# MATH 567: Lecture 23 (04/03/2025)

Today: * Heuristic algo for set cover
* variants of set cover problem

## Receiver location problem: Heuristic algorithm

coverage info for meter $i$ ; all coverage info

Let $C_i = \{j \in P \mid j \text{ covers } i\}$, and $C = \{C_i\}_i$.

set of poles ; set of meters ; $k \in \mathbb{Z}_{>0}$

HEURISTIC_SETCOVER $(P, M, C, k, \varepsilon)$ — tolerance for comparing score

INPUT: set of poles $P$, set of meters $M$, coverage $C$, parameter $k$, tolerance $\varepsilon$;

OUTPUT: $P' \subseteq P$ that covers all $i \in M$.

Initialization: $P' = \emptyset$, $M' = \emptyset$ (where $M' \subseteq M$ is the set of meters covered by $P'$).

while $M' \neq M$ do

PREPROCESS $(P \setminus P', M \setminus M')$; — set difference

↳ steps (i), (ii), (iii)

Compute Score$(\cdot)$ for $(P \setminus P', M \setminus M')$; — after preprocessing

Select $j \in P \setminus P'$ with $\text{Score}(j) = \max\limits_{l \in P \setminus P'} (\text{Score}(l)) \pm \varepsilon$;

$P' = P' \cup \{j\}$;

Update $M'$;

CLEANUP $P'$; ← could do CLEANUP only after every $10^{th}$ pole is selected, say.

end-while

# 4. IP formulation

We first look at the IP formulation to describe how to implement the steps in the heuristic algorithm (preprocessing, in particular)

We let $M := \{1,2,\dots,m\}$ denote the set of meters, and $P := \{1,2,\dots,p\}$ the set of poles. The coverage data is specified as $C_i = \{j \in P \mid j \text{ covers } i\}$ $\forall i \in M$.

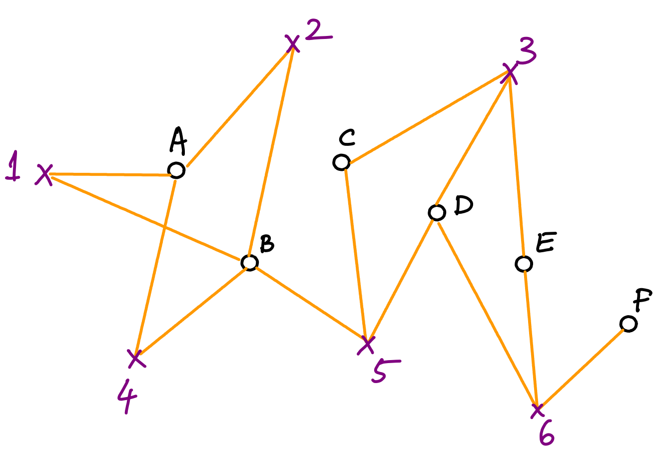
Let $x_j = \begin{cases} 1 & \text{if pole } j \text{ is selected (to locate a receiver)} \\ 0 & \text{otherwise} \end{cases}$

facility

Here is the IP:

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{p} x_j \\
\text{s.t.} \quad & \sum_{j \in C_i} x_j \geq 1 \quad \forall i \in M \qquad (\Delta) \\
& x_j \in \{0,1\} \quad \forall j \in P
\end{aligned}
$$

Illustration on the example:




$$
\begin{aligned}
\min \quad & x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \qquad \mathbb{1} \\
\text{s.t.} \quad & x_1 + x_2 + x_4 \geq 1 \quad (A) \\
& x_1 + x_2 + x_4 + x_5 \geq 1 \quad (B) \\
& x_3 + x_5 \geq 1 \quad (C) \\
& x_3 + x_5 + x_6 \geq 1 \quad (D) \\
& x_3 + x_6 \geq 1 \quad (E) \\
& x_6 \geq 1 \quad (F)
\end{aligned}
$$

$A\bar{x}$

$x_j \in \{0,1\},\ j = 1,\dots,6.$

$A\bar{x} \geq \mathbb{1}$ vector of ones

## Steps in Preprocessing

(i) (singleton rows): Set $x_6 = 1$, delete all rows containing $x_6$, i.e., (D) & (E)

$$
\begin{array}{llll}
 & & \text{(ii)} & \\
\min & x_1 + x_2 + x_3 + x_4 + x_5 + x_6 & & \\
\text{s.t.} & x_1 + x_2 \quad + x_4 & \geq 1 & (A) \\
 & \cancel{x_1 + x_2 \quad + x_4 + x_5} & \cancel{\geq 1} & (B) \quad \text{(iii)} \\
 & \quad x_3 \quad + x_5 & \geq 1 & (C) \\
 & \cancel{\quad x_3 \quad + x_5 + x_6} & \cancel{\geq 1} & (D) \\
A\bar{x} & \cancel{\quad x_3 \quad + x_6} & \cancel{\geq 1} & (E) \quad \text{(i)} \\
 & \quad 1 = x_6 & \geq 1 & (F)
\end{array}
$$

$$x_j \in \{0,1\}, \; j \in \{1, \ldots, 6\}$$

(ii) After step (i), $x_5$-column dominates $x_3$-column $\Rightarrow$ set $x_3 = 0$.

**Def** For $\bar{u}, \bar{v} \in \{0,1\}^n$, $\bar{u}$ (strictly) dominates $\bar{v}$ if $u_i \geq v_i \;\; \forall i$, and $u_j > v_j$ for at least one $j$.

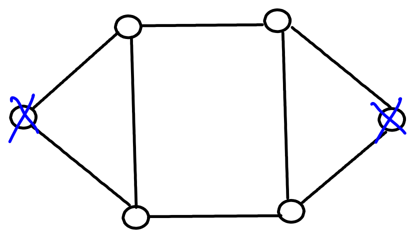
(iii) Now, row (B) dominates row (A) $\Rightarrow$ eliminate row (B).

Thus, all operations are reduced to matrix operations.

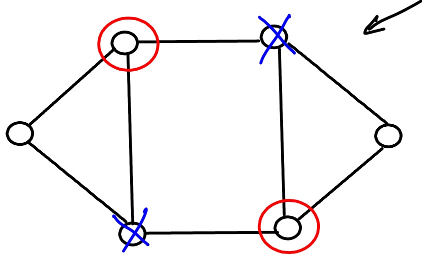
# Variants of Set Covering

**Variant 1:** 'o' = 'x' (facility and customer locations are same)

Consider the following instance:



Here is an optimal solution with size 2.



But this solution is better, as two customers are covered twice (indicated by ⊙).

This example motivates the following variant of the default set covering problem:

Objectives: (1) minimize total # facilities;

(2) among all optimal solutions to (1), pick one that maximizes # customers covered two or more times.

Define $y_i = \begin{cases} 1 & \text{if customer } i \text{ covered twice or more} \\ 0 & \text{o.w.} \end{cases}$

IP is **NOT**

$$\min \sum_{j=1}^{n} x_j - \sum_{i=1}^{m} \cancel{y_i}$$

$$\text{s.t.} \quad \sum_{j \in C_i} x_j - y_i \geq 1$$

$$x_j, y_i \in \{0,1\}$$

applies to the general set cover problem where 'x' and 'o' are distinct.

The objective function does not work quite correctly!

If $\bar{x}, \bar{y}$ is an optimal solution to this IP, we can find $\bar{x}', \bar{y}'$ with

$$\sum_{j \in C_i} x_j' = \sum_{j \in C_i} x_j + 1 \quad \text{and}$$

$$\sum_{i=1}^{m} y_i' \geq \sum_{i=1}^{m} y_i + 2.$$

Hence, we use the following objective function instead:

$$\min \quad \sum_{j=1}^{p} x_j - \varepsilon \underbrace{\sum_{i=1}^{m} y_i} \quad \text{where} \quad \varepsilon = \frac{1}{m+1}.$$

The contribution of the second sum term is hence $< 1$.

<u>Variant 2</u> Let $w_i$ = importance of covering customer $i$ ($w_i \geq 0$) and let $p_0$ be the maximum number of facilities (poles picked) ($p_0 \in \mathbb{Z}_{>0}$ represents a budget constraint).

Objective: Maximize sum of importance of covered customers.

Let $s_i = \begin{cases} 1 & \text{if customer } i \text{ is covered,} \\ 0 & \text{otherwise.} \end{cases}$ $i \in \{1, \dots, m\} = M.$

it is not required to cover all customers

IP:

$$\max \sum_{i=1}^{m} w_i s_i$$

$$\text{s.t.} \quad \sum_{j \in C_i} x_j \geq s_i \quad \forall i \in M$$

$$\sum_{j=1}^{p} x_j \leq p_0$$

$$x_j, s_i \in \{0, 1\}$$

Variant 3 (Existing and new facilities)

$P = \{1, 2, \ldots, p_{ex}, p_{ex}+1, \ldots, p_{ex}+p_{new}\}$   $p_{new} = \#$ new facilities
$p_{ex} = \#$ existing facilities

Objectives: (1) minimize total # facilities;
(2) among all optimal solutions to (1), pick one that minimizes # new facilities.

We want to use as many of the existing facilities first, before using new ones. We could also incorporate the costs for establishing the new facilities. It is not necessary to use all existing facilities — or, there could be a cost for using one.

$$\min \quad \sum_{j=1}^{p_{ex}} x_j + (1+\varepsilon) \sum_{j=p_{ex}+1}^{p_{ex}+p_{new}} x_j$$

$$\text{s.t.} \quad \sum_{j \in C_i} x_j \geq 1 \quad \forall\, i \in M$$

$$x_j \in \{0,1\},$$

where $0 < \varepsilon \leq \dfrac{1}{p_{new}+1}$

Need $0 < \varepsilon \leq \dfrac{1}{p_{new}+1}$, in the same way as in Variant 1, so that the contribution of the extra sum term is $< 1$.