图论与代数结构

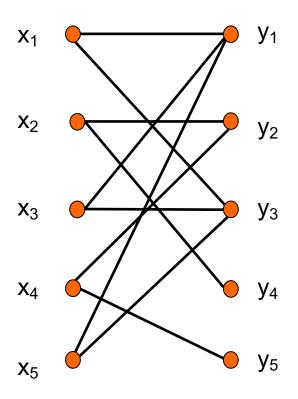
匹配(最大/完全/最佳)

崔 勇 清华大学计算机系 网络技术研究所

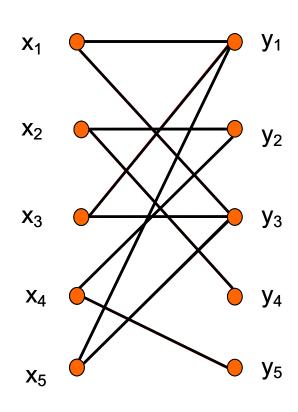
主要内容:匹配

- 二分图的最大匹配
 - 掌握概念、定理和匈牙利算法
- 完全匹配(特殊情况)
 - 掌握概念和定理及其应用
- 最佳匹配及其算法(最优)
 - 掌握K-M算法的基本思路

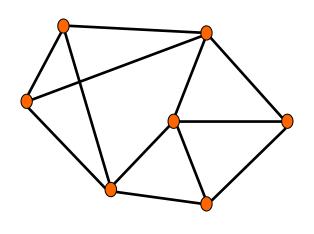
- 分配n个人来做m项工作
- 图论建模 , G=(V,E)
 - 边(x_i,y_j)表示x_i可以从事y_j
- 约束条件
 - 每个人最多从事其中一项
 - 每项工作只能由其中一个人来承担
- 问怎样才能让更多的人安排上工作?
 - 该问题的图论建模?



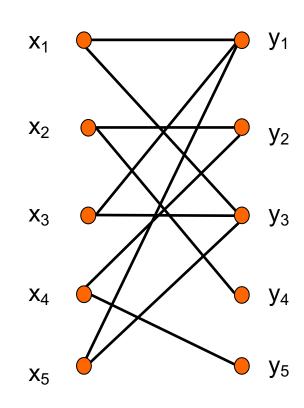
- 该图是一个二分图 (二部图)
 - 按照要求,如果x_i从事了y_j,就不允许再从事y_k, 同时y_i也不再允许其它人承担
- 如何描述呢?
 - 因此它相当于用一种颜色比如红色对G的边进行 着色,保证每个结点最多只与一条红色边相关联
 - 这种红色边的集合记为M,它就称为匹配
- 原问题就是计算G中包含边数最多的一个匹配M



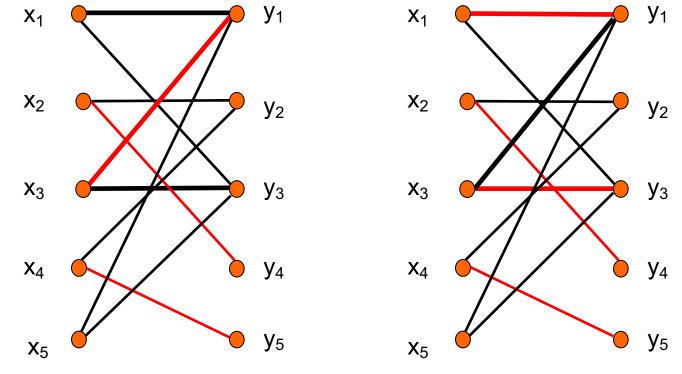
- 例5.1.2
 - 二战期间,盟军派飞行员到英国参加对德的空袭,每架飞机需要领航员和飞行员各1人
 - 有些人只会领航,一些人只会驾驶,也有两者均会的
 - 求最多的编队方案 (要求二人语言相通)
 - 图论建模
 - 以点表示人
 - 边表示二人语言相通并且一人可领航一人可驾驶
 - 那么最多的编队方案就是计算G的一个最大匹配



- 定义5.1.1: 匹配
 - 令M是图G的边集,若M中任意两条边都没有共同顶点,则称M为G的一个匹配
- 关键术语定义
 - 图中与M边相关联的结点称为<mark>饱和点</mark>,否则 称为非饱和点
 - 饱和点: x₁,x₂,x₃,x₄,y₁,y₃,y₄,y₅
 - 非饱和点: x₅,y₂
- 最多的任务分配是什么?

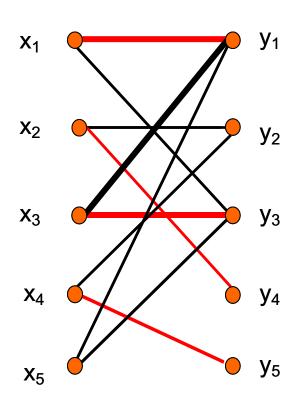


 定义5.1.2:设M是G=(V,E)中的一个匹配,如果对G的任意匹配M',都有 |M|≥|M'|,则称M为一个最大匹配



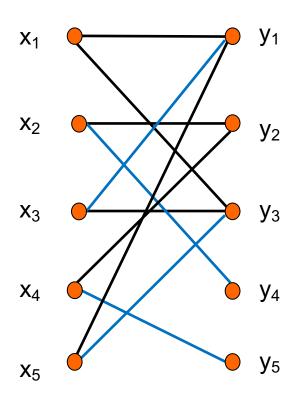
• 定义5.1.3:给定了G的一个匹配M,G中属于M与不属于M的边交替出现的 道路称为<mark>交互道路</mark>(x_1 - y_1 - x_3 - y_3)

- 定义5.1.4:可增广道路
 - 设P是G中关于匹配M的一条交互道路
 - 如果P的两个端点是关于M的非饱和点,那么它就称为可增广道路
 - 例如:x₁-y₁-x₃-y₃(黑-红-黑)
- 可增广道路的基本特点
 - 可增广道路P一定包含奇数条边,且其中不属于匹配 M的边比M中的边多一条
 - M'=P ⊕ M仍然是一个匹配
 - M'使P的两个端点变成了饱和点
 - |M'|=|M|+1,即M'是比M更大的匹配



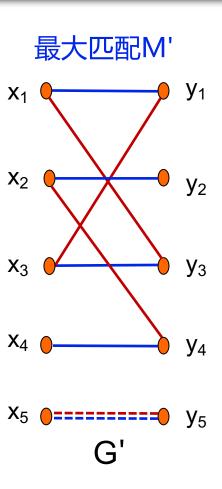
例

- 蓝色的边构成图的匹配M
- 关于该匹配的饱和点是{x₃, y₁, x₂, y₄, x₄, y₅, x₅, y₃}
- 非饱和点是{x₁, y₂}
- 道路 (x1, y1, x3, y3, x5)是—条交互道路,但 不是可增广道路

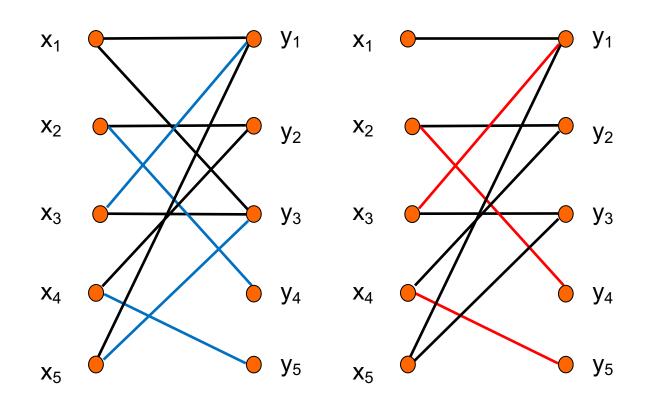


- 定理5.1.1(Berge 1957) (重点)
 - M是G的最大匹配当且仅当G中不存在关于M的可增广道路
- 证明
 - 必要性(反证法)
 - -若存在M的可增广道路P,则M⊕P=M'是G的一个新匹配,且 |M'|>|M|,与M是最大匹配矛盾
 - 充分性(反证法):要证M不是最大匹配,则存在M的可增广道路
 - 如果匹配M不是G的最大匹配,则存在一个最大匹配M',做G'=M'⊕M,我们逐一分析G'中三种可能的连通支

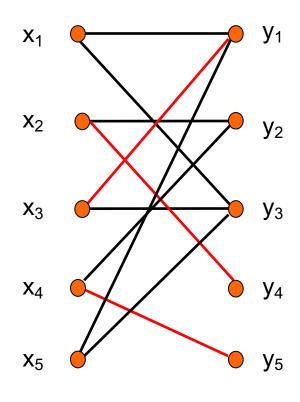
- 要证: M不是最大匹配则存在M的可增广道路;
 已有G'= M'⊕M,且M'是最大匹配,即|M'|>|M|
- 1. 孤立结点v_i (M'和M均为不连通的边集)
 - 关联v_i的边(v_i,v_i)∈ M'∩M,对M'和M的贡献相同
- 2. 交互回路(例如x₁-y₁-x₃-y₃)
 - 该回路必为偶回路,其中属于M'和M的边数相同
- 3. 交互道路
 - 若不存在增广道路,则|M'|=|M|,与假设矛盾
 - 若只存在M关于M'的增广路/红蓝红,与M'是最大匹配矛盾
 - 由于|M'|>|M|, 故必定存在M'关于M的可增广交互道路,例如 $x_4-y_4-x_2-y_2$,即G中存在关于M的可增广道路



- 例:
 - 蓝色边构成的匹配
 - 不存在可增广道路
 - 是最大匹配
 - 红色边构成的匹配
 - 不是最大匹配
 - 存在可增广道路

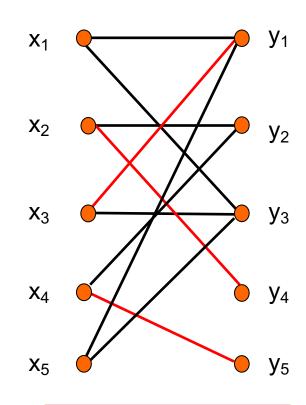


- 计算二分图的最大匹配算法匈牙利算法
 - 输入为二分图G=(X,Y,E);
 - 结点标记
 - 0:表示尚未搜索
 - 1:表示饱和点
 - 2:表示无法扩大匹配的点(仅对X中的结点)
 - 1. 任给初始匹配M,给饱和点"1"标记
 - 2. 判断X各结点是否都已有非零标记
 - 2.1 是, M是最大匹配, 结束
 - 2.2 否,找一个"0"标记点x₀∈X,开始本轮搜索令U← {x₀},为本轮搜索的X结点集令V ← Φ,为本轮已检查过的Y结点集



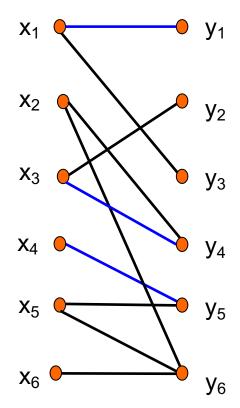
X₁y₃或 X₁y₁X₃y₃

- 3、判断集合U的邻接点集Γ(U)=V(访问过的Y)
 - 3.1 是, x₀无扩大匹配, 标x₀ "2", 转2
 - 3.2 否, 在Γ(U)-V中找一点y_i, 判断y_i的饱和性
 - (1)非饱和点(如y₃):
 存在从x₀到y_j的可增广路P
 令M ← M ⊕ P,给x₀,y_i标记1转2
 - (2)饱和点(如y₁):
 则存在x_i使得边(x_i, y_i) ∈ M
 令 U ← U + {x_i}, V ← V + {y_i}, 转3
 (如U= {x₁, x₃})

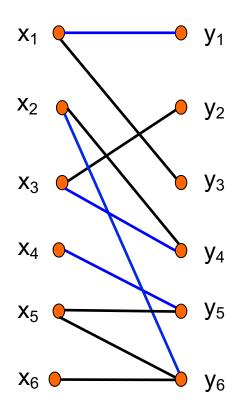


向右走非匹配边 向左则走匹配边

- 例5.1.3 设初始匹配: M={(x₁,y₁), (x₃,y₄), (x₄,y₅)}
- 用匈牙利算法求最大匹配
 - 进行节点标记
 - $U={x₂}$, V= Φ (为本次已访问过的Y结点集) Γ(U)={y₄,y₆} , y₆ ∈ Γ(U)-V , 且无标记
 - 得到增广路P=(x₂,y₆)
 - 更改节点标记
 - $-M = \{ (x_1,y_1), (x_3,y_4), (x_4,y_5), (x_2,y_6) \}$

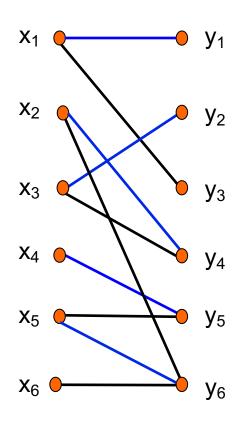


- U={x₅},访问过的y节点集V= Φ Γ(U)={y₅,y₆}, y₅∈Γ(U)-V,不可增广
- $U = \{x_5, x_4\}$, $V = \{y_5\}$ $\Gamma(U) = \{y_5, y_6\}$, $y_6 \in \Gamma(U) - V$
- $U = \{x_5, x_4, x_2\}$, $V = \{y_5, y_6\}$ $\Gamma(U) = \{y_5, y_6, y_4\}$, $y_4 \in \Gamma(U) - V$
- U={x₅,x₄,x₂,x₃}, V={y₅,y₆,y₄} Γ(U)={y₅,y₆,y₄,y₂}, y₂∈Γ(U)-V 且无标记
- ∴增广路P=(x_5,y_6,x_2,y_4,x_3,y_2) $M=M\oplus P=\{(x_1,y_1), (x_4,y_5), (x_5,y_6), (x_2,y_4), (x_2,y_3)\}$



- (3)U={ x_6 }, V= Φ Γ(U)={ y_6 }, y_6 ∈Γ(U)-V
- $U = \{x_6, x_5\}$, $V = \{y_6\}$ $\Gamma(U) = \{y_6, y_5\}$, $y_5 \in \Gamma(U) - V$
- $U = \{x_6, x_5, x_4\}$, $V = \{y_6, y_5\}$ $\Gamma(U) = \{y_6, y_5\}$, $\Gamma(U) = V$
- 给x₆标记2。结束。
- 因此其最大匹配是

 $M = \{ (x_1, y_1), (x_4, y_5), (x_5, y_6), (x_2, y_4), (x_3, y_2) \}$



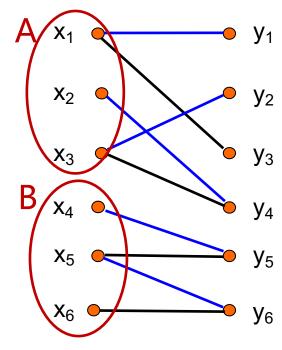
- 定理5.1.2
 - 最大匹配的匈牙利算法, 计算复杂度为O(mn), 其中n是二分图G中X的结点数
- 证明
 - 初始匹配可以是空匹配
 - 算法最多找n条增广路
 - 每找一条增广路时, 最多判断m条边
 - 因此其计算复杂性是O(mn)



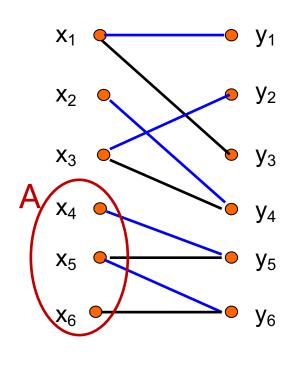
第五章 匹配与网络流

- 二分图的最大匹配
- 完全匹配
- 最佳匹配及其算法

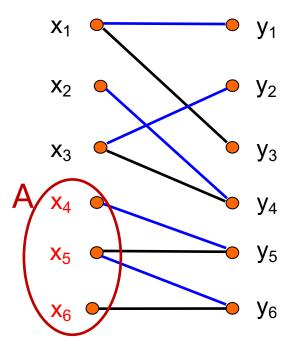
- 完全匹配
 - 二分图G=(X,Y,E)的最大匹配M包含的边数不会超过 |X| , 若|M|=|X| , 则称M是完全匹配
- 完美匹配
 - 如果|M|=|X|=|Y| , 则称M是完美匹配
- 呼唤定理:完全匹配的存在性?
- 定理5.2.1(Hall定理)
 - 在二分图G=(X,Y,E)中,X到Y存在完全匹配的充要 条件为:对于X的任意子集A,恒有|Γ(A)|≥|A|



- 证明(存在完全匹配的充要条件:任意子集A有|Γ(A)|≥|A|)
 - 必要性(反证法)
 - 若存在子集A⊆ X,使|A|>|Γ(A)|
 - •则A中的结点无法全部匹配
 - 因此X到Y不可能有完全匹配
 - 充分性(重点)
 - 要证:不存在完全匹配,则存在A满足|Γ(A)|<|A|)
 - 假定G的一个最大匹配M不是完全匹配
 - 一定存在结点 $x_0 \in X$ 是关于M的非饱和点(如 x_6)
 - 如果Γ(x₀)=Φ,则令A={x₀},于是|Γ(A)|< |A|,不满足条件

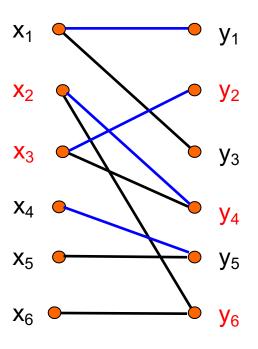


- 证(续) (要证存在子集|A| > |Γ(A)|)
 - 若Γ(x₀) ≠Φ , 对某y_i ∈ Γ(x₀)
 - 若 y_i 关于M为非饱和点,则存在增广路 (x_0, y_i) ,与M是最大匹配矛盾
 - 因此 y_i ∈ $\Gamma(x_0)$ 都是关于M的饱和点,但是 x_0y_i 为非匹配边
 - 这样可以寻找以xo为端点的相对于M的一切交互道路
 - 记交互道路中结点 y_i 的集合为 Y_1 , 结点 x_i 的集合为 X_1 (新增的x与y——对应)
 - 根据匹配的性质, Y_1 结点与 X_1 - x_0 的结点之间存在——对应,于是 $|X_1|>|Y_1|$,即 $|X_1|>|\Gamma(X_1)|$
 - 对非完全匹配,构造子集|A|,满足|A|>|Γ(A)|。
 - 证毕充分性



X到Y存在完全匹配, ⇔X的任意子集A, 恒有|Γ(A)|≥|A|

- 回顾: Hall定理
 - 在二分图G=(X,Y,E)中,X到Y存在完全匹配的充要条件为:对于X的任意子集A, 恒有|Γ(A)|≥|A|
- 推论5.2.1
 - 若二分图G=(X,Y,E)的每个结点 x_i ∈X,都有 $d(x_i)$ ≥k,每个结点 y_j ∈Y,都有 $d(y_j)$ ≤k,则X到Y存在完全匹配
- 证明
 - 对任意子集A⊆X
 - 设A共与m条边相关联,由d(x_i) ≥k且边不重复,于是m ≥ k |A|
 - 这m条边又与Y中的 $|\Gamma(A)|$ 个结点相关联,由d (y_j) ≤k,又有m ≤ k $|\Gamma(A)|$
 - 因此|Γ(A)| ≥ |A| , 由定理5.2.1即得



- 例5.2.1
 - 在一个舞会上男女各占一半,假定每位男士都认识k位女士,每位女士都 认识k位男士,那么一定可以安排得当,使每位都有认识的人作为舞伴
- 证明(图论建模)
 - 图论建模:用结点x_i表示每位男士,结点y_j表示每位女士,互相认识者用 边连之,则存在完美匹配
 - 于是得到二分图G=(X,Y,E),图中每个 x_i 结点有d(x_i) =k,每个 y_j 结点有 d(y_j) =k
 - 满足d(x_i) ≥k , d(y_i) ≤k
 - 由推论5.2.1, X到Y有完全匹配M(即完美匹配)
 - M就是一种安排方案

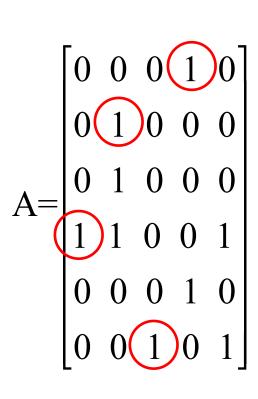
最大匹配是多少? 呼唤个定理?

- 定理5.2.2 (最大匹配数)
 - 在二分图G=(X,Y,E)中,X到Y的最大匹配边数是|X|-δ(G),其中δ(G) 为不能匹配的X节点数

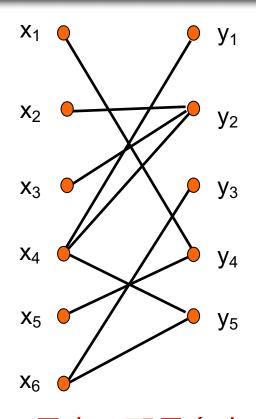
$$\delta(G) = \max_{A \subseteq X} \delta(A), \delta(A) = |A| - |\Gamma(A)|, \delta(A) \ge 0$$

- 例5.2.2
 - 10个人有10件不同的乐器,其中3人只会拉小提琴,其余7人每件乐器都会。若每人只用一件乐器,问最多有几人能同时登台演出?
 - 图论建模
 - 由定理5.2.2,最多只有8人能同时登台演出

- 二分图邻接矩阵
 - 简化为|X|×|Y|矩阵
- 最大匹配是多少?
 - A中不同行且不同列的非零元 的最大数目
 - 如果矩阵A是p×q矩阵,则: 最大匹配数≤min(p,q)



二分图邻接矩阵



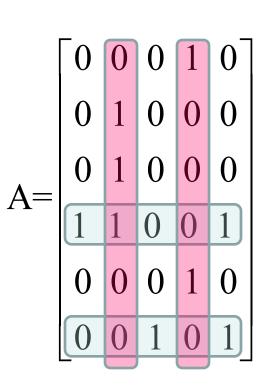
最大匹配是多少? 考虑 x_2 和 x_3 、 x_1 和 x_5

• 矩阵覆盖

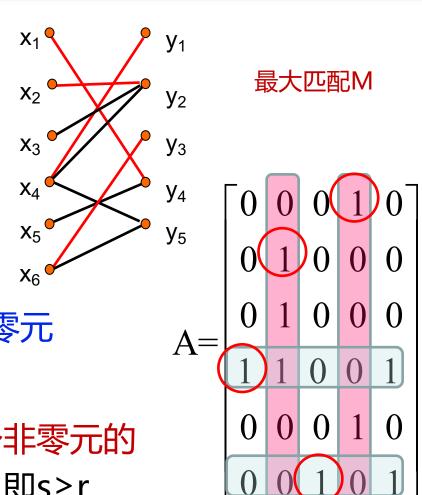
- 适当地选取A的某些行和列,使这些行和列能盖住A的全部 非零元,称为A的覆盖
- 覆盖数为所选取的行和列的个数
- 矩阵A中如果盖住第4、6行,第2、4列,就可以覆盖其全部非零元,覆盖数为4

• 最小覆盖

- 如果选取最少的行与列就能覆盖A的全部非零元,则称这样的覆盖为最小覆盖
- 在矩阵A的各种覆盖中,一定存在最小覆盖,其覆盖数为s,显然s<=min(p,q)
- 二分图:最大匹配数r P.K. 其邻接矩阵的最小覆盖数s

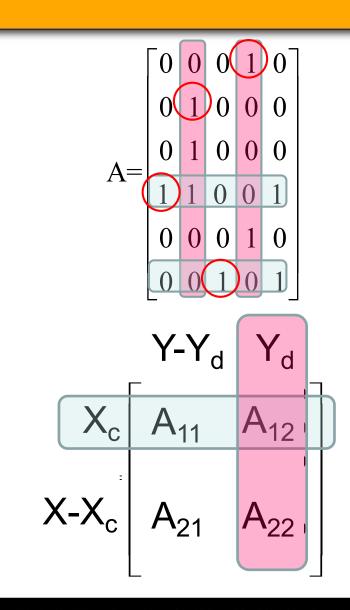


- 二分图的最大匹配数r,与其邻接矩阵的 最小覆盖数s相等
- 证明(重点):
 - 先证明:最小覆盖数s ≥最大匹配数r
 - 设M是二分图G的一个最大匹配
 - M在邻接矩阵中是r个既不同行也不同列的非零元
 - 一个覆盖必然把这r个非零元都盖住
 - 覆盖s中的任一行或列,都最多只能覆盖这r个非零元的 其中一个,因此覆盖s的行列数之和至少为r,即s≥r



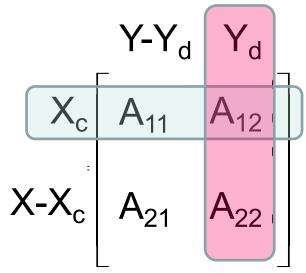
证明:

- 再证明:最大匹配数r≥最小覆盖数s
- 设G的一个最小覆盖由c行d列组成
- 最小覆盖数s=c+d
- 对邻接矩阵进行换行、换列操作, 将这c行d列交换至前c行和后d列
- 换行换列操作不改变矩阵的最小覆盖数
- 换行换列操作本质上是结点重新编号,因此不影响二分图的最大匹配数



证明:

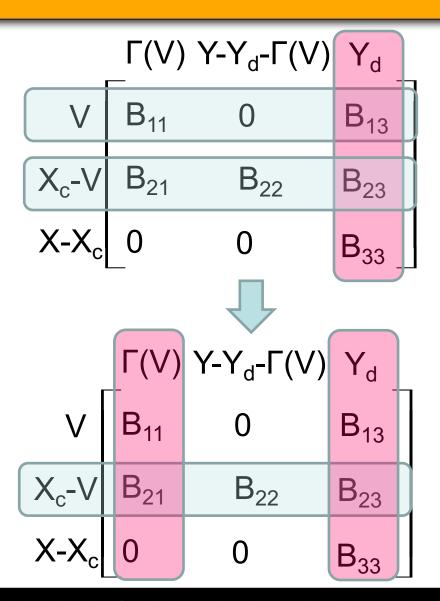
- 为证明最大匹配数r≥最小覆盖数s,需 找一个匹配M,使得|M| = s = c+d
- A₁₁, A₁₂, A₂₂都被覆盖了, A₂₁没 被覆盖,必有A₂₁=0
- A=
 \[
 \begin{bmatrix}
 0 & 0 & 0 & 1 & 0 \\
 0 & 1 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 & 1 \\
 0 & 0 & 1 & 0 & 1
 \end{bmatrix}
 \]
 \[
 \begin{bmatrix}
 X \\
 0 & 0 & 1 & 0 & 1 \\
 0 & 0 & 1 & 0 & 1
 \end{bmatrix}
 \]



- 后续证明思路:
 - X_c到Y-Y_d存在完全匹配(大小即为c), Y_d到X-X_c存在完全匹配 (大小即为d), 从而X到Y存在大小为c+d的匹配
- 考虑Hall定理(存在完全匹配的充要条件|Γ(A)|≥|A|)

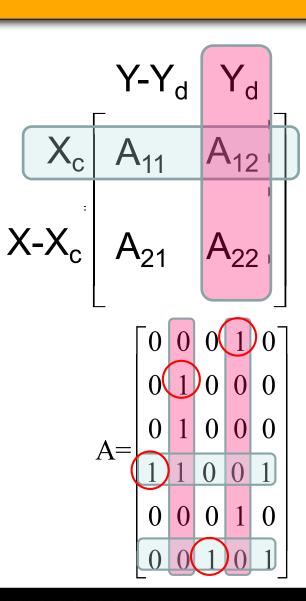
证明:

- 任取X_c的子集V,设V在Y-Y_d中的邻接点 集为Γ(V),证|V|≤|Γ(V)|
- V在矩阵中对应着|V|行, Γ(V)在矩阵中对应着|Γ(V)|列。这|V|行中的非零元均在这|Γ(V)|列或Y_d中
- 假设|Γ(V)|<|V|, 从覆盖这|V|行改为覆盖 |Γ(V)|列即可减少覆盖数,与当前是最小 覆盖矛盾
 即任意V有|V|≤|Γ(V)|, 对X_c, 有| X_c|≤|Γ(X_c)|



• 证明:

- 综上,由Hall定理知,存在一个 X_c 的完全匹配,即 X_c 与 $Y-Y_d$ 之间的匹配 M_1 ,使得 $|M_1|=|X_c|=c$
- 同理, Y_d 与 $X-X_c$ 也满足Hall定理的条件,故存在 一个 Y_d 与 $X-X_c$ 之间的匹配 M_2 ,使得 $|M_2|=|Y_d|=d$
- $\diamondsuit M = M_1 \cup M_2$,则M是X与Y之间的匹配,且 |M| = c + d = s
- 即X与Y的最大匹配数r ≥ |M| = 最小覆盖数s
- 结合先前证的s≥r,得r=s





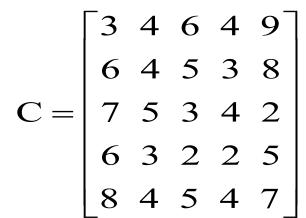
第五章 匹配与网络流

- 二分图的最大匹配
- 完全匹配
- 最佳匹配及其算法

- 从旅行商问题想到了什么?
 - 边权为1的H回路(存在性、唯一性)
 - 给定正权完全图,求最短H回路

• 最佳匹配

- 如果边权是非负实数,而且存在多个完全匹配,那么其中权和最大或最小的 匹配就叫最佳匹配(如最大权匹配)
- 例5.3.1
 - 5个码农要完成5个模块
 - i从事工作j的利润







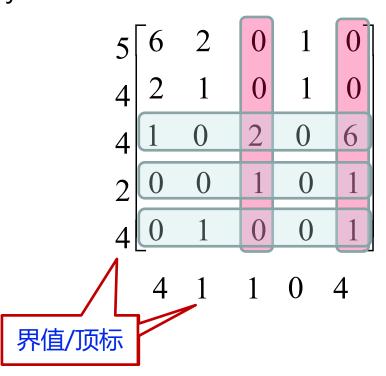
- 若C_{ij}表示i从事工作j的利润
 - 如果每个人只从事一项工作
 - 那么最大利润是 $\max \sum C_{ij}$
 - C_{ii}不在相同的行与列
- 若C_{ij}表示i从事工作j成本
 - 那么最小成本是 $\min \sum C_{ij}$
 - C_{ii}不在相同的行与列
- 此最佳匹配为二分图的最大权或是最小权匹配

最大匹配数r

最小覆盖数s

- 最大权匹配 (已知利润矩阵C): KM算法 (Kuh_Munkras)
 - -1.在C的每行中选一最大值作为本行的界值 $I(x_i)$,每列的界值 $I(y_j)=0$. 构造矩阵 $B=(b_{ij})_{n\times n}$,其中 $b_{ij}=I(x_i)+I(y_i)-c_{ij}$
 - 2.在B中对0元素进行最小覆盖,覆盖数为r
 - 2.1 若r=n , 转4
 - 2.2 更新B矩阵

在未覆盖的元素中选最小非零元 δ 若 x_i 行、 y_j 列均已覆盖,则 b_{ij} 一 b_{ij} + δ 若 x_i 行、 y_i 列均未覆盖,则 b_{ii} 一 b_{ii} - δ

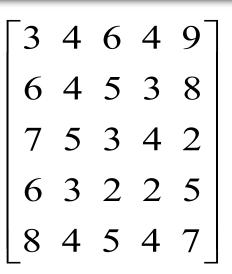


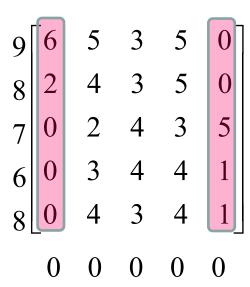
- 最大匹配算法(续)
 - 3. 修改界值

若 x_i 行没覆盖,则 (x_i) ← $I(x_i)$ - δ 若 y_j 列已覆盖,则 (y_j) ← $I(y_j)$ + δ 删除覆盖标记,转2

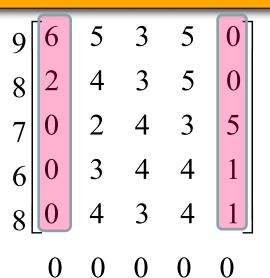
 $-4. \sum (l(x_i) + l(y_i)) = 29$ 即为最大权,结束

- 例5.3.2
 - 已知利润矩阵, 求二分图的最佳匹配(最大利润)
- 解
 - 计算界值(表的两旁标出)
 - 计算矩阵B
 - 求最小覆盖
 - 如1、5两列
 - 在未覆盖的元素中选最小非零元,即 $\delta=2$



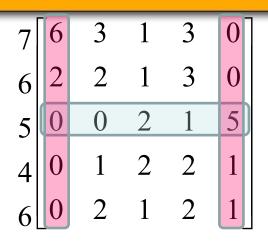


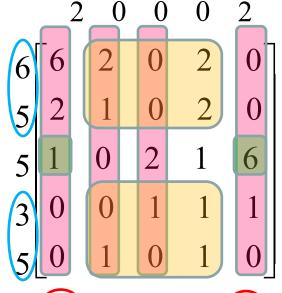
- 解(续)
 - 覆盖数r=2<n, δ=2
 - 产生新B:未覆盖的元素均减δ
 - 修改界值
 - 若x_i行没覆盖,则I(x_i) ← I(x_i)-δ
 - 若y_j列已覆盖,则I(y_j)←I(y_j)+δ
 - 重新求最小覆盖
 - 如第1、5两列和第3行
 - 在未覆盖的元素中选最小非零元,即 δ =1



<u> </u>	6	3	1	3	0
6	2	2	1	3	0
65	0	0	2	1	5
4	0	1	2	2	1
6	0	2	1	2	1

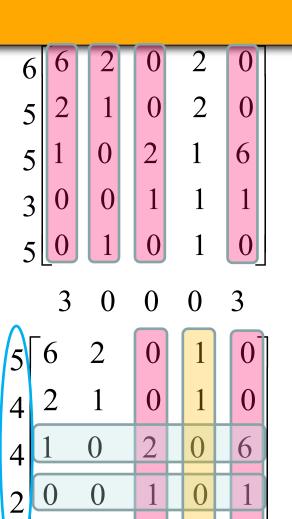
- 解(续)
 - 覆盖数r<n , δ=1
 - B中没覆盖的元素均减1
 - 双重覆盖元加1
 - 修改界值
 - 没覆盖的x_i行: I(x_i) ← I(x_i)-δ
 - 已覆盖的y_j列: I(y_j) ←I(y_j)+δ
 - 重新求最小覆盖
 - 如第1,2,3,5列
 - 在未覆盖的元素中选最小非零元,即 δ =1

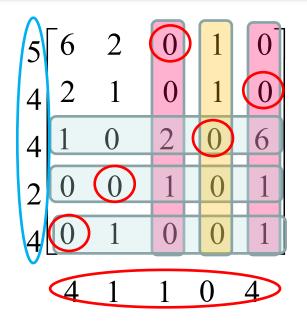


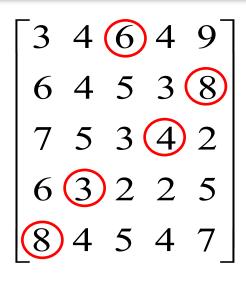


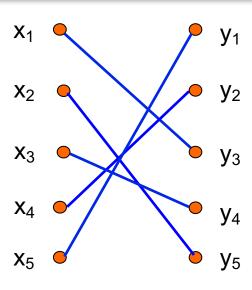
- 覆盖数r<n, δ=1
- B中没覆盖的元素均减1
- 修改界值
 - 没覆盖的x_i行: I(x_i) ← I(x_i)-δ
 - 已覆盖的y_i列: I(y_i) ←I(y_i) +δ
- 重新求最小覆盖
 - 如第3、4、5行和3、5列
- -最小覆盖数r=n
- 循环结束

肉眼求解最小覆盖









- 当前最小覆盖数r=n,停止扩充覆盖的循环
- 得到最大权 $\sum (l(x_i) + l(y_i)) = 29$
- 一个最大权匹配方案是{C₁₃, C₂₅, C₃₄, C₄₂, C₅₁}
- · 定理5.3.1:上述K-M算法(最大权匹配算法)的结果是矩阵C的最大权匹配
 - 证明略

总结

• 二分图的最大匹配

- 从细节定义开始:匹配,饱和点、非饱和点,神奇道路(交互/增广)
- 呼唤定理:可增广道路的存在性与最大匹配的关系
- 基于定理给出匈牙利算法:通过找可增广道路求解最大匹配
- 完全匹配
 - Hall定理:完全匹配的充要条件是任意子集A恒有|Γ(A)|≥|A|
 - 最大匹配数是|X|-δ(G) ⊗
 - 最大匹配数r=最小覆盖数s
- 最佳匹配及其算法
 - 从旅行商到最佳匹配(玩命化简)
 - 基于最小覆盖的最大权匹配算法(K-M算法)



• P119习题五

- 最大匹配:第1题

- 完美匹配:第5题(题目改为"完美匹配")

- 最佳匹配:第13题