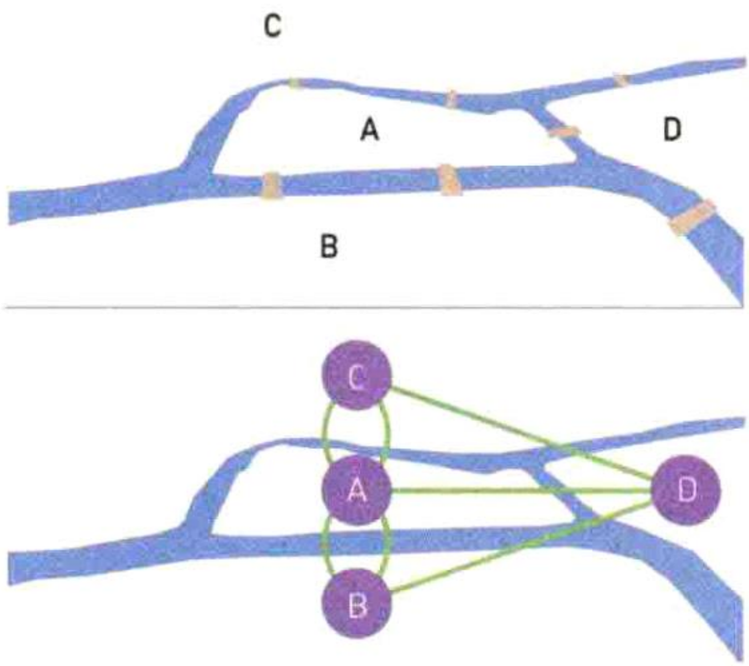
第一节 哥尼斯堡得到七桥问题

2022年12月12日　　8:53

1753年 著名的哥尼斯堡七桥问题被**欧拉**做出了严格的证明

七桥问题：有4块区域通过7座桥相连，那么能否存在一个路径使得一个人不重复的走过每一座桥。



欧拉使用字母来表示每一块区域，线来表示连接每一不同区域的桥梁，于是将上图的现实问题抽象成了字母、线段的图像问题。这极大地简化了我们求解的过程，屏蔽掉了一些复杂的干扰信息。

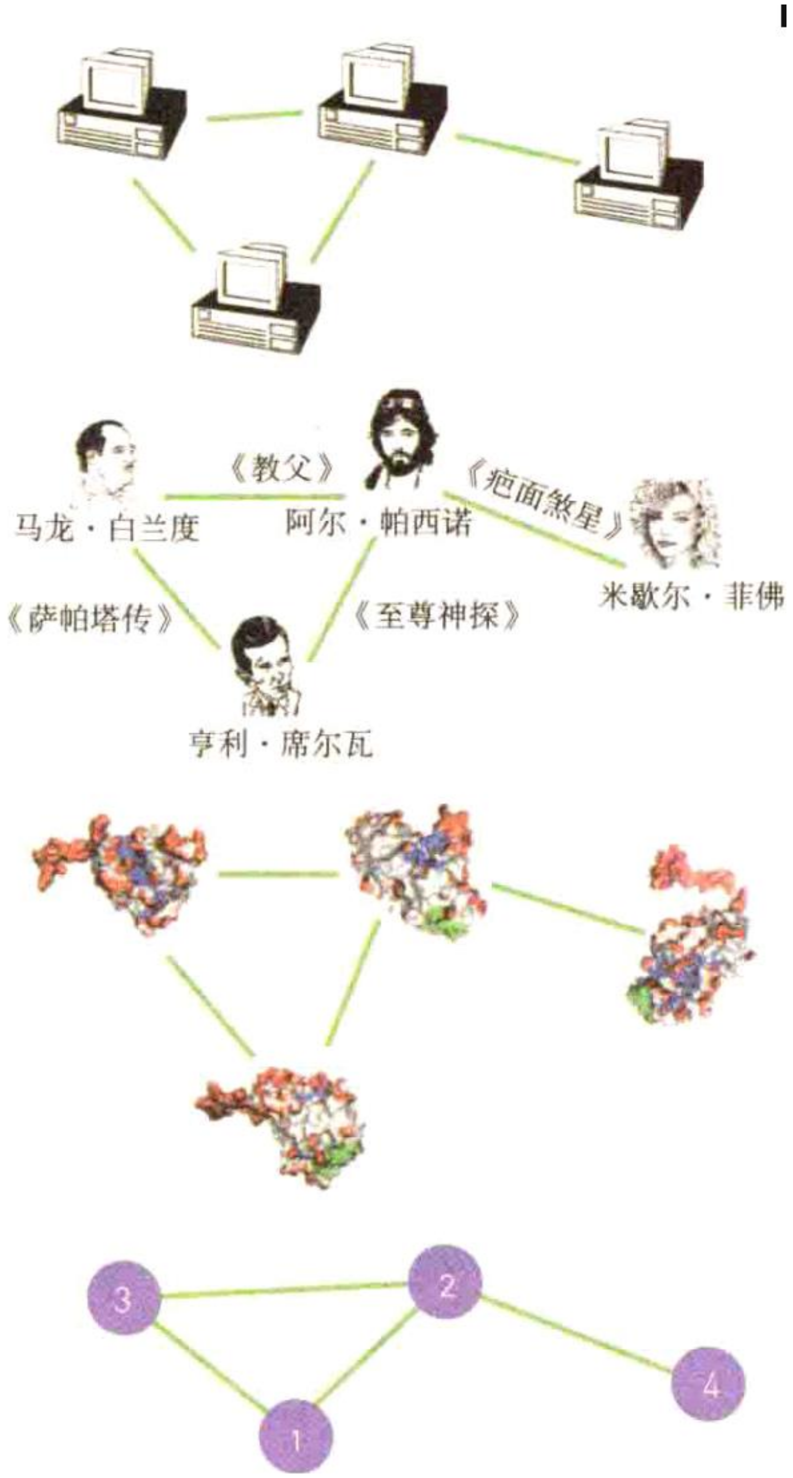
欧拉对于七桥问题给出了一个总结，如果一幅图画中*奇数连接点数超过2个*，那么这幅图就不存在一个路径能够不重复的走完每一个点，这也被我们称为**“一笔画”**问题。

同时，欧拉使用“图”的思想解决现实问题也给我们了一些启示：**图是一种能够简化现实问题，并且具备一定性质和内涵的一类工具**。这也是“图论”的起源。

第二节 网络和图

2022年12月12日 8:53

要理解复杂的系统，我们就需要去理解复杂系统之间的交互方式。网络作为复杂系统的一种表示方式，为我们研究复杂系统提供了公共语言。



那么我们首先从网络的基本概念入手：

1.**节点i**，表示网络中实体的数量，一般为数字、字母与圆圈的组合表示，例如最后一个图中节点1，节点2，节点3，节点4，当然也可以直接称呼节点名称。

2.**节点数N**，指的是一个网络中存在的节点数量，一般用来表示网络的大小。例如最后一个图，N=4。

3.**链接l**，表示网络实体之间的交互关系，一般用线将两个或多个实体相连来表示。
注：链接可以是有向的，也可是无向的。若一个网络中所有的连接都是有向的，则我们称其为有向网络或者有向图，同理，若都是无向链接，则称为无向图或无向网络。

4.**链接数L**：表示节点间交互关系的数量，例如最后一个图中L=4，也就是网络中线段的数量。

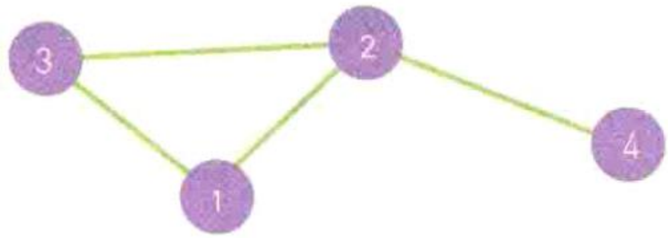
网络与图的细微差异：一般来说，我们对于一个现实社会中真实存在的网络结构称网络，而将数学中的抽象网络结构称为图，但是本质上并没有做细致的区分。图与网络在论文中也经常会当作同义词使用。

第三节 度、平均度和度分布

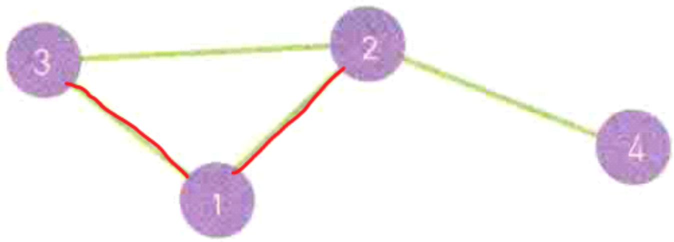
2022年12月12日 9:32

度是节点的关键属性，表示该节点和其他节点之间的链接数。

我们使用k_i网络中第i个节点的度。



我们可以看到的是$k_1 = 2$，因为节点1和其他节点有两个链接。
因此我们可以同样写出$k_2 = 3, k_3 = 2, k_4 = 1$



需要注意的是**节点的度和网络的链接数之间存在着关系**：我们可以看到在上图这样的无向图中，每一个节点相当于占有了一半的链接，例如：节点1和节点2,共用了一条链接，在没有权重比例的情况下，节点1和节点2平分这一条链接，因此我们可以总结，*在无权无向图中，节点度之和的一半就是链接数L。*

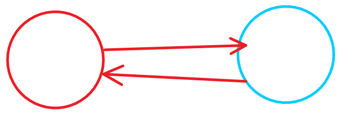
$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

平均度也是网络中的一个重要属性，在无向无权图中：

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

这是用来衡量网络中的每个节点平均占有多少的度。

有向网络中，度被分成出度和入度。在下图中我们能看到，红色节点的出度为1，入度也为1，蓝色节点出度入度也为1。



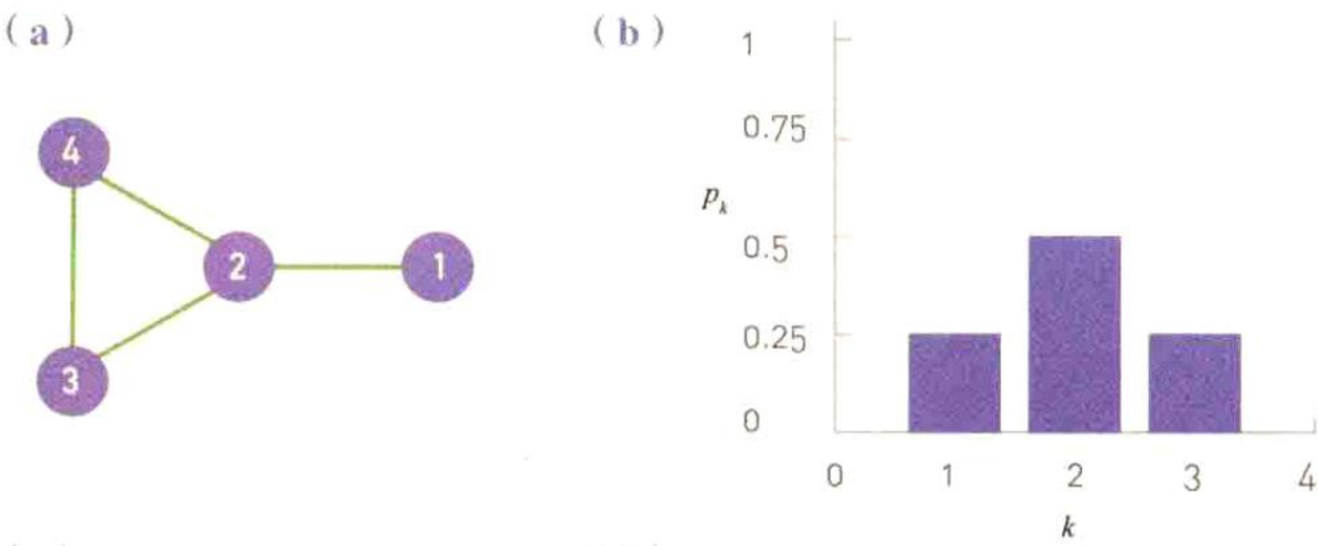
在这里，有向无权图的链接数与度数的关系为：

$$L=\sum_{i=1}^{N}k_i^{\text{in}}=\sum_{i=1}^{N}k_i^{\text{out}}$$

同样，平均度的表示为：

$$\langle k^{\text{in}}\rangle=\frac{1}{N}\sum_{i=1}^{N}k^{\text{in}}=\langle k^{\text{out}}\rangle=\frac{1}{N}\sum_{i=1}^{N}k^{\text{out}}=\frac{L}{N}$$

度分布用来表示网络中随机选择一个节点，其度为k的概率，用p_k表示：
我们可以看这样一个图例：

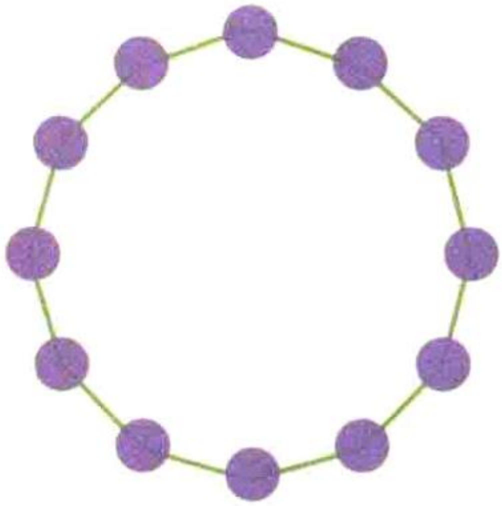


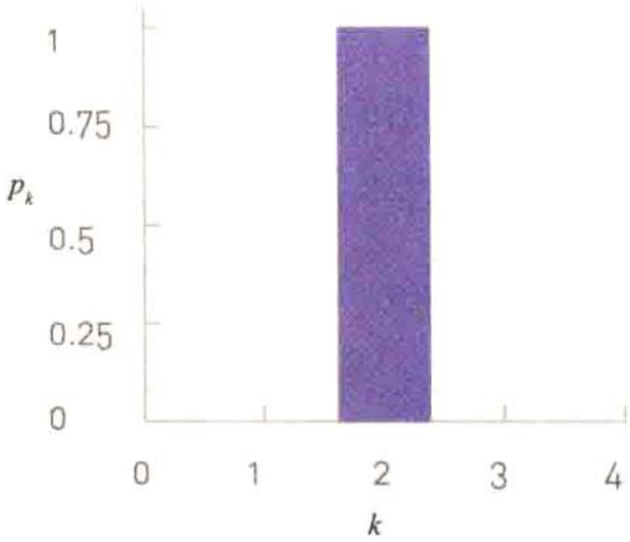
度分布的计算方式为 度为k的节点数/总的节点数：

$$p_k=\frac{N_k}{N}$$

度分布在网络中具有至关重要的作用，这主要是因为，**度分布可以找到不同网络存在的相似性，并且多数相同度分布的网络往往具备相同的结构或者相同的性质**，这是我们对网络研究想要了解的最重要内容。

而对于一些特殊的度分布网络，我们往往有特殊的名称来描述，比如：

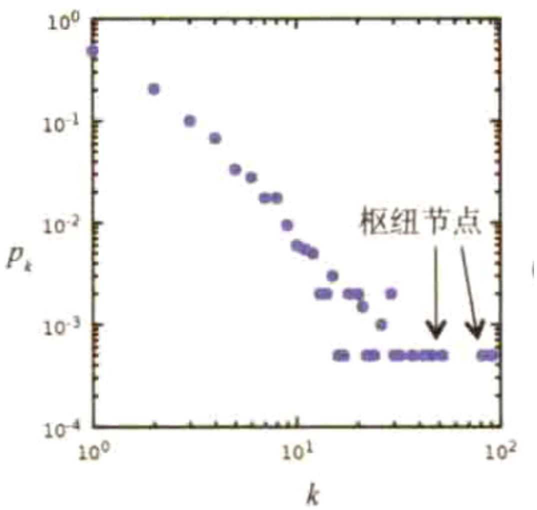




这样的网络，我们称为：***克罗内克的德尔塔函数度分布***

$$p_k = \delta(k-2)$$

同时，在真实网络中，网络中各个节点的度往往差异很大，有的可能就是单独的孤立节点，度为0，而有的则可能是枢纽节点，度非常大，而这个时候，我们**往往要通过度分布的方式去将各个节点分类画在直方图中**，从而能够让我们清楚看到整个网络大概有几个枢纽，几个孤立节点，从而能够更好的解决网络复杂性的问题。（如下图）



当然，**我们除了直接使用k和pk来衡量，我们还可以使用lnk和lnpk的双对数形式来衡量**。

第四节 邻接矩阵

2022年12月12日 10:16

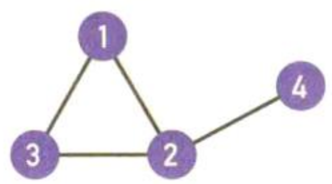
邻接矩阵可以作为网络的一种描述方式，一个有N个节点的网络，其应该存在一个N*N的邻接方阵

在邻接矩阵中，若节点i和节点j之间有链接，则在Aij位置写1,否则为0
我们的邻接矩阵的状态是这样的：

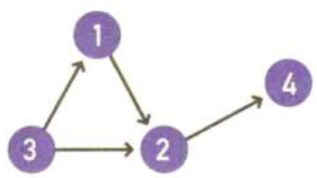
$$A_{ij} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix}$$

举一个具体的例子：

（b）无向网络



（c）有向网络



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad A_{ij} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

无向图的邻接矩阵是对称的，但是有向图就不一定。

第五节 真是网络的稀疏性

2022年12月26日 8:52

现实生活中的网络往往是稀疏的，例如我们举一个非方阵的例子，以用户和淘宝商品之间的关系来举例

如果你购买过某商品，那么在这一矩阵位置的值为1，否则为0，假设用户有800万，淘宝中的商品有100万中，
仔细思考，以个人的购买能力即使再强，也不会把这100万件商品都购买一遍，可能每个人也就浏览或者看过1000件商品，仅仅占1/1000的位置会被标记为1，其他都是0，那么这个关系或者邻接矩阵将是一个极大比例为0的矩阵，这是一个很稀疏的矩阵。

网络是巨大的，但是对于我们个人来说，我们远远用不完整个网络中的资源。

第六节 加权网络

2022年12月26日　8:58

现在我们主要讨论的都是没有加权的网络，但是现实生活中，很多网络是有权重的，比如社交关系网，你与同学A的关系和你与同学B的关系，以及你与父母的关系的密切程度或者重要程度是不一样的，所以，我们可以用不同的权重来表示不同的重要性。

对于加权网络而言，邻接矩阵的元素表示链接的权重，即

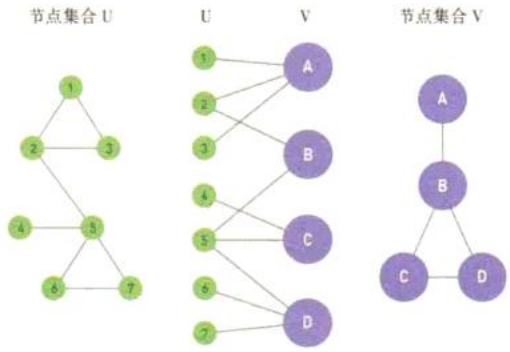
$$A_{ij} = w_{ij} \tag{1.13}$$

梅特卡夫定律：网络的价值与网络的节点个数的平方成正比（N^2）
但是在现实生活中，网络的实际价值往往不是平方级的变化，因为正如之前说的，网络是稀疏的，那么很多的资源是没有办法利用的。**一般情况下，网络的价值增加是随节点数N的线性变化。**
同时我们也要考虑网络权重对于网络价值的影响。

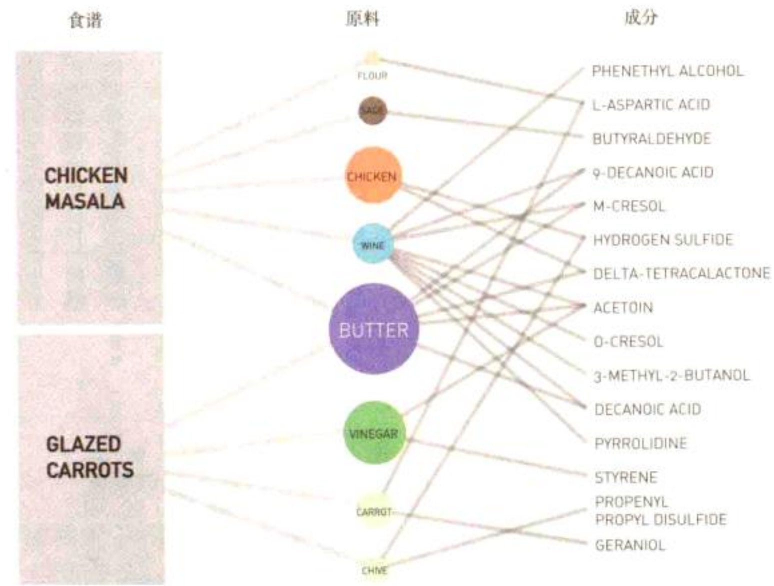
第七节 二分网络

2022年12月26日 9:08

二分网络：其节点可以分为两个节点集U和V。使得每一个链接中都有一个U中的节点和一个V中的节点。



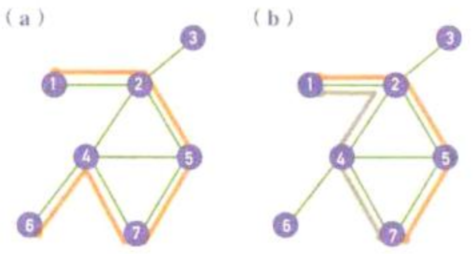
除了二分网络外，还可能存在三分网络或者多分网络。定义可以根据二分网络引申。



第八节 路径与距离

2022年12月26日　　9:12

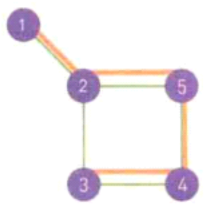
路径：表示为网络中链接行走的路线
路径的长度：路线中包含的链接数量



例如（a）中，橙色的路径为1->2->5->7->4->6，路径长度是5，因为包含了5段链接。

最短路径：从节点i到达节点j的包含连接数最少的路径
直径：最短路径的最大长度，也就是所有最短路径中最长的那一个路径的长度
欧拉路径：每条链接恰好经过一次的路径
汉密尔顿路径：每个节点恰好经过一次的路径

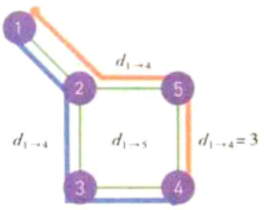
（a）



路径

一个节点序列，序列中每个节点有链接指向其后面紧邻的节点。每个路径包含 $n+1$ 个节点和 n 个链接。路径的长度是其包含的链接数，多次出现的链接重复计数。例如，橙色路径 $1\to2\to5\to4\to3$ 的路径长度为 4。

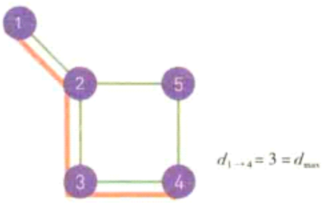
（b）



最短路径

两个节点间长度最短的路径。我们将两个节点之间最短路径的长度称为测地距离 d。注意，最短路径不一定是唯一的：节点 1 和节点 4 之间有两条最短路径，分别是 $1\to2\to3\to4$（蓝色）和 $1\to2\to5\to4$（橙色），它们的路径长度都为 $d_{1,4}=3$。

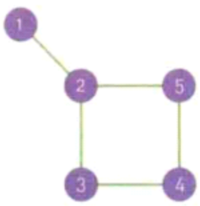
（c）



直径（d_{max}）

最短路径的最大长度，也是距离最远的节点对之间的距离。如图所示，节点 1 和节点 4 之间的距离最远，因此网络直径为 $d_{max}=3$。

（d）

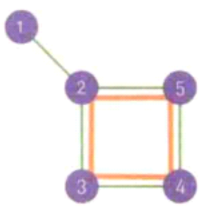


$\langle d\rangle=(d_{1\to2}+d_{1\to3}+d_{1\to4}+d_{1\to5}+$
$d_{2\to3}+d_{2\to4}+d_{2\to5}+$
$d_{3\to4}+d_{3\to5}+$
$d_{4\to5})/10=1.6$

平均路径长度（$\langle d\rangle$）

所有节点对之间最短路径的平均长度。如图所示，该网络的平均路径长度为 $\langle d\rangle=1.6$，计算过程列在了图的旁边。

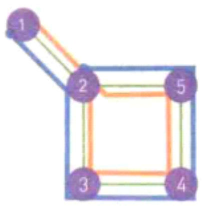
（e）



环

起始节点和终止节点相同的路径。如图所示，橙色路径是一个环。

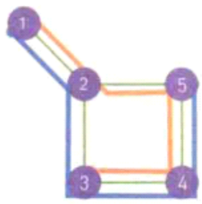
（f）



欧拉路径

经过每条链接恰好一次的路径。图中给出了两条欧拉路径，分别为橙色和蓝色的。

（g）



汉密尔顿路径

经过每个节点恰好一次的路径。图中给出了两条汉密尔顿路径，分别为橙色和蓝色的。

广度优先算法：

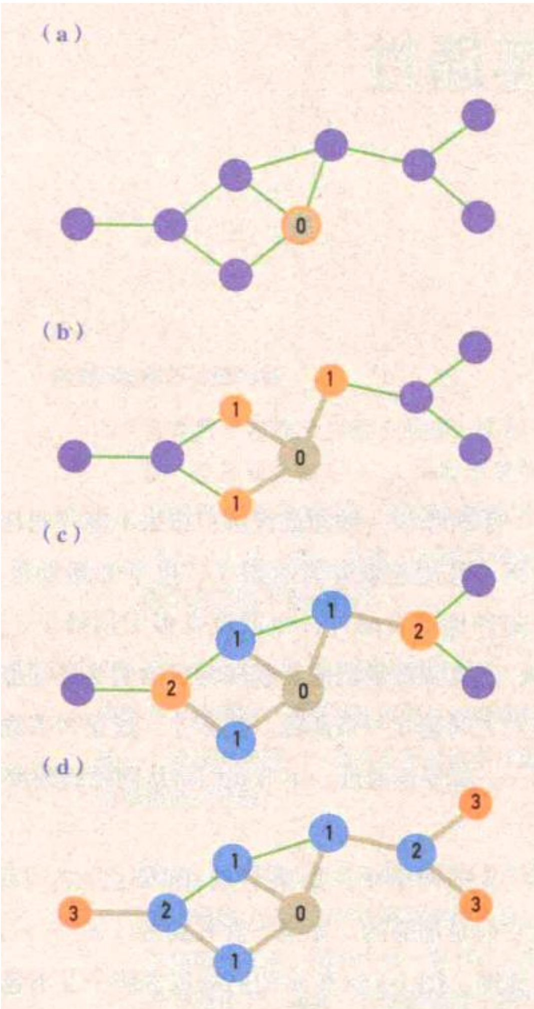
（1）从节点 i 出发，将其标记为 0；

（2）找出与节点 i 直接相连的节点，标记为距离 1，然后将它们放到一个队列中；

（3）从队列中取出最前面的节点（其距离标记假设为 n），找出和该节点相连且尚未被标记的节点，将它们标记为 $n+1$，然后把它们放在队列的后面；

（4）重复步骤 3，直到碰到目标节点 j 或者队列中没有节点了；

（5）节点 i 和节点 j 之间的距离为节点 j 的标记。如果节点 j 没有标记，则 $d_{ij}=\infty$



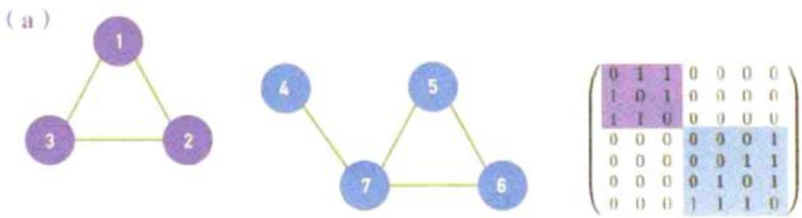
复杂度是(N+L),N为节点数，L为链接数。

第九节 连通性

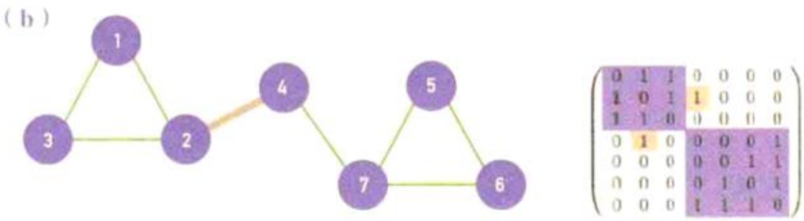
2022年12月26日 9:30

在无向图中，如果节点i与节点j之间存在路径，则我们称之为连通，若不存在路径则不连通。

对于整个网络而言，如果网络中所有节点都是连通的，则称这个网络是连通的，否则有至少一个不连通的节点，则网络不是连通的。对于不连通的网络来说，将网络的邻接矩阵重新排序后，我们可以使用线性代数中的分块对角矩阵来判别。



而想让两个不连通的网络重新连通，我们需要找到关键节点，将两个节点架桥就可以实现。



寻找连通分支：

寻找网络的连通分支

1. 从随机选择的一个节点 i 出发，执行边栏 1.5 描述的 BFS 算法。将从节点 i 出发可到达的所有节点标记为 $n=1$。

2. 如果已标记节点的个数等于 N，则网络是连通的。如果已标记节点的个数小于 N，则网络包含多个连通分支。找出这些连通分支需要继续执行步骤 3。

3. 标号增加 1，即 $n \to n+1$。选择一个未标记的节点 j，将其标记为 n。使用 BFS 算法找出从节点 j 出发可到达的所有节点，将它们都标记为 n。重复步骤 3，直到标记完所有的节点为止。

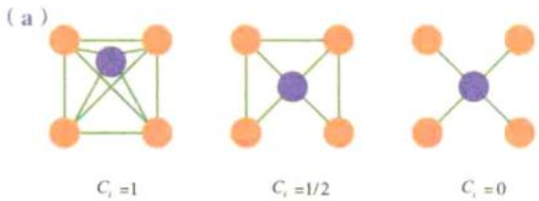
第十节 集聚系数

2022年12月26日 9:38

局部集聚系数：刻画了相邻节点之间连接的稠密程度
对于一个度为K的节点i的集聚系数计算为：

$$C_i = \frac{2L_i}{k_i(k_i-1)}$$

Li为与该节点连接的节点之间的连接数

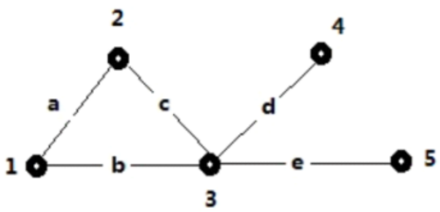


例如第二个，与紫球连接的是四个黄球，四个黄球之间的链接数是3，则L=3
集聚系数是一个0-1的值，若集聚系数为0，则表明该节点不连通，若为1，则表明这个节点与其余所有节点都连通。

全局集聚系数：主要用于计算网络中闭合三角形的数量

$$C_\Delta = \frac{3 \times 三角形个数}{连通三元组的个数}$$

例如下图{1，(2,3)}构成的triplet是封闭的，{3，(4,5）}构成的triplet是开放的



全局的Clustering coefficient比较简单，公式如下：Clustering coefficient(global) = number of closed triplet / number of triplet(closed+open)
以上图为例:
closed triplet ={1，(2,3)}，{2，(1,3)}，{3，(1,2)}
all triplet = {1，(2,3)}，{2，(1,3)}，{3，(1,2)}，{3， (2,4）}，{3， (4,5）}，{3， (1,5）}，{3， (2,5）}，{3， (1,4）}
number of closed triplet = 3
number of triplet = 8
number of triplet / number of triplet = 3/8=**0.375**

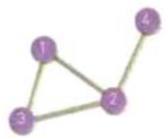
平均局部集聚系数：所有局部集聚系数的平均值，一般用来衡量网络的平均集聚情况。

第一章概念补充和总结

2022年12月26日　　9:47

(a)

无向



$$A_{ij}=\begin{pmatrix}0&1&1&0\\1&0&1&1\\1&1&0&0\\0&1&0&0\end{pmatrix}$$

$A_{ii}=0$　　$A_{ij}=A_{ji}$

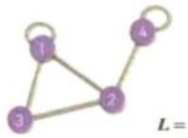
$L=\frac{1}{2}\sum_{i,j=1}^{N}A_{ij}$　　$<k>=\frac{2L}{N}$

无向网络

无向网络中链接没有方向。例如，互联网、电网、科学合作网络。

(b)

自环



$$A_{ij}=\begin{pmatrix}1&1&1&0\\1&0&1&1\\1&1&0&0\\0&1&0&1\end{pmatrix}$$

$\exists i, A_{ii}\neq 0$　　$A_{ij}=A_{ji}$

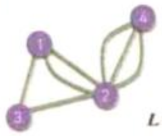
$L=\frac{1}{2}\sum_{i,j=1,i\neq j}^{N}A_{ij}+\sum_{i=1}^{N}A_{ii}$　　?

自环

在许多网络中，节点不和自己进行交互，因此其邻接矩阵的对角线元素是 0，即 $A_{ii}=0, i=1,\cdots,N$。在有些系统中，自我交互是允许的，在这样的网络中，自环表示节点的自我交互。例如，万维网、蛋白质相互作用网络。

(c)

多重
无向



$$A_{ij}=\begin{pmatrix}0&2&1&0\\2&0&1&3\\1&1&0&0\\0&3&0&0\end{pmatrix}$$

$A_{ii}=0$　　$A_{ij}=A_{ji}$

$L=\frac{1}{2}\sum_{i,j=1}^{N}A_{ij}$　　$<k>=\frac{2L}{N}$

多重网络 / 简单网络

在多重网络中，两个节点之间可以有多个平行的链接。因此，A_{ij} 可以是任意正整数。不允许多重链接存在的网络被称为简单网络。多重网络的一个例子是：区分朋友关系、家庭关系和同事关系的社会网络。

(d)

有向

$$A_{ij}=\begin{pmatrix}0&1&0&0\\0&0&1&1\\1&0&0&0\\0&0&0&0\end{pmatrix}$$

$A_{ij}\neq A_{ji}$

$L=\sum_{i,j=1}^{N}A_{ij}$　　$<k>=\frac{L}{N}$

有向网络

有向网络中，链接具有方向。例如，万维网、手机通话网络、引文网络。

(e)

加权
无向

$$A_{ij}=\begin{pmatrix}0&2&0.5&0\\2&0&1&4\\0.5&1&0&0\\0&4&0&0\end{pmatrix}$$

$A_{ii}=0$　　$A_{ij}=A_{ji}$

$<k>=\frac{2L}{N}$

加权网络

加权网络中，链接具有权重、强度或流量参数。对于权重为 w_{ij} 的链接，其邻接矩阵中对应的元素为 $A_{ij}=w_{ij}$。对于无权网络，邻接矩阵的元素仅表示链接是否存在（$A_{ij}=1$ 或 $A_{ij}=0$）。加权网络的例子有：手机通话网络、电子邮件网络。

(f)

完全图
无向

$$A_{ij}=\begin{pmatrix}0&1&1&1\\1&0&1&1\\1&1&0&1\\1&1&1&0\end{pmatrix}$$

$A_{ii}=0$　　$A_{i\neq j}=1$

$L=L_{max}=\frac{N(N-1)}{2}$　　$<k>=N-1$

完全图（团）

完全图或团中，任意两个节点彼此相连。例子包括出演同一部电影的演员网络，其是每两个演员都彼此相连。