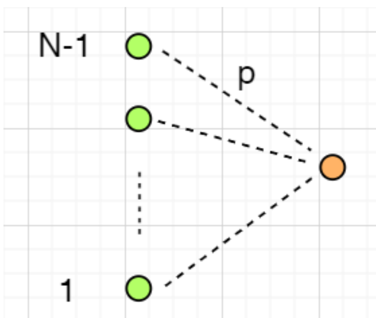
复杂网络建模 - 笔记

基础概念

Network Complexity

1. Structural complexity
2. Node-dynamics complexity
   - A node can be a dynamical system
   - A network may have different kinds of nodes
3. Mutual interactions among various complex factors

Random Graph Theory



- Average degree:

$$<k>= p(N-1) \approx pN$$

- Clustering coefficient(聚类系数):

$$C = <k>/N = p$$

- Degree distribution(一个结点度为 k 的概率):

$$P(k) = \frac{\mu^k}{k!}e^{-\mu} \quad 泊松分布(Poisson)$$

Clustering Coefficient

相关定义：设有结点 i，则：

- **E(i)** 是该结点所有 **相邻结点之间 实际存在** 的边的条数。
- **T(i)** 是该结点所有相邻结点(假设有 n 个)之间可能存在边数的最大值，即 **n(n-1)/2**。

那么结点 i 的聚类系数为：

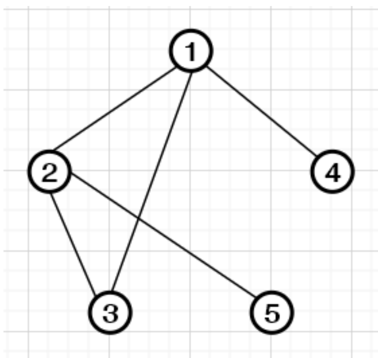
$$\begin{aligned} C(i) &= \frac{E(i)}{T(i)} \\ &= \frac{E(i)}{n(n-1)} \end{aligned}$$

整个网络的聚类系数为**所有结点聚类系数的平均数。**

另外，聚类系数还可以从另外一个方面定义：

$$C(i) = \frac{以\ \mathrm{i}\ 为角的闭合三角形个数}{以\ \mathrm{i}\ 为角的三角形个数}$$

例题：



- 结点 1 是一个闭合三角形的顶点（1,2,3），是三个三角形的顶点（1,2,3 1,3,4 1,2,4），所以 $C(1) = 1/3$
- 同样，$C(2) = 1/3$，$C(3) = 1$，$C(4) = C(5) = 0$

所以对于整个网络，$C = (1/3 + 1/3 + 1 + 0 + 0)/5 = 1/3$

Average Path Length

Distance：结点 i, j 之间的 **距离** $d_{i,j}$ 指的是能把这两个结点连起来所需经过的 **最少边数。**

Diameter：$\max\{d_{i,j}\}$

Average path length：$L = sum\{d_{i,j}\}/\mathrm{C}_n^2$

图理论

Simple Graph

简单图需要满足的条件：

1. unweighted
2. undirected
3. containing no loops
4. containing no nultiple edges

A simple graph **may be either connected or disconnected.**

Some Concepts

Handshaking Lemma

The total node degree of a graph is always an **even** number.

Isomorphism(同构)

如果两个图 G_1 和 G_2 满足如下条件，则称两图是同构的：G_1 和 G_2 的结点一一对应，且 G_1 中连接任意两个结点之间若存在一条边，那么在 G_2 中 **对应的两个结点之间同样存在一条边。**

Bipartite Graph(二分图)

在二分图中，每个环所经过的结点数为偶数。

Adjacency Matrices and Incidence Matrices(邻接矩阵和关联矩阵)

设图中有 N 个结点和 M 条边，则：

邻接矩阵是一个 $N \times N$ 矩阵，每个元素表示对应两个结点之间是否直接相连。

关联矩阵 是一个 N 行 M 列矩阵，元素表示对应的结点和边是否关联。

Laplacian Matrix

$$L_{ij} = \begin{cases} v_i\text{的度}, & if\ i = j \\ -1, & if\ i \neq j\ \&\ v_i\ v_j\text{邻接} \\ 0, & otherwise \end{cases}$$

- 对角线元素表示各个结点的度；
- 其他位置，如果结点 i,j 直接相连，则为 -1；
- 否则为 0。

Regular Graphs

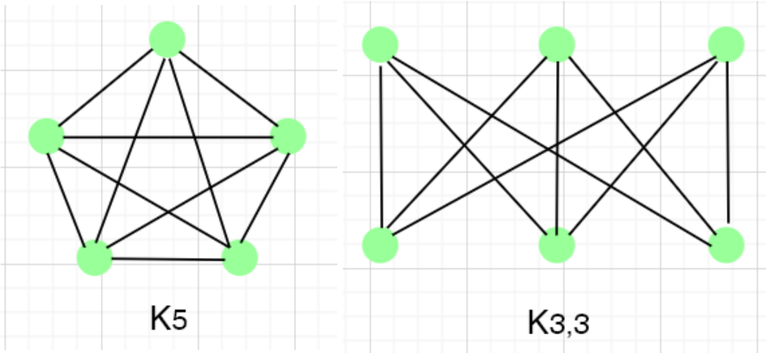
每个结点的度都相同的图叫做正则图。

【定理】：N 个结点的 r 度正则图的边数为 $rN/2$。

Planar Graph(平面图)

如果一个图能够画在平面上，而且 **没有交叉的边**，这个图就是 Planar graph。

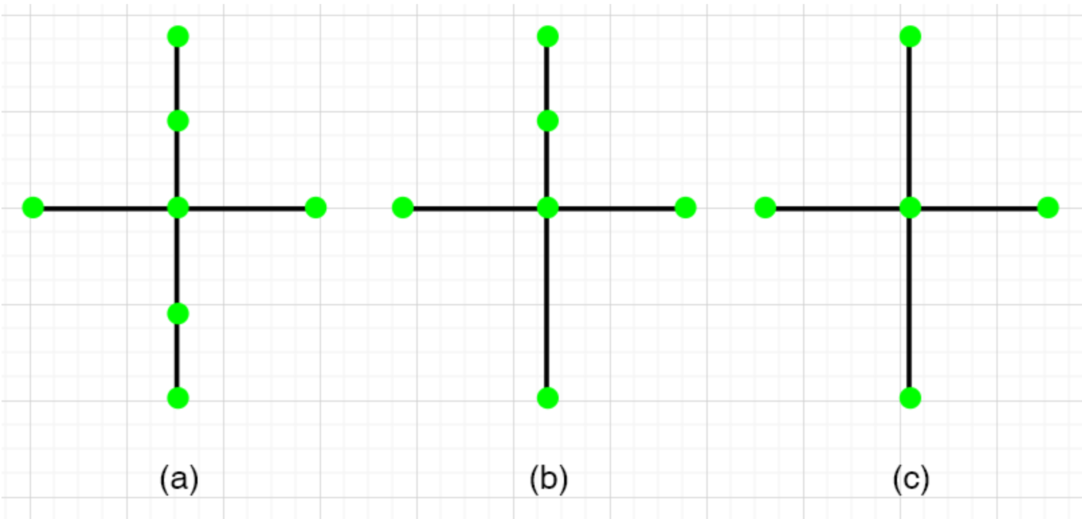
下面是两个典型的非平面图：



Homeomorphism(同胚图)

如果两个图可以从同一个图通过增加任意度为 2 的结点得到，则这两个图称为是 Homeomorphic 的。

例如，下图 (a) 和 (b) 都可以通过向 (c) 中边上增加结点得到，所以(a) (b) 同胚。



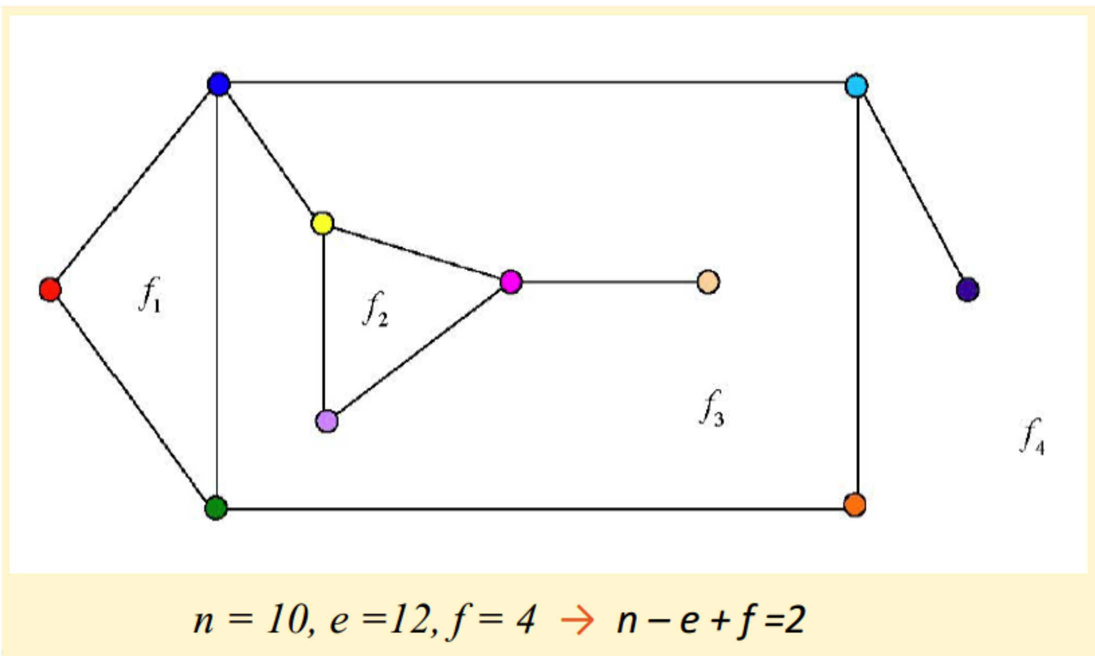
【定理】：一个图是平面图，当且仅当它不包含 K_5 和 $K_{3,3}$ 的同胚图。

Euler Formula for Plane Graphs

【定理】设一个连通平面图的结点数、边数、面数分别为 n, e, f, 则有下面的公式：

$$n - e + f = 2$$

例如：



> f4 是外部面。

【定理】如果一个简单图有 N 个结点，M 条边，K 个成员(components)，那么 M 满足：

$$N-K \le M \le (N-K+1)(N-K)/2$$

【推论】如果一个简单图有 N 个顶点，M 条边，且满足 $M > (N-1)(N-2)/2$，那么该图必定连通。

【引理】如果图中每个结点的度都不小于 2，那么改图中必定有环。

Some Concepts

Walk：一系列首尾相连的边的序列，可以有重复边。

Trail：没有重复边的 Walk。

Path：没有重复结点的 Trail，但是首尾结点可以相同，这时候叫做闭合 path，即 **circuit**。

【定理】设 T 是一个包含 N 个结点的图，那么下面这些说法等价：

1. T 是一棵树。
2. T 有 N-1 条边，且没有环。
3. T 有 N-1 条边，且连通。
4. T 是连通的，但是去掉任何一条边都会导致它不连通。
5. 任意两个结点之间只有一条路径。
6. T 没有环，但是添加任意一条边都会出现环。

Complementary Graph(补图)

$$G^c = 完全图 - G$$

G 的补图包含G 中所有结点，但是只包含G中没有的边。

Disconnecting Sets & Cut-Sets(不连通集 & 割集)

Disconnecting set: 一个 **边集**，若把这些边去掉，图将不连通。

Cut-set: **最小不连通集**，也就是说，其中的任意真子集都不是不连通集。

Bridges：**只包含一条边的割集**，即去掉这条边之后，图将不连通。

【定理】设 G 的生成树为 T，则：

1. G 的任何一个割集都和 T 有一条相同的边；
2. G 的任意一个环都和 T 的补图有一条相同的边。

找最小生成树的贪心算法

1. 选择最小权重的边 e_1；
2. 依次选出剩余边中权重最小的边 $e_2, \dots, e_{N-1}$，要求每次选出的边不能和已经选出的所有边构成回路。
3. 知道没有更多的边可选出。

Eulerian Graph(欧拉图)

欧拉图：如果一个图包含一条 **闭合 Trail(迹)**，能够 **遍历所有边一次且只遍历一次**，那么这个图叫做欧拉图。这条 Trail 叫做欧拉迹。

半欧拉图：图中包含将所有边遍历一次且仅遍历一次的迹（**无需是闭合迹**），那么叫做半欧拉图。

【定理】当且仅当一个连通图中 **每个结点的度都是偶数** 时，这个图是欧拉图。

【推论】寻找欧拉迹的方法：从任意结点开始，按照如下规则任意选择一条边来前进：

- 将 Bridge 留到无路可走时再走；
- 去掉已经走过的边；
- 去掉已经孤立的结点。

Hamiltonian Graph(汉密尔顿图)

汉密尔顿图：如果一个图包含一条能 **将所有结点遍历一次且仅遍历一次的闭合迹**，那么这个图叫做汉密尔顿图。这个闭合迹叫做汉密尔顿圈。

半汉密尔顿图：图中包含将所有结点遍历一次且仅遍历一次的迹（**无需是闭合迹**），那么叫做半汉密尔顿图。

> 欧拉图是从遍历 **边** 的角度定义的，汉密尔顿图是从遍历 **结点** 的角度定义的。

【定理】设图 G 有 N 个结点，如果对于任何一对不邻接的结点 v, u，它们的 **度之和都不小于** N，那么 G 是汉密尔顿图。

中国邮递员问题

问题要求：

- 将所有边遍历至少一遍，最后返回起点。
- 允许两个结点之间存在多条边。
- 求解最短路径。

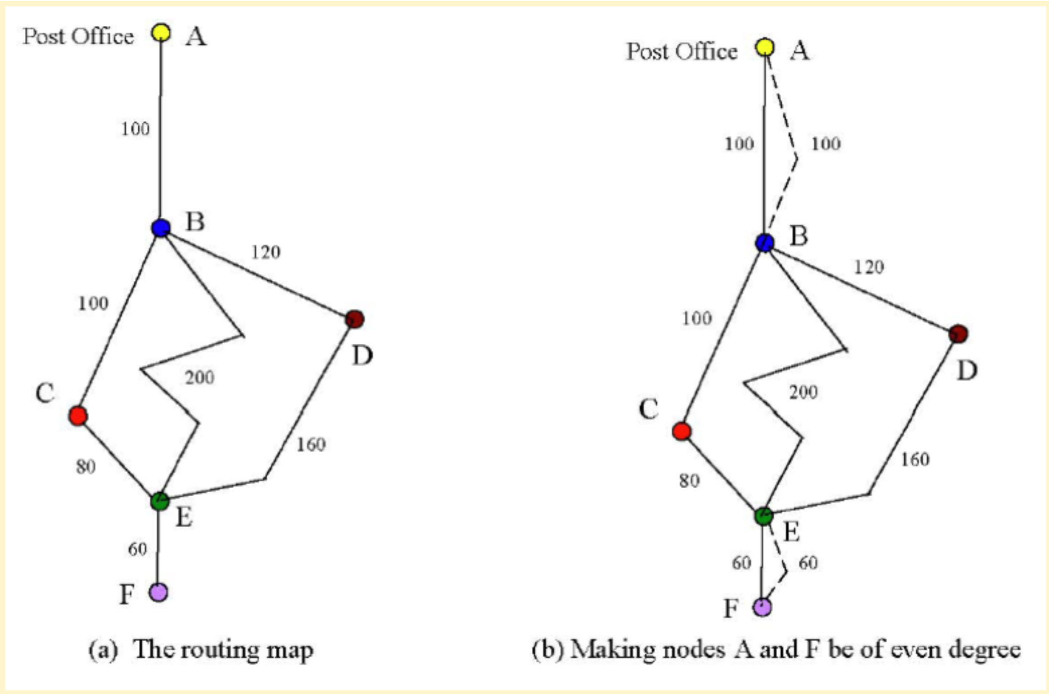
管梅谷算法：

1. 确定初始可行投递路线；
2. 对于每一个**度为奇数**的结点，添加一条边将它和任意一个邻接结点相连；
3. 重复步骤 2，直到图中没有奇数度结点；
4. 检查包含新增边的环——对于 **每一个** 这样的环，如果环中所有新增边的长度之和，**均** 不大于该环总长度的一半，那么保留这个环。

> 算法的另一种描述方法：
>
> 1. 确定初始可行投递路线：
>   1. 找出所有奇度顶点，将其一一 **配对**；
>   2. **找出每一对奇度顶点之间的任意一条路，将路上各边重复一次**，权保持不变；
>   3. 此时原图已经变成欧拉图，取其中一条欧拉环作为初始可行路线。
> 2. 检查图的每条边：若某条边的重复条数大于 1，则去掉偶数条重复边，使得此边至多有一条重复边。转到步骤 3.
> 3. 检查图中 **每个圈**，若每个圈上重复边的权重之和都不大于该圈的一半，则得出最优路线。
> 4. 否则去掉该圈上的重复边，而将该圈上未重复的边重复一次，转到步骤 2.

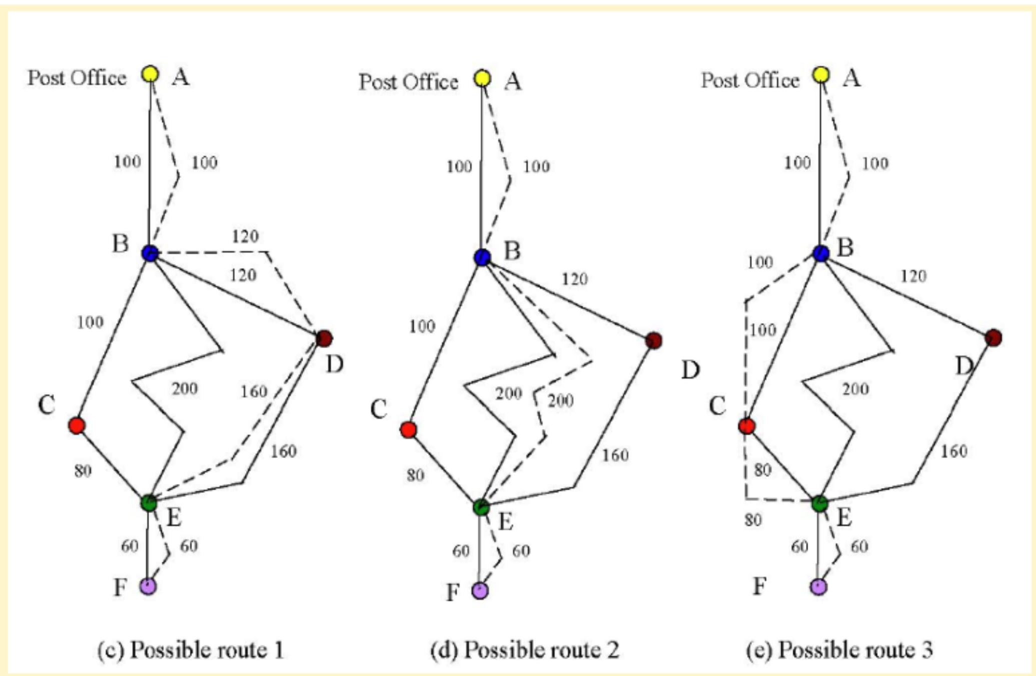
例如：

A 结点的度为奇数，F 结点的度为奇数，所以在这两个结点上添加边。



(a) The routing map (b) Making nodes A and F be of even degree

然后 B 结点和 E 结点的度变成了奇数， 先处理 B 结点，有三种情况，可以分别连在三个邻接结点上。然后继续往下调整，直到没有奇数度的结点。



(c) Possible route 1 (d) Possible route 2 (e) Possible route 3

出现了三种可能结果，一一来看。

1. 第一种情况中，新增边组成的环有 B-D-E-B、B-D-E-C-B 两个
    - B-D-E-B 环的总长度为 480
    - B-D-E-C-B 环的总长度为 460

    而新增边 BD 和 DE 的长度之和为 120+160=280，上面两个环长度的一半均小于该值，所以不符合。
2. 第二种情况中，新增边组成的环有两个：
    - B-C-E-B 长度为 380
    - B-E-D-B 长度为 480

    新增边 BE 长度为 400，大于 B-C-E-B 长度的一半，所以不符合。
3. 第三种情况中，新增边组成的环有两个：

- B-C-E-B 长度为 380
- B-C-E-D-B 长度为 460

新增边的长度为 BC + CE = 180，满足条件。所以一种走法为 A-B-C-E-F-E-D-B-E-C-B-A。

网络拓扑

Small World Network

小世界网络就是一个由大量顶点构成的图，其中 **任意两点之间的平均路径长度比结点数量小得多。** 最早由 Duncan Watts 和 Steven Strogatz 提出，将 **高聚类系数和低平均路径长度** 作为特征，一般称作 WS 模型。

- **平均路径长度**：也称为平均最短路径长度，指的是网络中所有结点两两之间最短路径长度的平均值：

$$\text{dist}_c = \frac{2}{N(N+1)}\sum_{i\leq N}\sum_{j\geq i}\text{dist(i, j)}$$

这里的 N 是节点数目，并定义结点到自身的最短路径长度为 0，如果不计算到自身的距离(少考虑 N 条边)，那么平均路径长度的定义就变成：

$$\text{dist}_c = \frac{2}{N(N-1)}\sum_{i\leq N}\sum_{j>i}\text{dist(i, j)}$$

- **聚类系数**：用来描述网络中结点之间集结成团的程度，具体来说，是一个点的邻接点之间相互连接的程度。

$$\text{一个结点 s 的聚类系数} = \frac{\text{与 s 相连的顶点相互之间所连边数}}{\text{这些顶点之间可以连接的最大边数}}$$

WS 模型

对于纯粹的规则网络，当其中连接数量接近饱和时，聚类系数很高，平均路径长度也很短，例如完全图中集聚系数和平局路径长度都是 1.如果考虑将节点排列成正多边形，各节点都只与距离最近的 2K 个结点相连，那么在 K 比较大但仍保证 $K < \frac{2}{3}N$ 时，集聚系数约为 $\frac{3}{4}$。

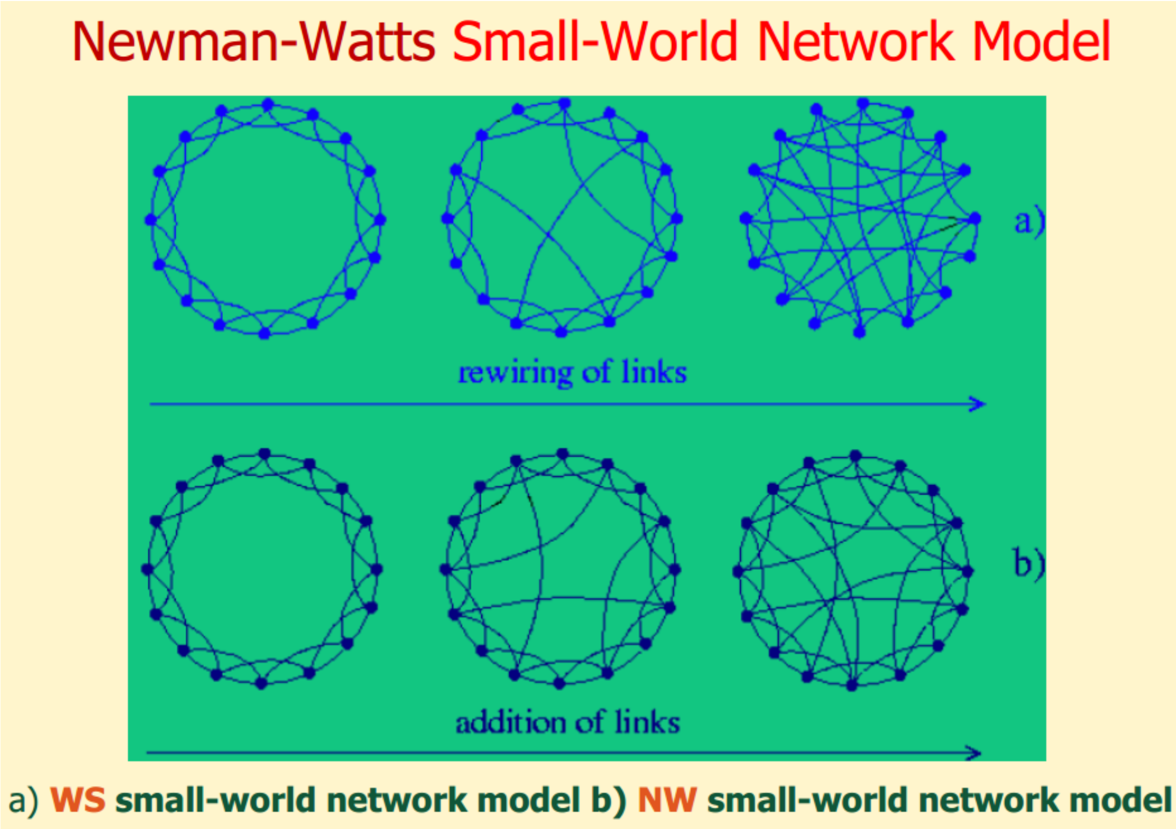
WS 模型基于两人的一个假设：**小世界模型是介于规则网络和随机网络之间的网络**，因此模型从一个完全的规则网络出发，以一定的概率将网络中的连接打乱重连。

WS 小世界网络模型的聚类系数：

$$C(p) = \frac{3(K-2)}{4(K-1)}(1-p)^3$$

小世界网络的两种模型：

- WS 模型（rewiring 算法）：从一个拥有 N 个节点的环状网络开始——这个网络中每个节点都和它相邻的 $2K$ 个节点相连——沿着顺时针/逆时针方向，对于每个节点，处理它的 K 条连线，对于每条连线，靠近邻居的那一端以某个概率 p 断开，然后连到随机的一个节点。
- NW 模型（增加边的算法）：和 WS 模型一样，从一个拥有 N 个节点的环状网络开始——这个网络中每个节点都和它相邻的 $2K$ 个节点相连——以某个概率 p 在每一对节点之间添加一条边。



Scale-free Network（无尺度网络）

无尺度网络的典型特征是网络中大部分节点只和很少节点连接，而有极少的节点与非常多的节点连接，这种关键的节点称作 **枢纽**，它们的存在使得无尺度网络对意外故障有强大的承受能力，而对协同性攻击则显得脆弱。

BA模型：是一个无尺度网络模型，基于两个假设：

1. 网络不断增长；
2. 新的节点在加入时会倾向于和度更大的节点相连。

BA 模型的具体构造过程为：

1. 增长：从一个较小的网络 G_0 开始（有 n_0 个节点和 E_0 条边），逐步加入新的节点；
2. 连接：假设原来网络已经有 n 个节点，在每次加入一个新节点时，从该节点向原有的 n 个节点连出 $m < n_0$条边；
3. 优先连接：优先考虑度数高的节点。

免疫策略：

- **随机免疫**：免疫阈值是 $g_c = 1 - \frac{\lambda_c}{\lambda}$，其中 $\lambda_c = \frac{<k>}{<k^2>}$ 表示有效流行阈（effective epidemic threshold）（λ 表示传播率，λ_c 表示 threshold），所以有 $g_c = 1 - \frac{1}{\lambda}\frac{<k>}{<k^2>}$，那么当节点数量 N 很大时，k 也会很大，那么 λ_c 就会趋近于 0，而免疫阈值 g_c 就会趋近 1，也就是说为了免疫一个无尺度网络（scale-free network），需要免疫几乎所有人才行。
- **针对性免疫**：免疫那些度最高的节点。BA 模型中，免疫阈值（immunization threshold）是 $g_c \sim e^{-\frac{2}{m\lambda}}$。因为针对性免疫的免疫门槛比随机免疫的要低很多（$\lambda_c$ 要小很多），所以有效传染率 λ 相对就小很多，那么 g_c 就随之降低很多，所以针对性免疫适合于无尺度网络。
- **熟人免疫**：随机挑一个节点 A，然后免疫它度最大的邻居节点 B，然后免疫 B 的最大度邻居节点 C ……

Coreness

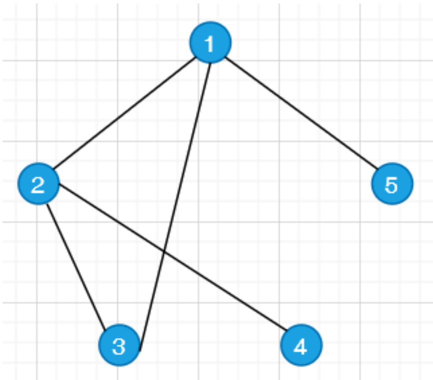
定义： K-core 是指删除图中 $degree \leq k-1$ 的所有结点之后，**剩下的子图**。在删除的过程中：

- 如果一个结点被删除，那么它的所有邻接边都要一并删除；
- $degree \leq k-1$ 的结点被删除之后，**剩下的** 度(可能比之前少 1)不大于 $k-1$ 的结点也要被删除。

结点的 coreness ： 如果一个结点属于一个图的 k-core，但是将在生成 (k+1)-core 前被删除，那么这个节点的 coreness 是 k。

图的 coreness： 图中最大的 coreness 就是图的 coreness。

例子：



- 1-core 是图本身（删除度为 0 的结点）
- 2-core 是三角形 1-2-3（删除度为 1 的结点，及剩下结点中度小于等于 1 的结点）
- 结点 1、2、3 的 coreness 为 2，结点 4、5 的 coreness 为 1，所以图的 coreness 为 2

Coreness 的现实意义

- coreness 高的子图是相对而言更重要的子图，coreness 高的结点是相对而言更重要的结点。
- core 的划分能够揭示网络的层级结构，coreness 更高的结点属于更高层次的网络。
- coreness 更高的网络具有更好的健壮性。

Betweeness

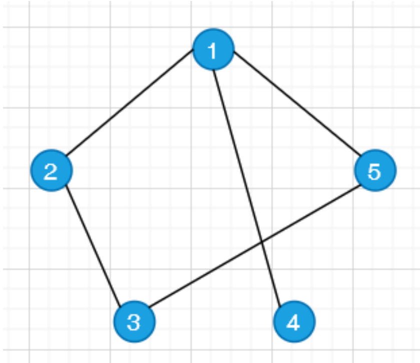
结点 betweeness 定义：

$$B(i) = \sum_{i \neq j \neq l} \frac{L_{jl}(i)}{L_{jl}}$$

其中：

- L_{jl} 是指结点 j, l 之间最短路径的条数。
- $L_{jl}(i)$ 是指结点 j, l 之间最短路径中经过结点 i 的条数。
- 规范化：除以 $(N-1)(N-2)/2$.

例子：



$$B(1) = \frac{0}{(3,2)} + \frac{(2,1,4)}{(2,1,4)} + \frac{(2,1,5)}{(2,1,5)+(2,3,5)} + \frac{(3,2,1,4)+(3,5,1,4)}{(3,2,1,4)+(3,5,1,4)} + \frac{0}{(3,5)} + \frac{(4,1,5)}{(4,1,5)}$$
$$= 0 + \frac{1}{1} + \frac{1}{2} + \frac{2}{2} + 0 + \frac{1}{1} = \frac{7}{2}$$

规范化：

$$\frac{7/2}{(N-1)(N-2)/2}$$

边 betweeness 定义： 和结点 betweeness 定义相似，但要求分母是 **经过指定边的最短路径的条数。**

同步（synchronization）

有害同步的例子：

- Synchronization occurs when separate TCP connections **share a common bottleneck router**
- Two processes can become synchronized simply if they **are both synchronized to the same external clock**
- Multiple clients can become synchronized as they **wait for services from a busy server**.

有用同步的例子：

- secure communications
- generation of harmonic oscillations (谐波振荡😵)
- language emergence and development: synchronization in conversations → common vocabulary
- biological systems

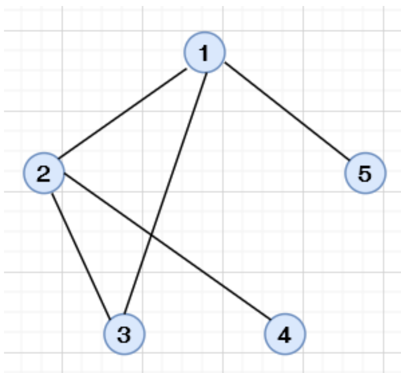
耦合矩阵（coupling matrices）：

一个包含 N 个节点的图，其耦合矩阵是一个 $N \times N$ 对称方阵，其中对角线上的元素 a_{ii} 的值为对应 **结点度数的相反数**；其他位置元素 a_{ij} 的值为 1 或者 0，表示对应节点是否直接相连。

耦合矩阵的特征值 $\lambda_N \le \cdots \le \lambda_2 < \lambda_1 = 0$

拉普拉斯矩阵和耦合矩阵的关系 $L = -A$。

例子：

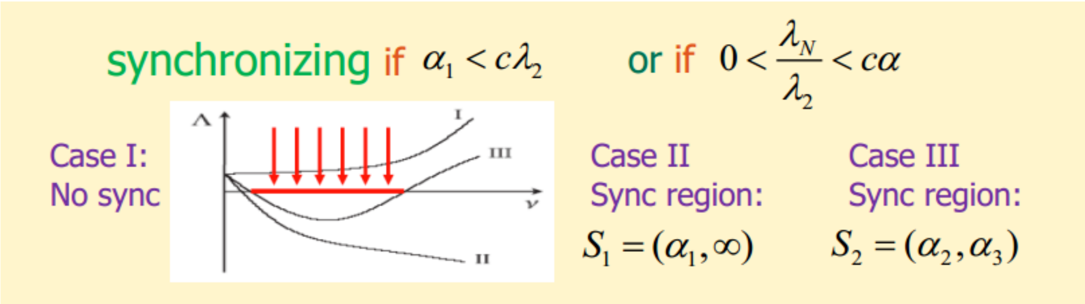


$$
A=\begin{bmatrix}
-3 & 1 & 1 & 0 & 1\\
1 & -3 & 1 & 1 & 0\\
1 & 1 & -2 & 0 & 0\\
0 & 1 & 0 & -1 & 0\\
1 & 0 & 0 & 0 & -1
\end{bmatrix}
$$

耦合矩阵的性质：对称（symmetrical）、diffusive、不可约（irreducible）、行/列和为 0（zero row-sum）。

同步性 synchronizability

拉普拉斯矩阵的特征值 $0=\lambda_1<\lambda_2\le\cdots\le\lambda_N$



> 上图中，纵轴小于零表示同步，横轴为 α 的取值。

同步性情况分析：

- case 2：设零点为 α_1，则同步范围为 $S_1=(\alpha_1,\infty)$。这种情况的同步性由拉普拉斯矩阵的 **最小非零特征值** λ_2 决定，即需要满足条件 $c\lambda_2>\alpha_1$。λ_2 **越大**，满足这个条件需要的 c 值就可以越小，这就意味着这个网络的 **同步性越强**。
- case 3：同步范围为 $S_2=(\alpha_2,\alpha_3)$，同步性由 λ_N/λ_2 决定，需要满足条件 $0<\frac{\lambda_N}{\lambda_2}<c\alpha$。这个 **比率越小，同步性越强**。

正则偶图的同步性：

- 全局耦合网络（globally coupled network）：（$\lambda_2=\cdots=\lambda_N=N$）通常，**不管耦合强度 c 多小**，全局耦合网络只要规模足够大，就**都会同步**。
- 本地耦合网络（locally coupled network）：**不管耦合强度 c 多大**，只要规模足够大，就 **都不会同步**。

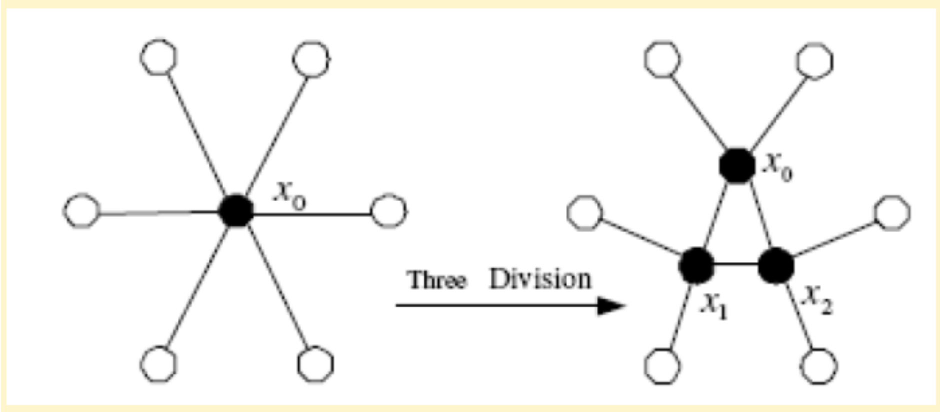
小世界网络的同步性： 通过增加很小比例的 **长距离连线**（long-range connections），小世界网络的同步性都能得到极大提高，这是小世界网络的一个优点。

影响同步性的几个因素：

- **建立连接的可能性 p**：**小世界网络** 中，p **越大**，则 λ_N/λ_2 越小，同步性就越好。
- **幂定律指数（power-law exponent）**：**无尺度网络** 中，这个**指数越大**，λ_N/λ_2 就越小，同步性就越好。
- **结点 betweenness**：小世界网络和无尺度网络中，**结点 betweenness 越小**，同步性越好。

增强同步性的途径：

- **To eliminate the maximal betweenness**：the node with the largest betweenness is **replaced by several connected nodes**, so that the shortest paths that passed through the original node will now only pass one or two new nodes , which will reduce the maximal betweenness dramatically.



- **To reduce the impact of the heterogeneity**(异质性) of degree and betweenness distributions.
- 除了 betweenness 大的节点，负载大的边也能引起网络上的流量拥堵（data-traffic congestion），所以，如果解除这些负载大的边的耦合（if such **heavily-loaded edges are decoupled**），数据流量会重新分布，从而网络更通畅。

增加边对同步性的影响：

- （引理 1）如果同步域（synchronized region）是 **无限** 的（case 2），那么 加边**不会**降低同步性，因为特征值的关系：$\lambda_i(G+e) \geq \lambda_i(G)$，就是说加边只会让特征值变大，那么同步性只会增加。
- 而在同步域有限（case 3）的情况中，不存在这样的关系。

补图的特征值：

给定一个图 G，补图为 G^c：

1. G 最大的特征值 $\lambda_N(G)$ 满足 $\lambda_N(G) \leq N$。
2. $\lambda_N(G) = N \Leftrightarrow G^c \text{不连通}$。
3. $\lambda_N(G^c) = N - \lambda_{N-i+2}(G)$。