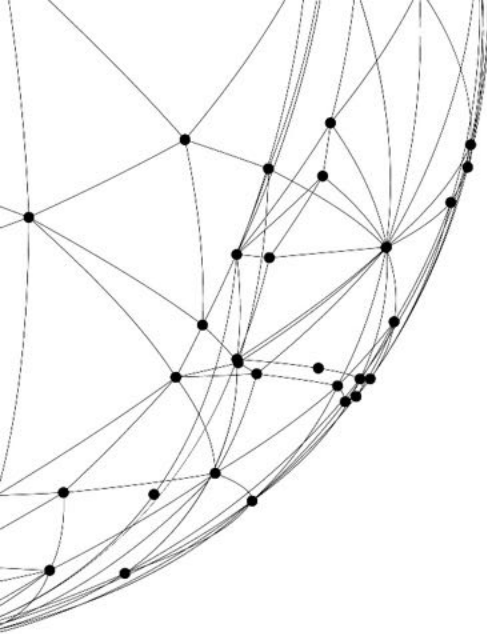
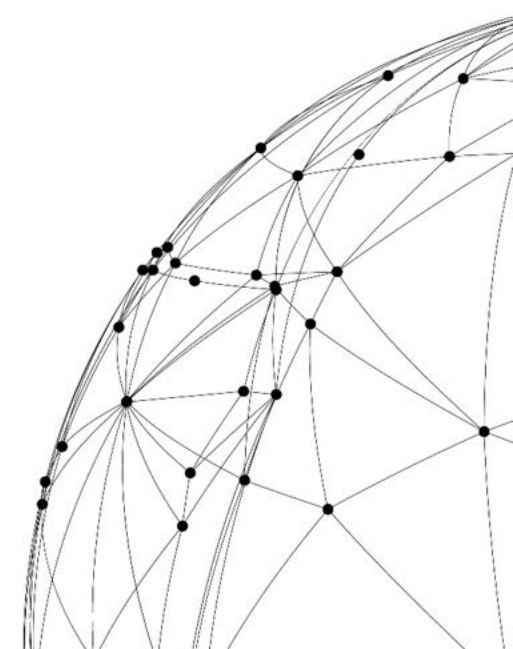




小世界网络

陶韵竹
地球与空间科学学院





目录

01

概念

02

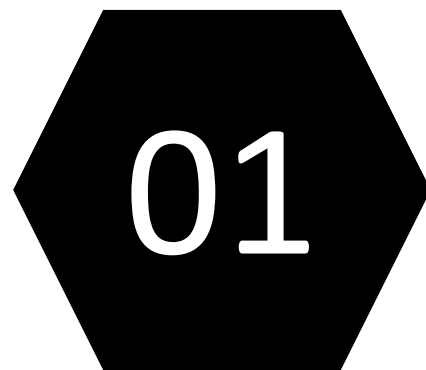
集聚系数

03

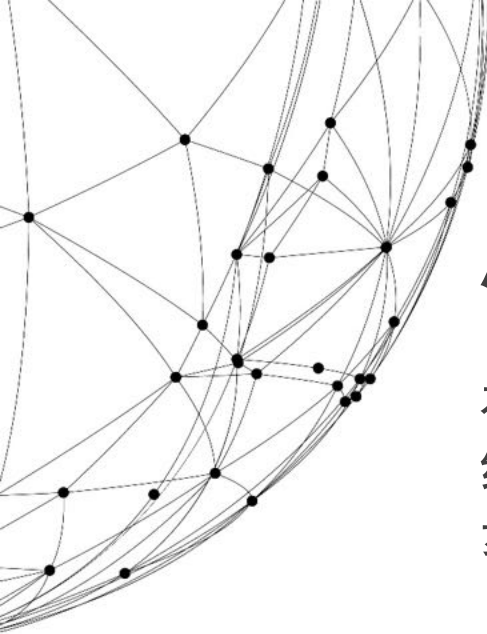
小世界网络的判断

04

实例



概念

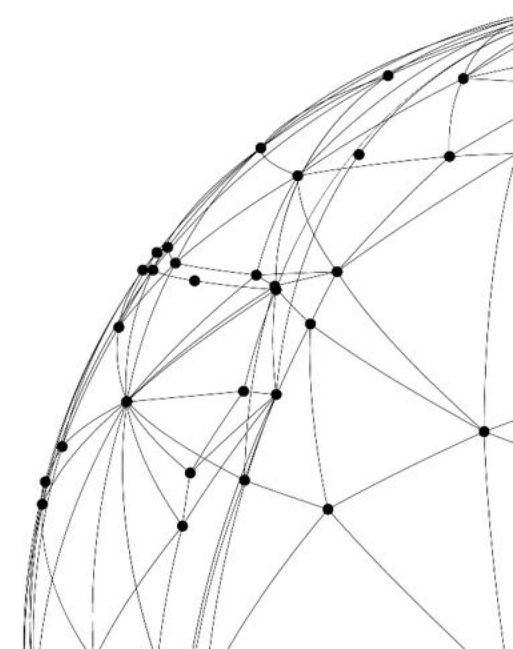


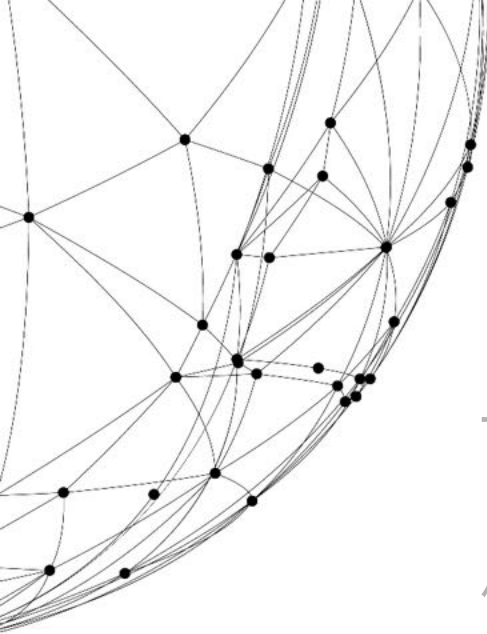
小世界网络

在网络理论中，**小世界网络**是一类特殊的复杂网络结构，在这种网络中大部分的节点彼此并不相连，但绝大部份节点之间经过少数几步就可到达。

在日常生活中，有时你会发现某些你觉得与你隔得很“遥远”的人，其实与你“很近”。小世界网络就是对这种现象（也称为小世界现象）的数学描述。

用数学中图论的语言来说，小世界网络就是相对于同等规模节点的随机网络，具有较短的平均路径长度和较大的聚类系数特征的网络模型。除了社会人际网络以外，小世界网络的例子在生物学、物理学、计算机科学等领域也有出现。



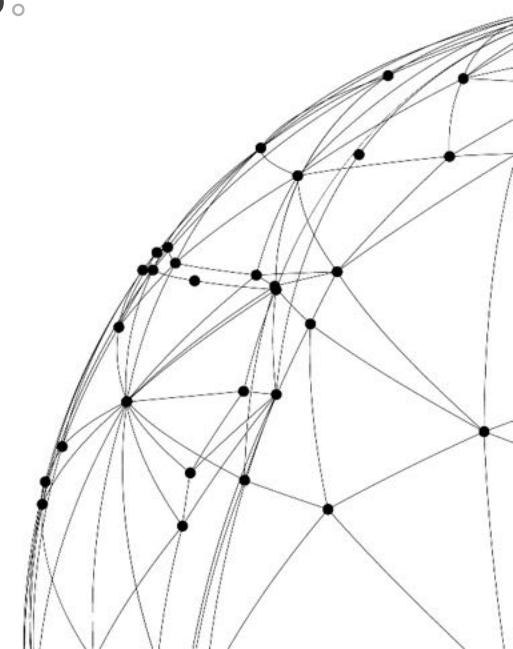


“六度分隔”理论

二十世纪60年代，美国哈佛大学社会心理学家斯坦利·米尔格伦 (Stanley Milgram) 做了一个连锁信实验。

他将一些信件交给自愿的参加者，要求他们通过自己的熟人将信传到信封上指明的收信人手里，他发现，20%的信件最终送到了目标人物手中。而在成功传递的信件中，平均只需要6.5次转发，就能够到达目标。也就是说，在社会网络中，任意两个人之间的“距离”是6。

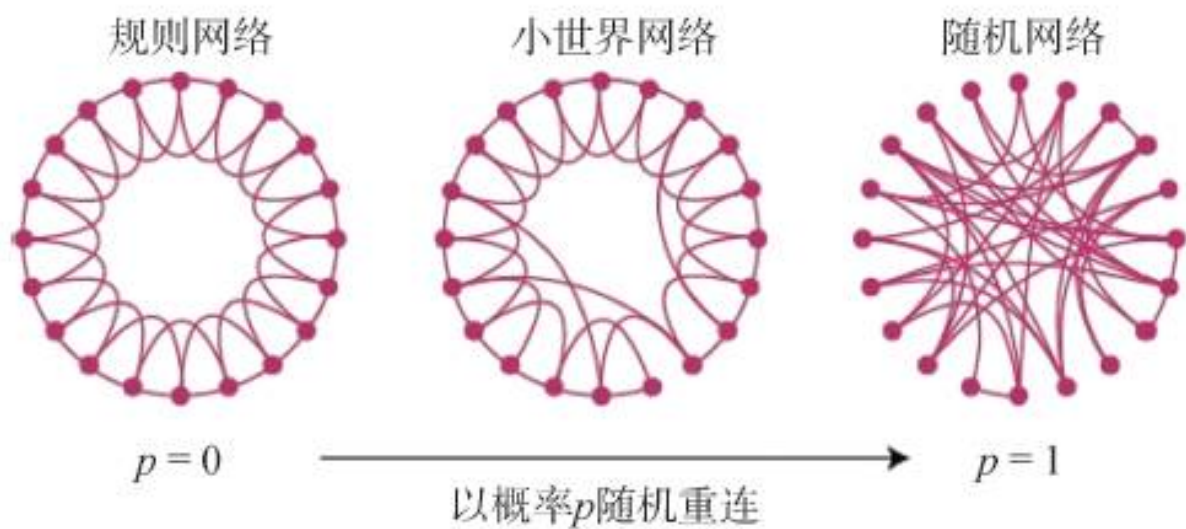
米尔格拉姆实验展示了大型社会网络的两个重要事实：大型社会网络一定包含了丰富的最短路径；人们只需要通过大型社会网络的局域信息，而不是全局信息，就可以找到这些最短路径。



小世界网络

以前，人们认为网络分为完全规则网和完全随机网，这两类网络具有各自的特征。

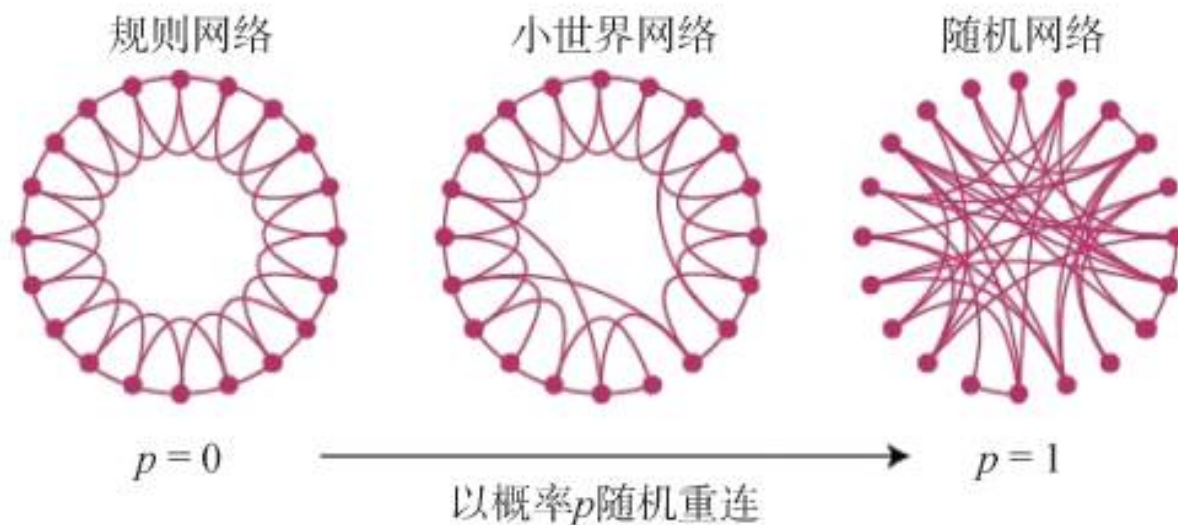
规则网具有较大的特征路径长度，聚类系数也较大，而随机网络具有较小的特征路径长度，但是聚类系数较小。真实世界的网络既非完全规则，也非完全随机，而是介于这两者之间。



瓦茨-斯特罗加茨 (WS) 模型

以一个规则网络为基础，然后以概率 p 对规则网络中的每条边进行随机重连（剔除节点的自我连接和重边连接），这些随机重连的边在规则网络中既引入了随机性，又引入了长程连接（或称为捷径）。

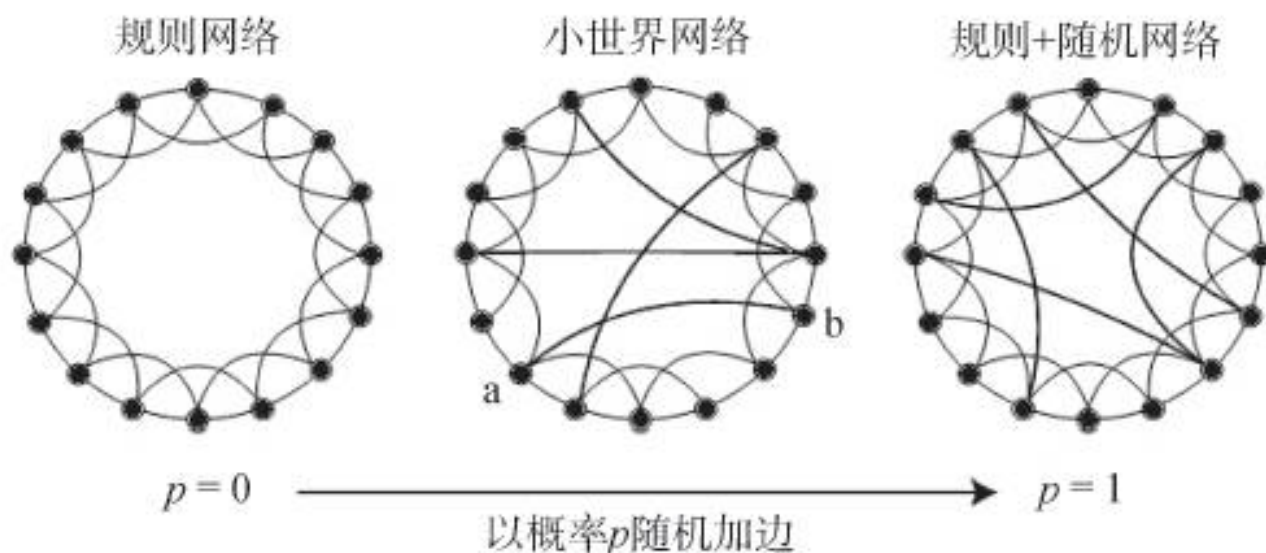
这些少量的、随机的长程连接大大地减小了网络节点之间的平均距离，却又保持了原有网络较高的群聚性，使网络具有了小世界特性。

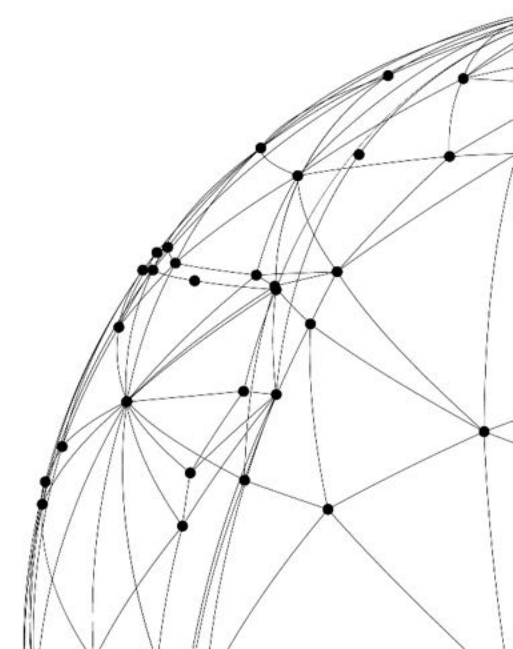
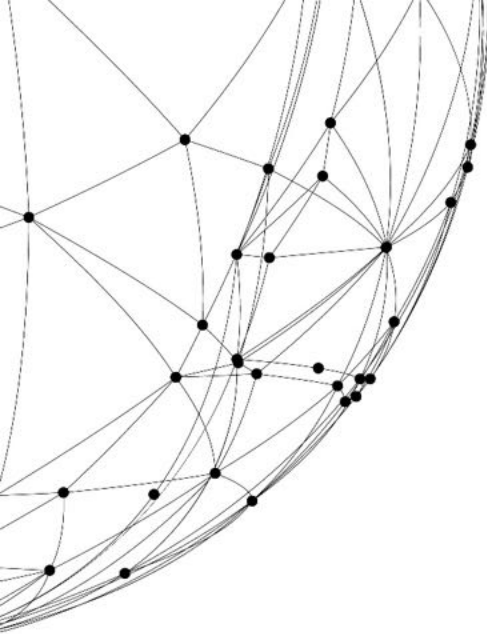


纽曼-瓦茨 (NW) 模型

对WS小世界网络模型作了改进，在不改变原规则网络中各节点之间的连边情况下，任选二个节点，由概率 p 决定是否在这2个节点之间增加一条边，如下图在a、b两节点之间就增加了一条随机连边。

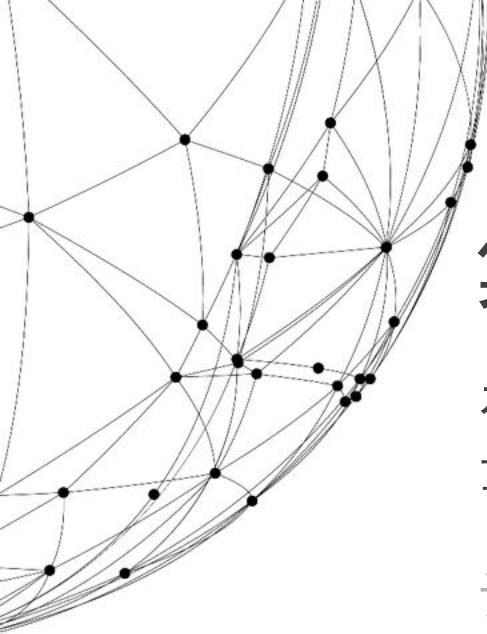
这样的操作同样在规则网络中既引入了随机性，又引入了长程连接，从而使网络具有了小世界特性。





02

集聚系数



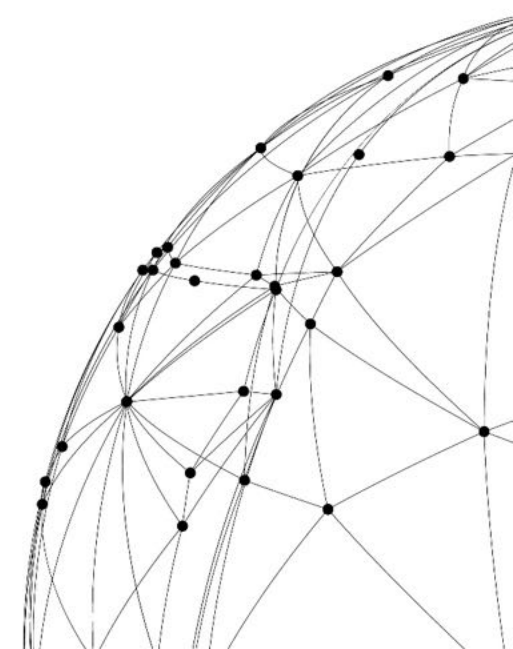
集聚系数

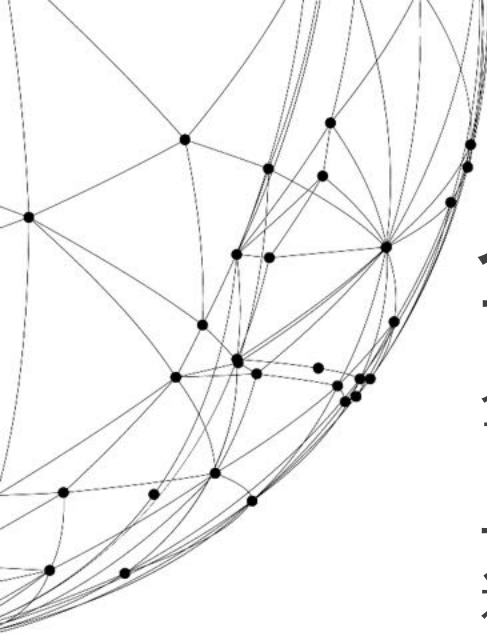
在图论中，**集聚系数**是图中的点倾向于集聚在一起的程度的一种度量。

证据显示：在多数实际网络以及特殊的社会网络中，结点有形成团的强烈倾向。

在实际网络中，这种可能性比随机生成的均匀网络的那种结点连接情况的可能性大。

这一度量有两种版本的方法：全局的和局部的。
全局的方法旨在度量整个网络的集聚性；
局部的方法给出了单个结点的嵌入性的度量。

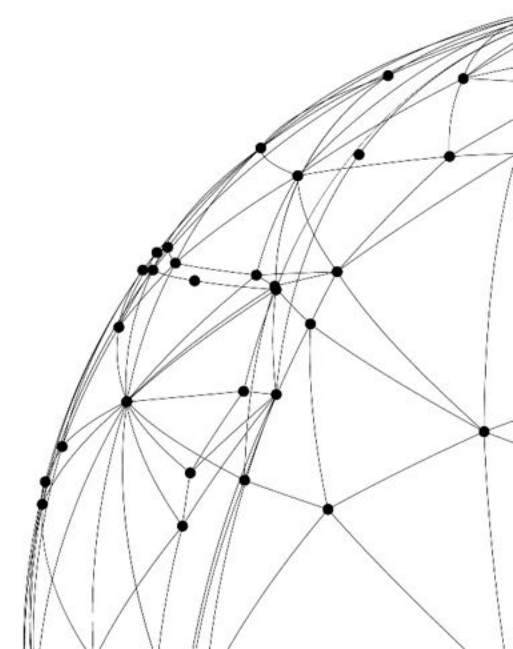


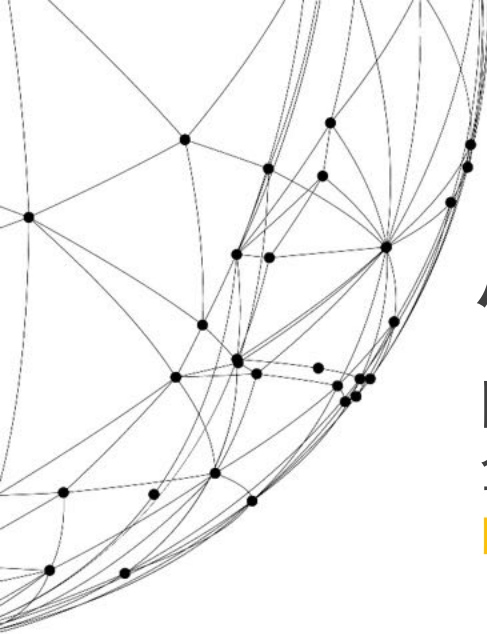


全局集聚系数

全局集聚系数是基于**结点三元组**的。

一个三元组是其中有两边（开三元组）或三边（闭三元组）无向边连接的三个结点。一个三角由三个闭三元组构成，三角集中在每一个结点上。**全局集聚系数是所有三元组（包括开和闭的）中闭三元组的数目。**

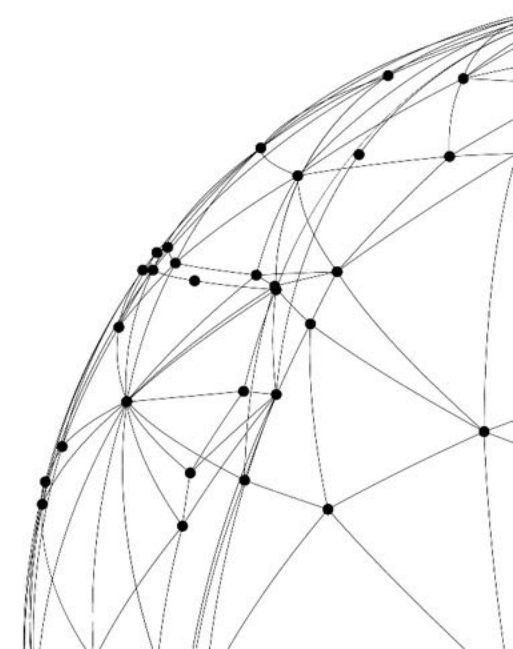
$$C = \frac{3 \times \text{三角形个数}}{\text{所有节点的连通三点组个数}} = \frac{\text{闭三点组个数}}{\text{所有节点的连通三点组个数}}$$


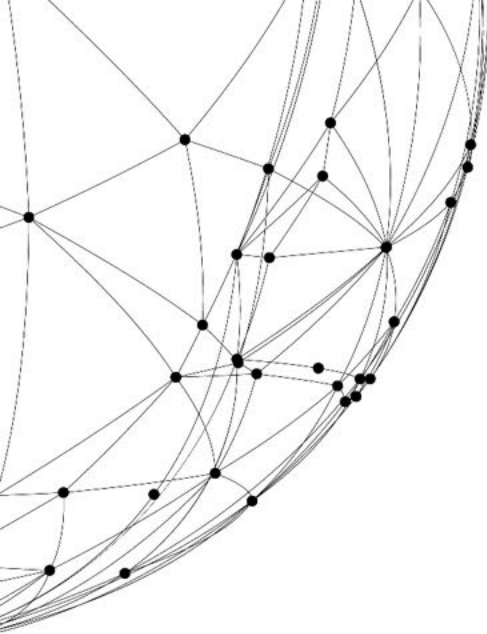


局部集聚系数

图中一个结点的局部集聚系数表示了它的相邻结点形成一个团（完全图）的紧密程度。结点 v_i 的局部集聚系数 C_i 是它的相邻结点之间的连接数与它们所有可能存在连接的数量的比值。

局部聚合系数是网络的局部特征，在社交网络中可以反映相邻两个人之间朋友圈子的重合度，即该节点的朋友之间也是朋友的程度。





局部集聚系数

$$G = (V, E)$$

e_{ij} : 连接结点 v_i 与结点 v_j 的边

$$N_i = \{v_j : e_{ij} \in E \cap e_{ji} \in E\}$$

k_i : v_i 相邻结点的数量

① 有向图的局部集聚系数

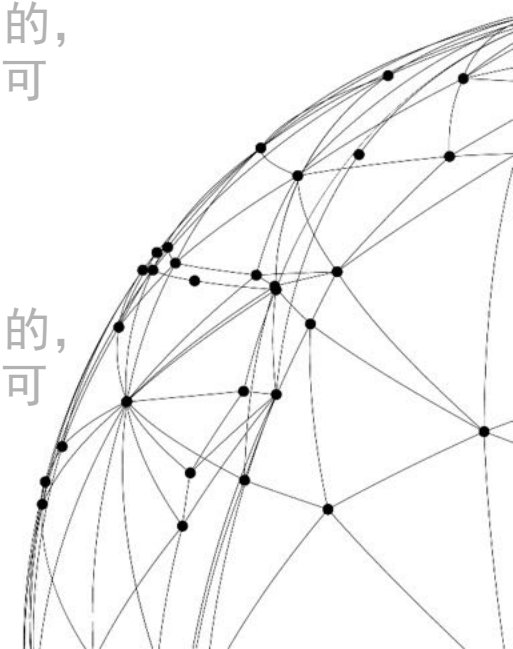
$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E$$

对于一个有向图， e_{ij} 与 e_{ji} 是不同的，因此每个邻结点 N_i 的邻结点之间可能存在有 $k_i(k_i-1)$ 条边。

② 无向图的局部集聚系数

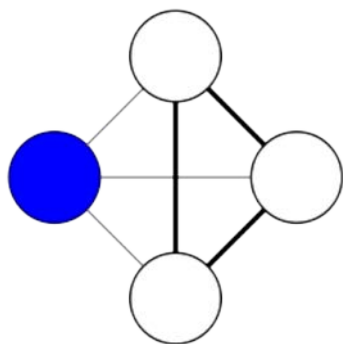
$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E$$

对于一个无向图， e_{ij} 与 e_{ji} 是相同的，因此每个邻结点 N_i 的邻结点之间可能存在有 $k_i(k_i-1)/2$ 条边。



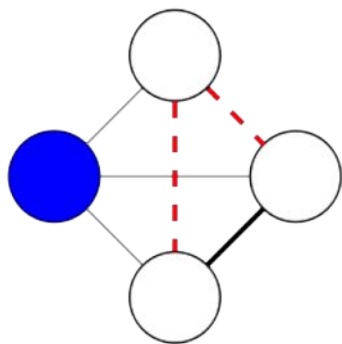
局部集聚系数

计算局部集聚系数的例子：（无向图）



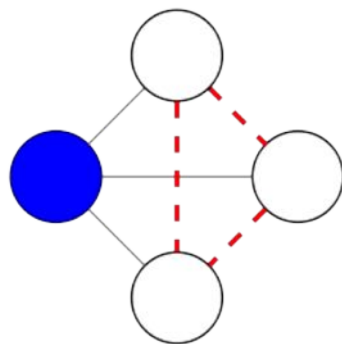
$$c = 1$$

$$\begin{array}{l} |\{e_{jk}\}| = 3 \\ k_i = 3 \end{array}$$



$$c = 1/3$$

$$\begin{array}{l} |\{e_{jk}\}| = 1 \\ k_i = 3 \end{array}$$

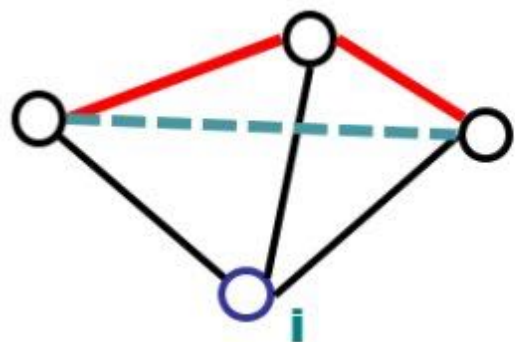


$$c = 0$$

$$\begin{array}{l} |\{e_{jk}\}| = 0 \\ k_i = 3 \end{array}$$

局部集聚系数

计算局部集聚系数的例子：（无向图）

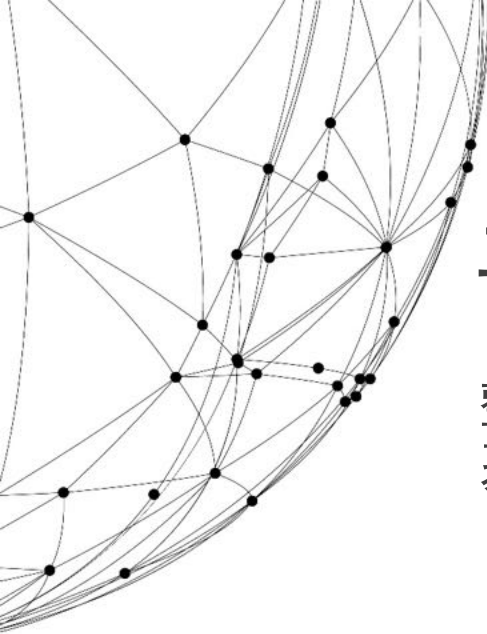


$$n_i = 3$$

节点 i 的邻居节点当前有2个连接

节点 i 的邻居节点最多可能有3个连接。

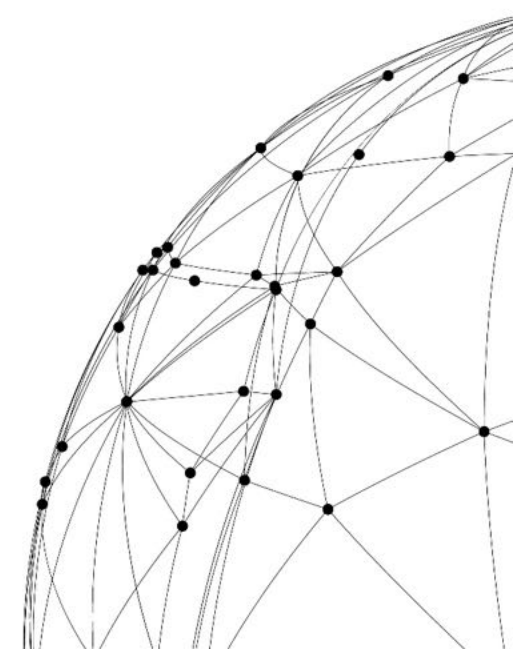
$$C_i = 2/3$$

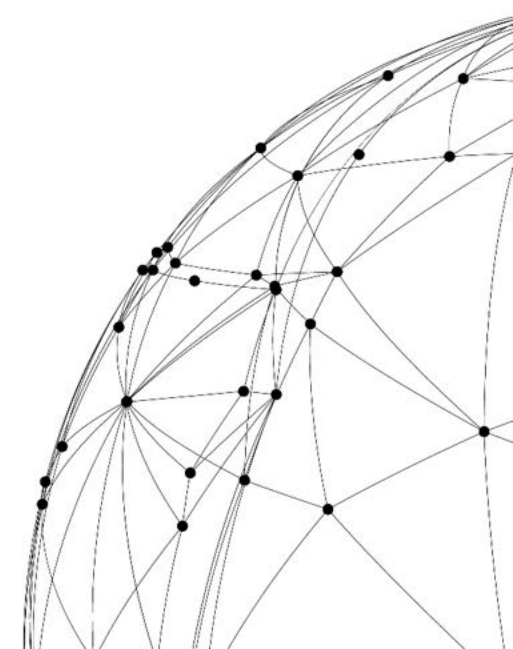
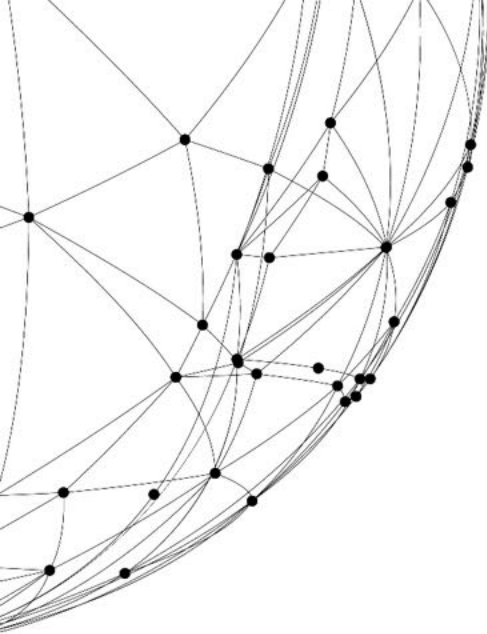


平均集聚系数

整个网络的集聚系数由Watts和Strogatz定义为所有结点n的局部集聚系数的均值：

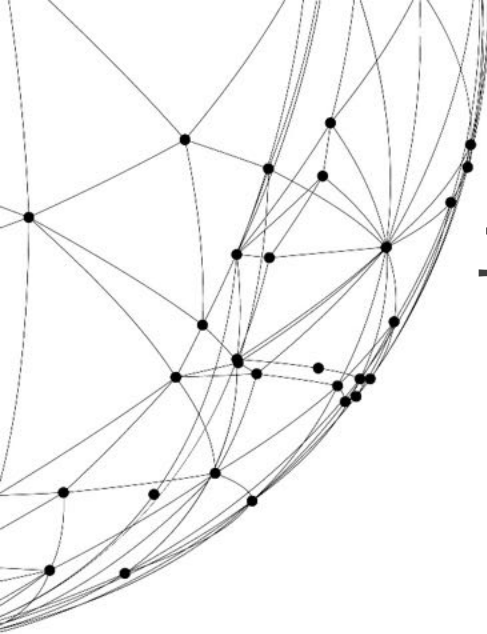
$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$





03

小世界网络的判断



平均最短路径

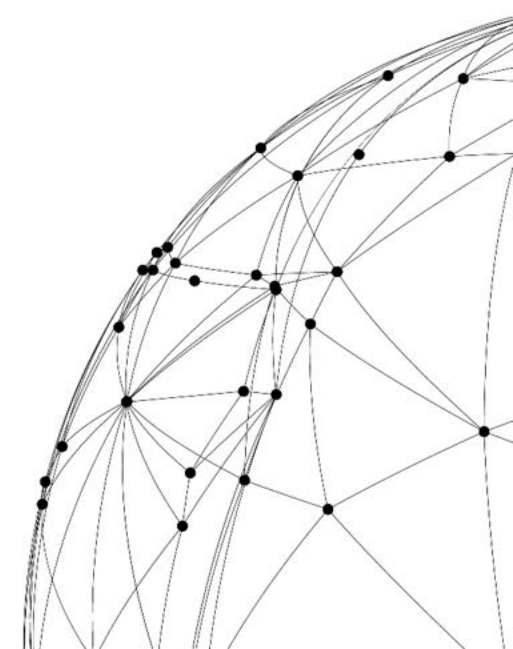
$$G = (V, E)$$

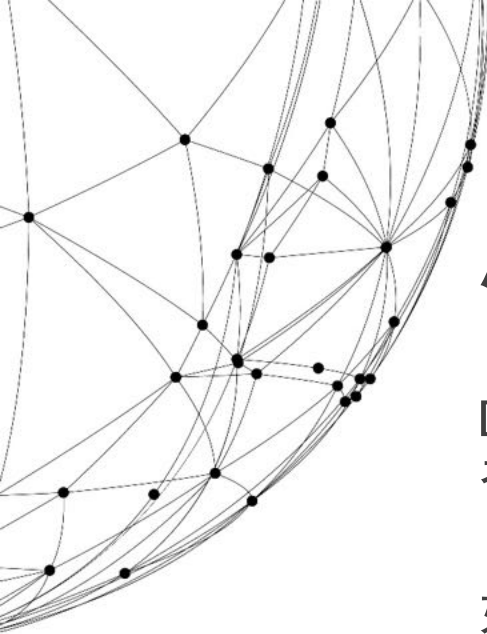
$$d(v_1, v_2) = \begin{cases} \text{结点 } v_1 v_2 \text{ 间的最短路径长度} & (\text{如果 } v_1 \text{ 到 } v_2 \text{ 是可达的}) \\ 0 & (\text{如果 } v_1 \text{ 到 } v_2 \text{ 是不可达的}) \end{cases}$$

平均最短路径 l_G 定义为

$$l_G = \frac{1}{n \cdot (n - 1)} \cdot \sum_{i \neq j} d(v_i, v_j),$$

n 是图 G 中顶点的个数

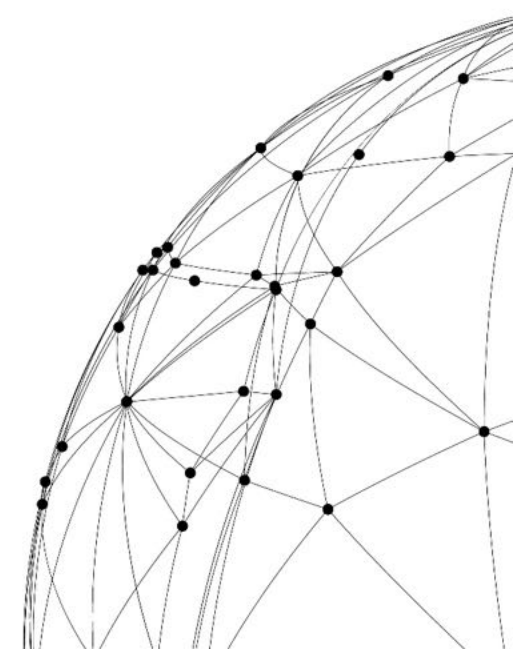


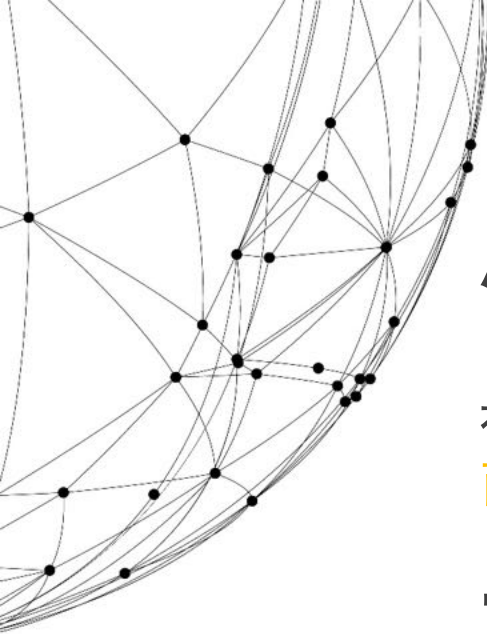


小世界网络的判断

Duncan J. Watts和Steven Strogatz在1998年引入了度量一个图是否是
小世界网络的方法。

如果一个图的平均集聚系数显著高于相同结点集生成的随机图，而且平均最短距离与相应随机生成的随机图相近，那么这个图被认为是小世界的。



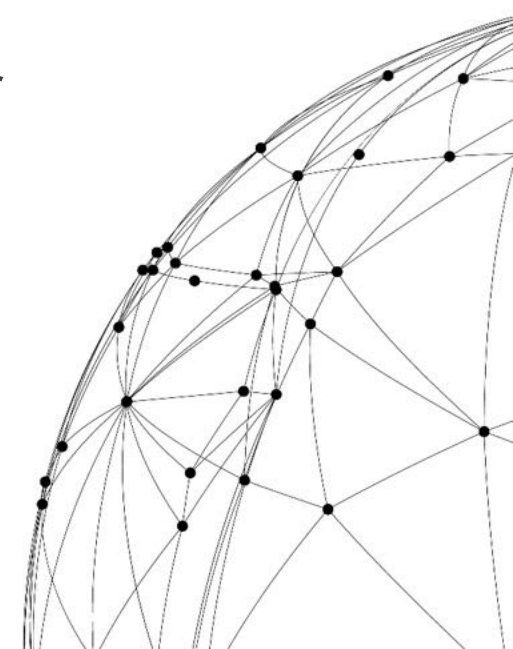


小世界网络的判断

社会学中描述网络是否呈现小世界特性的特征参数是所谓的小世界商 Q 。

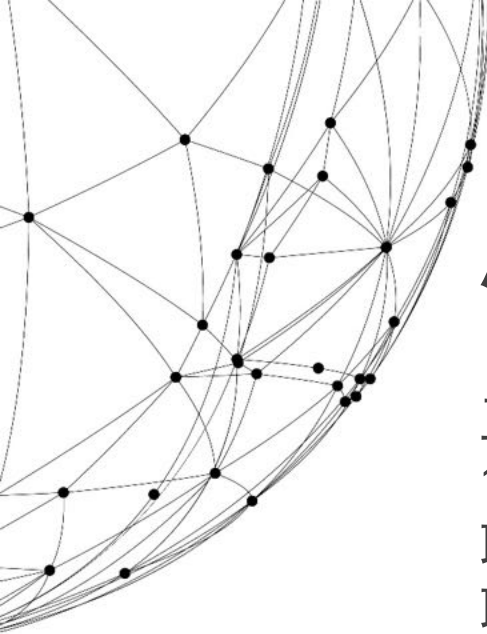
它是将实际网络的平均群聚系数 C 与平均路径长度 L 的比值 C/L ，去除以对应的具有相同节点数和边数的随机网络的平均群聚系数 C_0 与平均路径长度 L_0 的比值 C_0/L_0 ，这样的一个商值。

小世界商 Q 大于1，则表明该网络具有小世界特性。商值 Q 越大，网络的小世界特性越显著。



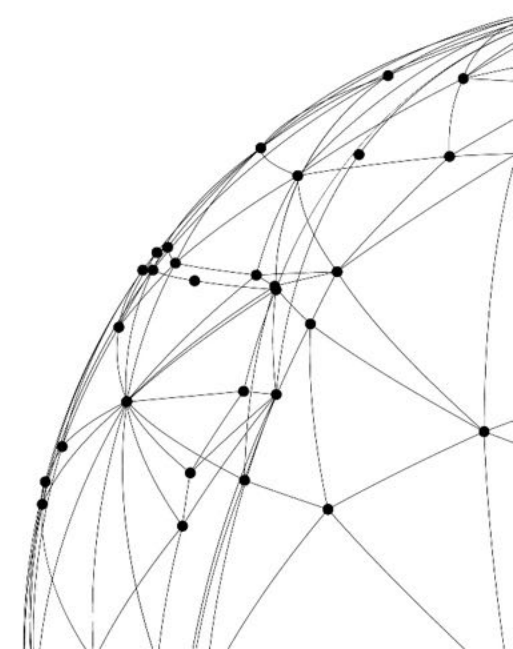


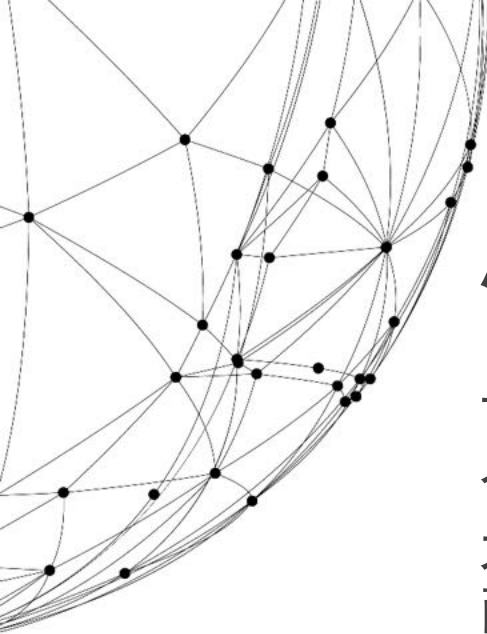
实例



小世界网络的实例

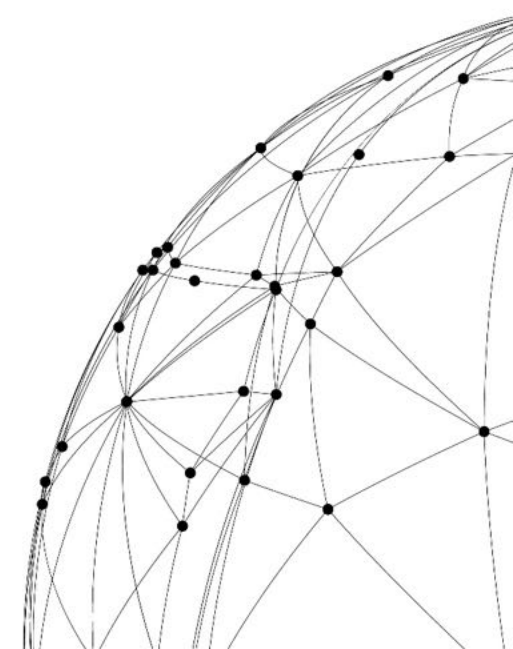
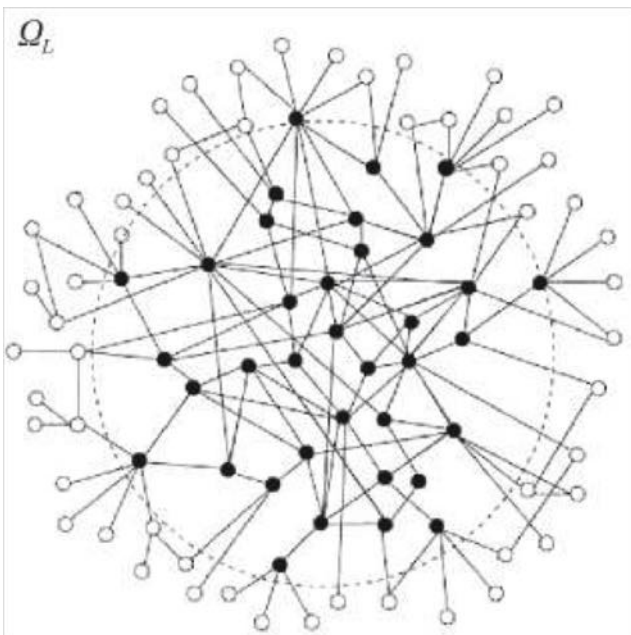
互联网的节点是各个路由器，连边则是连接各个路由器的光纤。在1995~1999 年对于互联网网站及路由器层次都进行了计算，发现互联网的平均路径长度是 $L=4.0$ ，这与ER随机图的 $L=10$ 相比明显小。互联网的群聚系数则是 $C=0.18\sim 0.3$ ，这与ER随机图的 $C=0.001$ 相比明显偏大，是一个小世界网络。

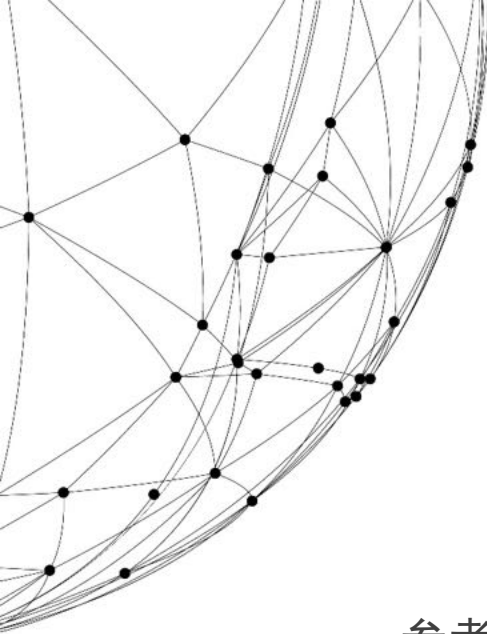




小世界网络的实例

下图所示的语言网络也是小世界网络。每一个单词是一个节点，两个单词相连接出现在一个句子中即为有连边。人脑可以记忆的单词大约有 $10^4 \sim 10^5$ 个。据计算，两个单词之间的平均距离是 $d=2 \sim 3$ ，即两个单词平均只要通过 $2 \sim 3$ 个中间单词的相连即可出现在一个句子中。因此，语言网络是典型的小世界网络。





THANK YOU

参考网站：

小世界网络浅介 http://www.sohu.com/a/125724380_372479

社交网络分析：小世界网络

http://blog.sina.com.cn/s/blog_72ef7bea0102v8g1.html

浅谈小世界网络

http://blog.csdn.net/database_zbye/article/details/8246889

Clustering coefficient (集聚系数)

http://blog.sina.com.cn/s/blog_439371b501012lgw.html

Average path length(from Wikipedia)

https://en.wikipedia.org/wiki/Average_path_length

