Constant Approximation for Facility Location:

Facility Location: there is a set D of client (such as stores) that required to be serviced by a facility (F & D are just a set of vertices of the graph which is a metric)

- there is a set of locations F at which we may build a facility (such as a warehouse), where the cost of building at location i is f_i.
- If a client at location j is assigned to a facility at location i, a cost of C_{ij} is incurred that is proportional to the distance between i and j. (we assume the distances between locations are non-negative, symmetric and satisfy triangle Ineq. thus it is a METRIC)
- The objective is to determine a set of locations at which to Open facilities so as to minimize the total facility and assignment costs. (we can have Capacitated or UnCapacitated versions where capacities are on the facilities).

Thm: metric uncapacitated facility location has constant factor approximation [FROM Shmoys, Tardos, Aardal [STOC 97]]

IP & LP for the problem: Minimize $\sum_{i\in F} f_i y_i + \sum_{i\in F}\sum_{j\in D} d_j C_{ij} x_{ij}$ (all demands are one for simplicity)

such that

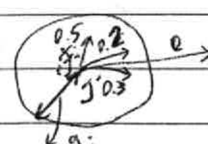
$\sum_{i\in F} x_{ij} = 1$ for each $j\in D$ (C_{ij} are metric i.e. $C_{ij} \le C_{ik} + C_{kj}$)

$x_{ij} \le y_i$ for each $i\in F, j\in D$

$x_{ij} \geq 0 \leftarrow \cancel{x_{ij} \in \{0,1\}}$ for each $i\in F, j\in D$

$y_i \geq 0 \leftarrow \cancel{y_i \in \{0,1\}}$ for each $i\in F$

- Given g_j, for each $j\in D$, we say a feasible solution (x, y) is g-close if it satisfies $x_{ij} > 0$ then $C_{ij} \le g_j$



- For a client $j\in D$, sort the facilities according to C_{ij} i.e., $C_{\pi(1)j} \le C_{\pi(2)j} \le \cdots \le C_{\pi(m)j}$. The α-point $C_j(\alpha) = C_{\pi(i^*)j}$ where $i^* = \min\{i' : \sum_{i=1}^{i'} x_{\pi(i)j} \geq \alpha\}$ (we assign α-fraction of the demand, we determine α later, $\alpha = \frac{1}{4}$ indeed) let $S = \{i : c_{ij} \geq C_j(\alpha)\}$ then by the definition of $C_j(\alpha)$, $\sum_{i \in S} x_{ij} \geq 1-\alpha$ (i^* was minimal)

Thus $\sum_{i \in F} c_{ij} x_{ij} \geq \sum_{i \in S} c_{ij} x_{ij} \geq (1-\alpha) \underbrace{C_j(\alpha)}_{\text{because of sort}}$ and thus $C_j(\alpha) \leq \frac{1}{1-\alpha} \sum_{i \in F} c_{ij} x_{ij}$

($S \subseteq F$)

- Thus $C_j(\alpha)$ is intuitively the average connection cost.

Lm 1: For α in $(0,1)$, we can charge a solution (x,y), we can find a $C(\alpha)$-close feasible solution $(\bar{x}, \bar{y})$ in polynomial time such that $\sum_{i \in F} f_i \bar{y}_i \leq \frac{1}{\alpha} \sum_{i \in F} f_i y_i$

Pf: for each $j \in D$, let $\alpha_j = \sum_{i \in F : c_{ij} \leq C_j(\alpha)} x_{ij}$. clearly $\alpha_j \geq \alpha$

- we set $\bar{x}_{ij} = \begin{cases} x_{ij}/\alpha_j & \text{if } c_{ij} \leq C_j(\alpha) \\ 0 & \text{otherwise} \end{cases}$ (by definition we are $C_j(\alpha)$-close)

- for each $i \in F$ we set $\bar{y}_i = \min\{1, \frac{y_i}{\alpha}\}$ (since $\bar{y}_i \leq \frac{1}{\alpha} y_i$, the facility cost blows up by $\frac{1}{\alpha}$)

feasibility because $\alpha_j \geq \alpha$ and $\sum_{i \in F} \bar{x}_{ij} = 1$ for all $j \in D$ ✓

So far we lost a factor α in facility costs. Now we show

Lm 2: given a feasible fractional g-close solution $(\bar{x}, \bar{y})$, we can find a feasible INTEGER $3g$-close solution $(\hat{x}, \hat{y})$ such that $\sum_{i \in F} f_i \hat{y}_i \leq \sum_{i \in F} f_i \bar{y}_i$.

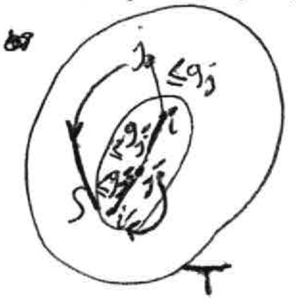
Pf: the algorithm iteratively converts this solution into a $3g$-close integer (not integer at the beginning) solution $(\hat{x}, \hat{y})$. without blowing up the facility cost, and without violating the feasibility. (initially $(\hat{x}, \hat{y}) = (\bar{x}, \bar{y})$)

let $\hat{F}$ be the set of partially opened facility, i.e. $\hat{F} = \{i \in F, 0 < \hat{y}_i < 1\}$.

let $\hat{D}$ be the set of clients assigned only to $\hat{F}$, i.e. $\hat{x}_{ij} > 0 \Rightarrow i \in \hat{F}$.

(essentially $\hat{D} = D$ initially except maybe those assigned fully to some facility. (we could sort D by g_j also)

In each iteration, pick $j' \in \hat{D}$ with smallest $g_{j'}$. Let $S = \{i \in \hat{F} : \hat{x}_{ij'} > 0\}$.



We assign j' to $i' \in S$ with smallest $f_{i'}$, set $y_{i'} = 1$, $y_{i \in S \setminus \{i'\}} = 0$.

let $T = \{j \in D : \exists i \in S \text{ such that } \hat{x}_{ij} > 0\}$

we assign each $j \in T$ to the facility opened at i', i.e. $\hat{x}_{i'j} = 1$ and $\hat{x}_{ij} = 0$ for $i \neq i'$

when $\hat{D}$ becomes empty, for each location $j \in D$, there exists i such that $\hat{x}_{ij} > 0$ and $\hat{y}_i = 1$, and so j can be assigned to i', i.e. $\hat{x}_{i'j} = 1$ and $\hat{x}_{ij} = 0$ for $i \neq i'$

(by the definition of $\hat{D}$, $\hat{F}$ this happens)

Now, we keep the folloing properties:

(P1) $(\hat{x},\hat{y})$ is a feasible fractional solution (we only assign clients to open facilities and if we set $y_i=0$, we also set all $x_{ij}=0$)

(P2) $\sum_{i\in F} f_i\hat{y}_i \leq \sum_{i\in F} f_i\bar{y}_i$

(P3) $\hat{x}_{ij}>0$ and $i\in\hat{F} \Rightarrow c_{ij}\leq g_j$ (we do not set a new variable x_{ij} in $(0,1)$ nor a new facility to $\hat{F}$

(P4) $\hat{x}_{ij}>0$ and $i\notin\hat{F} \Rightarrow c_{ij}\leq 3g_j$

(note that at the beginning, we have these properties and at the end, we prove the lemma)

For P4: Say we set $\hat{x}_{i'j}=0$ during the iteration (see the figure)

$$c_{i'j}\leq \underbrace{c_{i'j'}+c_{j'i}+c_{ij}}_{\substack{\text{Triangle Ineq \&}\\ c_{ij}\leq g_j\\ c_{ij'}\leq g_{j'}\\ c_{i'j'}\leq g_{j'}\\ \text{(see Figure)}}} \leq 2g_{j'}+g_j \underbrace{\leq}_{\substack{\text{since } g_{j'}\\ \text{was minimum}}} 3g_j$$

For P2: $f_{i'}=\min_{i\in S} f_i \leq \sum_{i\in S} f_i\hat{x}_{ij}$ since $\sum_{i\in S}\hat{x}_{ij}=1$

Since $\hat{x}_{ij}\leq\hat{y}_i$, $f_{i'}\leq\sum_{i\in S} f_i\hat{y}_i$ which says the facility cost $\hat{y}$ never increases throughout the algorithm and we have P2.

Finally when $\hat{D}$ is empty four properties hold (e.g. P_4 holds with g_j instead of $3g_j$) ✓

Now we start with (x,y), by lm 1, we obtain $(\bar{x},\bar{y})$ and by lm 2 $(\hat{x},\hat{y})$ which in integer.

The facility cost is $\sum_{i\in F} f_i\hat{y}_i \leq \sum_{i\in F} f_i\bar{y}_i \leq \frac{1}{\alpha}\sum_{i\in F} f_i y_i$. (by lm 1)

On the other hand, for each $j\in D$, it's unit assignment cost is $3g_j = 3\,c_j(\alpha) \leq \frac{3}{1-\alpha}\sum_{i\in F} c_{ij}x_{ij}$

Thus, the cost of integer solution $(\hat{x},\hat{y})$ is

$$\sum_{i\in F} f_i\hat{y}_i + \sum_{i\in F}\sum_{j\in D} d_j c_{ij}\hat{x}_{ij} \leq \frac{1}{\alpha}\sum_{i\in F} f_i y_i + 3\sum_{j\in D} d_j c_j(\alpha) \leq \max\left\{\frac{1}{\alpha},\frac{3}{1-\alpha}\right\}\left(\sum_{i\in F} f_i y_i + \sum\sum d_j c_{ij} x_{ij}\right)$$

that if we set $\frac{1}{\alpha}=\frac{3}{1-\alpha} \Rightarrow 1-\alpha=3\alpha \Rightarrow \alpha=\frac{1}{4}$, we obtain 4-approx ✓

By choosing α at random, we obtain 3.16 approximation.

The best current app. factor is 1.5 due to Byrka. There is a lowerbound 1.463 unless $NP\subseteq DTIME(n^{\log\log n})$. so almost tight. Also the capacitated one has constant approx.

now 1.48 due to shi

Connected Facility Location:

Def: given an undirected graph $G=(V,E)$ with non-negative edge costs c_e, a set $D \subseteq V$ of demands and a parameter $M \geq 1$. The objective is to identify a subset F of the vertices V as open facility and to build a steiner tree T connecting F to minimize
(no facility cost in this version)

$$\underbrace{\sum_{j \in D} d_j \cdot l(i(j), j)}_{\text{connection cost}} + \underbrace{M\, C(T)}_{\text{steiner cost}}$$

where $i(j)$ is the closest open facility to demand j, l is the shortest path distance (w.r.t. edge lengths c_e), d_j is the weight of demand j and $C(T)$ is the cost of the edges in the steiner tree T. (not a generalization of facility location) not obviouse

Note that it is more general than steiner tree (say $M=1$ (we buy each edge instead of rent))

We have the concept of Renting (connection cost C) and buying (steiner cost S) costs.
(this problem is also called single-sink, rent-or-buy network design problem)

Suppose we have root $r \in V$ which belongs to the tree always (w.l.o.g since we can try all)

we give a simple 3.39 ($=2+$ best steiner tree app. factor) randomized due to (Gupta, Kumar, Roughgarden STOC'03)

3.39 is the current best factor.

Simple ALG: 1– Mark each demand $j \in D$ with prob $\frac{1}{M}$ and let $D' \subseteq D$ be the set of Marked vertices

2– Construct a ρ_{ST} – app. steiner tree on $F = D' \cup \{r\}$ and buy the edges of tree
best factor

3– Assign each demand to its closest facility in F.

Thm: The algorithm Simple ALG is a $(2+\rho_{ST})$-approximation algorithm for connected Facility location.

Proof follows from the two lemmas below which bound the expected steiner cost and the expected connection cost separately.

Lm 1: The expected steiner cost is at most $\rho_{ST} \cdot Z^*$ (where $Z^* = C^* + S^*$ is the optimum)

Pf: It suffices to show that the expected cost of a min-cost steiner tree on the (random) set of facilities is at most Z^*. We use T^*, the steiner tree on F^* in OPT. We define the steiner tree T on F as the union of the edges of T^* and the edges on shortest j–$i^*(j)$ paths for all $j \in F \setminus \{r\} \subseteq D$, where j is assigned to $i^*(j)$ in OPT.

The cost of $T^* \subseteq T$ is deterministically S^*. For a demand $j \in D$, the cost incurred for buying the shortest $j - i^*(j)$ path is $M \cdot l(j, i^*(j))$ with prob. $\frac{1}{M}$ (if $j \in F$) and zero otherwise (if $j \notin F$). In the worst case, all of the bought $j - i^*(j)$ shortest paths are edge-disjoint; by linearity of expectation the lemma follows:

$$E[c(T)] \leq S^* + \sum_{j \in D} \left(\frac{1}{M}\right) M\, l(j, i^*(j)) = S^* + C^* = Z^*. \quad \square$$

Lm 2: The expected cost of Step 3 (connection cost) is at most $2Z^*$.

Pf: the expectation of connection costs is independent of the particular steiner tree in step 2. Thus w.l.o.g., we assume the steiner tree is given by the minimum spanning tree (in the graph of shortest-path distances) on D for the analysis.

Now using MST, we give a different view of the algorithm:
Instead of fliping coins all at once, the algorithm flips the coin in some order. With prob $\frac{1}{M}$ mark it and add it to F and joing it to the preexiting steiner tree, OR connect to some previously marked vertex in F.

To decide the order on the vertices, we maintain two sets. At step t, A_t is the set previously considered by the algorithm and $B_t \subseteq A_t$ those marked. Initially $A_1 = B_1 = \{r\}$. In step t, we pick the vertex $v_t \in V - A_t$ that is closest to B_t and flip the coin for it. With prob $\frac{1}{M}$, we define A_{t+1}, B_{t+1} by adding v_t to both the sets A_t, B_t and we update our steiner tree by buying the shortest path from v_t to its nearest neighbor in B_t. With prob $1 - \frac{1}{M}$ we set $A_{t+1} = A_t \cup \{v_t\}$ and $B_{t+1} = B_t$ and assign v_t to its nearest neighbor in B_t.

Indeed our algorithm for steiner tree T is PRIM'S MST in shortest-path metric in F. This new randomized process does not increase the connection cost. Let random variable X_t denote the cost of assigning v_t to its nearest neighbor in B_t minus the cost of adding v_t to B_t and connecting it to the existing steiner tree in step t. Let $X = \sum_i X_i$ denote the connection cost − the steiner cost. The expected value of X_t conditioning on the first $t-1$ coins so that v_t & B_t are deterministically known is $(1 - \frac{1}{M})\, l(v_t, B_t) - \frac{1}{M} M\, (l(v_t, B_t)) \leq 0$. This holds for any outcome of the first $t-1$ coins and thus unconditionally $E[X_t] \leq 0$ for all t. By linearity of expectation $E[X] \leq 0$. Thus the expected connection cost is at most the expected cost of MST on F which is $2Z^*$ by Lm 1, since MST is 2-app for steiner tree. $\square$

It can be generalized to general facility cost to get higher constant factor.