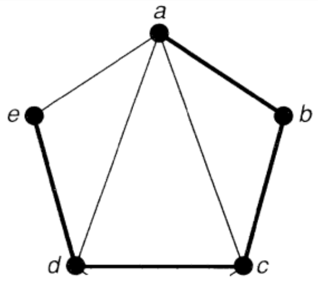
本次课程提纲：生成树

- 生成树的概念与性质
- 生成树的计数
- 最小生成树
- 根树

生成树的概念

- 若图 G 的一个生成子图 T 是树，称它为 G 的一棵生成树
  - 若 T 为森林，称它为 G 的一个生成森林
- 生成树的边称为树枝，G 中非生成树的边称为弦



- 在连通边赋权图 G 中总权值最小的生成树，称为最小生成树

最短路与生成树

- s 到所有节点的最短路构成一棵树
- 是生成树
- 不一定是最小生成树

生成树的性质

存在性

每个连通图 G 至少包含一棵生成树

证明

- 如 G 是树，则其本身是一棵生成树
- 若 G 有圈 C，则去掉 C 中一条边后得到的图仍然是连通的，这样不断去掉圈，最后得到一个 G 的无圈连通子图，它为 G 的一棵生成树
- 也可以用归纳法

- 证明实际上给出了生成树的求法，称为破圈法
- 生成树一般不唯一

推论

若 G 是 (n, m) 连通图，则 $m \geq n - 1$

生成树的计数：Cayley 递推

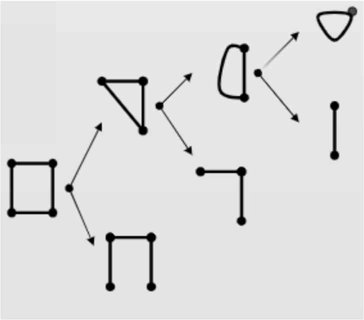
图 G 中，删掉边 e 后，把 e 的两个端点重合，得到的图记为 $G.e$

生成树个数递推定理

记图 G 中生成树的个数为 $\tau(G)$，有 $\tau(G) = \tau(G - e) + \tau(G.e)$

证明

G 的生成树中包含 e 的棵数为 $\tau(G.e)$，不包含 e 的棵数为 $\tau(G - e)$



Cayley

- Cayley (1821—1895): 剑桥大学数学教授，著名代数学家
- 发表论文数仅次于 Erdos, Euler, Cauchy
- 著名成果是 1854 年定义了抽象群，并且得到著名定理：
  - 任意一个群都和一个变换群同构
- 同时，是一名出色的律师，作律师 14 年期间，发表 200 多篇数学论文，
- 生成树递推计数公式是他在 1889 年建立的

生成树的计数：矩阵树定理

矩阵树定理

设 G 是顶点集合为 $V(G) = \{v_1, \cdots, v_n\}$ 的图，设 $A = (a_{ij})$ 是 G 的邻接矩阵，$C = (c_{ij})$ 是 n 阶方阵，其中：

$$c_{ij} = \begin{cases} d(v_i) & i = j \\ -a_{ij} & i \neq j \end{cases} \tag{1}$$

则 G 的生成树棵数为 C 的任意一个余子式的值

- 矩阵树定理的证明很复杂，在此略去证明
- 定理中的矩阵 C 又称为图的 Laplace 矩阵，定义为 $C = D(G) - A(G)$
- 其中：$D(G)$ 为度对角矩阵，即主对角元为对应顶点度数，其余元素为 0；$A(G)$ 是邻接矩阵

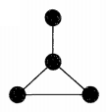
Kirchhoff

- 该定理是由物理学家 Kirchhoff 提出的
- Kirchhoff：1824 年出生于普鲁士的哥尼斯堡
- 1845 年因宣布著名的 Kirchhoff 电流电压定律而闻名，
- 1847 年大学毕业时发表了生成树计数文章，给出了矩阵树定理
- 一生主要花在实验物理上，担任过德国柏林数学物理会主席

生成树的计数：矩阵树定理

习题

利用矩阵树定理求下图生成树的棵数



解答

图的 Laplace 矩阵为

$$C = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

一行一列对应的余子式为

$$(-1)^{1+1} \begin{vmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 1 \end{vmatrix} = 3$$

生成树的计数

习题

证明 $\tau(K_n) = n^{n-2}$

解答

K_n 的 Laplace 矩阵为

$$C = \begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & \cdots & n-1 \end{pmatrix}$$

一行一列对应的余子式为

$$(-1)^{1+1} \begin{vmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & \cdots & n-1 \end{vmatrix}$$

可以求出上式等于 n^{n-2}

生成树的计数

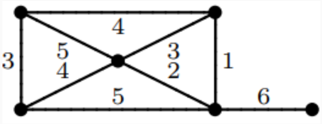
- 上题证明有好几种不同方法。用矩阵树定理证明是最简单的方法
- 1967 年，加拿大的 Moon 用了 10 种不同方法证明，之后有人给出了更多证明方法
- Moon 的学术生涯主要是对树和有向图问题进行研究
- 正如大多数科学家一样，他对音乐也很感兴趣
- 他还认为：当一个人发现了新事物，而且很难对非数学工作者解释该发现时，会产生一种满足喜悦感

最小生成树 Kruskal 算法

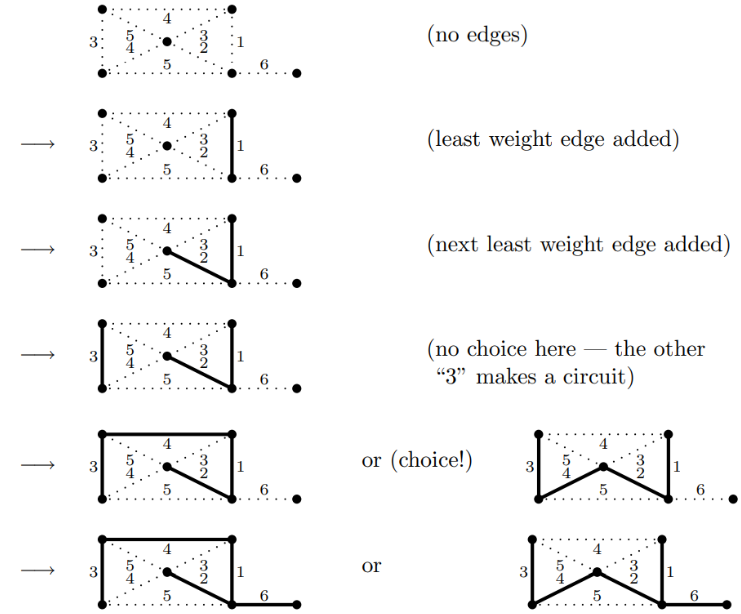
- 算法思想：从最小边开始，进行避圈式扩张
- 选择权值最小边 e_1
- 若已经选定边 $e_1, e_2, \cdots, e_k$, 则从 $E-\{e_1, e_2, \cdots, e_k\}$ 中选择最小边 e_{k+1}, 使 $e_1, e_2, \cdots, e_k$ 无圈
- 不能增加边时，停止

习题

用 Kruskal 算法求下图的最小生成树



最小生成树 Kruskal 算法



Kruskal

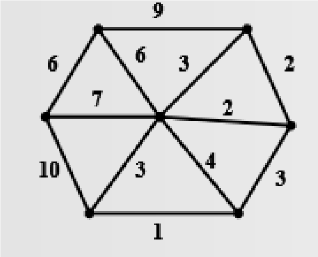
- 1928 年生，一家 3 弟兄都是数学家，
- 1954 年在普林斯顿大学获博士学位，导师是 Erdos
- 他大部分研究工作是数学和语言学，主要在贝尔实验室工作
- 1956 年发表包含 Kruskal 算法论文，名声大振

最小生成树管梅谷破圈法

- 从任意圈开始，去掉圈中权值最大的一条边
- 不断破圈，直到 G 中没有圈为止

习题

用破圈法求下图最小生成树



管梅谷

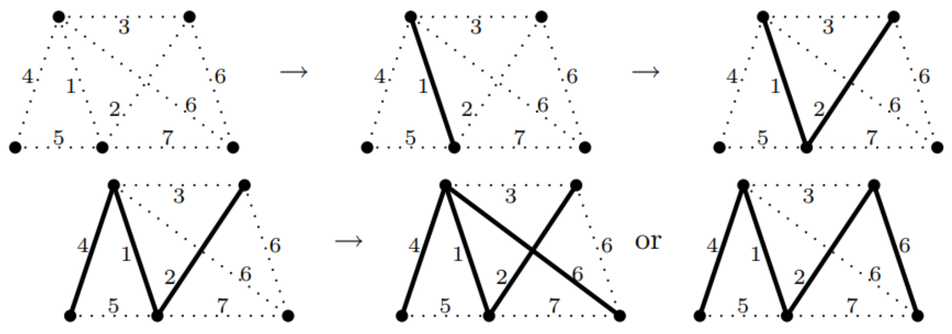
- 中国著名数学家，曾任山东师范大学校长
- 从事运筹学及其应用的研究，对最短投递路线问题的研究取得成果，冠名为中国邮路问题
- 致力于城市交通规划的研究

最小生成树 Prim 算法

- 任选一个顶点 u，选择与 u 关联的权值最小的边作为最小生成树的第一条边 e_1
- 在与已经选取边只有一个公共端点的边中，选取权值最小的边
- 直到选取 $n-1$ 条边为止

最小生成树 Prim 算法

- 任选一个顶点 u，选择与 u 关联的权值最小的边作为最小生成树的第一条边 e_1
- 在与已经选取边只有一个公共端点的边中，选取权值最小的边
- 直到选取 $n-1$ 条边为止



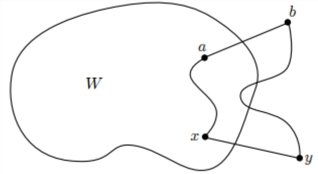
最小生成树 Prim 算法

习题

Prim 算法得到的任何生成树一定是最小生成树

证明：反证法

- 记 Prim 算法得到的树为 $T = e_1, \cdots, e_{n-1}$，记 U 为最小生成树中和 T 相同前缀最长的，设 $e_i = \{xy\}$ 是第一个不在 U 中的边
- $e_i \notin U$，故存在边 ab，考察树 $T' = U + \{xy\} - \{ab\}$
  - 若 $w_{ab} = w_{xy}$，$w(T') = w(U)$，而 T' 和 T 有更长的共同前缀，矛盾
  - 若 $w_{ab} < w_{xy}$，Prim 算法应先选择 ab，矛盾
  - 若 $w_{ab} > w_{xy}$，$w(T') < w(U)$，矛盾

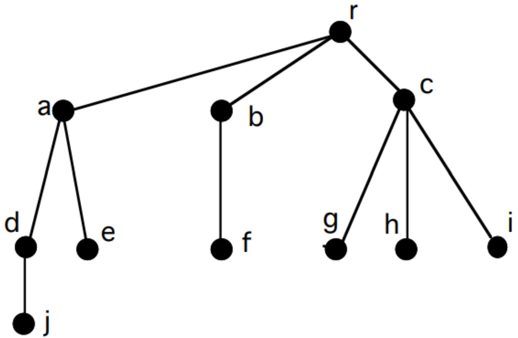


根树

- 每条边都有一个方向的树称为有向树
  - 以 v 为终点和起点的边数称为 v 的入度和出度
  - 入度与出度之和称为点 v 的度
- 一棵有向树，如果恰有一个顶点的入度为 0，其余顶点入度为 1，这样的的树称为根树
  - 入度为 0 的点称为树根
  - 出度为 0 的点称为树叶
  - 入度为 1，出度大于 1 的点称为内点
  - 内点和树根统称为分支点
  - 顶点 v 到树根的距离称为 v 的层数
  - 最大层数称为树高

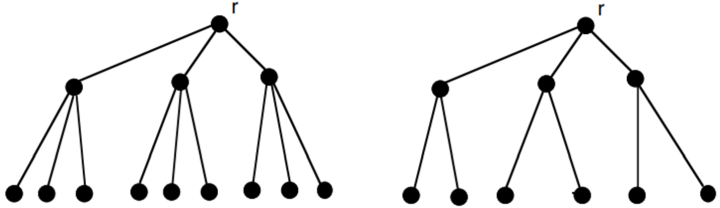
Example The height of this tree is 3. Also,

- r, a, b, c, and d are the internal vertices;
- vertices e, f, g, h, i, and j are the leaves;
- vertices g, h, and i are siblings;
- vertex a is an ancestor of j; and
- j is a descendant of a.



有序树与完全树

- 对于根树 T，若规定了每层顶点的访问次序，这样的根树称为有序树
  - 一般次序为从左至右
- 对于根树，由其中节点及其后代导出的子图，称为子根树
- 对于根树，若每个分支点至多 m 个儿子，称为 m 元根树；若每个分支点恰有 m 个儿子，称它为完全 m 元树



Two 3-ary trees: one complete, one incomplete.

m 元树

定理

在完全 m 元树中，若树叶数为 t , 分支点数为 i , 则

$$(m-1)i = t-1$$

证明

- 由树的性质得： $m(T) = i + t - 1$
- 由握手定理得： $2m(T) = t + m + (i-1)(m+1)$
- 联立得 $(m-1)i = t-1$

m 元树

习题

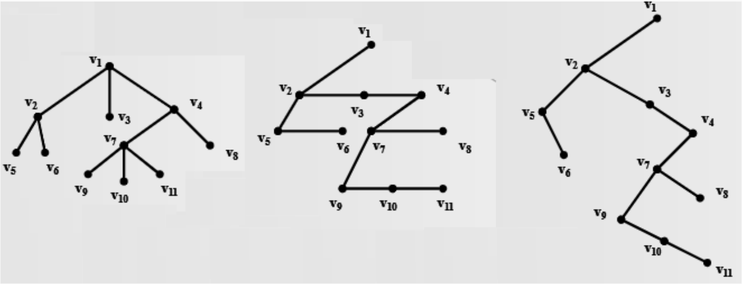
一条加法指令可以计算 3 个数的和。9 个数的和至少需要执行多少次加法指令?

解答

- 用 3 个顶点表示 3 个数，用一个父结点表示 3 个数的和
- 问题转化为求一棵有 9 个叶点的完全 3 元树的分支点数
- 即：$m = 3, t = 9, i =?$
- $(3-1)i = 9-1$：$i = 4$

二元树（二叉树）

- m 元树中，应用最广泛的是二元树，原因是它在计算机中容易处理
- 有序树转化为二元树的方法
  - 从根开始，保留每个父亲同其最左边儿子的连线，撤销与别的儿子的连线
  - 兄弟间用从左至右的有向边连接
  - 直接位于给定结点下面的儿子，作为左儿子，对于同一水平线上与给定结点右邻的结点，作为右儿子



二叉树遍历

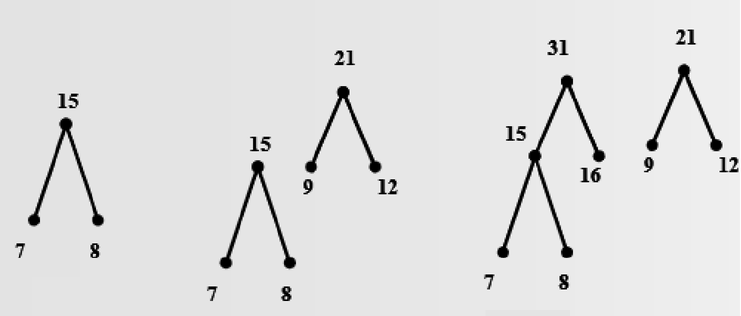
- 二元树的遍历：找到一种方法，每个结点恰好访问一次
- 有三种常用遍历方法
  - 先根次序遍历
    - 访问根
    - 按先根次序遍历根的左子树
    - 按先根次序遍历根的右子树
  - 中根次序
  - 后根次序

最优二叉树

- 设 T 是二元树，若对所有 t 片树叶赋权 w_i $(1 \leq i \leq t)$，权值为 w_i 的树叶层数记为 $L(w_i)$，称 $W(T) = \sum_{i=1}^{t} w_i L(w_i)$ 为该二元树的权。
- $W(T)$ 最小的二元树称为最优二元树
- Huffman 算法
  - 令 $S = \{w_1, \cdots, w_t\}$
  - 从 S 中取出两个权值最小者 w_i, w_j
  - 画结点 v_i, v_j，权值 w_i, w_j，画 v_i, v_j 的父亲 v，权值 $w_i + w_j$
  - $S \leftarrow (S - \{w_i, w_j\}) \cup \{w_i + w_j\}$
  - 如果 S 只含一个元素，停止，否则转第二步

习题

求权值为 $7, 8, 9, 12, 16$ 的最优二叉树



最优二叉树

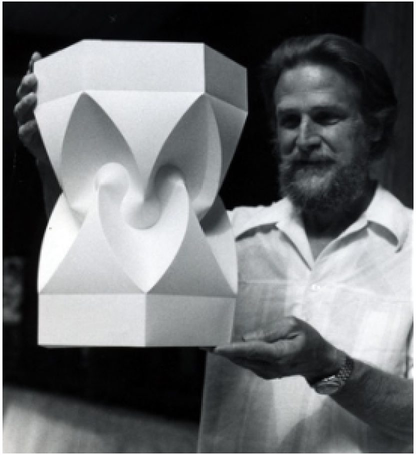
Huffman 算法的最优性

Huffman 算法生成的是最优二叉树

证明

- 设 x, y 是权值最小的 2 个节点，最优二叉树中，它们是最深的 2 个兄弟
- 令 S' 为 S 去掉 x, y 加入权值为 $x+y$ 的节点 z 的点集
- 设 T, T' 是基于 S, S' 构建的 Huffman 树
    - $W(T) = W(T') + x + y$
- 对节点数做归纳，假设对 S'，T' 是最优树
- 若对 S，T 不是最优树：存在树 R 使得 $W(R) < W(T)$
- 记 R' 为 R 去掉 x, y 加入权值为 z 对应的树：$W(R) = W(R') + x + y$
- $W(R) = W(R') + x + y < W(T) = W(T') + x + y$，故 $W(R') < W(T')$：矛盾

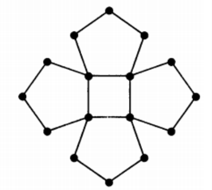
David Huffman



David Huffman in 1978 and in 1999 (both photos courtesy of University of California, Santa Cruz)

课后练习与思考题

- 下图是 2000 年在 Kalamazoo 召开的图论会议的 logo，它有多少棵生成树



- 考察 n 个消息，每个出现的概率为 p_i，已知 p_i 可以写成 2^{-j}，证明（1）至少存在两个消息出现的概率相等，（2）用 Huffman 树对消息编码后，平均码长为 $-\sum_i p_i \log p_i$，1948 年 Shannon 得出了最短码长