Can we do better?

Not if we compare ourselves to the value of an optimum LP-solution.

Definition 7 (Integrality Gap)

The integrality gap for an ILP is the worst-case ratio over all instances of the problem of the value of an optimal IP-solution to the value of an optimal solution to its linear programming relaxation.

Note that the integrality is less than one for maximization problems and larger than one for minimization problems (of course, equality is possible).

Note that an integrality gap only holds for one specific ILP formulation.

Lemma 8

Our ILP-formulation for the MAXSAT problem has integrality gap at most $\frac{3}{4}$.

$$\begin{array}{llrcl} \max & & \sum_j w_j z_j & & \\ \text{s.t.} & \forall j & \sum_{i\in P_j} y_i + \sum_{i\in N_j}(1-y_i) & \geq & z_j \\ & \forall i & y_i & \in & \{0,1\} \\ & \forall j & z_j & \leq & 1 \end{array}$$

Consider: $(x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_2)$

- any solution can satisfy at most 3 clauses
- we can set $y_1 = y_2 = 1/2$ in the LP; this allows to set $z_1 = z_2 = z_3 = z_4 = 1$
- hence, the LP has value 4.

Facility Location

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

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

Facility Location

Integer Program

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

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

Facility Location

Dual Linear Program

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

Facility Location

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 $c_{ij} \le v^*_j$.

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 $c_{ij} \le v^*_j$.

Hence, the total assignment cost is

$$\sum_j c_{i_j j} \le \sum_j v^*_j \le \mathrm{OPT} \enspace ,$$

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.

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

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

$$f_{i_k} = f_{i_k} \sum_{i \in N(j_k)} x^*_{ij_k} \le \sum_{i \in N(j_k)} f_i x^*_{ij_k} \le \sum_{i \in N(j_k)} f_i y^*_i \ .$$

Summing over all k gives

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

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

```
C ← D// unassigned clients
k ← 0
while C ≠ 0 do
    k ← k + 1
    choose j_k ∈ C that minimizes v*_j
    choose i_k ∈ N(j_k) as cheapest facility
    assign j_k and all unassigned clients in N^2(j_k) to i_k
    C ← 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 $c_{ij} \le v_j^*$.
- Let $\ell \in N^2(j)$ and h (one of) its neighbour(s) in $N(j)$.

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

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

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

We know something stronger namely

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

Observation:

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

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

  for client j_k. (In the previous algorithm we estimated this by $v_{j_k}^*$).
- 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 $x_{ij_k} \le y_i$ (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

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

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

$$\sum_i c_{ij} x^*_{ij_k} + c_{hj} + c_{h\ell} \leq C_j^* + v_j^* + v_\ell^* \leq C_\ell^* + 2v_\ell^*$$

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

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

Hence, it is at most $2\mathrm{OPT}$ plus the total assignment cost in an optimum solution.

Adding the facility cost gives a 3-approximation.

Lemma 12 (Chernoff Bounds)

Let $X_1, \ldots, X_n$ be n independent 0-1 random variables, not necessarily identically distributed. Then for $X = \sum_{i=1}^n X_i$ and $\mu = E[X]$, $L \leq \mu \leq U$, and $\delta > 0$

$$\Pr[X \geq (1+\delta)U] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^U ,$$

and

$$\Pr[X \leq (1-\delta)L] < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L ,$$

Lemma 13

For $0 \leq \delta \leq 1$ we have that

$$\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^U \leq e^{-U\delta^2/3}$$

and

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^L \leq e^{-L\delta^2/2}$$