选取一对没有边相连的不同的节点，并在这对节点之间添加一条边。
  - step 2 重复 step 1，直到在  $M$  对不同的节点对之间各添加了一条边。

# 具有固定连边概率的 ER 随机图 $G(N, p)$

## 算法 6-2 ER 随机图 $G(N, p)$ 构造算法

- ① 初始化：给定  $N$  个节点以及连边概率  $p \in [0, 1]$ 。
- ② 随机连边：
  - step 1 选择一对没有边相连的不同的节点。
  - step 2 生成一个随机数  $r \in (0, 1)$ 。
  - step 3 如果  $r < p$ , 那么在这对节点之间添加一条边；否则就不添加边。
  - step 4 重复 step 1 ~ step 3, 直到所有的节点对都被选择过一次。

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

## 边数分布

- 给定网络节点数  $N$  和连边概率  $p$ , 生成的随机图恰好具有  $M$  条边的概率为标准的二项分布:

$$P(M) = \binom{\binom{N}{2}}{M} p^M (1-p)^{\binom{N}{2}-M}$$

- 边数分布的平均值:

$$\langle M \rangle = \sum_{M=0}^{\binom{N}{2}} M P(M) = \binom{N}{2} p = pN(N-1)/2$$

# 边数分布

- 边数分布的方差:

$$\sigma_M^2 = \langle M^2 \rangle - \langle M \rangle^2 = p(1-p) \frac{N(N-1)}{2}$$

- 边数分布的变异系数:

$$\frac{\sigma_M}{\langle M \rangle} = \sqrt{\frac{1-p}{p} \frac{2}{N(N-1)}} \approx \frac{1}{N}$$

- 随机图的稀疏性: 如果连边概率  $p$  与  $1/N$  同阶, 即  $p = O(1/N)$ , 那么有

$$\langle M \rangle = pN(N-1)/2 \sim O(N)$$

# 度分布

- 网络中任一给定节点的度为  $k$  的概率同样服从二项分布:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- 度分布的平均值:

$$\langle k \rangle = p(N-1)$$

## 度分布

- 度分布的方差:

$$\sigma_k^2 = p(1-p)(N-1)$$

- 度分布的变异系数:

$$\frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{1-p}{p} \frac{1}{N-1}} \approx \sqrt{\frac{1}{N-1}}$$

- 泊松分布: 当  $N$  很大且  $p$  很小时, 二项分布可以近似为泊松分布, 即有

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

# 聚类系数与平均路径长度

- ER 随机图的聚类系数:

$$C = p = \langle k \rangle / (N - 1)$$

- 直观上, 由于 ER 随机图的聚类系数很小, 意味着网络中的三角形数量相对很少。
- 对于 ER 随机图中随机选取的一个点, 网络中大约有  $\langle k \rangle$  个其他的点与该点之间的距离为 1; 大约有  $\langle k \rangle^2$  其他节点与该点之间的距离为 2; 以此类推, 由于网络总的节点数为  $N$ , 设  $D_{ER}$  是 ER 随机图的直径, 大体上应该有  $N \sim \langle k \rangle^{D_{ER}}$ 。

# 聚类系数与平均路径长度

- 因此，网络的直径和平均路径长度满足

$$L_{ER} \leq D_{ER} \sim \ln N / \ln \langle k \rangle$$

- 小世界特征！

# 聚类系数与平均路径长度

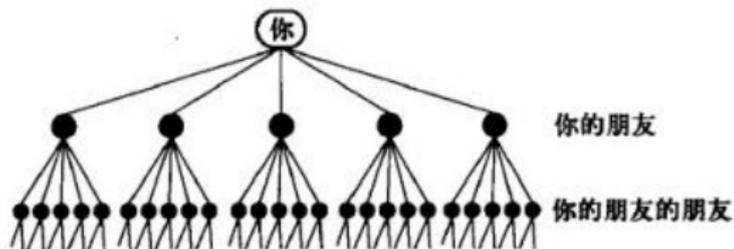


图 6-6 朋友数的指数增长

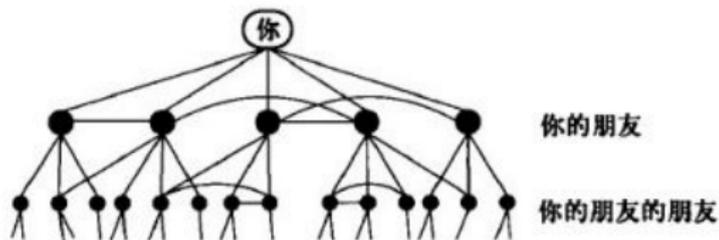


图 6-7 实际社会网络中朋友数的增长

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 随机图的演化

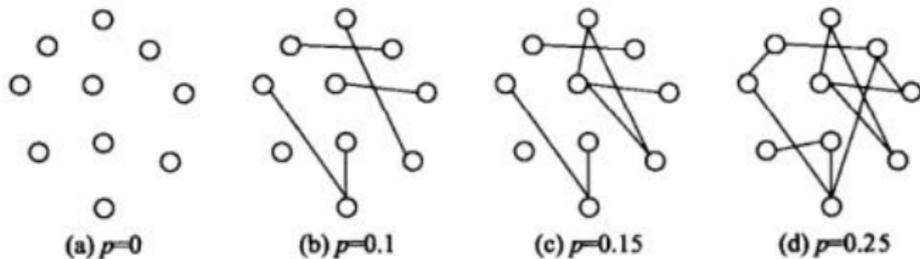


图 6-8 具有不同连接概率的随机图示例

现在的问题是：当连接概率  $p$  从 0 开始逐渐增加到 1 时，最大连通片的规模是如何具体变化的？特别地，当  $p$  多大时才会出现包含网络中一定比例节点的巨片（即最大的连通片）？

## 巨片的涌现

许多实际的大规模复杂网络都是不连通的，但是往往会存在一个特别大的连通片，它包含了整个网络中相当比列的节点，这一连通片称为巨片 (Giant component)

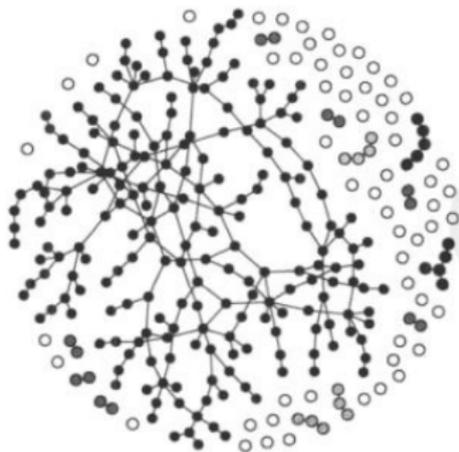


图 6-9 随机图中的巨片

# 巨片的涌现

Erdos 和 Renyi 系统性地研究了当  $N \rightarrow \infty$  时 ER 随机图的性质（包括巨片的出现）与概率  $p$  之间的关系。特别地，他们发现 ER 随机图具有如下的**涌现或相变**性质：

ER 随机图的许多重要性质都是突然涌现的：对于任一给定的连边概率  $p$ ，要么几乎每一个  $G(N, p)$  的实例都具有某个性质  $Q$ ，要么几乎每一个这样的图都不具有性质  $Q$ 。

这里，如果当  $N \rightarrow \infty$  时产生一个具有性质  $Q$  的 ER 随机图的概率为 1，那么就称几乎每一个 ER 随机图都具有性质  $Q$ 。

# 巨片的涌现

当  $N \rightarrow \infty$  时 ER 随机图的巨片的相对规模  $S \in [0, 1]$  定义为巨片中所包含的节点数占整个网络节点的比例，亦即为网络中一个随机选择的节点属于巨片的概率。 $u = 1 - S$  为不属于巨片的节点所占的比例，即一个随机选择的节点不属于巨片的概率。显然，存在如下两种可能：

- 网络中不存在巨片，即  $S = 0, u = 1$ ;
- 网络中存在巨片，即  $S > 0, u < 1$ 。

# 巨片的涌现

网络中一个随机选择的节点  $i$  如果不属于巨片，那么就说明它也没有通过其他任一节点与巨片相连，也即对于网络中的任一其他节点  $j$ ，必然有如下两种情形之一：

- 节点  $i$  与节点  $j$  之间没有边相连：此情形发生的概率为  $1 - p$ 。
- 节点  $i$  与节点  $j$  之间有边相连，但是节点  $j$  不属于巨片：此情形发生的概率为  $pu$ 。

## 巨片的涌现

因此, 节点  $i$  没有通过任一节点与巨片相连的概率为

$$u = (1 - p + pu)^{N-1} = \left[ 1 - \frac{\langle k \rangle}{N-1}(1-u) \right]^{N-1}$$

对上式两边取对数有

$$\begin{aligned} \ln u &= (N-1) \ln \left[ 1 - \frac{\langle k \rangle}{N-1}(1-u) \right] \\ &= -(N-1) \frac{\langle k \rangle}{N-1}(1-u) \\ &= -\langle k \rangle(1-u) \end{aligned}$$

从而

$$u = e^{-\langle k \rangle(1-u)}$$

于是可以得到巨片中节点的比例  $S = 1 - u$  满足

$$S = 1 - e^{-\langle k \rangle S} \quad (1)$$

# 巨片的涌现

尽管式 (1) 看上去简单，却不存在简单的解析解。我们可以用图示的方法求得数值解。

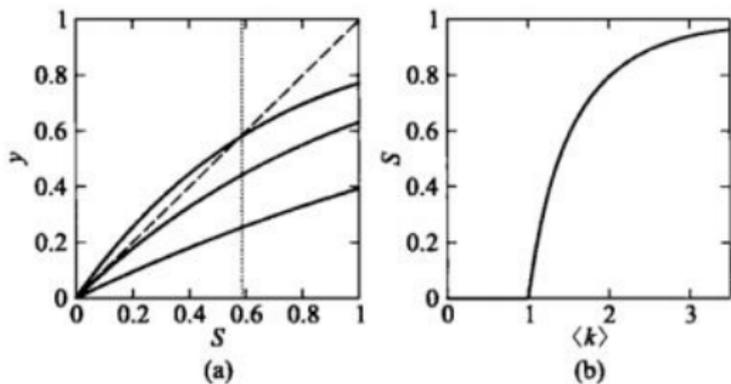


图 6-10 ER 随机图的平均度  $\langle k \rangle$  和巨片规模  $S$  的关系

# 巨片的涌现

可以看到, 在  $\langle k \rangle < 1$  时,  $S = 0$  意味着不存在巨片; 在  $\langle k \rangle > 1$  时,  $S > 0$  意味着涌现巨片。由于 ER 随机图的平均度是  $\langle k \rangle = p(N - 1) \approx pN$ , 从而产生巨片的连边概率  $p$  的临界值为

$$p_c = \langle k \rangle_c / (N - 1) \approx 1/N$$

即当  $p > p_c$  时, 几乎每一个随机图都包含巨片。

# 巨片的涌现

进一步地，可以推得，当  $p > \ln N/N$  (对应于  $\langle k \rangle > \ln N$  时)，ER 随机图几乎总是连通的。下图显示了  $N = 100$  时，ER 随机图的几种状态。

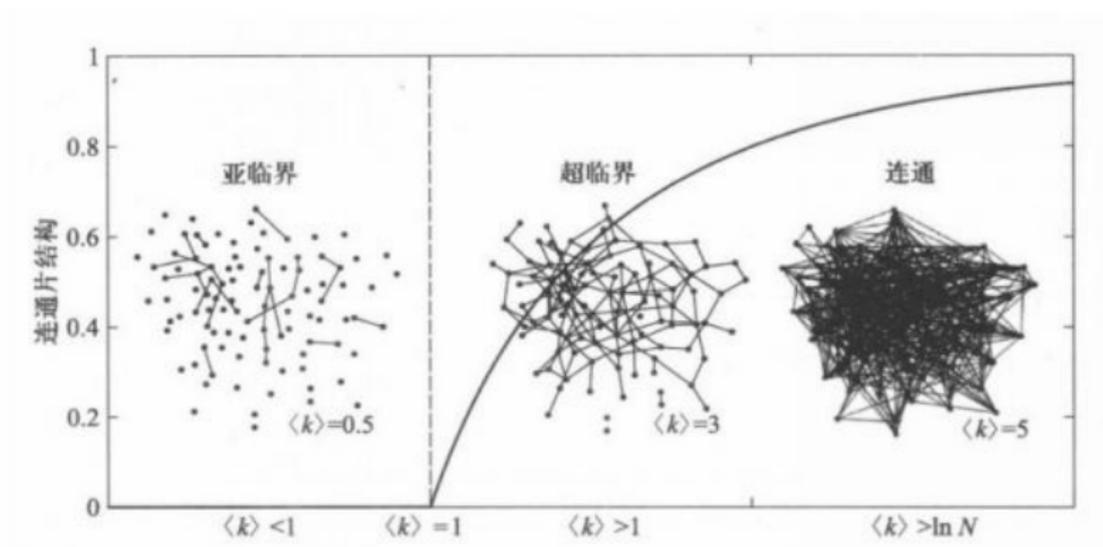


图 6-11 ER 随机图中连通片结构的演化

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 随机图与实际网络的比较

ER 随机图和许多实际网络相比具有如下一些共性特征：

- 稀疏性。实际网络往往是稀疏的，而当连边概率  $p$  与网络规模的倒数同阶 ( $p \sim O(1/N)$ ) 时，ER 随机图是一个边数与网络规模同阶的稀疏图，( $M \sim O(N)$ )。
- 有巨片。实际网络往往存在巨片，当  $p > p_c \sim 1/N$  时，ER 随机图具有一个包含网络中相当比例节点的巨片。
- 小世界。ER 随机图的平均距离大体上是网络规模的对数函数， $L \sim \ln N / \ln \langle k \rangle$ ，而实际网络往往也具有与相同规模和密度的 ER 随机图相近的平均距离（见下表）。

# 随机图与实际网络的比较

Network	Size	$\langle k \rangle$	$\ell$	$\ell_{rand}$	$C$	$C_{rand}$	Reference	Nr.
WWW, site level, undir.	153, 127	35.21	3.1	3.35	0.1078	0.00023	Adamic 1999	1
Internet, domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001	Yook <i>et al.</i> 2001a, Pastor-Satorras <i>et al.</i> 2001	2
Movie actors	225, 226	61	3.65	2.99	0.79	0.00027	Watts, Strogatz 1998	3
LANL coauthorship	52, 909	9.7	5.9	4.79	0.43	$1.8 \times 10^{-4}$	Newman 2001a,b	4
MEDLINE coauthorship	1, 520, 251	18.1	4.6	4.91	0.066	$1.1 \times 10^{-5}$	Newman 2001a,b	5
SPIRES coauthorship	56, 627	173	4.0	2.12	0.726	0.003	Newman 2001a,b,c	6
NCSTRL coauthorship	11, 994	3.59	9.7	7.34	0.496	$3 \times 10^{-4}$	Newman 2001a,b	7
Math coauthorship	70, 975	3.9	9.5	8.2	0.59	$5.4 \times 10^{-5}$	Barabási <i>et al.</i> 2001	8
Neurosci. coauthorship	209, 293	11.5	6	5.01	0.76	$5.5 \times 10^{-5}$	Barabási <i>et al.</i> 2001	9
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner, Fell 2000	10
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner, Fell 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya, Solé 2000	12
Silwood park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya, Solé 2000	13
Words, cococcurrence	460,902	70.13	2.67	3.03	0.437	0.0001	Cancho, Solé 2001	14
Words, synonyms	22, 311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> 2001	15
Power grid	4,941	2.67	18.7	12.4	0.08	0.005	Watts, Strogatz 1998	16
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts, Strogatz 1998	17

Figure: 一些实际网络与随机图的聚类系数和平均距离的比较

# 随机图与实际网络的比较

但是，ER 随机图也具有一些与实际网络显著不同的特征：

- 聚类特性的差异。对于固定的网络密度，当  $N \rightarrow \infty$  时，ER 随机图的聚类系数  $C_{rand} = \langle k \rangle / (N - 1) \rightarrow 0$ ，意味着 ER 随机图没有聚类特性。
- 度分布的差异。ER 随机图的度分布近似服从均匀的泊松分布，意味着网络中节点的度基本都集中在平均度  $\langle k \rangle$  附近。

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 配置模型

人们可以从多个角度对 ER 随机图进行扩展以使其更接近实际网络。其中一个自然的推广就是具有任意给定度分布、但在其他方面完全随机的**广义随机图** (Generalized random graph)。到目前为止研究最多的广义随机图模型是**配置模型** (Configuration model)。在配置模型中事先给定的是网络的度序列  $\{d_1, d_2, \dots, d_N\}$  其中非负整数  $d_i$  为节点  $i$  的度。

## 配置模型

显然，度序列并不能完全任意给定，否则有可能无法生成符合度序列的简单图。两个显而易见的必要条件是：

- 由于网络中所有节点的度值之和等于网络中所有边数之和的两倍， $\sum_{i=1}^N d_i$  必须为偶数并且有

$$\sum_{i=1}^k d_i \leq N(N-1)$$

- $d_i \leq N-1, i = 1, 2, \dots, N$ , 等号只有当一个节点与其他所有的节点都相连时才能成立。

# 配置模型

在上述条件的基础上少许加以推广，就可得到如下的充要条件：

## Theorem

一个非负整数序列  $\{d_1, d_2, \dots, d_N\}$  是某个简单图的度序列的充要条件为：

- $\sum_{i=1}^n d_i$  为偶数
- 对于每个整数  $k, 1 \leq k \leq N$ , 均有

$$\sum_{i=1}^k d_i \leq k(k-1) + \sum_{j=k+1}^N \min(k, d_j) \quad (2)$$

## 配置模型

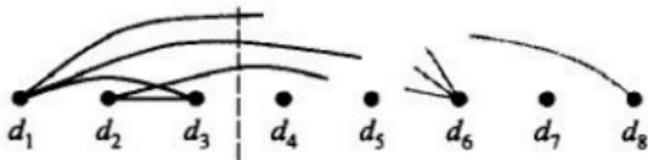


图 6-12 公式(6-24)的几何说明

上述定理最早是由 Erdos 和 Gallai 在 1960 年提出并加以证明的。对于前  $k$  个节点，式 (2) 左端  $\sum_{i=1}^k d_i$  为这  $k$  个节点的度值之和，右端第一项  $k(k-1)$  这  $k$  个节点之间的连边对它们的度的最大可能的贡献，第二项  $\sum_{j=k+1}^N \min(k, d_j)$  为这  $k$  个节点与其余节点之间的连边对这  $k$  个节点的度值的最大可能的贡献。

# 配置模型

## Example

验证整数序列  $\{6, 6, 5, 4, 4, 2, 1\}$  是否为某个简单图的度序列的步骤如下:

- $\sum_{i=1}^7 d_i = 28$  为偶数;
- $k = 1$  时, 式 (2) 显然成立;  $k = 2$  时, 式 (2) 不再成立, 因为

$$\sum_{i=1}^k d_i = 6 + 6 = 12$$

$$k(k-1) + \sum_{j=k+1}^N \min(k, d_j) = 2 \times 1 + (2 + 2 + 2 + 2 + 1) = 11$$

所以, 整数序列  $\{6, 6, 5, 4, 4, 2, 1\}$  不可能为某个简单图的度序列。

# 配置模型

生成具有给定度序列的广义随机图的配置模型的算法：

## 算法 6-3 配置模型构造算法

- ① 初始化：根据给定度序列确定  $N$  个节点的度值。
- ② 引出线头：从度为  $k_i$  的节点  $i$  引出  $k_i$  个线头。共有  $\sum_{i=1}^N k_i = 2M$  个线头， $M$  为网络的边数。
- ③ 随机配对：完全随机地选取一对线头，把它们连在一起，形成一条边；再在剩余的线头中完全随机地选取另一对线头连成一条边；以此进行下去，直至用完所有的线头。

## 配置模型

关于配置模型算法的几点说明：

- 度序列应满足相应条件。
- 生成具有给定度分布的网络。
- 等概率随机配对。
- 生成模型的不唯一性。
- 有可能产生自环和重边。

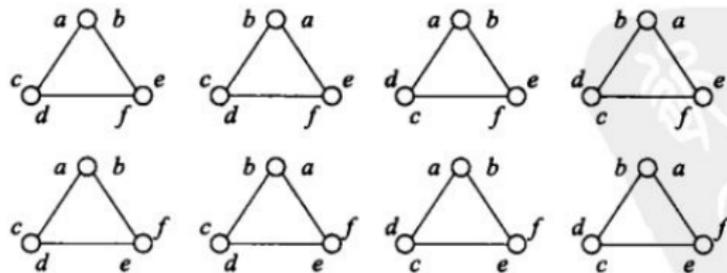


图 6-14 包含 3 个节点的相同的配置方案

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 余平均度

- 网络的平均度  $\langle k \rangle$

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

- 节点  $i$  的余平均度  $\langle k_{nn} \rangle_i$

$$\langle k_{nn} \rangle_i = \frac{1}{k_i} \sum_{j=1}^{k_i} k_{ij}$$

- 网络的余平均度  $\langle k_n \rangle$

$$\langle k_n \rangle = \frac{1}{N} \sum_{i=1}^N \langle k_{nn} \rangle_i = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{k_i} \sum_{j=1}^{k_i} k_{ij} \right)$$

# 网络的平均度 $\langle k \rangle$ 与网络的余平均度 $\langle k_n \rangle$ 的关系

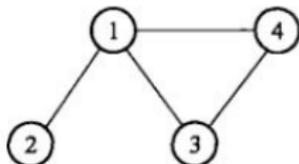


图 6-16 包含 4 个节点的简单网络

$$\langle k \rangle = \frac{1}{4} \sum_{i=1}^4 k_i = \frac{1}{4} (3 + 1 + 2 + 2) = 2$$

$$\begin{aligned} \langle k_n \rangle &= \frac{1}{4} \sum_{i=1}^4 \langle k_{nn} \rangle_i \\ &= \frac{1}{4} \left( \frac{1+2+2}{3} + \frac{3}{1} + \frac{3+2}{2} + \frac{3+2}{2} \right) \\ &= 2 + \frac{5}{12} \end{aligned}$$

$$> \langle k \rangle$$

## 余度分布

- 我们知道, 网络节点的平均度与度分布之间有如下关系:

$$\langle k \rangle = \sum_{k=0}^{\infty} kP(k)$$

其中  $P(k)$  是网络中度为  $k$  的节点的概率。

- 类似地, 余平均度也可以通过余度分布来计算:

$$\langle k_n \rangle = \sum_{k=0}^{\infty} kP_n(k) = \sum_{k=0}^{\infty} kq_k$$

其中, 余度分布  $P_n(k) \equiv q_k$  定义为网络中随机选取的一个节点的随机选取的一个邻居节点的度为  $k$  的概率。

## 余度分布

- 计算配置模型的余度分布:

$$P_n(k) = Np_k \times \frac{k}{2M-1} = \frac{Nkp_k}{2M} = \frac{kp_k}{\langle k \rangle} \quad (3)$$

- 式 (3) 意味着, 在给定网络平均度的情形下, 从网络中一个随机选择的节点出发, 沿着一条边到达一个度为  $k$  的邻居节点的概率与  $kp_k$  而不是与  $p_k$  成正比。也就是说, 到达的可能是比一个典型节点的度更高的节点。
- 基于式 (3), 配置模型中随机选取的一个节点的邻居节点的平均度为

$$\langle k_n \rangle = \sum_{k=0}^{\infty} kP_n(k) = \sum_{k=0}^{\infty} k \frac{kp_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

- 于是有

$$\langle k_n \rangle - \langle k \rangle = \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} = \frac{\sigma^2}{\langle k \rangle} \geq 0$$

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 零模型

ER 随机图可以视为阶数最低的零模型。有时我们需要具有更多约束条件的零模型。按照约束条件从少到多，可以定义如下不同阶次的零模型：

- 0 阶零模型：即与原网络具有相同节点数  $N$  和边数  $M$  的随机化网络。
- 1 阶零模型：即与原网络具有相同节点数  $N$  和度分布  $P(k)$  的随机化网络。通常的做法是每个节点的度值都保持不变（即度序列保持不变）。

## 零模型

- 2 阶零模型：即与原网络具有相同节点数  $N$  和二阶度相关特性（即联合度分布） $P(k, k')$  的随机化网络。有时也考虑与原网络具有相同同配系数的随机化网络。
- 3 阶零模型：即与原网络具有相同节点数  $N$  和三阶度相关特性（即联合边度分布） $P(k_1, k_2, k_3)$  的随机化网络。考虑到 3 个节点构成的连通三元组包括图 6-17 所示的两种情形，因此三阶度相关特性  $P(k_1, k_2, k_3)$  是由  $P_{\wedge}(k_1, k_2, k_3)$  和  $P_{\Delta}(k_1, k_2, k_3)$  共同组成的。

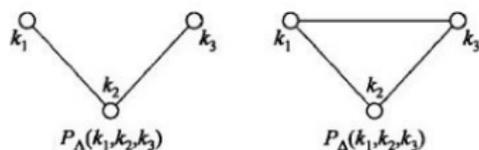


图 6-17 连通三元组的两种情形

以此类推，我们还可以定义更高阶的零模型。

## 零模型

显然，对于任一给定网络  $G$  和任意两个自然数  $d_1 < d_2$ ，具有与网络  $G$  相同的  $d_2$  阶分布的模型集合一定是与网络  $G$  具有相同的  $d_1$  阶分布的模型集合的子集，而网络  $G$  的  $d$  阶零模型的性质就取为与网络  $G$  具有相同的  $d$  阶分布的模型集合的性质的平均。随着  $d$  的增大， $d$  阶零模型将越来越接近给定网络  $G$ 。

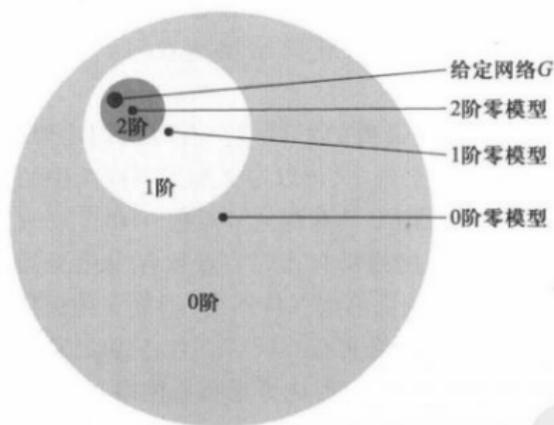


图 6-19 各阶模型之间的关系

# 目录

- 1 常见规则网络及其基本拓扑性质
- 2 随机图
  - 模型描述
  - 拓扑性质
  - 巨片的涌现与相变
  - 随机图与实际网络的比较
- 3 广义随机图
  - 配置模型
  - 配置模型的理论分析
- 4 随机重连与零模型
  - 零模型
  - 随机重连
- 5 基于零模型的拓扑性质分析

# 随机重连算法

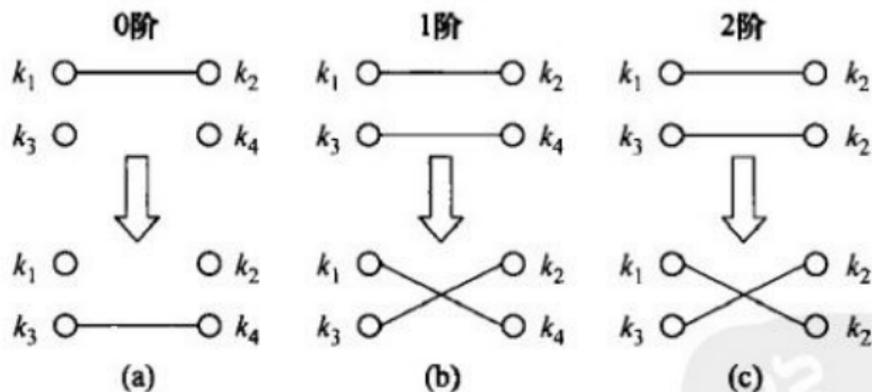


图 6-20 生成 0 阶、1 阶和 2 阶零模型的随机重连过程

# 随机重连算法

- 随着零模型阶次的增加，约束条件进一步加强，重连的可能性会不断减小，生成网络的随机化程度也逐渐降低。
- 上述随机重连算法可以推广到有向网络的情形。例如，可以通过保持每个节点的入度和出度不变的随机重连而生成 1 阶零模型。在此情形，图 6-20(b) 中的随机选取的两条无向边变为两条有向边  $k_1 \rightarrow k_2$  和  $k_3 \rightarrow k_4$ 。

# 随机重连算法

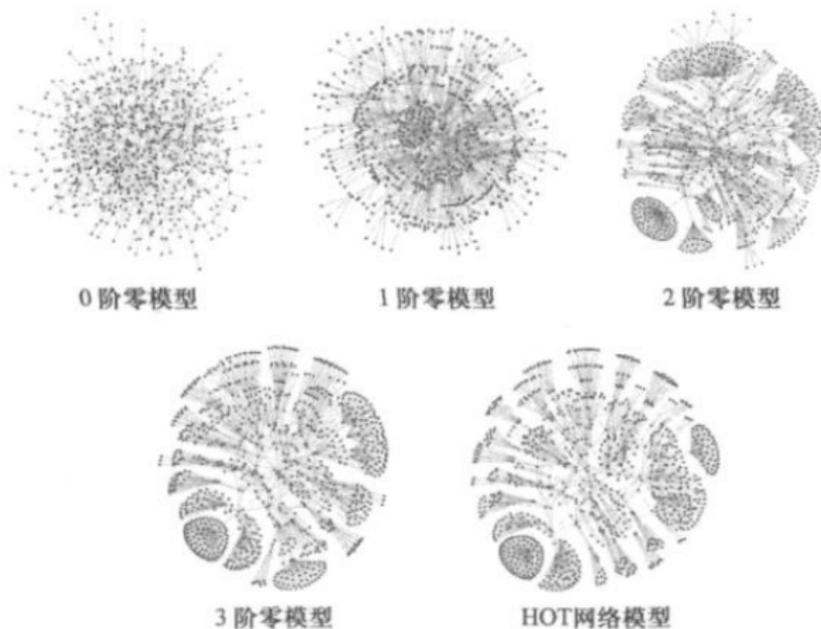


图 6-21 HOT 模型及相应的零模型

# 基于零模型的拓扑性质分析

- 比较判据
- 度相关性分析
- 模体分析
- 同配性质分析

## 比较判据

一般而言，基于零模型研究网络特征时需要明确两点：

- 确定零模型。根据所要研究的特征，确定合适的保持低阶特征不变的零模型。
- 确定比较方法。把实际网络的特征与相应零模型的特征做恰当比较。具体地说，假设某种拓扑特征（例如，某个子图）在一个实际网络中出现的次数为  $N(j)$ ，在相应的随机化网络中出现次数的平均值为  $\langle N_r(j) \rangle$ ，那么可以计算如下比值：

$$R(j) = \frac{N(j)}{\langle N_r(j) \rangle}$$

如果  $R(j) > 1$  ( 或  $R(j) < 1$  ), 那么就意味着实际网络的设计或者演化过程促进（或抑制）了该拓扑特征的出现。

## 比较判据

- 如果要进一步刻画某个拓扑模式在实际网络中出现的频率与相应随机化网络中出现的频率的差异是否显著 ( 即统计重要性), 那么可以采用统计学中的  $Z$  检验方法。具体地说, 拓扑特征  $j$  的统计重要性可用如下的  $Z$  值来刻画:

$$Z_j \triangleq Z(j) = \frac{N(j) - \langle N_r(j) \rangle}{\sigma_r(j)}$$

其中  $\sigma_r(j)$  为随机化网络中拓扑特征  $j$  的出现次数  $N_r(j)$  的标准差。 $Z$  值的绝对值越大就表示差异越显著。

## 比较判据

- 通常的做法是在平面上绘制出比值  $R$  和  $Z$  值的图形, 称为相关性剖面 ( Correlation profile)。
- 为了便于比较不同规模的网络, 通常对  $Z$  租做归一化处理, 得到重要性剖面 ( Significance profile,SP) :

$$SP_j = \frac{Z_j}{(\sum Z_i^2)^{1/2}}$$

## 度相关性分析

- 网络的联合概率分布可以表示为（见式 (4-2)）

$$P(k_1, k_2) = m(k_1, k_2) \mu(k_1, k_2) / (2M)$$

其中  $m(k_1, k_2)$  是度为  $k_1$  的节点和度为  $k_2$  的节点之间的连边数。如果  $k_1 = k_2$ , 那么  $\mu(k_1, k_2) = 2$ ; 否则  $\mu(k_1, k_2) = 1$ 。

- 我们可以通过比较一个实际网络的  $m(k_1, k_2)$  及其相应的 1 阶零模型所对应的均值  $\langle m_r(k_1, k_2) \rangle$  来分析实际网络的度相关性。具体地说, 可以计算并绘制如下的相关性剖面:

$$R(k_1, k_2) = \frac{m(k_1, k_2)}{\langle m_1(k_1, k_2) \rangle}, \quad Z(k_1, k_2) = \frac{m(k_1, k_2) - \langle m_r(k_1, k_2) \rangle}{\sigma_1(k_1, k_2)}$$

# 度相关性分析

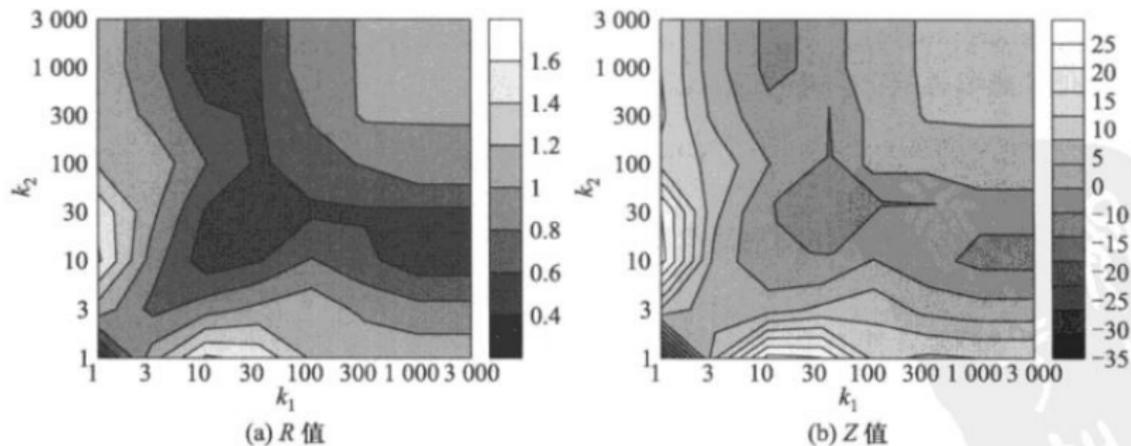


图 6-22 双对数坐标下的互联网拓扑的相关性剖面(取自文献[11])

# 模体分析

- 实际网络可能包含各种各样的子图，其中一些子图所占的比例明显高于相应的零模型中这些子图所占的比例，这些子图称为**模体 (Motif)**，辨识出模体有助于识别网络的典型的局部连接模式。

# 模体分析

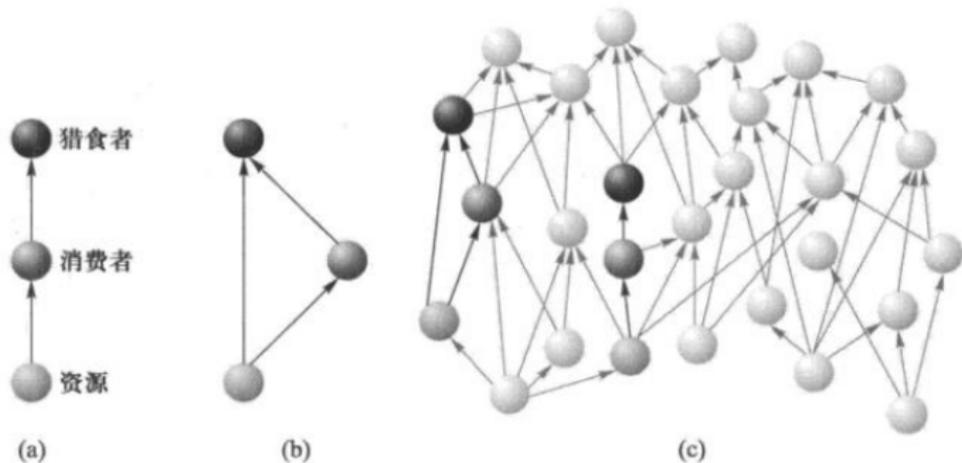


图 6-24 生态网络中常见的两种模体(取自文献[17])

# 模体分析

- 为了判断实际网络中的一个子图  $j$  是否为模体，可以比较该子图在实际网络中出现的次数  $N(j)$  与在相应的随机化网络中出现次数的平均值  $\langle N_r(j) \rangle$ ，一般要求

$$R(j) = \frac{N(j)}{\langle N_r(j) \rangle} > 1.1$$

# 模体分析

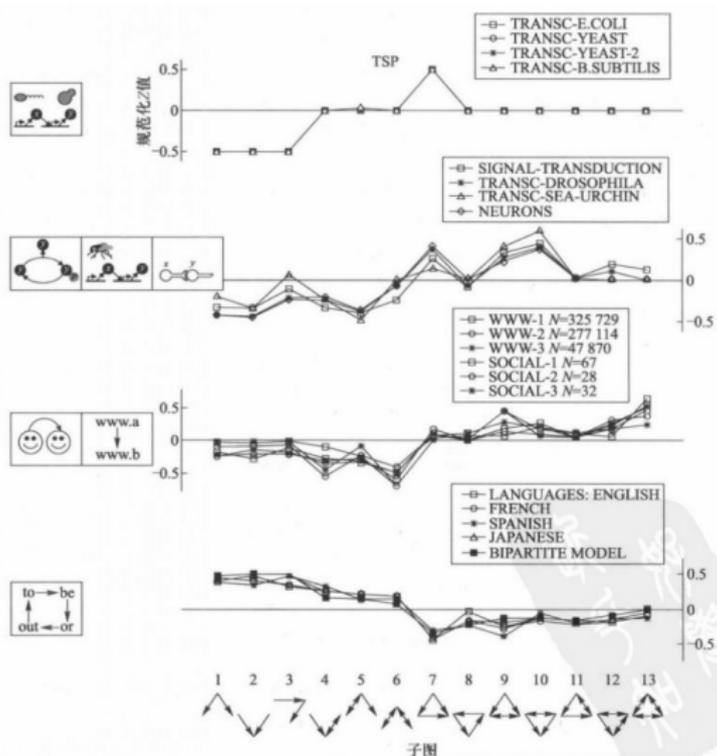


图 6-26 不同网络的三元组重要性剖面(取自文献[20])

## 同配性质分析

- 无向网络的度同配性质反映了度值相近的节点之间互相连接的倾向性, 它可以用如下的同配系数 (即 Pearson 相关系数) 来表征 (参见第 4 章):

$$r = \frac{\text{cov}(k_i, k_j)}{\text{Var}(k_i, k_j)}$$

如第 4 章所述,  $r > 0$  对应于同配,  $r < 0$  对应于异配。

- 一般而言, 同配系数的大小是与网络规模和密度相关的。因此, 一种更为合理的评价是把一个实际网络的同配系数与相应的零模型的同配系数做比较以判断网络的同配或异配程度。

## 同配性质分析

- 下面针对一般的有向网络加以说明。在有向网络情形, 边的方向有可能对网络的同配性质产生重要影响。一个有向网络的同配性可以有如下 4 种度量:  $r(out, in)$ ,  $r(in, out)$ ,  $r(out, out)$  和  $r(in, in)$ 。其中  $r(out, in)$  量化的是高出度的节点有边指向高入度的节点的倾向性程度, 其余三种的定义类似。

## 同配性质分析

- 我们用  $\alpha, \beta \in \{ \text{in}, \text{out} \}$  标记出度或者入度类型, 并把有向边  $i$  的源节点和目标节点的  $\alpha$  度和  $\beta$  度分别记为  $j_i^\alpha$  和  $k_i^\beta$ 。有向网络的一组同配系数可以用如下的 Pearson 相关系数刻画:

$$r(\alpha, \beta) = \frac{M^{-1} \sum_i (j_i^\alpha - \langle j^\alpha \rangle)(k_i^\beta - \langle k^\beta \rangle)}{\sqrt{M^{-1} \sum_i (j_i^\alpha - \langle j^\alpha \rangle)^2} \sqrt{M^{-1} \sum_i (k_i^\beta - \langle k^\beta \rangle)^2}}$$

其中  $M$  为网络边数,  $\langle \cdot \rangle$  为均值。这里规定每种情形下, 边都是从  $\alpha$  标度的节点指向  $\beta$  标度的节点。

# 同配性质分析

- 每个相关性  $r(\alpha, \beta)$  的统计重要性可通过  $Z$  值来刻画:

$$Z(\alpha, \beta) = \frac{r(\alpha, \beta) - \langle r_r(\alpha, \beta) \rangle}{\sigma_r(\alpha, \beta)}$$

其中  $\langle r_r(\alpha, \beta) \rangle$  和  $\sigma_r(\alpha, \beta)$  分别为一阶零模型的同配系数的均值和标准差。

## 同配性质分析

- 通常网络规模越大  $Z$  值也越大, 但我们可以对  $Z$  值作归一化处理以消除网络规模的影响, 得到如下定义的同配重要性剖面 (Assortativity significance profile, ASP) :

$$ASP(\alpha, \beta) = \frac{Z(\alpha, \beta)}{\left(\sum_{\alpha, \beta} Z(\alpha, \beta)^2\right)^{1/2}}$$

- $ASP(\alpha, \beta) > 0$  表明实际网络比具有相同度序列的零模型更为同配, 此时称网络是  $Z$  同配的 ;  $ASP(\alpha, \beta) < 0$  则表明实际网络比相应的零模型更为异配, 此时称网络是  $Z$  异配的。

## 同配性质分析

- 我们选取图 6-26 中显示的属于同一超家族的 3 个有向网络 WWW-1, Social -1 和 Social-3, 列于表 6-2。图 6-28 表明这 3 个网络具有不同的同配性质。

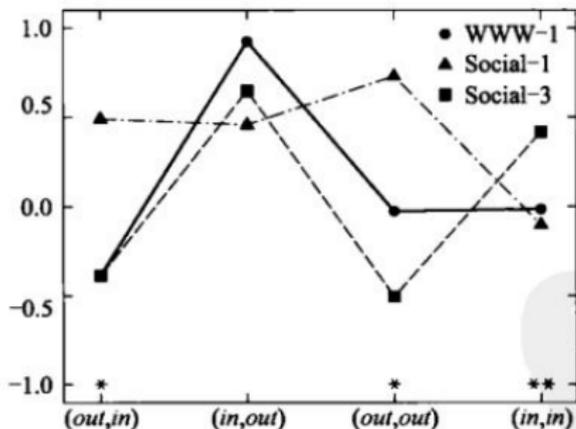


图 6-28 属于同一超家族的 3 个有向网络的不同的同配性质

# 同配性质分析

表 6-2 属于同一超家族的 3 个有向网络

WWW-1	Social-1	Social-3
( <i>out, in</i> ) <i>Z</i> 异配	( <i>out, in</i> ) <i>Z</i> 同配	( <i>out, in</i> ) <i>Z</i> 异配
( <i>in, out</i> ) <i>Z</i> 同配	( <i>in, out</i> ) <i>Z</i> 同配	( <i>in, out</i> ) <i>Z</i> 同配
( <i>out, out</i> ) 无明显倾向	( <i>out, out</i> ) <i>Z</i> 同配	( <i>out, out</i> ) 异配
( <i>in, in</i> ) 无明显倾向	( <i>in, in</i> ) <i>Z</i> 异配	( <i>in, in</i> ) 同配

Thanks for listening!