《网络科学导论》第二章

作　者：汪小帆、李翔、陈关荣
报告人：李进之

Contents

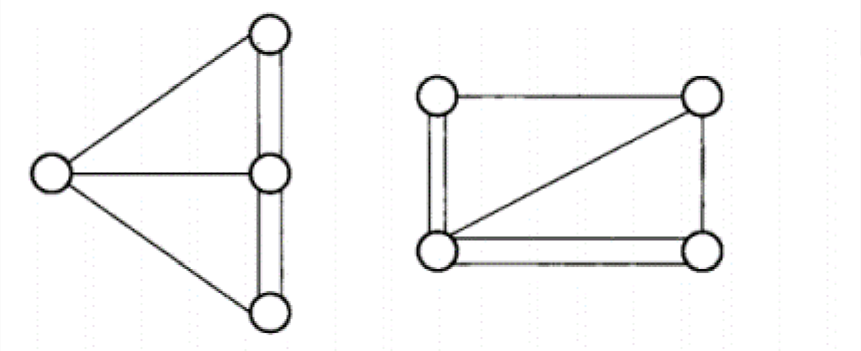
PART 1

图模型导引

1.1 图模型

- **模型**：简化、保留本质——拓扑性质
- **拓扑性质**：**连续变形**下保持不变的性质。

  连续变形包括伸缩和扭曲，但不允许割断和黏合。



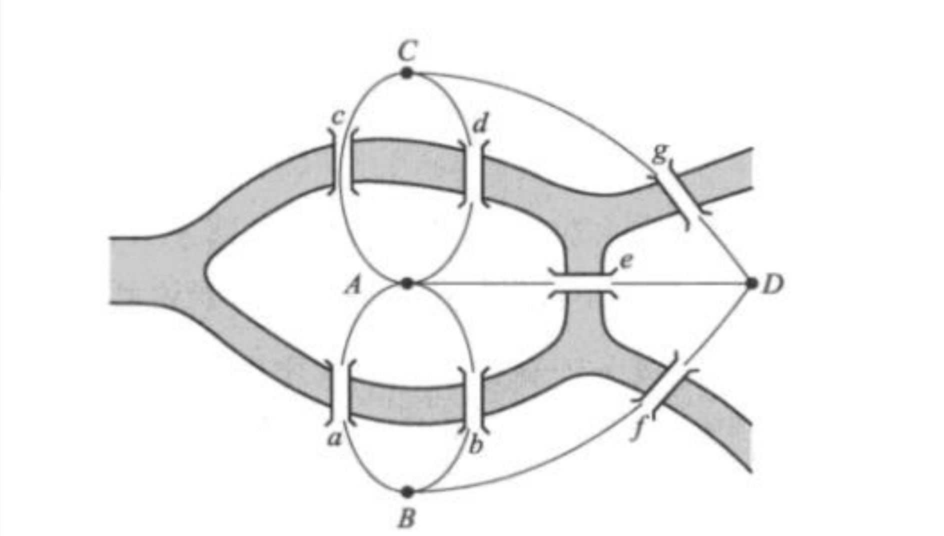
- **图模型**：有哪些节点，节点之间是否有边连接。

1.1 图模型

- **图的定义**：
  - 顶点集合V，边集合E
  - 图G=（V，E）
  - N=|V|，M=|E|
- **限定讨论范围**：
  - 无重边：从顶点i到j只允许至多一条边
    - 特别问题可以特别处理：“**七桥问题**”
  - 无自环：不允许存在顶点i到自己的边

1.1 图模型

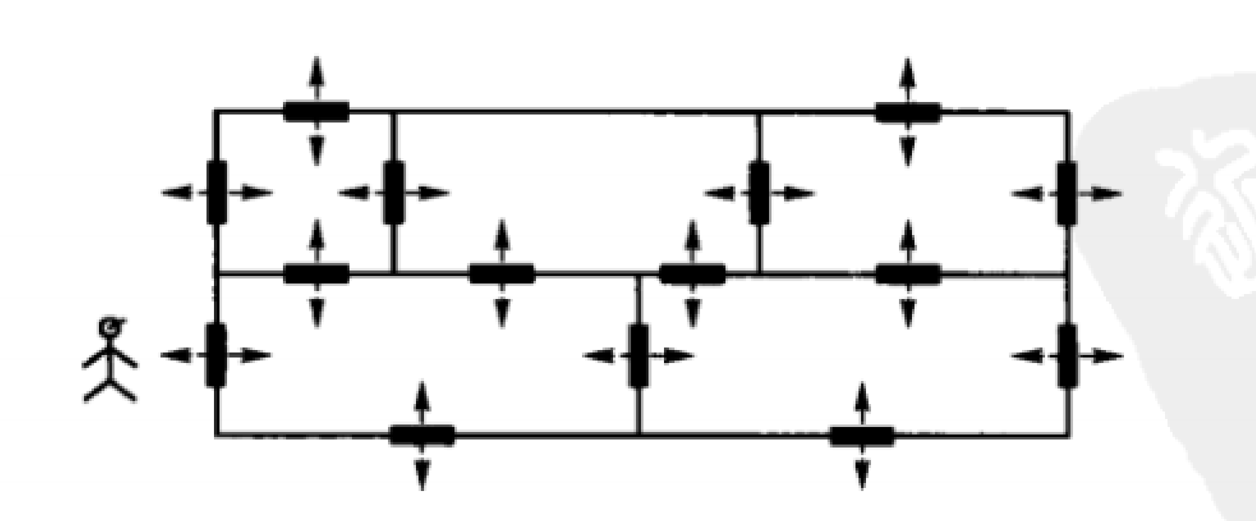
- 七桥问题:



  - 穷举...?

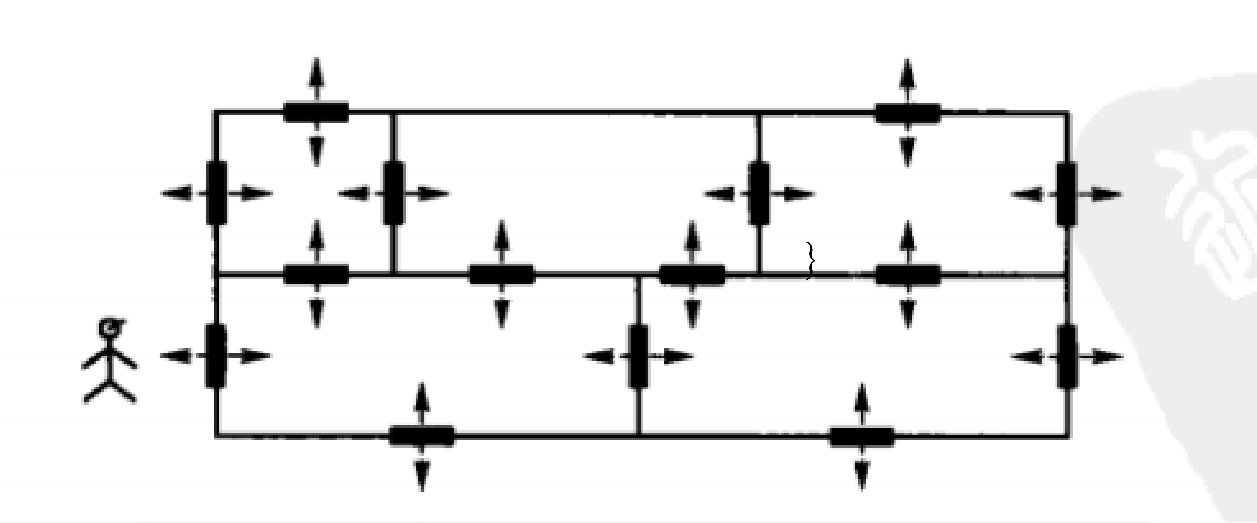
1. 1　图模型

- 案例1：七桥问题:



1.1 图模型

- **案例1：七桥问题:**



  - 只有起点和终点可以有奇数条边
  - 一条路径只能有至多一个起点、一个终点
  - **因此至多2个奇数条边的点，否则无法找到这样一条路径**

1.2　图的类型

- **按顶点分类**:
  - 顶点“全同” 与“非全同”
    - 异顶点：**二分图**
- **按边分类**:
  - 有权与无权
  - 有向与无向
- **按边与顶点相对数目**:
  - 稠密与稀疏
    - 稠密：完全图
    - 稀疏：空图、**树**

1.2　图的类型

- 图类型之间的转化:

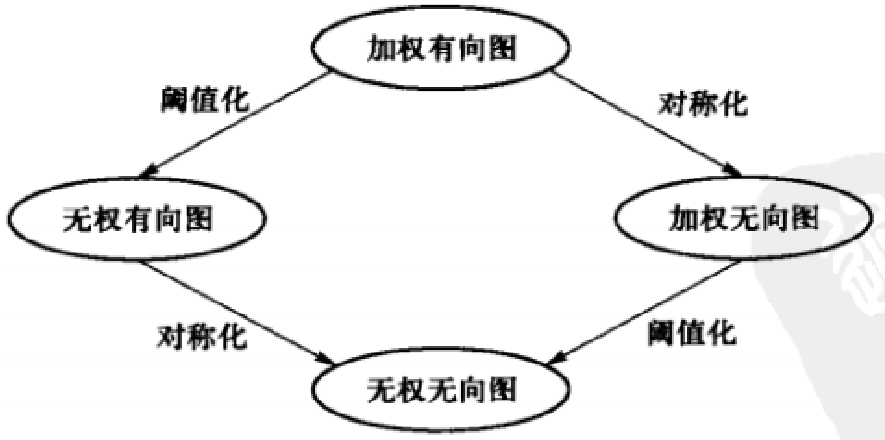


图 2-2　四种类型的图之间的关系

- 有些时候却也不那么容易…
  - 如:**引文网络**

1.3 图的表示

- 邻接矩阵：

$$a_{ij}=\begin{cases} w_{ij} & i\text{有指向}j\text{权值为}w_{ij}\text{的边} \\ 0 & i\text{无指向}j\text{的边} \end{cases}$$

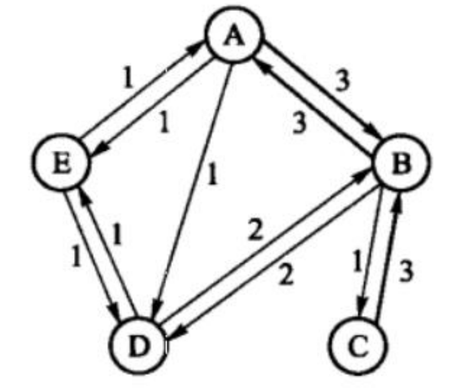


图 2-3 加权有向的朋友关系图

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

- 邻接链表：

1	2	4	5
2	1	3	4
3	2		
4	2	5	
5	1	4	

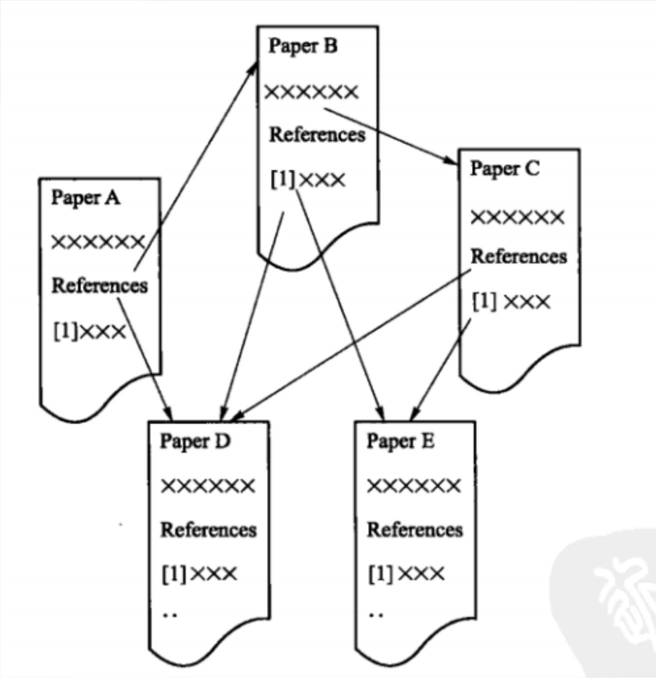
- 三元组：

1	2	3
1	4	1
1	5	1
2	1	3
2	3	1
2	4	2
3	2	3

4	2	2
4	5	1
5	1	1
5	4	1

1.4　引文网络：

- **案例2：引文网络**:
  - 天然就具有方向性



$$A=\begin{Bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{Bmatrix}$$

  - 文献i与j的有多相关?

1.4　引文网络：

- 案例2：引文网络:
  - 共引网络

$a_{ki} \cdot a_{kj} = 1: i$和$j$共同被$k$引用

$$c_{ij} = \begin{cases} \sum_{k=1}^{N} a_{ki} \cdot a_{kj} & i \neq j \quad \text{共引文献数} \\ \sum_{k=1}^{N} a_{ki}^2 & i = j \quad \text{引用文献数} \end{cases}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$C = A^T A$$

1.4 引文网络：

- 案例2：引文网络:
  - 文献耦合网络

$a_{ik} \cdot a_{jk} = 1$: i和j同时引用k

$$b_{ij} = \begin{cases} \sum_{k=1}^{N} a_{ik} \cdot a_{jk} & i \neq j \quad \text{共被引文献数} \\ \sum_{k=1}^{N} a_{ik}^2 & i = j \quad \text{被引用文献数} \end{cases}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B = AA^T$$

PART 2

路径与连通性

2.1 路径

- **路径：**顶点序列 $P = v_1v_2...v_k$，其中每一对相邻顶点有 $v_i \rightarrow v_{i+1}$ 边。
- **简单路径：**顶点序列中各个顶点互不相同的路径。
- **圈（circle）：**
  - P中k>2
  - P中前$k-1$个顶点互不相同，而 $v_{k-1} = v_k$
- **回路（circuit）：**
  - P中$v_{k-1} = v_k$

2.2　连通性

- 无向图：每一对顶点之间至少存在一条路径
- 有向图：
  - 强连通性：每一对顶点(i,j)中既有i到j的路径，也有j到i的路径。
  - 弱连通性：每一对顶点(i,j)中，i到j的路径或j到i的路径至少有一条存在

2.2 连通性

- 距离：
  - 从i到j的最短路径的长度
- 连通片：
  - 不连通图的连通子图
  - 且任意不属于该子图的顶点与该子图中的任意顶点都没有路径连通
- 最大连通片：
  - 包含顶点数最多的连通片

2.3 连通性判据

- 如何判断i到j是否有距离为2的路径?

$$a_{ik} \cdot a_{kj} = 1: \text{存在} i - k - j \text{的路径}$$

$$N_{ij}^{(2)} = AA = A^2$$

$$N_{ij}^{(k)} = A^k$$

- 连通性判据一:
  - 一个网络是连通的，当且仅当:

$$(I + A + A^2 + \ldots + A^{N-1})_{ij} > 0, \forall i, j$$

2.3　连通性判据

- 不连通网络的邻接矩阵有何特点?

$$\boldsymbol{A}=\begin{bmatrix} \square & 0 & \cdots \\ 0 & \square & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

- 称方阵$\boldsymbol{A}$可约：
  - 如果$\boldsymbol{A}$存在某种行和列的置换：$(1,2,...,N)\Rightarrow(i_1,i_2,...i_N)$，使得置换后的矩阵$\bar{\boldsymbol{A}}$满足：

    $\bar{a}_{ij}=0,\ \text{for any } i=\mu+1,\mu+2,...,N,\ \text{and any } j=1,2,...,\mu$。$\mu$ 是某个小于$\boldsymbol{N}$的正整数。
  - 即：$$\bar{A}=\begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{bmatrix}$$

2.3 连通性判据

- 方阵A可约数学定义：
  - 若存在顺列矩阵（每行仅一个元素为1，其他元素都为0的正交矩阵）U使得：

$$U^T A U = \overline{A} = \begin{bmatrix} \overline{A}_{11} & \overline{A}_{12} \\ 0 & \overline{A}_{22} \end{bmatrix}$$

称方阵A可约。

$$eg: \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- 连通性判据二：
  - 一个网络是连通的，当且仅当其邻接矩阵是不可约的。

2.4 连通性的鲁棒性

- 若想使得顶点s和t分属不同连通片，至少需要从图G中去除多少个顶点（多少条边）？

- *Menger*定理：

定理 2-1（Menger 定理） 1）点形式：设顶点 s 和顶点 t 为图 G 中两个不相邻的顶点，则使顶点 s 和顶点 t 分别属于不同的连通片所需去除的顶点的最少数目等于连接顶点 s 和顶点 t 的独立的简单路径的最大数目。

2）边形式：设顶点 s 和顶点 t 为图 G 中两个不同的顶点，则使顶点 s 和顶点 t 分别属于不同的连通片所需去除的边的最少数目等于连接顶点 s 和顶点 t 的不相交的简单路径的最大数目。

2.4 连通性的鲁棒性

- $Menger$定理：

定理 2-1(Menger 定理) 1) 点形式：设顶点 s 和顶点 t 为图 G 中两个不相邻的顶点，则使顶点 s 和顶点 t 分别属于不同的连通片所需去除的顶点的最少数目等于连接顶点 s 和顶点 t 的独立的简单路径的最大数目。

2) 边形式：设顶点 s 和顶点 t 为图 G 中两个不同的顶点，则使顶点 s 和顶点 t 分别属于不同的连通片所需去除的边的最少数目等于连接顶点 s 和顶点 t 的不相交的简单路径的最大数目。

- 不相邻的顶点：没有直接相连的边
- 独立的路径：不共享顶点
- 不相交的路径：可以共享顶点，但是不共享边

2.4 连通性的鲁棒性

- $Menger$定理：

定理 2-1（Menger 定理） 1）点形式：设顶点 s 和顶点 t 为图 G 中两个不相邻的顶点，则使顶点 s 和顶点 t 分别属于不同的连通片所需去除的顶点的最少数目等于连接顶点 s 和顶点 t 的独立的简单路径的最大数目。

2）边形式：设顶点 s 和顶点 t 为图 G 中两个不同的顶点，则使顶点 s 和顶点 t 分别属于不同的连通片所需去除的边的最少数目等于连接顶点 s 和顶点 t 的不相交的简单路径的最大数目。

- 点割集：为了使得顶点s和t分属不同连通片，所需要去除点的集合
- 边割集：为了使得顶点s和t分属不同连通片，所需要去除边的集合

PART 3

树与生成树

3.1　树的定义

- 树的定义

一般地，一个包含 N 个顶点的无向图 G 称为一棵树，如果它满足如下任意一个条件：

① 图 G 是连通的并且有 $N-1$ 条边；

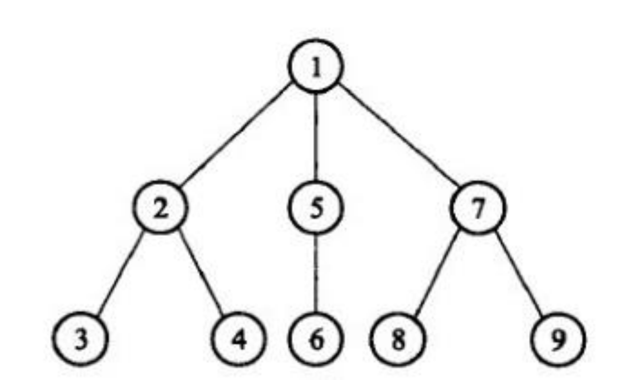
② 图 G 是连通的并且不包含圈；

③ 图 G 不包含圈并且有 $N-1$ 条边；

④ 图 G 中任意两个顶点之间有且只有一条路径；

⑤ 图 G 中任意一条边都是桥，即去掉图 G 中任意一条边都会使图变得不连通。

- 根树：树的层次表示



3.2 广度优先搜索

- 搜索方式：

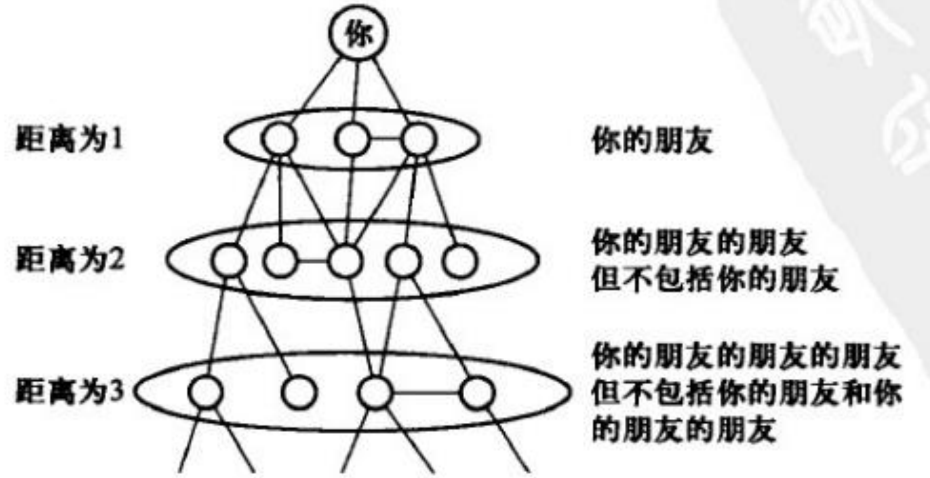


图 2-21 朋友关系层次示意图

- 例：

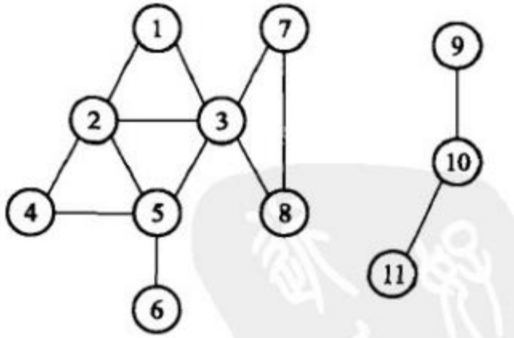
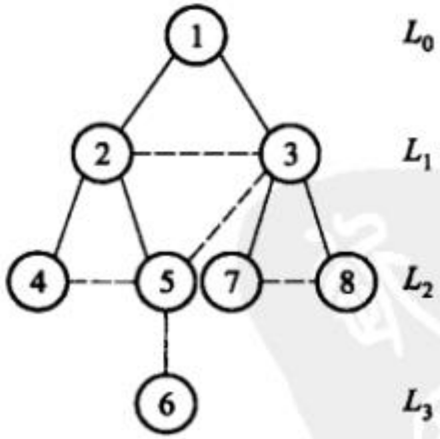


图 2-17 包含两个连通片的
一个不连通图

⇒



(c)

3.3 生成树

- 生成树定义：

  图$\boldsymbol{G}(\boldsymbol{V},\boldsymbol{E})$的子图，包含$\boldsymbol{G}$中所有顶点，且是一棵树。

- 生成树数目巨大：

  完全图$\boldsymbol{K_N}$的生成树总数：（$\boldsymbol{Caylay}$公式）

$$\tau(K_N) = N^{N-2}$$

3.4 最小生成树

- 有权无向图：生成树的边权值之和不同。

- 最小生成树（MST）：

  当$G(V,E,W)$中所有权值都不相等时，必有唯一的最小生成树，否则不一定唯一。

- 定理：

定理 2-2 连通图 $G=(V,E,W)$ 的一个权值之和最小的连通的子图 $\bar{G}=(V,\bar{E},\bar{W})$ 一定是图 G 的一个最小生成树。

3.4 最小生成树

- 最小生成树的性质：

定理 2-3 给定连通图 $G=(V,E,W)$ 并假设所有边的权值互不相同，则有

1）最小生成树的割性质：令 S 是图 G 的任意顶点子集，$S\neq\varnothing$（空集），$S\neq V$。令边 e 是一个端点在 S 中，另一端点在 $V-S$ 中的最小权值边。那么图 G 的最小生成树必然包含边 e。

2）最小生成树的圈性质：令 C 是图 G 中的一个圈，并假设边 e 是圈 C 中权值最大的一条边，那么图 G 的最小生成树必然不包含边 e。

3.4 最小生成树

- $Prim$算法：$O(M+NlogN)$

Prim 算法：初始时，$U=\{u_0\}$为图 G 中任一顶点，$TE=\varnothing$为空集。在所有 $u\in U, v\in V-U$ 的边中，选择一条权值最小的边(u,v)加入边集 TE，同时将顶点 v 加入点集 U。重复这一操作，直到 $U=V$，即包含图 G 中的全部顶点为止。

- $Kruskal$算法： $O(MlogN)$

Kruskal 算法：初始时，$U=V, TE=\varnothing$。将图 G 中的边按权值从小到大的顺序依次选取，若选取的边使生成树不形成圈，则把它加入 TE 中；若选取的边使生成树形成圈，则将其舍弃，如此进行下去，直到 TE 中包含 $N-1$ 条边为止。

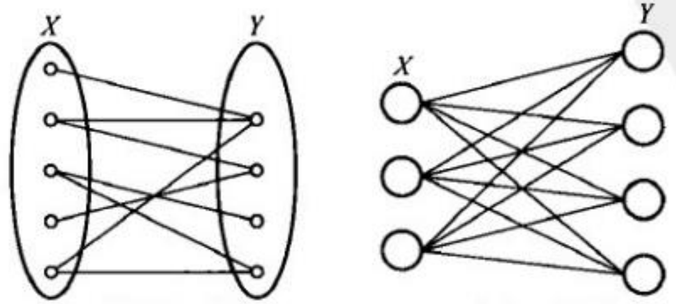
- 应用：
  - 稠密网络：$Prim$算法
  - 稀疏网络：$Kruskal$算法

PART 4

二分图与匹配问题

4.1 二分图

- 二分图定义：
  - 如果顶点集V可分为互不相交的非空子集X和Y，并且边集E中每条边(i,j)中顶点i和j分属于不同子集，则称图G为一个二分图。
  - 当二分图子集X任一点i和子集Y中任一点j之间都有边(i,j)相连，则称其为完全二分图。



(a) 非完全二分图　　(b) 完全二分图

图 2-27　二分图的例子

4.2 二分图到单分图的投影

- 二分图对集合X投影得到单分图：
  - 包含X中所有顶点
  - 在原来的二分图中，如果集合X中顶点i和j都与集合Y中顶点k有边相连，则单分图中，存在边(i,j)。

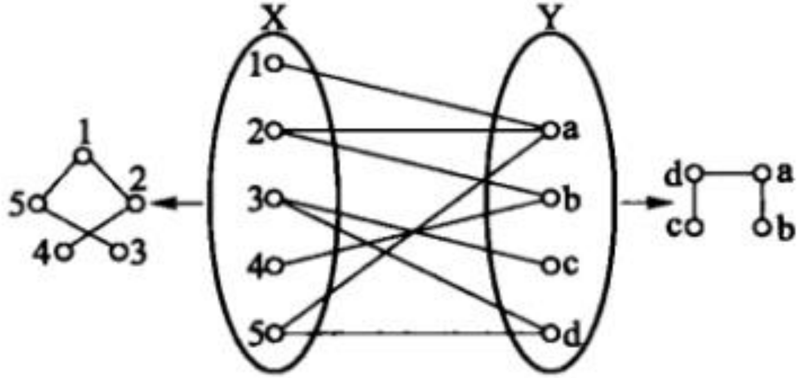


图 2-30　一个二分图可以投影为两个单分图

4.3　二分图匹配问题

- 匹配：
  - $G=(X,E,Y)$为二分图，F为边集E的一个子集，若F中任意两条边都没有公共端点，则称F为G的一个匹配。
- 最大匹配：G所有中边数最多的匹配
- X完全匹配 / Y完全匹配 / 完全匹配：

  集合X（或/和集合Y）中任一顶点均为匹配F中的端点

4.4 稳定匹配

- 假设X、Y中每个顶点都有自己的匹配优先序列：
  - 匹配稳定条件：
    - X中任何一个顶点想要更换更优先的匹配对象，Y中没有顶点愿意接受
    - Y中任何一个顶点想要更换更优先的匹配对象，X中没有顶点愿意接受
- 稳定匹配：
  - 满足稳定匹配条件的完全匹配

4.4 稳定匹配

- 稳定匹配的求解：$Gale-Shapley$算法（以研究生选择导师为例）

  1.随机选择一位还没有导师的学生，让该学生按照自己的意愿表依次和最心仪的导师沟通；

  2.每位导师在与学生沟通时，如果尚未带学生，则直接接受；

  若已经有沟通好的学生，则将两者进行比较，选择更为心仪的学生。若原沟通好的学生未能胜出，则他重新变为没有导师的状态。

  3.重复1、2步骤，直到所有学生都有导师。

- 定理2-4：$G-S$算法在至多N^2次迭代后终止，且算法终止时所得到的集合是一个完全匹配。
- 定理2-5：$G-S$算法终止时所得到的集合是一个稳定匹配。

4.4 稳定匹配

- 稳定匹配的公平性：
  - 问题1：$G-S$算法中是否越早选上的学生会更有利?
  - 问题2：$G-S$算法是否整体对于学生不利?
- 基本事实：
  - 事实1：教师从第一次有学生去找他，就会一直有学生，且他的学生只会越变越好。
  - 事实2：学生可能会在有导师和没导师的状态之间交换，且他的导师只会越来越差。

4.4 稳定匹配

- 稳定匹配的公平性：
  - 问题1：$G-S$算法中是否越早选上的学生会更有利?

    回答：$G-S$算法不论按照什么顺序选择学生，每次获得的结果都是一样的。

  - 问题2：$G-S$算法是否整体对于学生不利?

    回答：（定理2-6）$G-S$算法执行获得的结果总是对学生最理想、对导师最不理想的稳定匹配。

4.5 完全匹配存在条件

- 现实情况下导师与学生可能都会自带一些限制：
  - 比如某位导师只想选择某类背景学生；某名学生只想选择某个方向导师。

    完全匹配还存在吗?

- 抑制集：

  X（或Y）中一个顶点集S，将S中所有顶点在Y（或X）中的邻居顶点组成的集合记为$N(S)$，若S中顶点数目严格大于$N(S)$中顶点数目，则称S是抑制的。

- 定理2-7（匹配定理）：

  一个左右节点数相同的二分图存在完全匹配的充要条件是它不包含抑制集。

THANKS!