

Facility Location

Given a set L of (possible) locations for placing facilities and a set D of customers together with cost functions $s : D \times L \rightarrow \mathbb{R}^+$ and $o : L \rightarrow \mathbb{R}^+$ find a set of facility locations F together with an assignment $\phi : D \rightarrow F$ of customers to open facilities such that

$$\sum_{f \in F} o(f) + \sum_c s(c, \phi(c))$$

is minimized.

In the **metric facility location** problem we have

$$s(c, f) \leq s(c, f') + s(c', f) + s(c', f') .$$

Integer Program

$$\begin{array}{ll} \min & \sum_{i \in F} f_i y_i + \sum_{i \in F} \sum_{j \in D} c_{ij} x_{ij} \\ \text{s.t.} & \forall j \in D \quad \sum_{i \in F} x_{ij} = 1 \\ & \forall i \in F, j \in D \quad x_{ij} \leq y_i \\ & \forall i \in F, j \in D \quad x_{ij} \in \{0, 1\} \\ & \forall i \in F \quad y_i \in \{0, 1\} \end{array}$$

As usual we get an LP by relaxing the integrality constraints.

Dual Linear Program

$$\begin{array}{ll} \max & \sum_{j \in D} v_j \\ \text{s.t.} & \forall i \in F \quad \sum_{j \in D} w_{ij} \leq f_i \\ & \forall i \in F, j \in D \quad v_j - w_{ij} \leq c_{ij} \\ & \forall i \in F, j \in D \quad w_{ij} \geq 0 \end{array}$$

Definition 9

Given an LP solution (x^*, y^*) we say that facility i neighbours client j if $x_{ij} > 0$. Let $N(j) = \{i \in F : x_{ij}^* > 0\}$.

Lemma 10

If (x^, y^*) is an optimal solution to the facility location LP and (v^*, w^*) is an optimal dual solution, then $x_{ij}^* > 0$ implies $c_{ij} \leq v_j^*$.*

Follows from slackness conditions.

Suppose we open set $S \subseteq F$ of facilities s.t. for all clients we have $S \cap N(j) \neq \emptyset$.

Then every client j has a facility i s.t. assignment cost for this client is at most $c_{ij} \leq v_j^*$.

Hence, the total assignment cost is

$$\sum_j c_{i_j j} \leq \sum_j v_j^* \leq \text{OPT} ,$$

where i_j is the facility that client j is assigned to.

Problem: Facility cost may be huge!

Suppose we can partition a subset $F' \subseteq F$ of facilities into neighbour sets of some clients. I.e.

$$F' = \bigcup_k N(j_k)$$

where j_1, j_2, \dots form a subset of the clients.

Now in each set $N(j_k)$ we open the **cheapest** facility. Call it f_{i_k} .

We have

$$f_{i_k} = f_{i_k} \sum_{i \in N(j_k)} x_{ij_k}^* \leq \sum_{i \in N(j_k)} f_i x_{ij_k}^* \leq \sum_{i \in N(j_k)} f_i y_i^* .$$

Summing over all k gives

$$\sum_k f_{i_k} \leq \sum_k \sum_{i \in N(j_k)} f_i y_i^* = \sum_{i \in F'} f_i y_i^* \leq \sum_{i \in F} f_i y_i^*$$

Facility cost is at most the facility cost in an optimum solution.

Problem: so far clients j_1, j_2, \dots have a neighboring facility.
What about the others?

Definition 11

Let $N^2(j)$ denote all neighboring **clients** of the neighboring facilities of client j .

Note that $N(j)$ is a set of facilities while $N^2(j)$ is a set of clients.

Algorithm 1 FacilityLocation

- 1: $C \leftarrow D$ // unassigned clients
- 2: $k \leftarrow 0$
- 3: **while** $C \neq \emptyset$ **do**
- 4: $k \leftarrow k + 1$
- 5: choose $j_k \in C$ that minimizes v_j^*
- 6: choose $i_k \in N(j_k)$ as cheapest facility
- 7: assign j_k and all unassigned clients in $N^2(j_k)$ to i_k
- 8: $C \leftarrow C - \{j_k\} - N^2(j_k)$

Facility cost of this algorithm is at most OPT because the sets $N(j_k)$ are disjoint.

Total assignment cost:

- ▶ Fix k ; set $j = j_k$ and $i = i_k$. We know that $c_{ij} \leq v_j^*$.
- ▶ Let $\ell \in N^2(j)$ and h (one of) its neighbour(s) in $N(j)$.

$$c_{i\ell} \leq c_{ij} + c_{hj} + c_{h\ell} \leq v_j^* + v_j^* + v_\ell^* \leq 3v_\ell^*$$

Summing this over all facilities gives that the total assignment cost is at most $3 \cdot \text{OPT}$. Hence, we get a 4-approximation.

In the above analysis we use the inequality

$$\sum_{i \in F} f_i y_i^* \leq \text{OPT} .$$

We know something stronger namely

$$\sum_{i \in F} f_i y_i^* + \sum_{i \in F} \sum_{j \in D} c_{ij} x_{ij}^* \leq \text{OPT} .$$

Observation:

- ▶ Suppose when choosing a client j_k , instead of opening the cheapest facility in its neighborhood we choose a random facility according to x_{ijk}^* .
- ▶ Then we incur connection cost

$$\sum_i c_{ijk} x_{ijk}^*$$

for client j_k . (In the previous algorithm we estimated this by v_{jk}^*).

- ▶ Define

$$C_j^* = \sum_i c_{ij} x_{ij}^*$$

to be the connection cost for client j .

What will our facility cost be?

We only try to open a facility once (when it is in neighborhood of some j_k). (recall that neighborhoods of different j'_k s are disjoint).

We open facility i with probability $x_{ij_k} \leq y_i$ (in case it is in some neighborhood; otw. we open it with probability zero).

Hence, the expected facility cost is at most

$$\sum_{i \in F} f_i y_i .$$

Algorithm 1 FacilityLocation

- 1: $C \leftarrow D$ // unassigned clients
- 2: $k \leftarrow 0$
- 3: **while** $C \neq \emptyset$ **do**
- 4: $k \leftarrow k + 1$
- 5: choose $j_k \in C$ that minimizes $v_j^* + C_j^*$
- 6: choose $i_k \in N(j_k)$ according to probability x_{ij_k} .
- 7: assign j_k and all unassigned clients in $N^2(j_k)$ to i_k
- 8: $C \leftarrow C - \{j_k\} - N^2(j_k)$

Total assignment cost:

- ▶ Fix k ; set $j = j_k$.
- ▶ Let $\ell \in N^2(j)$ and h (one of) its neighbour(s) in $N(j)$.
- ▶ If we assign a client ℓ to the same facility as i we pay at most

$$\sum_i c_{ij} x_{ijk}^* + c_{hj} + c_{h\ell} \leq C_j^* + v_j^* + v_\ell^* \leq C_\ell^* + 2v_\ell^*$$

Summing this over all clients gives that the total assignment cost is at most

$$\sum_j C_j^* + \sum_j 2v_j^* \leq \sum_j C_j^* + 2\text{OPT}$$

Hence, it is at most 2OPT plus the total assignment cost in an optimum solution.

Adding the facility cost gives a 3-approximation.