College of Computer & Information Science — Fall 2007
Northeastern University — 21 September 2007
CS G399: Algorithmic Power Tools I — Scribe: San Tan

**Lecture Outline:**

- Facility location
  - Variants & applications
  - Rounding algorithm
  - combinational "greedy" algorithm

This lecture introduces facility location problems, relates them to set cover, and considers greedy and randomized rounding algorithms.

# 1 Facility location

**Problem 1.** Given $n$ demand locations (or clients) $V$ and $m$ facility locations $\widetilde{F}$, define

$$c : \text{cost /distance between locations, } (V \bigcup \widetilde{F}) \times (V \bigcup \widetilde{F}) \rightarrow Q^+,$$

$$f_j : \text{cost of opening facility at } j \in \widetilde{F},$$

$$d_i : \text{ demand at client } i \in V.$$

The goal is to determine locations $F \subseteq \widetilde{F}$ where to open facilities and $\sigma : V \rightarrow F$ to minimize $\sum_{j \in F} f_d + \sum_{i \in V} c_{i\sigma_i} d_i$.

There are different variants of problem 1

- Metric or non-metric: metric means the cost function of problem 1 satisfies triangle inequality,

$$c_{ij} = c_{ji};$$

$$c_{ij} \leq c_{ik} + c_{kj};$$

$$c_{ii} = 0.$$

- Capacity of facility: There is a bound on the total demand (or number of clients) that a facility can serve
- There is a bound on the number of facilities (or cost of facilities) that can be opened
- Facility costs

We will look at

- uncapacited metric facility location (UFL)

- uncapacited metric $k$-median

Example applications of facility locations include

- Hub and spoke scheduling;
- Locating concentrators in a routing network;
- Locating servers in a content-delivery network.

**Theorem 1.** *UFL is NP-Hard.*

*Proof.* Reduction from set cover. Use notations defined in set cover problem and problem 1. Construct a UFL problem as follows from set cover.

- $\mathcal{U} \to V$,
- $S \to \widetilde{F}$,
- $c_{ij} = \begin{cases} 1 & \text{if } e_i \in s_j; \\ 3 & \text{otherwise;} \end{cases}$
- $f_j =$ constant number $M \leq 2$,
- $d_i = 1$.

We now argue that there exists a set cover of cost $C$ if and only if there exists a facility location solution of cost at most $C \cdot M + n$.

One direction is trivial. If there is a set cover of cost $C$, then the sets form the facilities, with a total cost of $C \cdot M + n$.

We now consider the other direction. Suppose we have a facility location solution of cost $C \cdot M + n$.

- case 1: The service cost is equal to $n$. The selected facilities yield the collection of sets of cost $C$ for the set cover problem.
- case 2: The service cost is greater than $n$. For each client $i$ that is paying service cost at least 3, add a facility $j$ such that $j \notin$ {selected facilities} and $c_{ij} = 1$. Since $M + c_{ij} \leq 3$, the total cost doesn't increase. Then go to case 1.

□

A possible approximation algorithm for facility location is to reduce it to set cover problem, then solve it using LP-rounding. For each facility, $(2^{|V|} - 1)$ sets are generated, with cost of $c = x_{js}, s \subseteq V$, in which $x_{js} = f_j + \sum_{i \in s} c_{ij}$. However, this approach has exponential-complexity. It's not hard to reduce the solution to polynomial time by identifying only those sets that could possibly be in an optimal solution.

LP for UFL. For any $j \in \widetilde{F}$, define

$$y_j = \begin{cases} 1 & \text{if facility is opened at } j, \\ 0 & \text{otherwise.} \end{cases}$$

Obviously clients should be assigned to nearest facility. For $i \in V$, $\sigma(i) =$ nearest $j$ such that $y_j = 1$. Define

$$x_{ij} = \begin{cases} 1 & \text{if } i \text{ is served by } j, \\ 0 & \text{otherwise.} \end{cases}$$

We have the LP form of UFL problem.

$$\begin{aligned} min \quad & \sum_j y_j f_j + \sum_i x_{ij} c_{ij} d_i \\ s.t. \quad & \sum_j x_{ij} \geq 1, \forall i \\ & y_j - x_{ij} \geq 0; \end{aligned}$$

Suppose that LP gives solution $\{x^*\}$ and $\{y^*\}$.

- open facility at $j$ with probability $y_j^*$. so
$$E[\text{facility cost so far}] = \sum_j y_j^* f_j.$$
- if distance of $i$ to nearest open facility is at most $r_i = 2\sum_j x_{ij}^* c_{ij}$, then assign $i$ to nearest open facility; otherwise don't assign.

**Lemma 1.** $\forall i \in V, p_i[i \textit{ is assigned}] \geq 1 - \frac{1}{\sqrt{e}}$.

*Proof.* Supose $B_L$ is a ball centering $i$ with radius $2\sum_j x_{ij}^* c_{ij}$. We have

$$\begin{aligned} \sum_{j \in B_L} y_j^* &\geq \sum_{j \in B_L} x_{ij}^* \\ &\geq \frac{1}{2}, \end{aligned}$$

because

$$\begin{aligned} \sum_{j \notin B_L} x_{ij}^* &= \frac{r_i \sum_{j \notin B_L} x_{ij}^*}{r_i} \\ &\leq \frac{\sum_j x_{ij}^* c_{ij}}{r_i} \\ &= \frac{\frac{1}{2} r_i}{r_i} \\ &= \frac{1}{2}. \end{aligned}$$

So

$$\begin{aligned} p_i[i \text{ is assigned}] &= 1 - \prod_{j \in B_L} (1 - y_j^*) \\ &\geq 1 - (1 - \frac{1}{2^{|\widetilde{F}|}})^{|\widetilde{F}|} \\ &\geq 1 - \frac{1}{\sqrt{e}}. \end{aligned}$$

□

Then, if we repeat the above round step $t$ times and combine the rounding result in a way as we did in set cover, we have

$$\forall \text{ client } i, \Pr[\text{client } i \text{ is not assigned}] \leq (\frac{1}{\sqrt{e}})^t.$$

Set $t = log_{\sqrt{e}}(4n) = O(logn)$, and obtain

$$\Pr[\text{client } i \text{ is not assigned}] \leq \frac{1}{4n};$$

$$\Pr[\text{some client is not assigned}] \leq \frac{1}{4};$$

$$E[\text{cost of solution}] \leq 2t \cdot OPT_{LP}.$$

By Markov's inequality, $\Pr[\text{cost } \geq 8t \cdot OPT_{LP}] \leq \frac{1}{4}$.

Putting these together with $t = log_{\sqrt{e}}(4n) = O(logn)$, we obtain that with probability at least 1/2, we have a feasible facility location solution that has cost $O(logn)OPT_{LP}$. As we did for the set cover problem, we can improve our success probability by repeating the above process until we have a solution with the desired approximation ratio.