

第二章 无权网络零模型的构造及应用

第二章无权网络零模型的构造及应用	1
本章简介	2
2.1 引言	3
2.2 不同阶数网络零模型的定义	4
2.2.1 0 阶零模型	5
2.2.2 1 阶零模型	5
2.2.3 2 阶零模型	6
2.2.4 2.25 阶零模型	6
2.2.5 2.5 阶零模型	6
2.2.6 3 阶零模型	7
2.3 基于 ER 随机图的 0 阶零模型	8
2.4 基于配置模型的零模型	9
2.4.1 基于配置模型的 1 阶零模型	10
2.4.2 基于配置模型的 2 阶零模型	11
2.4.3 基于配置模型的 3 阶零模型	13
2.5 基于随机断边重连的零模型	15
2.5.1 基于随机断边重连的 0 阶零模型	16
2.5.2 基于随机断边重连的 1 阶零模型	17
2.5.3 基于随机断边重连的 2 阶零模型	21
2.5.4 基于随机断边重连的 2.5 阶零模型	23
2.6 检验富人俱乐部效应的局部断边重连零模型	25
2.7 检验匹配特性的有倾向性断边重连零模型	28
2.8 无权有向网络的零模型	32
2.9 本章小结与展望	36
参考文献	38

作者：崔丽艳，毋亚男，许小可 (xiaokeie@gmail.com)

大连民族大学信息与通信工程学院

本章简介

静态无权网络是目前最常见的复杂网络形式，这种网络零模型被研究的最广泛和最深入。在本章中我们将无权网络分成了无权无向网络和无权有向网络两种形式，分别研究了这两类网络的零模型构造及应用，其中重点是无权无向网络。首先根据 dk 随机图的分布情况阐述了无权无向网络由低到高各阶零模型的定义，接着描述了使用 ER 随机图、配置模型和基于断边重连等方式构造不同阶数零模型的过程及相关应用，其中最重要的就是断边重连方式有无倾向性的随机断边重连、有倾向性的同配或异配断边重连，以及检测网络是否具有富人俱乐部性质的局部断边重连这几种。同时，本章也讨论了无权有向网络的 1 阶零模型以及如何基于该模型检测网络中存在的 4 种出入度匹配特性。本文不仅从理论上总结了各种无权网络零模型的具体实现过程，同时也画出交换过程的原理图、提供了生成网络零模型的代码和例子程序，有利于感兴趣的读者进行学习和实践。

2.1 引言

本章将主要围绕无权网络模型展开，无权网络包括无权无向网络和无权有向网络两种。无权无向网络是指网络中的连边既没有权重，也没有方向，如图 2-1 (a)所示。无权有向网络是指网络中的连边没有权重但有方向，如图 2-1 (b)所示。无权有向网络的连边可以是单向的，也可以是双向的，通过连边方向来区分节点之间关系的发出方和接受方。

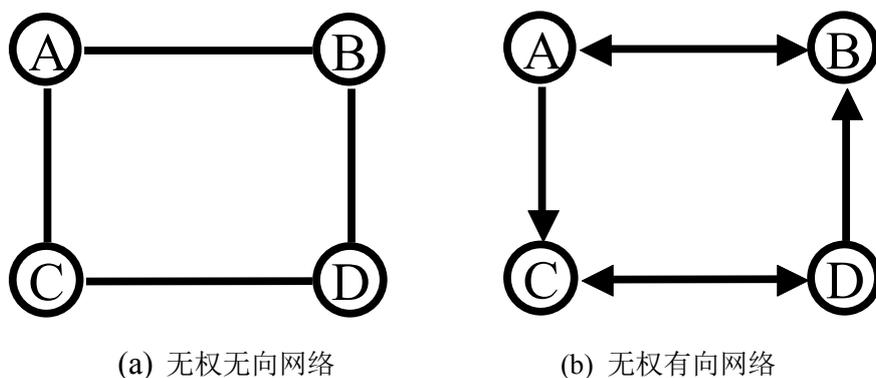


图 2-1 无权网络的两种形式

图 2-2 列出了无权网络零模型的分类以及零模型的产生过程。无权无向网络的零模型产生方法有两种：一种是基于网络模型的方法，基于网络模型生成的零模型网络可分为 ER 随机图和配置模型两种方式，其中 ER 随机图一般只能产生 0 阶零模型，而配置模型可产生 1 阶、2 阶或更高阶零模型。另一种是将原始网络进行置乱(断边重连)的方法，原始网络断边重连既可以进行随机断边重连，也可以使用有倾向性的断边重连。随机断边重连可以产生 0、1、2、2.5 阶网络零模型，其中 1 阶零模型可以用来检测原始网络的度匹配特性。有倾向性断边重连有两方面的应用：一是可用来检测网络是否具有富人俱乐部特性，二是可以用来检测网络的匹配特性。无权有向网络的零模型只有一种构造方式，即将原始网络断边重连，一般使用该方法来产生的 1 阶零模型，适用于检测有向网络中节点与节点之间的出入度匹配特性。

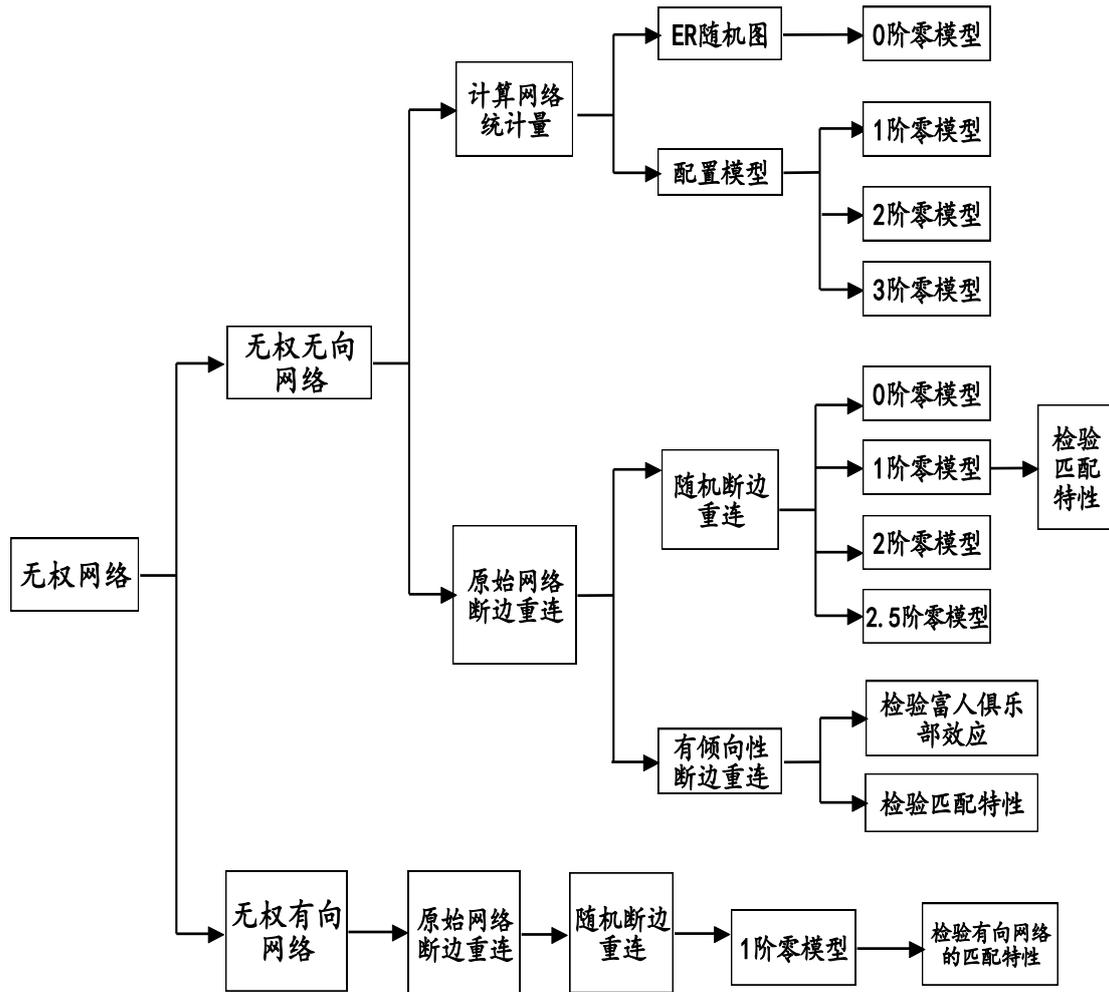


图 2-2 无权网络的分类及其零模型的产生过程

2.2 不同阶数网络零模型的定义

使用网络零模型主要的目的是检测实证网络的非凡特性，这就需要从粗糙到精确逐步逼近原始网络的性质。Mahadeven 等为了逐步逼近实证网络，根据原始网络的基本特性，可将网络中的基本组成单元分成不同的小模块如图 2-3(a)所示，最左端为原始网络，右端从右到左依次介绍了产生不同阶数的零模型所需要原始网络的基本性质，从而引入了不同阶数网络零模型的概念^[1,2]，如图 2-3(b)所示。不同阶数网络零模型之间并不是独立的，按照约束条件从少到多，存在一种包含关系，即 $0K \supseteq 1K \supseteq 2K \supseteq \dots \supseteq (n-1)K \supseteq nK$ ，任何一个 n 阶零模型都会包含 $(n-1)$ 阶零模型的性质^[3]。

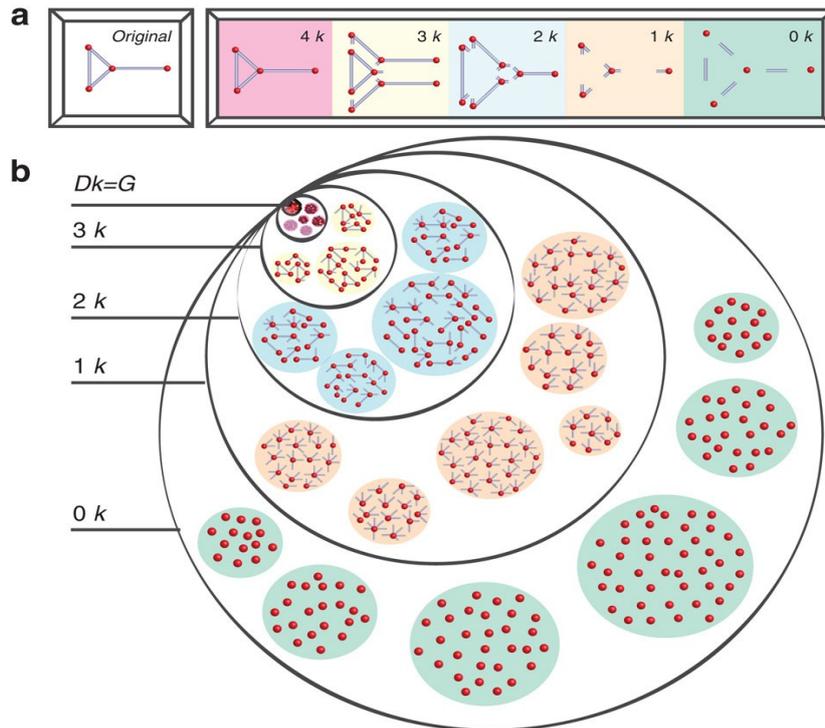


图 2-3 不同阶数网络零模型之间的层次关系图(取自文献[4])

2.2.1 0阶零模型

0阶零模型是最简单也是最随机化的网络零模型，只需要与原始网络具有相同的节点数和平均度。平均度 $\langle k \rangle$ 是指原始网络中所有节点的度的平均值，设 m 为原始网络连边的数量， n 为原始网络节点的数量，则 $\langle k \rangle = 2m/n$ 。

2.2.2 1阶零模型

1阶零模型与原始网络不但具有相同的节点数、相同的平均度，更重要的是具有相同的节点度分布 $p(k)$ ，度分布是指原始网络中节点度的概率或数目的分布。若 $n(k)$ 为原始网络中 degree of k 的节点的数目，则度分布为 $p(k) = n(k)/n$ 。1阶零模型主要是保证与原始网络具有相同的节点度分布。

2.2.3 2 阶零模型

2 阶零模型与原始网络具有相同的联合度分布 $p(k_1, k_2)$ ，联合度分布是指每条边两端连接节点的度值数目(概率)。若 $m(k_1, k_2)$ 为度为 k_1 和 k_2 的节点之间连边的总数，则

$$p(k_1, k_2) = m(k_1, k_2) \mu(k_1, k_2) / 2m \quad (2.1)$$

其中，若 $k_1 = k_2$ 则 $\mu(k_1, k_2) = 2$ ，否则 $\mu(k_1, k_2) = 1$ 。

2.2.4 2.25 阶零模型

2.25 阶零模型是根据原始网络中的联合度分布和平均聚类系数这两个物理属性构造的，要求与原始网络具有相同的联合度分布 $p(k_1, k_2)$ 和平均聚类系数 \bar{C} 。假设网络中的一个节点 i 有 m_i 条边将它和其他节点相连，这 m_i 个节点就称为节点 i 的邻居节点。显然，在这 m_i 个节点之间最多可能有 $\frac{m_i(m_i - 1)}{2}$ 条边。 m_i 个节点之间实际存在的边数 M_i 和总的可能边

数 $\frac{m_i(m_i - 1)}{2}$ 之比就定义为节点 i 的聚类系数 C_i ，即

$$C_i = \frac{2M_i}{m_i(m_i - 1)} \quad (2.2)$$

平均聚类系数 \bar{C} 定义为在原始网络中，所有节点的聚类系数的平均值，即

$$\bar{C} = \frac{1}{N} \sum_i C_i \quad (2.3)$$

2.2.5 2.5 阶零模型

2.5 阶零模型与原始网络具有相同的联合度分布 $p(k_1, k_2)$ 和度相关的平均聚类系数 $\bar{C}(k)$ ，即

$$\bar{C}(k) = \frac{1}{|N_k|} \sum_{i \in N_k} C_i \quad (2.4)$$

其中， N_k 表示度为 k 的节点的集合。

2.2.6 3 阶零模型

3 阶零模型与原始网络具有相同的联合边度分布 $p(k_1, k_2, k_3)$ ，联合边度分布考虑到三个节点之间的相互连接性，主要有两种情况，一种是开三角形如图 2-4(a)所示，即 3 个节点用 2 条边来连接称为 $p_{\wedge}(k_1, k_2, k_3)$ ；另一种是闭三角形如图 2-4(b)所示，3 个节点形成一个环称为 $p_{\Delta}(k_1, k_2, k_3)$ 。与原始网络具有相同的联合边度分布 $p(k_1, k_2, k_3)$ ，就要求零模型网络和原始网络具有相同的开三角形和闭三角形分布。

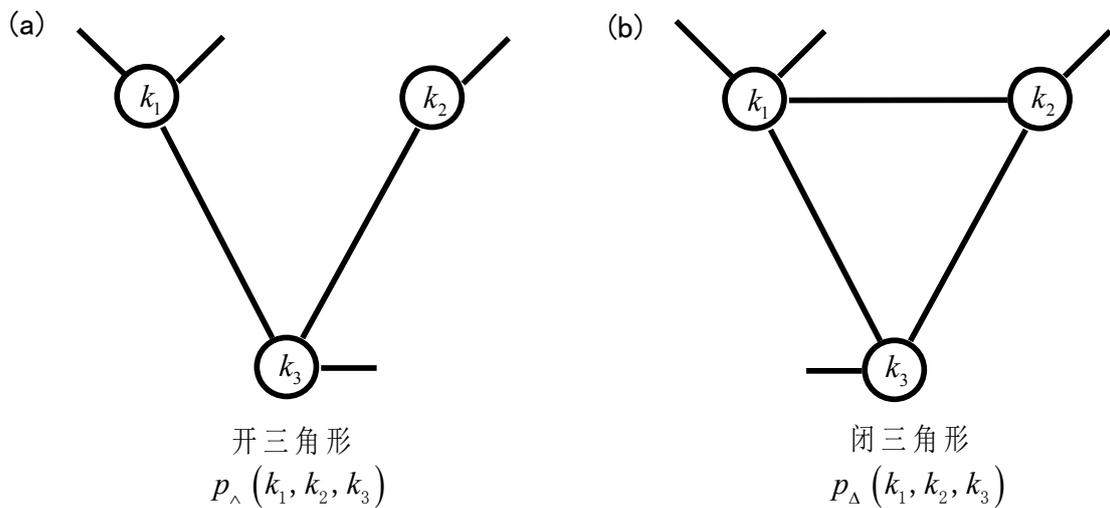


图 2-4 三个节点相互连接的两种可能形式

随着生成零模型的约束条件逐步增加，构造的高阶零模型比低阶模型更接近于原始网络，因此精确的 3 阶零模型比 0, 1, 2, 2.25 和 2.5 阶零模型更接近于原始网络。以此类推下去，理论上我们还可以定义更高阶的零模型^[3]。随着约束条件越多，可以被置乱的连边数也就越少，生成的网络零模型越接近于原始网络，但是这样进行连边置乱也就越困难。对于 3 阶网络零模型，该零模型受到的约束条件已经很多，很多情况下我们不能在一个有限的计算时间内生成相应的零模型，因此不具备实用性，2.5 阶零模型是现在能够实现的最高阶的实用零模型^[1,2]。因此，在本文中我们重点介绍可以将原始网络进行较高程度随机化的 0-2.5 阶网络零模型。

2.3 基于 ER 随机图的 0 阶零模型

构造静态无权网络零模型的方法主要有基于模型的 ER 随机图、配置模型和基于随机断边重连这 3 种。

20 世纪 60 年代匈牙利的 Erdos 和 Renyi 这两位数学家对随机图理论做出了重要贡献，以他们名字简写的 ER 随机图是最为大家熟知的随机网络。ER 随机图可以由两种方式构造出来。一种是给定 n 个节点和 m 条连边，随机的从 n 个节点中选择两个没有相连的节点用连边连接起来，重复上述过程直到连边数达到 m 条连边后停止选择，构造出没有重边和自环的简单图。另一种是给定 n 个节点，以概率 p 任意连接两个不同的节点，连通概率 p 介于 0 和 1 之间，重复上述过程直到相连的边数的概率达到 p ^[5]。尽管 ER 随机图在很长一段时间内被大家广泛使用，但该模型是一种完全随机的网络，度分布是泊松的、聚类系数接近于零，这些性质和实际复杂网络中幂律度分布和高聚类特性相差很大^[6]。

通过将实际网络与它对应的 ER 随机图进行比较，可以证实实际网络是否具有小世界特性。如表 2-1 中，Barabasi 等构造了 WWW、电影演员网络、电力网这三个网络的 ER 随机图，并将这 3 个随机图中的聚类系数和平均距离这两个统计量与实际网络中相应的统计量进行定量的比较，可以看到在实际网络和 ER 随机图具有相同的节点数和平均度的情况下，它们的平均距离长度几乎相同。同时，实际网络中的聚类系数远远高于相同规模的 ER 随机图，因此我们可以说聚类特性是原始网络所特有的属性而不能通过 ER 随机图产生，即实证网络具有小的平均路径长度和高的聚类系数，也就是我们常说的“小世界特性”。

表 2-1 实际网络与 ER 随机图之间的统计量比较（取自文献[7]）

网络	N	$\langle k \rangle$	L	L_{ER}	C	C_{ER}
WWW	153127	35.20	3.10	3.40	0.11	0.00023
电影演员	225226	61.00	3.60	2.90	0.79	0.00027
电力网	4941	2.70	18.70	12.40	0.08	0.005

此外，我们也可以采用随机断边重连的方式产生 0 阶零模型，但是无论采用哪种方式，0 阶零模型只和原始网络具有相同的节点数和边数，生成的网络结构随机性很强，性质和原始网络相差较大。ER 随机图的生成过程可以使用基于 Python 语言的 NetworkX 包来实现，具体语句为：

```
import networks as nx
```

```
G = nx.random_graphs.erdos_renyi_graph(n, p)
```

其中，输入的参数 n 为原始网络的节点个数，指定生成边的概率 p ， p 值应为原始网络的连边数 m 与节点数 n 平方的比值，即 $p = 2m / (n * n)$ 。图 2-5 为可视化的 HOT 原始网络模型及相应的 ER 随机图。

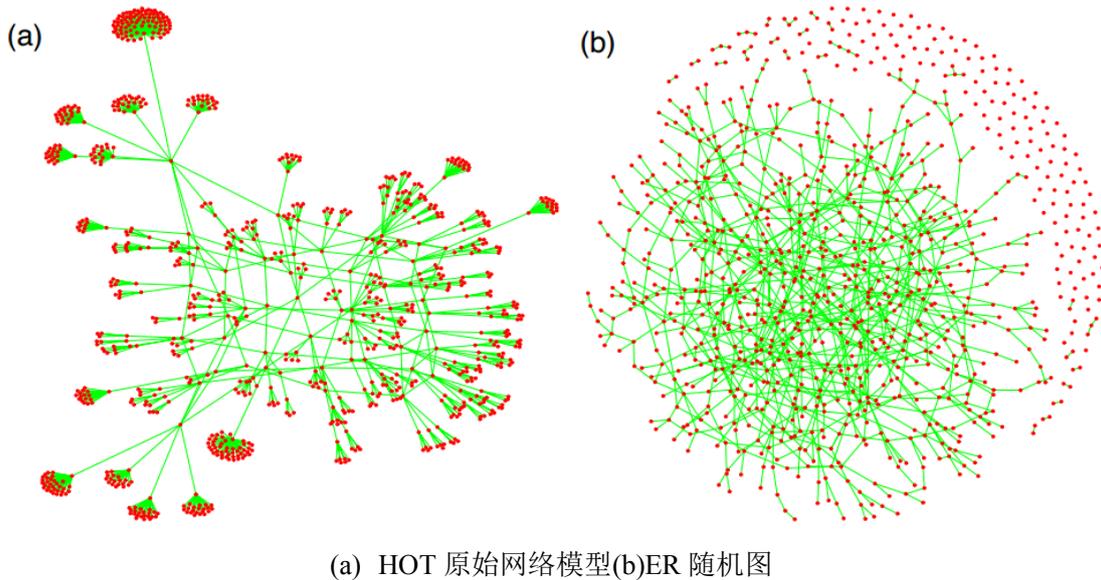


图 2-5 HOT 原始网络模型及相应的 ER 随机图

2.4 基于配置模型的零模型

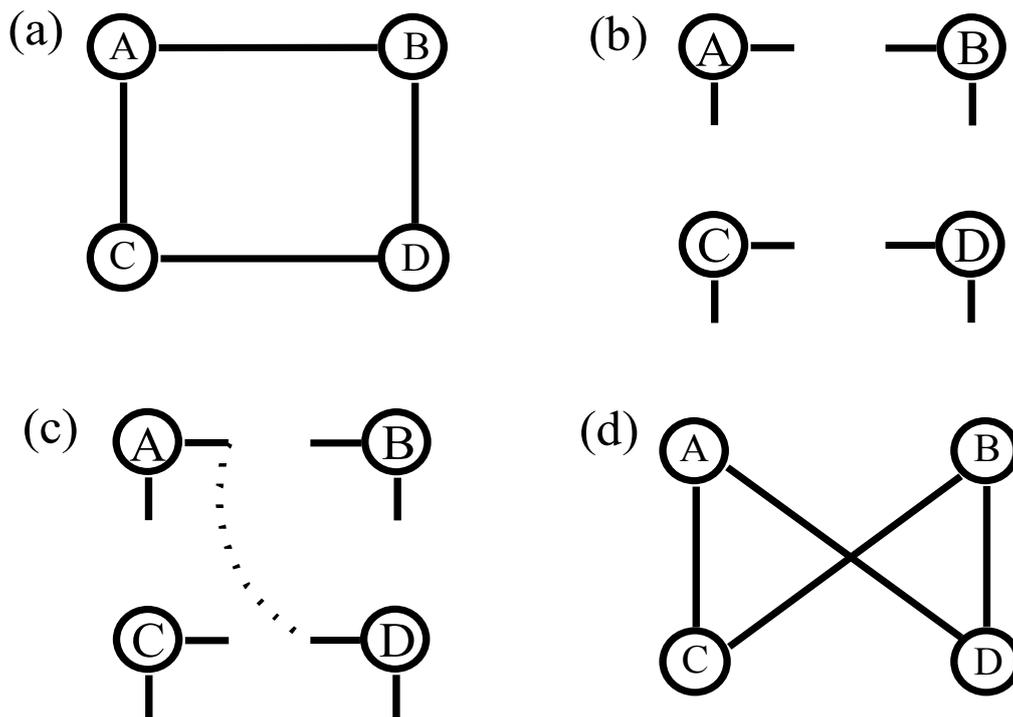
配置模型是在 ER 随机图的基础之上进行扩展产生的一种广义随机图。本节主要介绍基于配置模型构造网络零模型的方法，由配置模型生成零模型的基本思想是：根据产生不同阶数零模型所需要的基本特性，将原始网络拆成相应的含有某种特性的基本子图，并通过不同的约束条件将基本子图重新随机组合到一起的过程。

在本节中主要介绍基于配置模型生成 1 阶、2 阶、3 阶零模型网络的过程。在 1 阶配置模型中通过给定原始网络的度序列来构造零模型网络，2 阶配置模型通过给定原始网络中的联合度分布序列来构造零模型网络，3 阶配置模型更具体的给定了每个节点中边角的分布情况，即原始网络中的“线”和“角”序列来构造零模型网络^[3]。下面将逐一介绍这 3 种不同阶数的配置模型。

2.4.1 基于配置模型的 1 阶零模型

ER 随机图是在保证与原始网络中节点数和边数相同的情况下，采用随机连边方式生成的零模型。它仅仅保证了与原始网络具有相同的平均度分布，是随机性最强的一种网络。为了比 ER 随机图更精确地拟合原始网络，构造出与真实网络度分布相同的零模型，Molloy 等提出了广义随机图，也就是我们熟知的配置模型^[8,9]。

在 1 阶配置模型中，首先给定了原始网络的度序列 $\{k_i\}$ ，在此前提下使用“配置”连边的方式生成了新的零模型网络，下面我们阐述一下具体的构造过程。如图 2-6 所示：将原始网络图 2-6(a)拆分成 A、B、C、D 四个独立的节点，在拆分的过程中保持每个节点的度值不变，从每个节点中引出与度值数目相同的“线头”，如图 2-6 (b)所示，短线即为引出的“线头”。然后每次随机的选择一对线头连接在一起形成一条边，如图 2-6(c)，所有的线头都被连完所形成的网络(如图 2-6(d)所示)即为配置模型的 1 阶随机网络。1 阶配置模型给定了真实网络的度序列，最终可以生成具有与真实网络具有相同度序列的随机网络^[10-12]。



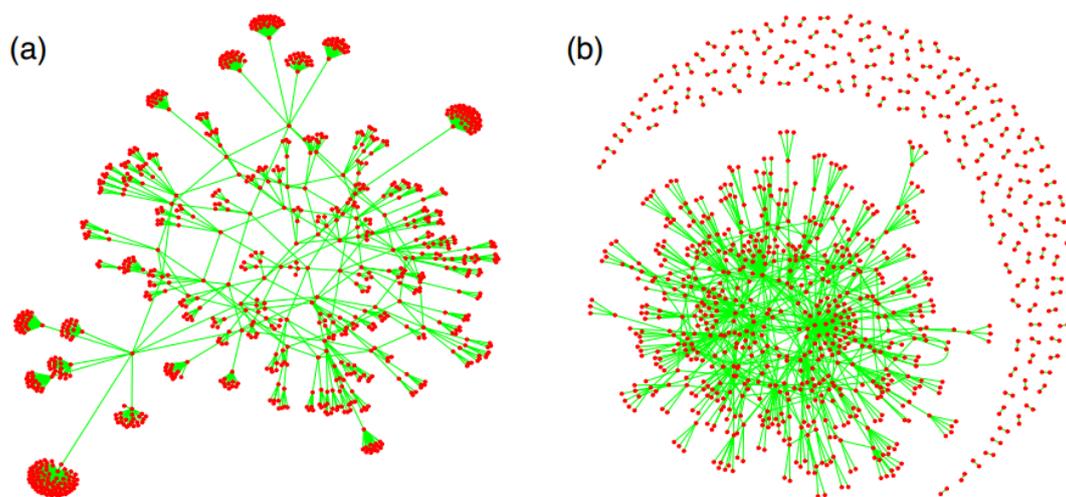
(a)原始网络(b)节点度分布图(c)节点之间连接的生成过程(d)原始网络的 1 阶配置模型

图 2-6 原始网络和使用配置模型构造的 1 阶零模型网络

NetworkX 算法包中已经实现了网络的 1 阶配置模型，假定原始网络为 G ，应首先调用 $G.degree().values()$ 方法得到原始网络的度序列，然后根据该度序列生成配置网络图 G 。具体代码如下：

```
import networkx as nx  
  
degree_seq = list(G.degree().values()) # 得到原始网络 G 的度序列  
  
G1 = nx.configuration_model(degree_seq) # G1 为原始网络 G 的 1 阶配置模型
```

图 2-7 显示的是 HOT 原始网络模型和使用配置模型构造的 1 阶零模型。

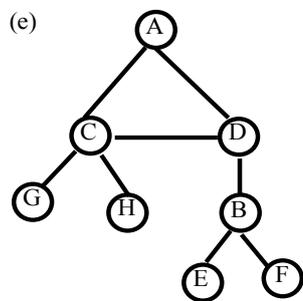
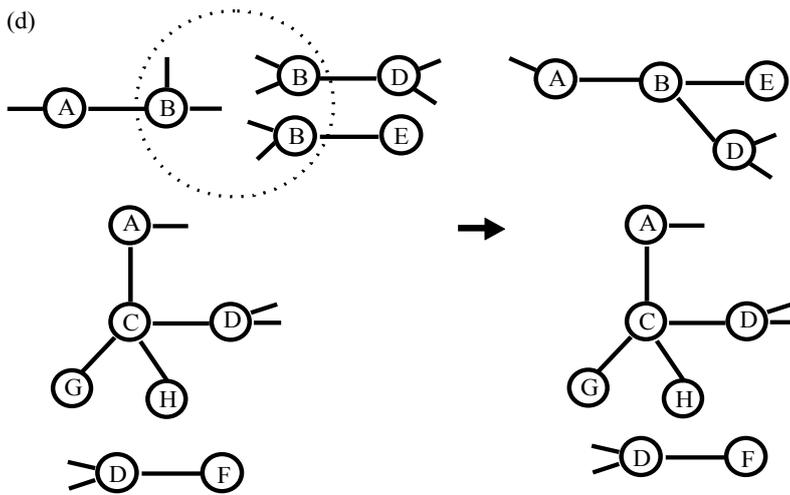
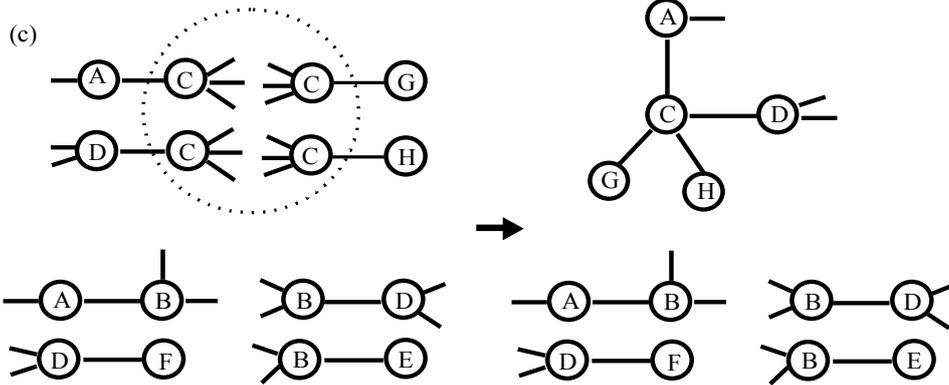
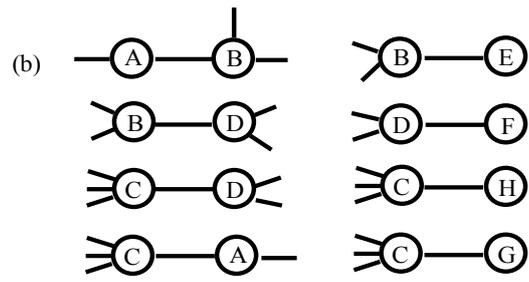
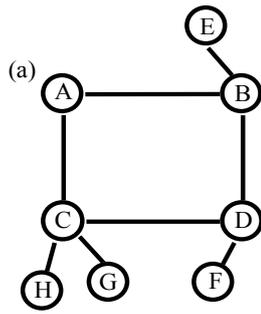


(a)HOT 原始网络模型(b)配置模型的 1 阶零模型

图 2-7HOT 原始网络模型和使用配置模型构造的 1 阶零模型

2.4.2 基于配置模型的 2 阶零模型

2 阶零模型网络要求与原始网络具有相同的联合度分布，使用配置模型也可以生成原始网络的 2 阶零模型。假定原始网络有 m 条边，首先我们将该网络拆成 m 个子图，每个子图为两个节点 $\{i, j\}$ 配对的形式。然后随机选取一个子图，并将子图中度值为 k_i 的节点 i 视为起始节点。随机的在网络中挑选 $k_i - 1$ 个子图，并且每个子图中至少含有一个度值为 k_i 的节点，最后将这 k_i 个节点重新组合成一个新的度为 k_i 的节点。持续随机选取一个分支节点作为起始节点重复上述过程，直到将所有的子图都重新组合就完成了 2 阶零模型的生成。



(a)原始网络(b)联合度分布图(c)(d)节点之间连接的生成过程(e)基于配置模型的 2 阶零模型

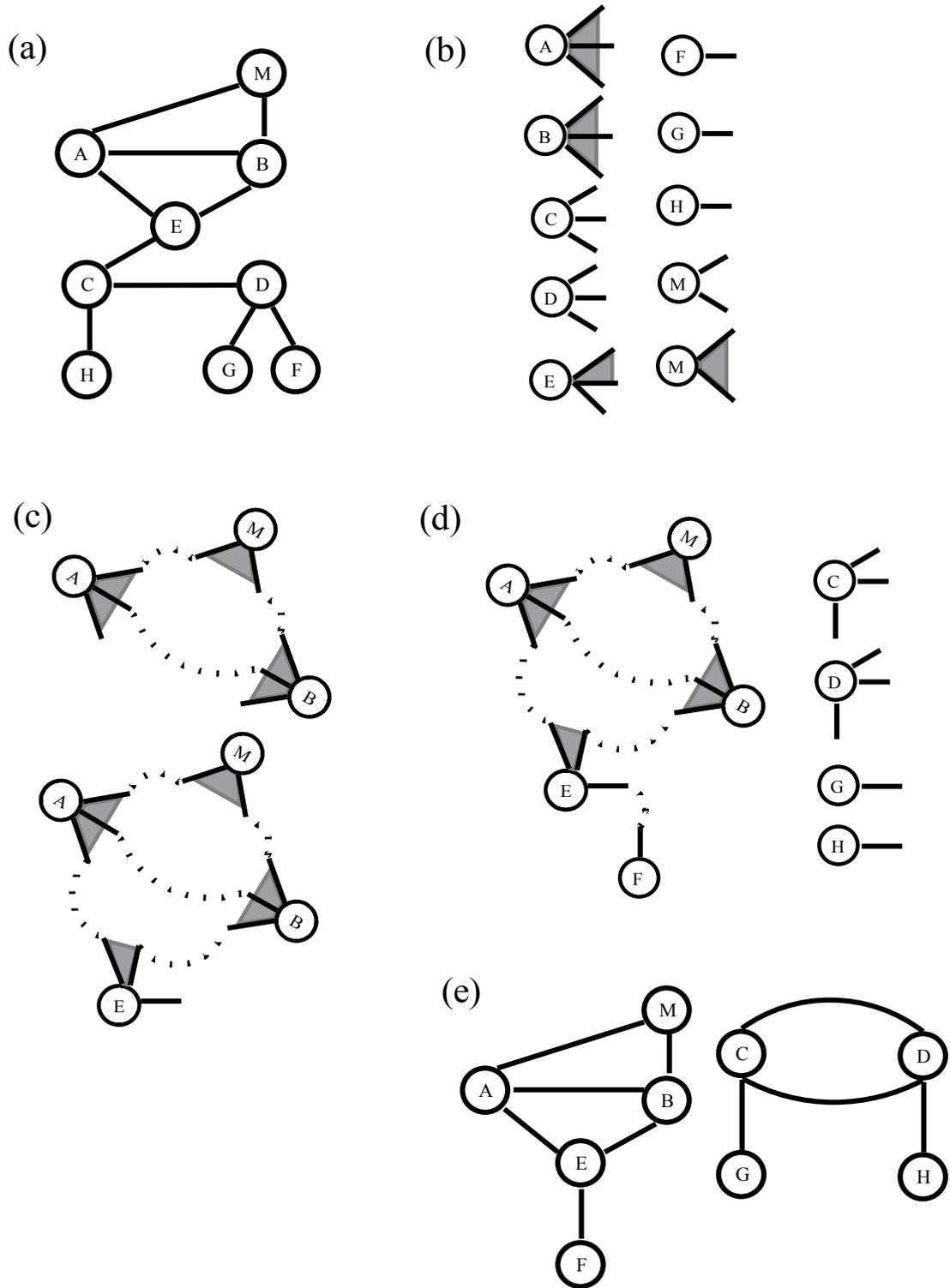
图 2-8 原始网络和使用配置模型构造的 2 阶零模型网络

为了让读者更好理解 2 阶配置模型的构造过程，我们以图 2-8 为例进行更细致的说明。原始网络图 2-8(a)中有 8 条边，因此原始网络可根据边数拆成如图 2-8(b)所示的 8 个子图，随机选取其中的一个 2 阶子图（如含 C 和 G 节点的子图），将子图中的节点 C 视为起始节点，由于节点 C 的度为 4，因此需要从图(b)中再随机的选取 3 个子图，每个子图中至少有一个节点的度值为 4，图 2-8 (c)显示了这 4 个度值为 4 的节点重新随机组合成一个新子图的过程，进行组合后形成大子图的结果为图 2-8 (c)所示。接着，可以从图 2-8(d)中随机选取一个有分支的新节点作为起始节点，重复聚合子图的过程，直到把所有 2 阶子图的分支都得到链接最终生成了如图 2-8 (e)所示的 2 阶零模型。由于在子图的聚合过程中，每条边两端节点的联合度都不会发生变化，因此保证了与真实网络具有相同的联合度分布。

2.4.3 基于配置模型的 3 阶零模型

许多实际网络都具有明显的聚类特性，而实际网络的高聚类性却没有在低阶配置模型，如 0 阶、1 阶和 2 阶模型中体现出来。为了解决这一缺陷，Newman 等提出了能够保持实际网络聚类特性的零模型，这也就是下面要介绍的 3 阶配置模型^[13]。

在 3 阶零模型中将原始网络分成“线”和“角”两种基本元素，这种形式的拆分保证了网络中各节点的聚类系数不变。我们定义 s_i 为与节点 i 相连的不构成三角形的单条边的数目， t_i 为节点 i 所参与的三角形中角的数目。因此，对于每个节点都可以得到它的联合“线”和“角”序列，即 $\{s_i, t_i\}$ ，每一个线角序列也可以称为一个线角子图。在给定网络联合线角序列的前提下，可以随机选取一对线连接它们形成一条连边，也可以随机选择不是同一节点的三个角连接它们形成完整的三角形。上述线和角聚合的过程可以同时进行，也可以随机进行，重复聚合过程用完所有的线角子图后就生成了 3 阶零模型。



(a)原始网络 (b)联合度序列图(c)(d)生成过程图(e)配置模型的 3 阶图
图 2-9 原始网络和使用配置模型构造的 3 阶零模型网络

为了让大家更好理解 3 阶配置模型的构造过程，下面以图 2-9 为例进行更细致的说明。根据前面的联合边角序列的定义，可以得到图 2-9(a)中所有节点的联合边角序列值： $A(0, 2)$ 、 $B(0, 2)$ 、 $C(3, 0)$ 、 $D(3, 0)$ 、 $E(1, 1)$ 、 $F(1, 0)$ 、 $G(1, 0)$ 、 $H(1, 0)$ 、 $M(0, 1)$ 。根据该联合边角序列，可以将原始网络拆成如图 2-9(b)所示的 9 个边角子图，子图中的灰色阴影部分代表角，黑色实线代表单独存在的边。可以在这些联合边角序列图中随机的选取一对黑线连接它们形成一条连边，也可以随机的选择不是同一节点的三个角连接它们形成完整的三角形，进行聚合后生成的大子图和剩余的小子图如图 2-9(c)(d)所示。不断重复角聚合过程或边聚合过程，用完所有的线角子图后就生成了 3 阶零模型如图 2-9(e)所示。

此外，依据 3 阶配置模型的生成原理，也可以生成和原始网络具有相同高阶子图分布的配置模型，大大拓展了配置模型的使用范围^[13, 14]。但需要注意的是，尽管配置模型可以精确展现真实网络的度序列、联合度分布、甚至是网络的聚类特性和子图分布，但配置模型也有如下缺点：一是算法比较复杂，因此针对加权网络和时变网络很难使用；二是配置模型产生的网络为了保证连边配对的随机性，可能会存在自环和重边。针对第二个缺陷，可以采用移除自环和重边这种做法，但这种方式也有一些待商榷之处，因为这样生成的网络不是从所有可能的匹配集中均匀选择网络，这意味着模型的很多性质不再能够利用解析方法进行计算，至少无法利用当前已知的方法计算^[15]。同时由于存在自环和重边，配置模型的连通性等重要指标和原始网络差异较大，不是具有较强的实用价值，因此在本书中仅我们介绍了 2 阶及 3 阶配置模型的构造方法，没有提供具体的编程实现供大家参考和使用。

2.5 基于随机断边重连的零模型

用 ER 随机图或配置模型的方法构造的零模型是从无到有生成新网络的过程，而用置乱算法构造的零模型则是将原始网络随机化的过程^[16, 17]。在静态无权网络中常用的置乱算法就是随机断边重连，断边重连的方法主要是在原始网络的基础上将网络中原有的连边随机的断开重连，使原始网络模型尽可能的随机化。在静态无权网络中，相对于配置算法，连边随机重连更简单、更容易操作，不需要理解和运用复杂的数学公式、也不会产生自环和重边现象，却能精确保持真实网络的一些物理属性，因此被广泛应用到实际各种类型的网络分析中。

2.5.1 基于随机断边重连的 0 阶零模型

ER 随机图可以作为网络的 0 阶零模型，本节采用随机断边重连的方式也可以生成网络的 0 阶零模型，它们的共同点是都与原始网络具有相同的平均度分布。不同的是，ER 随机图是根据原始网络的节点数和边数生成的新网络，而随机断边重连的 0 阶零模型是在原始网络的基础上进行随机断边重连得到的。

基于随机断边重连的 0 阶零模型构造过程如图 2-10 所示：先随机的从原始网络中删除一条连边 AC，再随机选择网络中两个不相连的节点，如图 2-10(a)所示中 A、D 节点，在 A 与 D 之间添加一条连边，重连后的结果如图 2-10(b)所示。为了使网络充分随机化，一般应根据网络的规模将上述断边重连过程重复多次，直到原始网络足够随机化后才认为生成了对应的零模型。

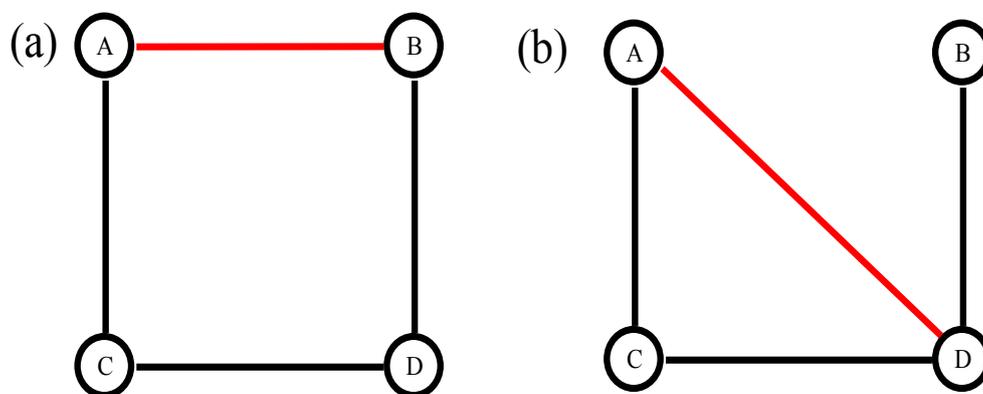
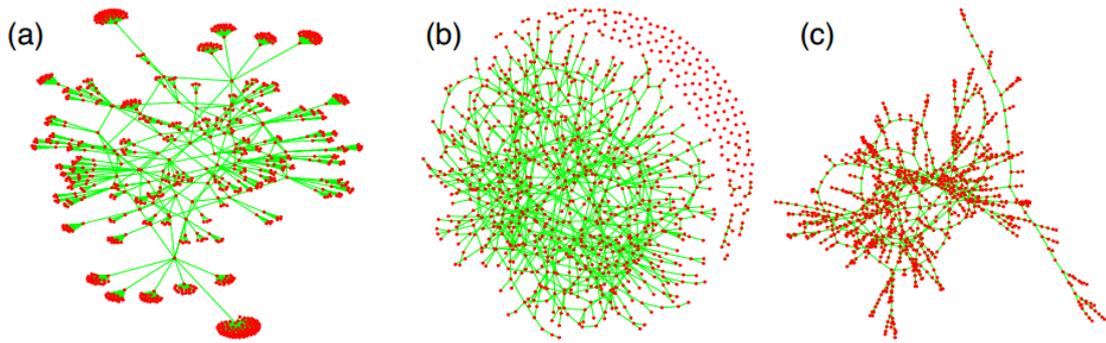


图 2-10 原始网络和使用断边重连算法构造的 0 阶零模型网络

随机断边重连的 0 阶零模型的程序实现见附录程序，输入参数为原始网络 G_0 ，交换次数 n_{swap} 和最大尝试次数 max_tries ，输出为基于随机断边重连的 0 阶零模型 G 。上述生成的网络零模型没有保证网络的连通性，如果原始网络是一个全联通网络，我们可以采用程序约束生成的零模型网络也是全联通的。图 2-11 为 HOT 原始网络模型和在无连通性及有连通性情况下使用断边重连构造的 0 阶零模型。



(a)HOT 原始网络模型(b)0 阶零模型(c)保证联通性情况下构造的 0 阶零模型

图 2-11HOT 原始网络模型和在无连通性及有连通性情况下使用断边重连构造的 0 阶零模型
 增进群等人在保证与大众生产合作网络具有相同节点数边数的前提下,对合作网络随机断边重连构造了 0 阶零模型,并计算了原始网络与零模型网络的集聚系数和平均最短路径长度等多个静态统计量^[18],如表 2-2 所示。

表 2-2 实证网络与 0 阶零模型网络统计量对比表 (数据取自文献[18])

静态参数	0 阶零模型网络	实证网络
节点数	36527	36527
平均度	13.3	13.3
集聚系数(CC)	3×10^{-5} (平均值)	0.928
平均最短距离(PL)	4.351 (平均值)	4.438
度分布	泊松分布	幂律分布

通过分析表 2-2 的多个统计量,可以看出实证网络与 0 阶零模型网络的平均最短距离长度相差无几,这说明平均最短距离较小不一定是实际网络所特有的性质。实际网络中的集聚系数远远高于相同规模 0 阶零模型的集聚系数,因此我们可以说集聚特性是原始网络所特有的属性而非随机造成的,同样通过对比可以得出实际网络符合幂律分布而非 ER 图等随机模型的泊松分布。通过由零模型作参照物比较原始网络和 0 阶零模型的集聚系数、平均最短距离及度分布的参数指标证明了开源社区网络具有小世界特性和无标度特性。

2.5.2 基于随机断边重连的 1 阶零模型

从 0 阶零模型到 1 阶零模型,随着零模型阶次的升高,约束条件也会进一步加强,导致了对原始网络断边重连的要求升高,原始网络断边重连的可能性降低,因而生成的网络随机

化程度也降低。1 阶零模型的生成条件要高于 0 阶零模型的生成条件，1 阶零模型需要在度分布不变的前提下进行随机断边重连。

1 阶零模型算法的连边置乱过程如图 2-12 所示。如果网络中有节点 A 与 B 相连，C 与 D 相连，且节点 A 与 D 不相连，B 与 C 不相连，具体情况如图 2-12(a)所示。我们就将连边 AB 和 CD 切断，再使节点 A 与 D 相连，B 与 C 相连，最终生成图 2-12(b)。根据网络规模 and 实际需要不断重复上述过程就会使网络中的连边关系被随机置乱，但是置乱前后每个节点的度均保持不变，保证了置乱前后网络的度序列相同。从图 2-12 可以看出，连边置乱算法破坏了网络拓扑结构，因此可以用来研究网络拓扑结构，并研究网络拓扑结构变化对网络功能的影响。

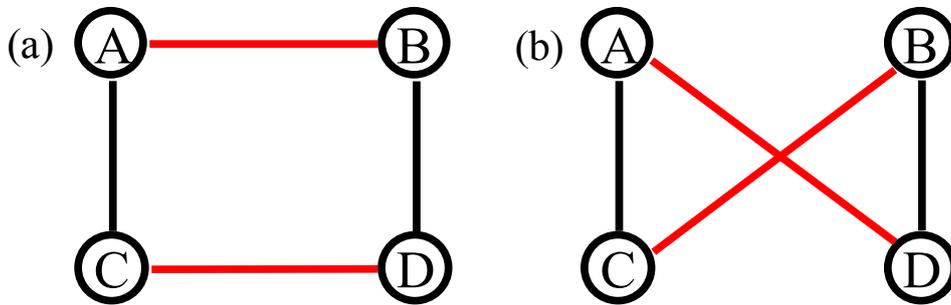
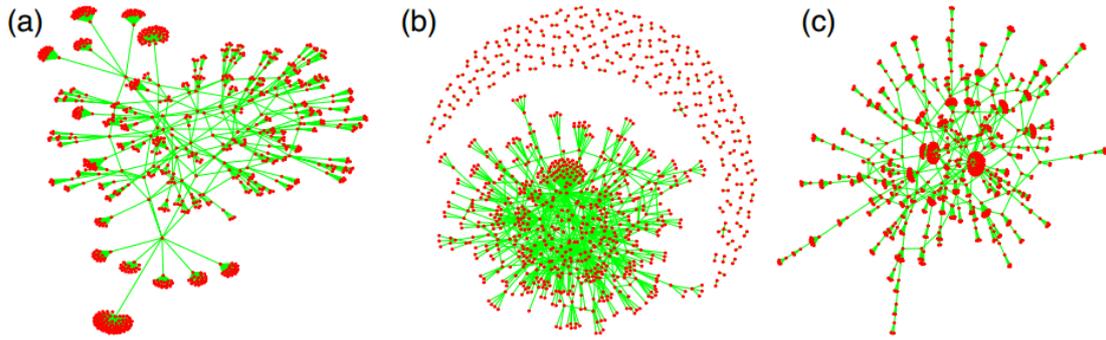


图 2-12 原始网络和使用断边重连算法构造的 1 阶零模型网络

由于在 networkX 算法包中已经编写了 `nx.double_edge_swap()` 方法完成从原始网络中任选两条边进行断边重连，因此基于随机断边重连生成 1 阶网络零模型比较容易，具体程序如下所示：

```
G = nx.double_edge_swap(G0, nswap, max_tries)
```

输入参数为原始网络 `G0`，交换次数 `nswap` 和最大尝试次数 `max_tries`，输出为基于随机断边重连的 1 阶零模型 `G`。上述生成的网络零模型没有保证网络的连通性，如果原始网络是一个全联通网络，我们可以采用程序约束生成的零模型网络也是全联通的。在程序清单中输入参数为具有连通性的原始网络 `G0`，交换次数 `nswap` 和最大尝试次数 `max_tries`，输出为基于随机断边重连的 1 阶零模型且保持网络连通性的网络 `G`。复用 `nx.double_edge_swap()` 方法的部分代码从原始网络中随机选两条边，然后断开该连边并将不相连两节点相连，若网络不保持连通性，则撤销该操作。图 2-13 为 HOT 原始网络模型和在无连通性及有连通性情况下使用断边重连构造的 1 阶零模型。



(a)HOT 原始网络模型(b)1 阶零模型(c)保证联通性的情况下构造的 1 阶零模型

图 2-13HOT 原始网络模型和在无连通性及有连通性情况下使用断边重连构造的 1 阶零模型

1 阶零模型是原始网络非常重要的参照物。为了确定某种非平凡特性是否来自原始网络，需要确定关于这一特性的原始网络本身或其生成的相关指数是否显著的异于零模型和零模型生成的相关参数^[19]。假设某种特性所构成的子图（模体）在一个实际网络中出现的次数为 $N(j)$ ，而在零模型中出现的次数的平均值为 $\langle N_r(j) \rangle$ ，则两者的比值为：

$$R(j) = \frac{N(j)}{\langle N_r(j) \rangle} \quad (2.5)$$

在实际应用中，如果 $R(j) > 1$ ，那么意味着实际网络的设计或者演化过程促进了该拓扑特征的出现。相反，如果 $R(j) < 1$ ，则意味着实际网络的设计或者演化过程抑制了该拓扑特征的出现。

在统计学中 Z 检验方法(平均值差异检验方法)是用标准正态分布的理论来推断差异发生的概率，从而比较两个平均数的差异是否显著。因此我们可以用 Z 检验方法进一步检验某种特性在实际网络中出现的频率和在零模型中出现的频率差异的显著性。具体地说拓扑特征 j 的统计重要性可以用公式(2.6)来刻画：

$$Z(j) = \frac{N(j) - \langle N_r(j) \rangle}{\sigma_r(j)} \quad (2.6)$$

其中， $\sigma_r(j)$ 是零模型中某个拓扑特征 j 出现的平均次数 $\langle N_r(j) \rangle$ 的标准差。Z 值的绝对值越大就表示差异越显著，越能说明拓扑特征 j 产生于原始网络，而非随机造成的。

Z 值与网络的规模有关，网络的节点越多 Z 值就越高。因此不同的网络之间也就没有了可比性^[19]。为此通常的做法是将不同网络上的评判结果统一到一个尺度上，对 Z 值做归一化处理使其标准化后就得到了如公式(2.7)所示的重要性剖面（Significance Profile, SP）。

$$SP_j = \frac{Z_j}{\left(\sum Z_i^2\right)^{1/2}} \quad (2.7)$$

无向网络的断边重连 1 阶零模型网络最早被 Maslov 等运用于蛋白质网络的研究中^[28]，在蛋白质网络中节点代表蛋白质，实验表明两个蛋白质之间如果能够形成化学键就相互连接。在对网络子图结构的研究中，Shen-Orr 等利用连边置乱算法构造了网络零模型，发现一些特定网络子图在原始网络中出现的频率远高于在零模型网络中出现的频率^[20, 21]。Milo 等还对比了原始网络与连边置乱零模型中所有子图出现的频率，提出了重要性剖面算法将来自不同领域的 19 个网络分成了四个超家族^[22]。

实际网络中可采用相关性剖面来量化度相关特性。设 $m(k_0, k_1)$ 设为度为 k_0 和 k_1 的节点之间相连边的总数，可以通过比较一个实际网络的 $m(k_0, k_1)$ 和其 1 阶零模型所对应的均值 $\langle m_r(k_0, k_1) \rangle$ 来分析实际网络的度相关特性。具体的做法是通过在平面上绘制出相关性剖面（Correlation Profile, CP）来刻画实际网络的度相关性，具体的 R 值和 Z 值相关性剖面计算方式如下所示：

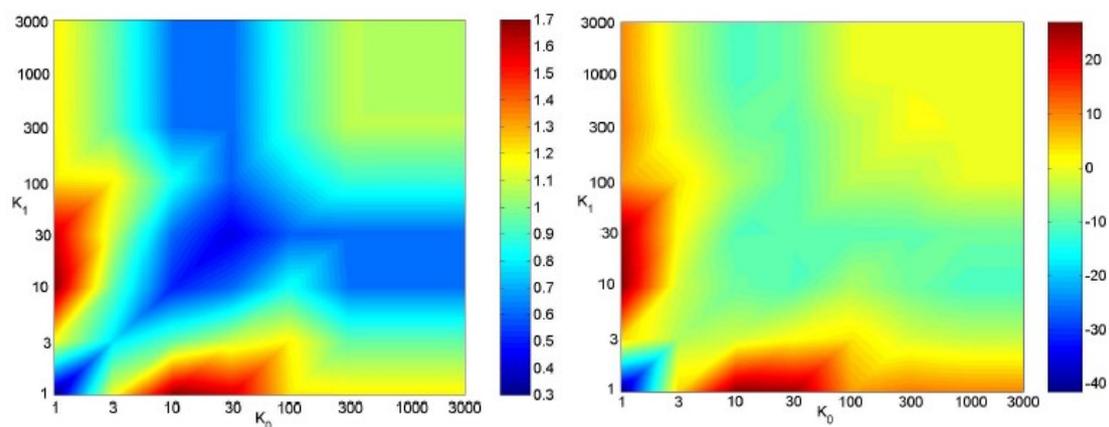
$$R(k_0, k_1) = \frac{m(k_0, k_1)}{\langle m_r(k_0, k_1) \rangle} \quad (2.8)$$

$$Z(k_0, k_1) = \frac{m(k_0, k_1) - \langle m_r(k_0, k_1) \rangle}{\sigma_r(k_0, k_1)} \quad (2.9)$$

Maslov 等人将 AS 网络和利用连边置乱生成的零模型进行对比，通过量化网络相关性剖面的新方法检验了 Internet 网络中的度度相关特性^[23]。图 2-14 显示的是 AS 网络相关性剖面图，图 2-14(a)绘制的是 R 值分布图，竖条表示的是 $R(k_0, k_1)$ 值数值分布对照图。当 $R(k_0, k_1) > 1$ 时表明在度为 (k_0, k_1) 的节点之间的连接强度比相应的随机网络更强，当 $R(k_0, k_1) < 1$ 时表明在度为 k_0, k_1 的节点之间的连接强度比相应的随机网络更弱。图 2-14(b)绘制的是 Z 值分布图，

竖条是 $Z(k_0, k_1)$ 值数值分布对照图, Z 的绝对值越大就表示差异越显著。总的来说, R 值是为了检测原始网络度度相关程度是受到了促进还是抑制, Z 值进一步检测这些连接强度(促进或抑制)的强烈程度。

通过图 2-14 可以看出: 当节点的度值分布在 $3 \geq K_0, K_1 \geq 1$ 时, 通过公式(2.8)计算得出的 $R(k_0, k_1)$ 值从图 2-14(a)中可以看出在 0.3 附近, 表明了连接强度比相应的随机网络更弱, 为了进一步刻画拓扑特征的显著性差异是否明显, 因此用公式(2.9)计算得出了 $Z(k_0, k_1)$ 值的绝对值并通过图 2-14(b)可以看出为 35, 通过 Z 值的刻画可以说明原始网络的度度相关性与随机网络相比差异非常显著, 因此可以说明当节点的度在 $3 \geq K_0, K_1 \geq 1$ 范围时连边受到很强的抑制。也可以说是度值小的节点倾向于和度值小的节点相连的这种度相关性受到了很强的抑制。



(a) $R(k_0, k_1)$ (b) $Z(k_0, k_1)$

图 2-14AS 网络的相关性剖面图(取自文献[23])

2.5.3 基于随机断边重连的 2 阶零模型

静态无权网络中最常用的就是使用随机断边重连来产生 1 阶复杂网络零模型, 为了更精确的刻画网络的拓扑结构, 可以用随机断边重连的方法构造 2 阶和更高阶次的零模型^[2, 3]。高阶零模型的随机重连算法可以看作 1 阶随机断边重连算法的拓展, 2 阶零模型的随机断边重连算法是保持联合度分布不变的连边置乱, 其具体步骤与前述的连边置乱一样, 只是多了

一个限制条件：即要求交换前后两条边连接节点的度值是不变的。因此如图 2-15 所示，只有节点 B 和 D(或 A 和 C)具有相同的度值时，连边交换才可以进行，以保证网络的联合度分布不变。随着零模型阶次的增加约束条件也越来越多，可以被随机断开再重连的连边就越来越少，因此网络拓扑被破坏的也越来越小，所选的零模型也越来越精确。

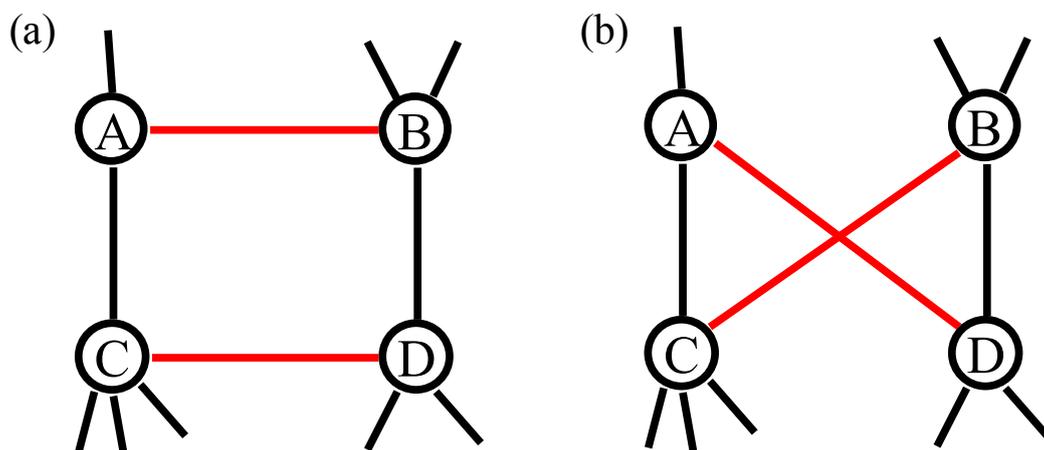
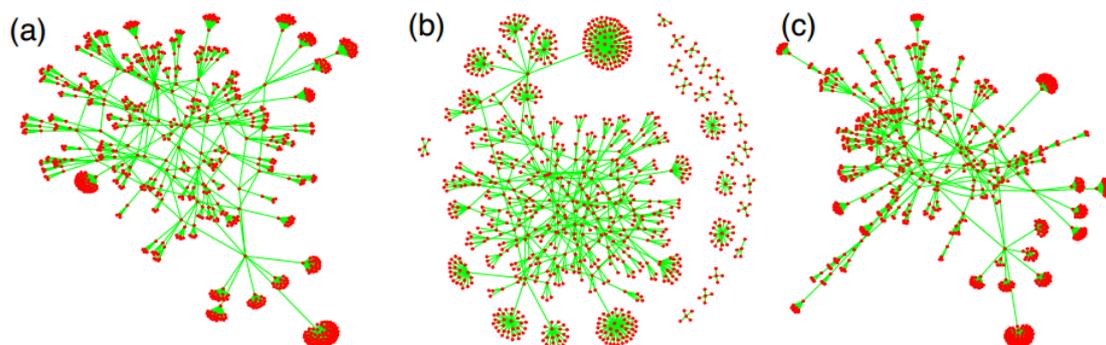


图 2-15 原始网络和使用断边重连算法构造的 2 阶零模型网络

该零模型使用程序实现的基本过程和一阶零模型很相似，但是需要每次从原始网络中任选两条有相同度节点的边进行断边重连。上述生成的网络零模型没有保证网络的连通性，如果原始网络是一个全联通网络，我们可以采用如下的程序约束生成的零模型网络也是全联通的。在上面程序的基础上进行修改，增加网络连通性的判断，若网络不保持连通性，则撤销断边重连操作。图 2-16 为 HOT 原始网络模型和在无连通性及有连通性情况下使用断边重连构造的 2 阶零模型。



(a)HOT 原始网络模型(b)2 阶零模型(c)保证连通性的情况下构造的 2 阶零模型

图 2-16HOT 原始网络模型和在无连通性及有连通性情况下使用断边重连构造的 2 阶零模型

在网络的高阶零模型研究中，Mahadevan 做了非常重要的系统性工作。一方面她系统性的研究了不同阶次复杂网络零模型的度相关特性和基于随机断边重连方式的构造算法^[3]。另一方面，在保持原始网络拓扑特性的前提下，提出了构造原始网络不同规模零模型的算法并提供了软件 Orbis 供研究者参考和使用^[2]。此外，Gjoka 等讨论了如何结合不同网络抽样算法生成网络 2.25 阶(与原始网络具有相同的联合度分布 $p(k_1, k_2)$ 和平均聚类系数 \bar{C})和 2.5 阶(与原始网络具有相同的联合度分布 $p(k_1, k_2)$ 和度相关的聚类系数 $C(k)$)零模型的理论算法与实现过程^[1]。

2.5.4 基于随机断边重连的 2.5 阶零模型

随着零模型阶数的增加，约束条件也会变得愈加苛刻，因此由断边重连方式产生的零模型要求也越来越高，生成的方式也越来越困难。2.5 阶零模型不但需要在随机断边重连的过程中需要保证 2 阶零模型的基本性质不变，而且也需要保证在网络断边前后度相关的聚类系数的值不变。Gjoka 等人结合网络抽样算法生成 2.5 阶零模型网络，我们这里也尝试使用随机断边重连的方式生成 2.5 阶零模型。

基于随机断边重连的 2.5 阶零模型构造的基本步骤是：在原始网络中随机选取两条边，首先要保证联合度分布不变即 2 阶零模型的基本性质不变。其次，断边重连之前计算两条连边两端节点度相关的聚类系数的值，经过断边重连之后计算新生成的连边两端节点之间的聚类系数的值。若断边重连前后所选边上的节点中度相关的聚类系数的值不变则证明断边过程是成功的，若度相关的聚类系数的值不一致，则证明所选的边不符合条件，需要重新再选。

我们以图 2-17 为例来具体阐述由原始网络生成 2.5 阶零模型网络的过程，在生成 2.5 阶零模型之前首先要计算原始网络中随机选取的两条边上 4 个节点的聚类系数，然后得出这 4 个节点的度相关的聚类系数分布情况。在图 2-17(a)中随机选取两条边 AB 和 CD，其中度值为 2 的 A 节点的聚类系数值为 $C_A = 1$ ，在这 4 个节点中经计算可得节点的度为 2 的平均聚类系数值为 $\bar{C}(2) = 1$ 。节点 B 的聚类系数 $C_B = \frac{1}{3}$ ，节点 C、D 的聚类系数的值都为 0，且 B、C、D 节点的度都为 3，节点的度为 3 的平均聚类系数的值为 $\bar{C}(3) = \frac{1}{9}$ 。此时，我们就

得到了在断边重连之前度值为 2, 3 节点的平均聚类系数分布情况: $\bar{C}(2) = 1$, $\bar{C}(3) = \frac{1}{9}$ 。将选取的两条边 AB 和 CD 切断, 再使节点 A 与 C 相连, B 与 D 相连。具体的生成过程如图 2-17(b)所示, 在经过断边重连后的网络中 A 节点的度为 2, 经计算可得 $\bar{C}(2) = 1$, B、C、D 节点的度都为 3, 节点的度为 3 的平均聚类系数的值为 $\bar{C}(3) = \frac{1}{9}$, 所以在经过断边重连之后度值为 2, 3 节点的平均聚类系数分布情况: $\bar{C}(2) = 1$, $\bar{C}(3) = \frac{1}{9}$ 。通过对断边重连前后随机选取的两条边上 4 个节点的度相关的聚类系数的度值分布情况的对比可以看出是不变的, 因此可以说图 2-17(b)是经过随机断边重连后生成的 2.5 阶零模型, 此模型保证了网络中的联合度分布及度相关的聚类系数值不变。图 2-18 为 HOT 原始网络模型和在无连通性及有连通性情况下断边重连构造的 2.5 阶零模型。

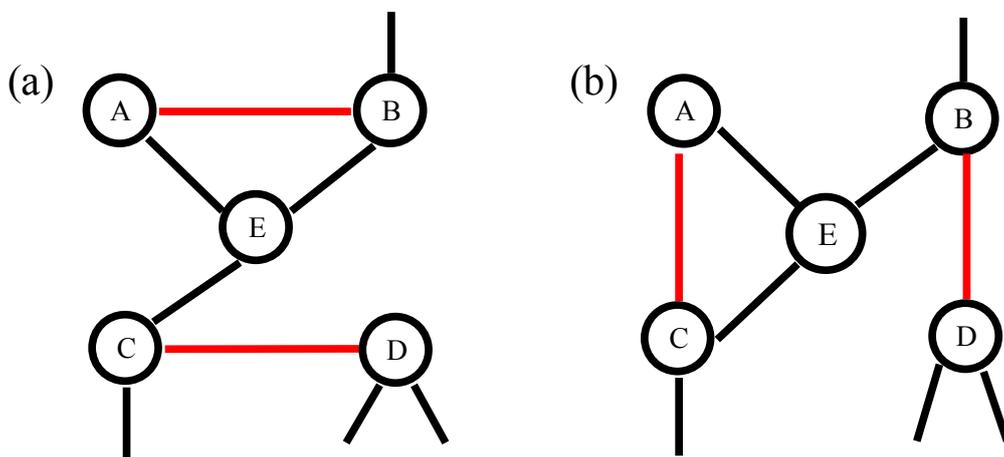
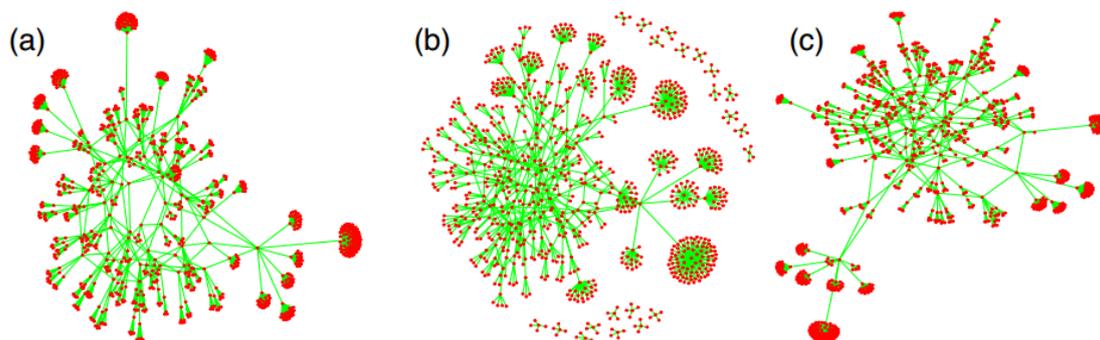


图 2-17 原始网络和使用断边重连算法构造的 2.5 阶零模型网络



(a)HOT 原始网络模型(b)2.5 阶零模型(c)保证连通性的情况下构造的 2.5 阶零模型

图 2-18HOT 原始网络模型和在无连通性及有连通性情况下断边重连构造的 2.5 阶零模型

Gjoka 等通过对节点的采样生成了联合度分布和度相关的聚类系数的有效估计量并成功构建了 2 阶零模型网络和 2.5 阶零模型网络，图 2-19 显示的是新奥尔良城市 Facebook 网站中九种不同的性质分布图。从图 2-19 中可以明显的看出 9 个性质图的分布情况，在 9 个不同性质的分布图中 2 阶、2.5 阶零模型与原始网络中相应的性质做对比，可以清楚地看出 2.5 阶图更接近原始网络。上述案例有效地说明了约束条件越多，生成的网络模型越接近于原始网络。

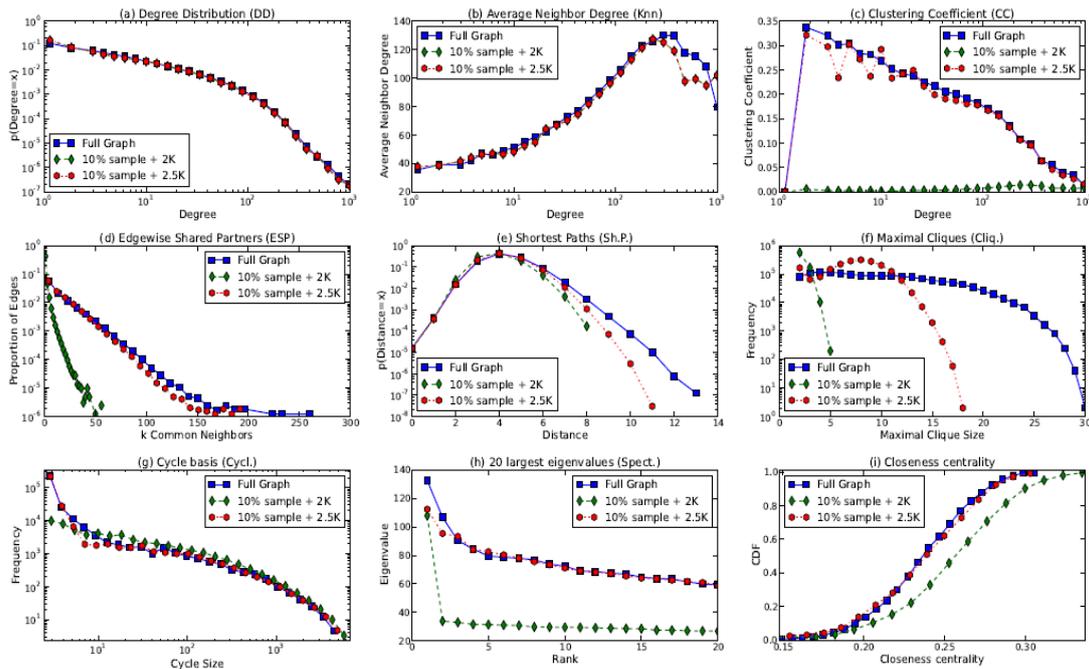


图 2-19 Facebook 局部网络的 2 阶零模型与 2.5 阶零模型性质比较(取自文献[1])

2.6 检验富人俱乐部效应的局部断边重连零模型

幂律分布也被称做无标度分布，具有幂律分布的网络一般被称为无标度网络。在一个无标度网络中，绝大部分的节点的度相对很低，但存在少量的节点的度相对较高，这样的节点被称为富节点或中枢(Hub)节点，Hub 节点在网络中占有比较重要的位置，这样的节点由于拥有大量的社会连接会对信息扩散产生影响。并且在变更传播过程中比较容易被影响，变更后的影响范围比较大。很多复杂网络都具有幂律形式的度分布，例如，通常我们写一篇新文章总是倾向于引用一些引用次数较多的文献。在新产品的开发推广中具有 Hub 节点性质的人往往能起到扩大推广的作用。

为了研究 Hub 节点之间的连接特性以及这种连接特性对于整个网络性质的影响, Zhou 等提出了复杂网络的富人俱乐部系数, 他们在 Internet 网络研究中没有使用该网络的零模型 仅仅使用了几种常用网络做基准模型, 发现 Internet 网络具有富人俱乐部特性^[24]。Colizza 等不仅仅计算了多种网络的富人俱乐部系数绝对值, 还计算了每个网络 1 阶零模型的富人俱乐部系数, 通过使用统计方法比较后他们得出了和 Zhou 等截然相反的结论: Internet 网络不具有富人俱乐部特性^[25]。

针对 Internet 网络是否具有富人俱乐部特性引起的诸多争论^[25, 26], Zhou 等提出了改变网络富人俱乐部系数的局部连边随机重连算法^[27], 再次有力证明了 Internet 网络确实存在富人俱乐部特性。具体的算法构造过程如图 2-20 所示, 其中网络中节点 A 和 B 是富节点, 重连以前这两个富节点之间无连接, 没有形成富人俱乐部。首先在网络中任意选取两个非富节点 C 和 D(如图 2-20(a)所示), 这两个节点分别连接到富节点 A 和 B 上且相互之间没有连边。然后采用上文提到的连边重连算法将连边 AC 和 BD 切断, 再使节点 A 与 B 相连, C 与 D 相连, 这样富节点 A 与富节点 B 相连后网络就有了富人俱乐部性质(如图 2-20(b)所示)。上述算法描述的是如何使网络构造具有强的富人俱乐部特性, 如果将这一过程反过来运用, 也可构造出一个几乎没有富人俱乐部特性的零模型网络。

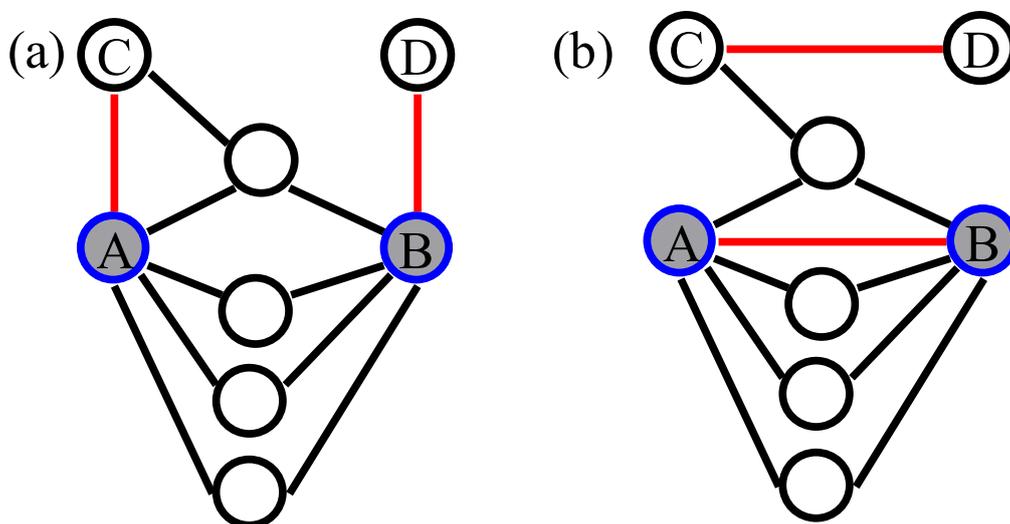
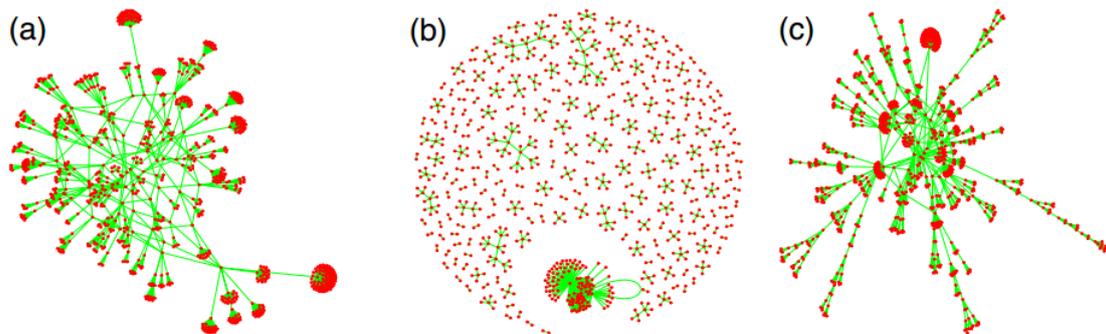


图 2-20 原始网络和使用连边置乱算法构造的具有富人俱乐部特性网络

下面程序的输入参数为原始网络 G_0 , 富节点的最小度 k (大于 k 的节点为富节点), 输出为基于随机断边重连构造的具有富人俱乐部特性的网络 G 。该函数同样也是先使用

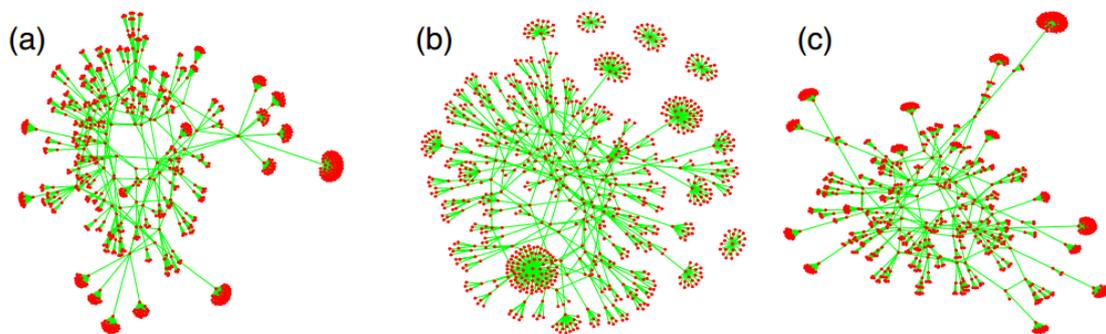
copy.deepcopy(G0)对原始网络进行深拷贝，然后从原始网络中任选两个富节点并找与之分别相连的两个非富节点，若富节点间无连边且非富节点间无连边则断开找到的两条边，并将富节点和富节点相连，非富节点和非富节点相连。重复该操作直到富节点之间所有可能的连接均相连为止。



(a)H0T 原始网络模型(b)具有富人俱乐部特性的网络(c)保证联通性的情况下构造的富人俱乐部特性的网络

图 2-21H0T 原始网络和在不连通性及有连通性情况下断边重连构造的具有富人俱乐部特性网络零模型

上述代码是生成具有富人俱乐部特性的网络，将上述过程反过来进行，就可以生成不具有富人俱乐部特性的网络。输入参数为具有连通性的原始网络 G_0 ，富节点的最小度 k （大于 k 的节点为富节点），输出为基于随机断边重连没有富人俱乐部特性的网络 G 。该函数主要是将富节点间的连边断开，保证网络中富节点和富节点间无连边。遍历节点列表将度大于 k 的节点存于列表 `hubs0` 中，然后将节点全是富节点的边存于列表 `hubedges`，节点全是非富节点的边存于列表 `nothubedges`。随机选一条富边和一条非富边，若存在富节点和非富节点间无连边则进行断边重连。



(a)HOT 原始网络模型(b)不具有富人俱乐部特性网络(c)保证联通性的情况下构造的不具有富人俱乐部特性网络

图 2-22HOT 原始网络和在无连通性及有连通性情况下断边重连构造的不具有富人俱乐部特性网络零模型

局部连边随机重连算法几乎可以任意改变中枢节点之间的连接特性,通过将完全没有富人俱乐部特性的网络、原始网络以及具有极强富人俱乐部特性的几种网络进行比较,就可以分析出富节点之间的连接特性对于整个网络性质的影响。

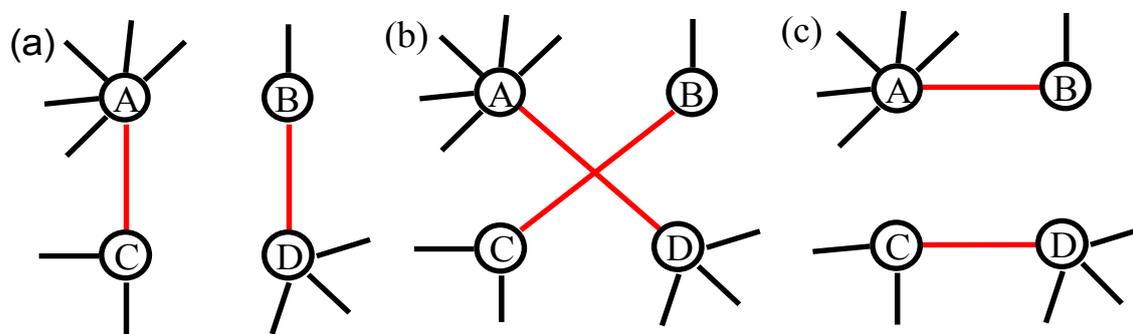
在营销领域,富节点对于产品的采纳和扩散至关重要,拥有大量社会连接的富节点可以显著影响产品的市场占有率。对于口碑营销,如果能将富节点识别出来,他们可以作为口头传播的有效信息源目标,从而导致采纳人数的迅速增长和市场规模的大幅增加。许小可等通过这一方法分析了多种网络中富人俱乐部特性对于网络聚类系数、匹配系数和平均路径长度的影响,并讨论了在不同度分布网络中富人俱乐部特性的影响力大小^[28]。

2.7 检验匹配特性的有倾向性断边重连零模型

在各种实证网络中,如果度大节点倾向于和度大节点相连,度小节点倾向于和度小节点相连,这种倾向性即为度匹配的正相关性;或者度小节点倾向于和度大节点相连,这种倾向性即为度匹配的负相关性。连边随机重连算法不仅破坏了网络拓扑,也扰乱了网络本身的度相关性,使连边不再具有同配或异配特性。Maslov 等利用连边置乱算法构造的零模型与原始网络进行对比,发现在蛋白质相互作用网络和基因调控网络中,度大与度大节点之间的连边受到了系统抑制,而度大与度小节点之间的连边出现的可能性则更大^[17]。此外, Maslov 等还将原始网络和利用连边置乱生成的零模型进行对比,通过相关性剖面的新方法检验了 Internet 网络中的度度相关特性^[23]。以上研究都说明了度度相关特性对于网络的重要性,那么有没有办法构造出满足某种度匹配模式的零模型呢?

如果想改变原始网络的度度匹配特性,就需要有选择的进行断边重连来构造同配或异配网络。如图 2-23 所示,原始网络(a)中有 A、B、C、D 四个节点,其中节点 A 与 C 相连, B 与 D 相连,相连节点之间的度有一定差距但相差不大,属于中性网络。如果将原始网络的度大节点 A 和度大节点 D 相连、度小节点 B 和度小节点 C 相连(如图 2-23(b)所示),那么就增强了网络的同配特性。将这一过程反复进行,就可以生成强同配性的网络。相反,将原

始网络的度最大的节点 A 和度最小的节点 B 相连，其余节点相连(如图 2-23(c)所示)，那么就增强了网络的异配特性，反复进行就可以生成强异配性的网络。



(a)中性匹配网络(b)同配网络(c)异配网络

图 2-23 复杂网络中三种不同的度匹配模式

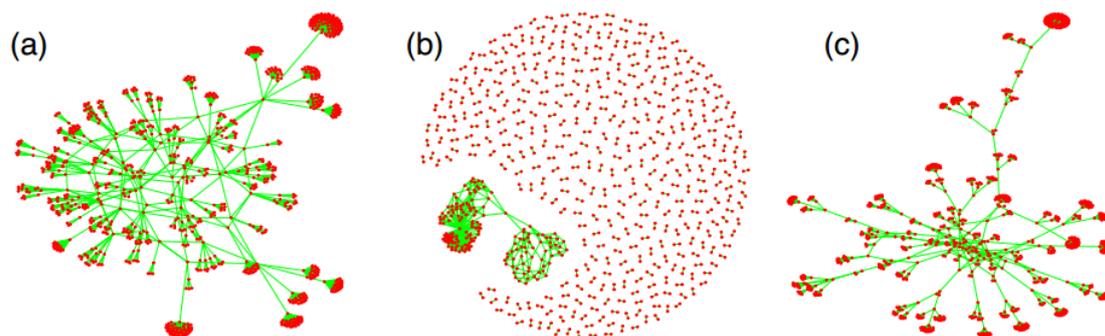


图 2-24HOTA 原始网络和 在无连通性及有连通性情况下断边重连构造的具有同配特性网络零模型

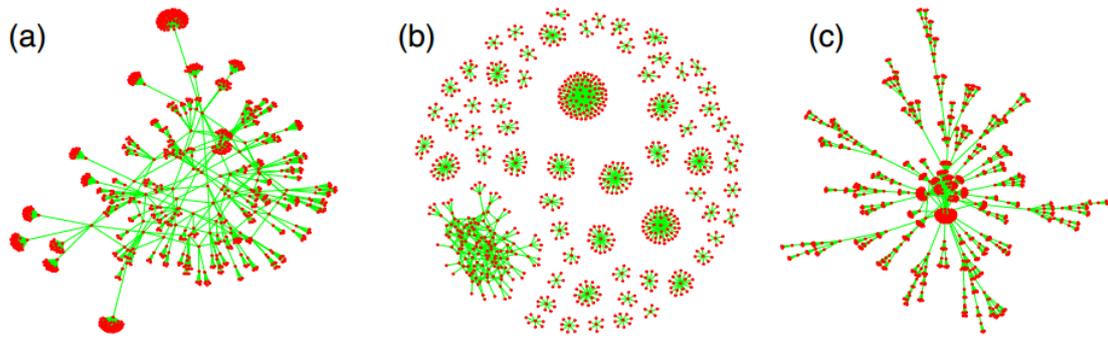


图 2-25 H0T 原始网络和在无连通性及有连通性情况下断边重连构造的具有异配特性网络零模型

使用上述有导向的连边重连算法生成强同配网络和异配网络可以研究网络拓扑结构对于度度相关特性的约束。Newman 定义的匹配系数是实际复杂网络中最常用的衡量网络匹配特性的统计量^[29, 30]，具体公式如下所示：

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (2.10)$$

其中， M 为网络中的边数， j_i 、 k_i 分别为第 i 条边的两端节点的度数。如果 $r > 0$ ，表明网络是同配性质的，如果 $r < 0$ ，表明网络是异配性质的。

理论上讲实际网络最大同配的时候该值应该接近于 1，最大异配的时候该值应该接近于 -1，但实证研究表明并非如此。图 2-26 给出了三个网络—因特网、蛋白质相互作用网络、科学合作网络的匹配系数分布图。Zhou 等人在保证度分布不变、度分布与其度为 k 的富节点连通性不变的两种情况下，对这三个原始网络分别进行随机重排、同配性重排、和异配性重排，每个原始网络都生成了相应的随机网络、强同配网络、强异配网络。图中的实线为在保持度分布不变的情况下按要求对网络进行的重排，虚线为在保持度分布与其度为 k 的富节点连通性不变的情况下按要求对网络进行的重排，并将重排后的网络用公式(2.10)量化得到三个网络的同配系数 r 值的范围，如图 2-26 所示。

在保持度序列不变的情况下，在对蛋白质相互作用网络、科学合作网络进行随机重排的过程中发现重排后网络的匹配系数在 0 附近。在对上述的两种网络分别进行强同配性重排和

强异配性重排后所产生的 1 阶强同配零模型和 1 阶强异配零模型的同配系数 r 值在 1 到 -1 之间，但是蛋白质相互作用网络同配系数 r 值的范围比科学合作网络的 r 值范围更窄。同时，Zhou 等发现实际的 Internet 网络，在保证度序列相同的情况下无论怎么进行连边重连，它的匹配系数总是在一个非常小的范围内且总是负值^[27]，这说明在对实际网络进行最大同配重排的时候该值 r 并非都会接近于 1，最大异配重排的时候该值 r 并非都会接近于 -1。在保持度序列与富节点连通性不变的情况下，分别对网络进行上述相同的随机重排，同配性重排，和异配性重排。我们可以看到相对应的匹配系数 r 值的范围变得更窄了，这说明富节点之间的连通性会进一步影响 r 值的分布情况。

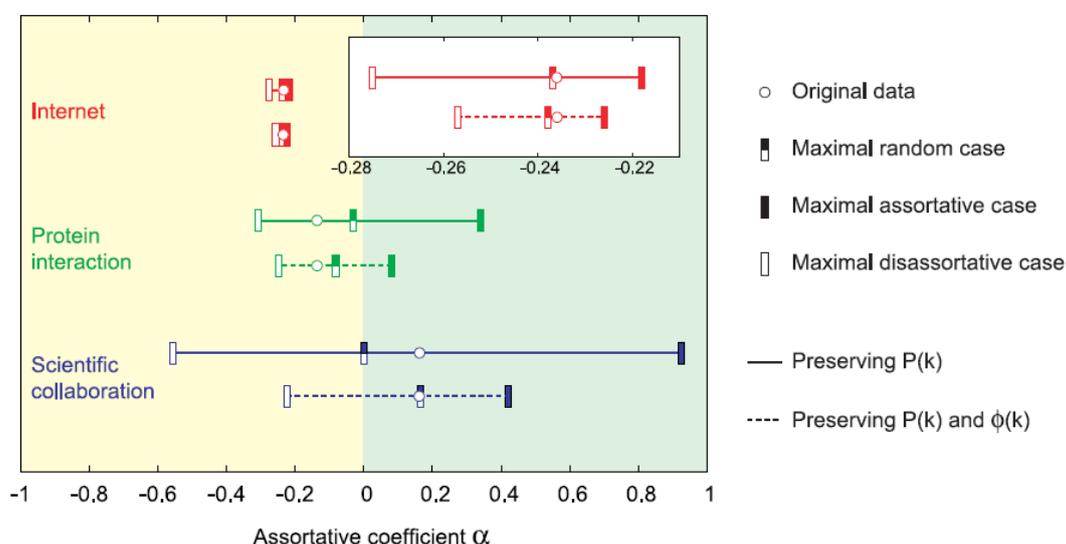


图 2-26 不同网络的匹配系数 r 值分布情况图(取自文献[27])

针对匹配系数这一缺陷，许小可等通过比较原始网络和它的零模型，分析了 Internet 网络中超级节点引起网络匹配系数异常的原因并给出了一个简单的修正方式^[31]。如果一个网络中存在度远远大于其他节点的节点，那么该节点可以视为网络中的“超级节点”，超级节点和其他节点之间悬殊的度差别导致它们和其他节点连接会导致负的匹配特性，因此为了避免超级节点的干扰使整个网络的匹配特性度量不准确，比较简单的方法就是先从网络中将超级节点过滤掉，再计算网络的匹配系数。假定排除了超级节点后网络中的连边数为 M_c ，则计算匹配系数的公式可简单修改为：

$$r_c = \frac{M_c^{-1} \sum_i j_i k_i - \left[M_c^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M_c^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M_c^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (2.11)$$

其中, M_c 为网络中排除超级节点后的边数, j_i 、 k_i 分别为第 i 条边的两端节点的度数。如果 $r_c > 0$, 表明网络是同配性质的, 如果 $r_c < 0$, 表明网络是异配性质的。

度相关性在网络的扩散现象中担当重要角色, 它会对信息传播的规模和速度产生重要影响。对于无标度网络, 网络中的度相关性会影响谣言扩散的阈值和最终的扩散规模, 增加网络的度相关性可以显著地抑制扩散。无标度网络中存在大量度值小的节点, 增强度相关性意味着度值小的节点倾向于连在一起, 而非连向度值大的节点, 这使得度值大的节点在信息扩散过程中很可能变得孤立, 从而抑制了扩散。此外, 增强度相关性也会抑制随机网络中的谣言扩散, 但由于随机网络的度分布较为均匀, 这种影响比无标度网络中的作用要小得多。

2.8 无权有向网络的零模型

与无向网络不同的是有向网络中节点的度分为出度和入度两种。节点的出度是指以该节点为起点有方向性的指向另一个节点, 而节点的入度则表示其他节点有方向性的指向该节点。在有向网络中每个节点的出度和入度可能并不相同, 但是每条有方向性的边是一个节点出度的同时又是另一个节点的入度, 因此从网络的整体来看, 整个网络的出度总数总是等于入度总数。以微博用户关注为例, 普通人对名人的关注可以视为普通人的出度同时又可视为名人的入度, 在微博所构成的社交网络来说出入度的总数总是相等的。

对无向网络采用断边重连的方式生成零模型能够研究无向网络的度相关性^[27], 同理, 对有向网络进行断边重连可以来研究网络的出入度相关性, 即有向网络中相连节点之间的出度和出度、入度和入度、入度和出度、以及出度和入度之间有何关系^[32]。有向网络的断边重连与无向网络类似, 不同的是可以分别考虑将节点的出边或者入边断边重连^[20]。

以 1 阶零模型为例该算法的构造过程如图 2-27 所示, 如果在有向网络中有节点 A 有出度指向 B, C 有出度指向 D; 且节点 A 与 D 不相连, B 与 C 不相连(如图 2-27(a)所示)。我们就将连边 AB 和 CD 切断, 再使节点 A 的出度指向 D, C 的出度指向 B(如图 2-27(b)所示)。

根据网络规模和实际需要不断重复上述过程就会使网络中的连边关系被随机置乱,但是置乱前后需要保证每个节点的出度和入度均保持不变,保证了置乱前后网络的度序列相同。

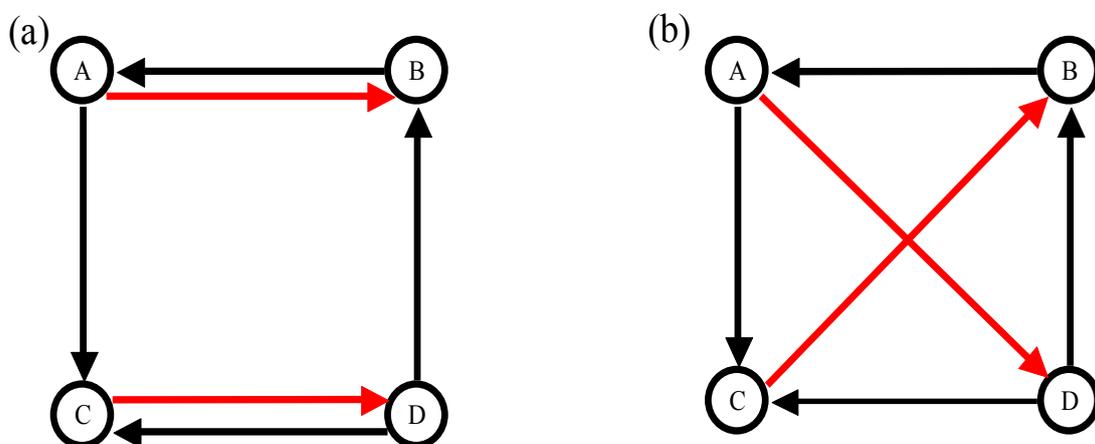
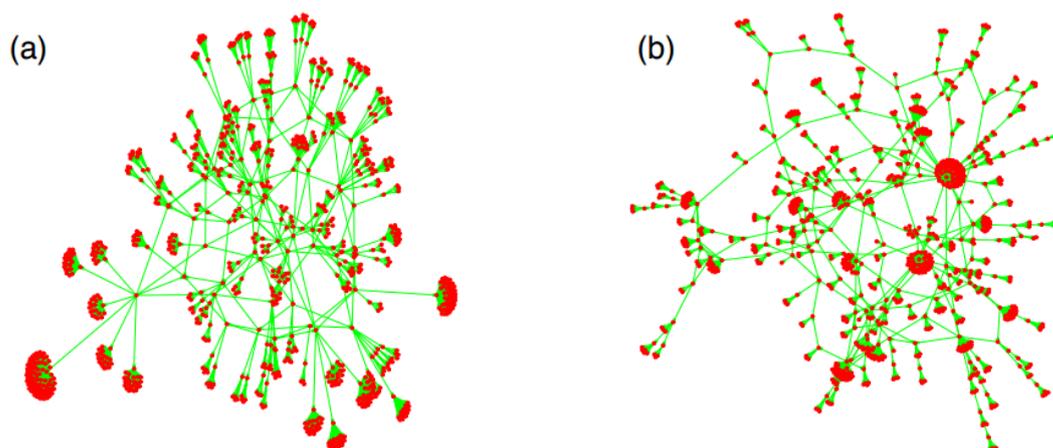


图 2-27 原始有向网络和使用断边重连算法构造的 1 阶零模型网络



(a)HOT 原始有向网络模型(b)1 阶有向网络零模型

图 2-28 原始有向网络和使用断边重连算法构造的 1 阶有向网络零模型网络

在无向网络中,可通过计算同配系数是否大于 0 来判断网络中的同异配性质。而在有向网络中出度和入度可能会对同配性质产生重要影响,在有向网络中存在如下图 2-29 所示的 4 种匹配特性的情况: $r(out,in)$, $r(in,out)$, $r(out,out)$, $r(in,in)$ 。其中 $r(out,in)$ 表示高出度的节点有边指向高入度的节点的倾向性程度, $r(in,out)$ 表示高入度的节点有边指向

高入度的节点的倾向性程度, $r(out, out)$ 表示高出度的节点有边指向高出度的节点的倾向性程度, $r(in, in)$ 表示高入度的节点有边指向高入度的节点的倾向性程度^[32]。

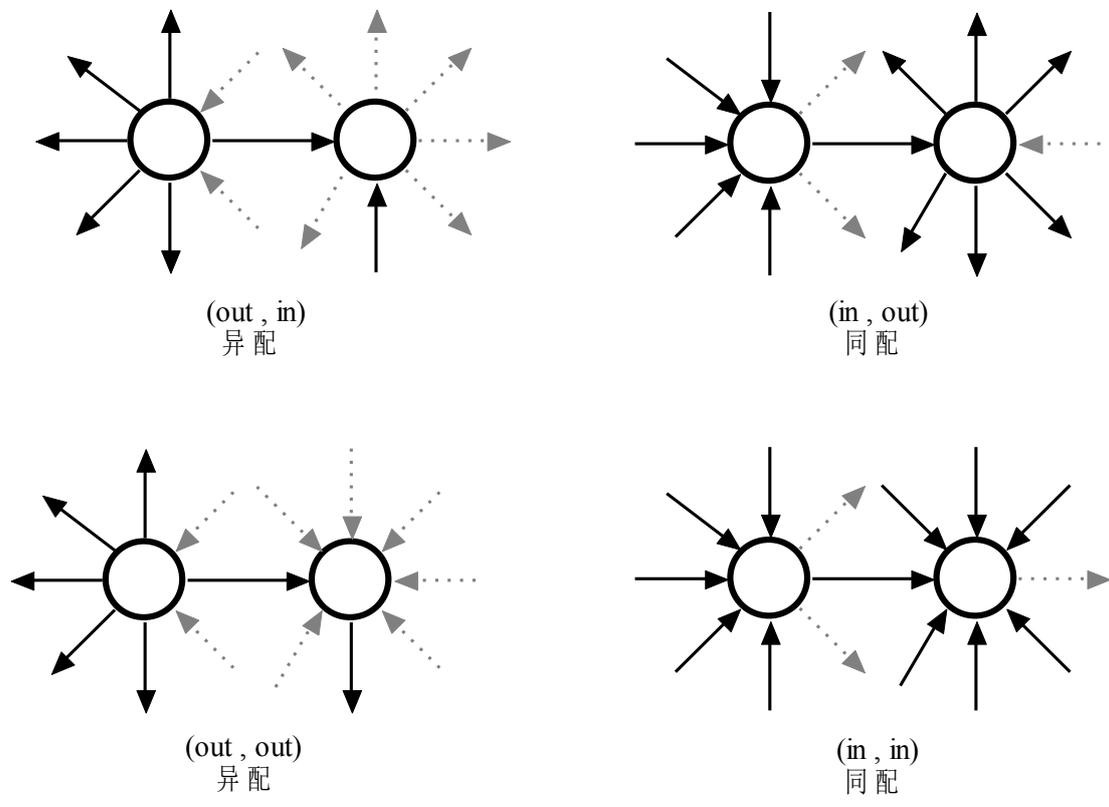


图 2-29 有向网络的 4 种度相关性(取自文献[32])

在有向网络的研究中, Foster 等利用原始网络和零模型中度相关性的对比, 提出了同配重要性剖面算法在该算法中同配系数可以用下面的公式来表示:

$$r(\alpha, \beta) = \frac{M^{-1} \sum_i (j_i^\alpha - \langle j^\alpha \rangle)(k_i^\beta - \langle k^\beta \rangle)}{\sqrt{M^{-1} \sum_i (j_i^\alpha - \langle j^\alpha \rangle)^2} \sqrt{M^{-1} \sum_i (k_i^\beta - \langle k^\beta \rangle)^2}} \quad (2.12)$$

其中, α 表示出度类型, β 表示入度类型, M 为网络的边数, j_i^α 表示有向边 i 的出节点的度值, k_i^β 表示有向边 i 的入节点的度值, $\langle j^\alpha \rangle$ 表示在零模型中出节点出现次数的均值, $\langle k^\beta \rangle$ 表示在零模型中入节点出现次数的均值。 $r(\alpha, \beta)$ 的统计重要性可以用 Z 值来刻画:

$$Z(\alpha, \beta) = \frac{r(\alpha, \beta) - \langle r_r(\alpha, \beta) \rangle}{\sigma_r(\alpha, \beta)} \quad (2.13)$$

其中， $\langle r_r(\alpha, \beta) \rangle$ 是 1 阶零模型中同配系数的均值、 $\sigma_r(\alpha, \beta)$ 是同配系数的标准差。通过对上面的 $Z(\alpha, \beta)$ 进行归一化处理可以得到同配重要性剖面 (Assortativity Significance Profile, ASP) :

$$ASP(\alpha, \beta) = \frac{Z(\alpha, \beta)}{\left(\sum_{\alpha, \beta} Z(\alpha, \beta)^2\right)^{1/2}} \quad (2.14)$$

$ASP(\alpha, \beta)$ 可用于检测有向网络的匹配特性，当 $ASP > 0$ 时，表明原始网络比零模型更具有同配性；当 $ASP < 0$ 时，表明原始网络比零模型更具有异配性。 ASP 绝对值的大小反映了原始网络同配或异配的强弱程度，并可用于比较不同的有向网络之间匹配特性的强弱程度。Foster 等人在保证原始网络度序列不变的前提下，对 WWW, Wikipedia, Political Blogs 三个有向网络进行随机重排生成了相应的 1 阶零模型，并根据公式(2.14)计算了三个有向网络的同配重要性剖面的值，生成了图 2-30 所示的匹配性质网络对比图。图 2-30 中的横坐标表示有向网络匹配性质的 4 种度量，纵坐标表示 ASP 的值。以 $r(out, in)$ 度量情况为例，通过对比三个网络的 ASP 值发现，对应于纵坐标可以看出三个网络的 ASP 值都小于 0，因此我们可以确定这三个原始网络都具有异配性。通过对 ASP 的绝对值对比可以看出三个网络之间匹配特性的强弱程度为：Political Blogs>Wikipedia>WWW。

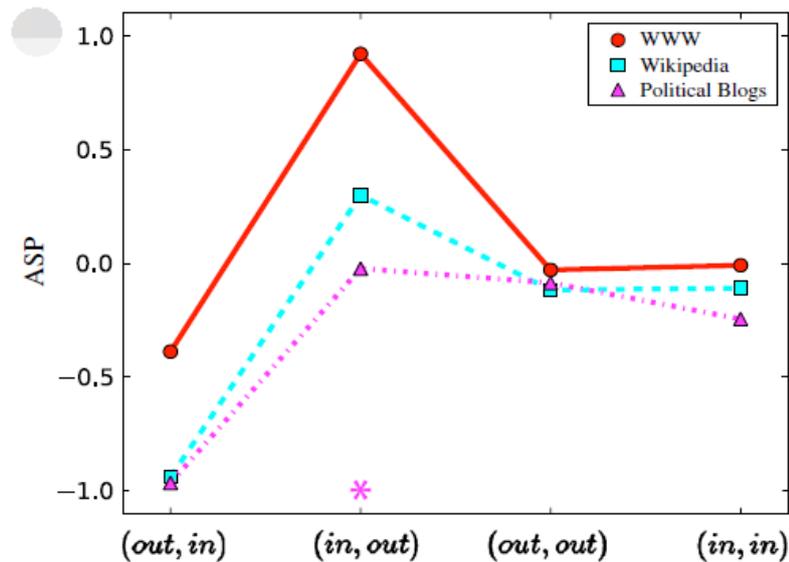


图 2-303 个有向网络的不同的同配性质(取自文献[32])

将同一个网络的 $r(out, in)$, $r(in, out)$, $r(out, out)$, $r(in, in)$ 进行横向比较, 可以发现同一个有向网络的不同处入度关系之间可能具有多种同配特性, 在所给的三个有向网络中并不存在单一的同配或者异配性质。在图 2-30 中, $r(out, in)$ 的 ASP 值发都小于 0, 因此我们可以确定这三个原始网络都具有异配性; 在 $r(in, out)$ 度量的情况下三个网络之间匹配特性的强弱程度为: WWW > Wikipedia > Political Blogs, 其中 WWW、Wikipedia 网络比零模型更具有同配性; 在 $r(out, out)$ 、 $r(in, in)$ 两种度量的情况下, 我们可以认为这三个网络既没有同配倾向性, 也没有异配倾向性。

2.9 本章小结与展望

综上所述, 本章对分别研究无权无向网络和无权有向网络的零模型构造及应用。对于无权无向网络, 重点总结了 0 阶的 ER 随机图, 1 阶、2 阶、3 阶的配置模型, 以及基于随机断边重连零模型的构造过程和它的实际应用。对于无权有向网络, 重点总结了基于随机断边重连的 1 阶零模型。这些零模型具有非常广泛的应用, 如可以帮助我们检测网络的随机性强弱, 网络是否具有富人俱乐部特性。

目前零模型中应用最广泛的是根据原始网络中度序列产生的 1 阶零模型, 约束条件的较

为苛刻的 2 阶及 2.5 阶零模型的还未得到广泛的应用。导致这一情况有两方面原因，一方面是高阶零模型的构造比 1 阶模型要复杂得多，另一方面是很多研究者对应该在何种情况下使用那一阶零模型还不是很清楚。对于实际网络的 2 阶和 2.5 阶零模型网络的具体实现，Mahadevan、Gjoka 和 Orsini 分别提供了几种开源实现的方法，这几种方法可以供研究者使用和借鉴。在本章中我们基于断边重连提出了新的 2.5 阶零模型产生方法，并以 HOT 网络为例进行了可视化。对于何种情况下使用何种零模型网络这一问题，一般而言应该遵循由低阶到高阶的顺序来使用。越是高阶的零模型，它的性质越接近于原始网络。如通过与相应的 2 阶零模型对比可以清楚地发现，2.5 阶零模型的各种性质比 2 阶零模型更接近于原始网络。

参考文献

- [1] Gjoka M, Kurant M, Markopoulou A. 2.5K-Graphs: From Sampling to Generation[M]//IEEE INFOCOM. New York: IEEE, 2013:1968-1976.
- [2] Mahadevan P, Hubble C, Krioukov D, et al. Orbis: Rescaling Degree Correlations to Generate Annotated Internet Topologies[J]. ACM SIGCOMM Computer Communication Review, 2007,37(4):325-336.
- [3] Mahadevan P, Krioukov D, Fall K, et al. Systematic Topology Analysis and Generation Using Degree Correlations: Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Pisa, Italy, 2006[C]. ACM.
- [4] Orsini C, Dankulov M M, Colomer-de-Simon P, et al. Quantifying randomness in real networks[J]. Nature Communications, 2015,6.
- [5] Bollobas B. Random graphs[M]. England: Cambridge university press, 2001.
- [6] Newman M. The Structure and Function of Complex Networks[J]. SIAM Review, 2003,45(2):167-256.
- [7] Barabási A, Albert R. Statistical Mechanics of Complex Networks[J]. Reviews of Modern Physics, 2002,74(1):47-97.
- [8] Molloy M, Reed B. The Size of the Giant Component of a Random Graph with a Given Degree Sequence[J]. Combinatorics, Probability and Computing, 1998,7(03):295-305.
- [9] Molloy M, Reed B. A Critical Point for Random Graphs with a Given Degree Sequence[J]. Random Structures and Algorithms, 1995,6(2-3):161-180.
- [10] Newman M E J, Strogatz S H, Watts D J. Random Graphs with Arbitrary Degree Distributions and Their Applications[J]. Physical Review E, 2001,64(2):26118.
- [11] Chung F, Lu L. Connected Components in Random Graphs with Given Expected Degree Sequences[J]. Annals of Combinatorics, 2002,6(2):125-145.
- [12] Chung F, LU L. The Average Distances in Random Graphs with Given Expected Degrees[J]. Proceedings of the National Academy of Sciences, 2002,99(25):15879-15882.
- [13] Newman M E J. Random Graphs with Clustering[J]. Physical Review Letters, 2009,103(5):58701.
- [14] Newman M E J, Karrer B. Random Graphs Containing Arbitrary Distributions of Subgraphs[J]. Physical Review E, 2010,82(6):66118.
- [15] Newman M E J. Networks: An Introduction[M]. 郭世泽, 陈哲, 译. 北京: 电子工业出版社, 2010.
- [16] Holme P. Network Reachability of Real-world Contact Sequences[J]. Physical Review E, 2005,71(4):46119.
- [17] Maslov S, Sneppen K. Specificity and Stability in Topology of Protein Networks[J]. Science, 2002,296(5569):910-913.
- [18] 曾进群, 杨建梅, 陈泉, 等. 基于零模型的开源社区大众生产合作网络结构研究[J]. 华南理工大学学报(社会科学版), 2013(02):29-34.

- [19] 陈泉, 杨建梅, 曾进群. 零模型及其在复杂网络研究中的应用[J]. 复杂系统与复杂性科学, 2013(01):8-17.
- [20] Milo R, Shen-Orr S, Itzkovitz S, et al. Network Motifs: Simple Building Blocks of Complex Networks[J]. Science, 2002,298(5594):824-827.
- [21] Shen-Orr S S, Milo R, Mangan S, et al. Network Motifs in the Transcriptional Regulation Network of Escherichia Coli[J]. Nature Genetics, 2002,31(1):64-68.
- [22] Milo R, Itzkovitz S, Kashtan N, et al. Superfamilies of Evolved and Designed Networks[J]. Science, 2004,303(5663):1538-1542.
- [23] Maslov S, Sneppen K, Zaliznyak A. Detection of Topological Patterns in Complex Networks: Correlation Profile of the Internet[J]. Physica A: Statistical Mechanics and its Applications, 2004,333:529-540.
- [24] Zhou S, Mondragon R J. The Rich-club Phenomenon in the Internet Topology[J]. IEEE Communications Letters, 2004,8(3):180-182.
- [25] Colizza V, Flammini A, Serrano M A, et al. Detecting Rich-club Ordering in Complex Networks[J]. Nature Physics, 2006,2(2):110-115.
- [26] Nunes Amaral L A, Guimera R. Complex Networks: Lies, Damned Lies and Statistics[J]. Nature Physics, 2006,2(2):75-76.
- [27] Zhou S, Mondragon R J. Structural Constraints in Complex Networks[J]. New Journal of Physics, 2007,9(6):173.
- [28] Xu X K, Zhang J, Small M. Rich-club Connectivity Dominates Assortativity and Transitivity of Complex Networks[J]. Physical Review E, 2010,82(4):46117.
- [29] Newman M E J. Assortative Mixing in Networks[J]. Physical Review Letters, 2002,89(20):208701.
- [30] Newman M E J. Mixing Patterns in Networks[J]. Physical Review E, 2003,67(2):26126.
- [31] Xu X K, Zhang J, Sun J, et al. Revising the Simple Measures of Assortativity in Complex Networks[J]. Physical Review E, 2009,80(5):56106.
- [32] Foeter J G, Foster D V, Grassberger P. Edge Direction and the Structure of Networks[J]. Proceedings of the National Academy of Sciences, 2010,107(24):10815-10820.