Facility Location Problem & A New Greedy Approach by

Karmal Jain [1], Mohammad Mahdian [2], Amin Saberi [3]

Microsoft Research.

Laboratory for Computer Science M.I.T.

College of Computing Georgia Tech.

Outline

Introduction...

Given:

- A set $\mathcal{F}$ of n_f *facilities*
- A set $\mathcal{C}$ of n_c *cities*
- For every facility $i \in \mathcal{F}$, f_i is the given opening cost
- For every facility $i \in \mathcal{F}$ and every city $j \in \mathcal{C}$ the connection cost c_{ij} is given

Objective:

- To open a subset of facilities in $\mathcal{F}$ and connect each city to at least one facility so that the total cost is minimized.
- Here we consider the metric variant of the problem.

Given:

- A set $\mathcal{F}$ of n_f *facilities*
- A set $\mathcal{C}$ of n_c *cities*
- For every facility $i \in \mathcal{F}$, f_i is the given opening cost
- For every facility $i \in \mathcal{F}$ and every city $j \in \mathcal{C}$ the connection cost c_{ij} is given

Objective:

- To open a subset of facilities in $\mathcal{F}$ and connect each city to at least one facility so that the total cost is minimized.
- Here we consider the metric variant of the problem.

Given:

- A set $\mathcal{F}$ of n_f *facilities*
- A set $\mathcal{C}$ of n_c *cities*
- For every facility $i \in \mathcal{F}$, f_i is the given opening cost
- For every facility $i \in \mathcal{F}$ and every city $j \in \mathcal{C}$ the connection cost c_{ij} is given

Objective:

- To open a subset of facilities in $\mathcal{F}$ and connect each city to at least one facility so that the total cost is minimized.
- Here we consider the metric variant of the problem.

Given:

- A set $\mathcal{F}$ of n_f *facilities*
- A set $\mathcal{C}$ of n_c *cities*
- For every facility $i \in \mathcal{F}$, f_i is the given opening cost
- For every facility $i \in \mathcal{F}$ and every city $j \in \mathcal{C}$ the connection cost c_{ij} is given

Objective:

- To open a subset of facilities in $\mathcal{F}$ and connect each city to at least one facility so that the total cost is minimized.
- Here we consider the metric variant of the problem.

Given:

- A set $\mathcal{F}$ of n_f *facilities*
- A set $\mathcal{C}$ of n_c *cities*
- For every facility $i \in \mathcal{F}$, f_i is the given opening cost
- For every facility $i \in \mathcal{F}$ and every city $j \in \mathcal{C}$ the connection cost c_{ij} is given

Objective:

- To open a subset of facilities in $\mathcal{F}$ and connect each city to at least one facility so that the total cost is minimized.
- Here we consider the metric variant of the problem.

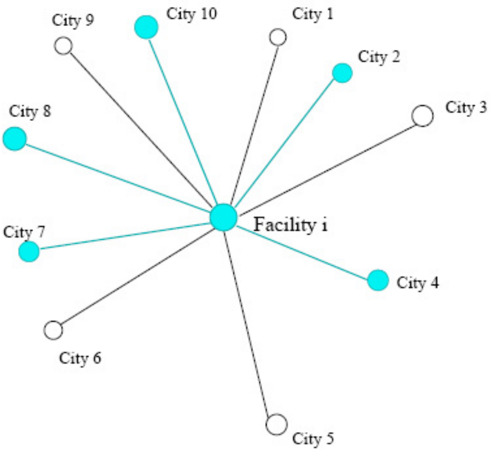
Given:

- A set $\mathcal{F}$ of n_f *facilities*
- A set $\mathcal{C}$ of n_c *cities*
- For every facility $i \in \mathcal{F}$, f_i is the given opening cost
- For every facility $i \in \mathcal{F}$ and every city $j \in \mathcal{C}$ the connection cost c_{ij} is given

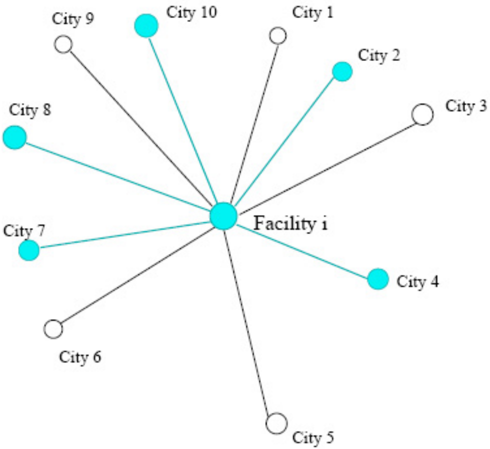
Objective:

- To open a subset of facilities in $\mathcal{F}$ and connect each city to at least one facility so that the total cost is minimized.
- Here we consider the metric variant of the problem.

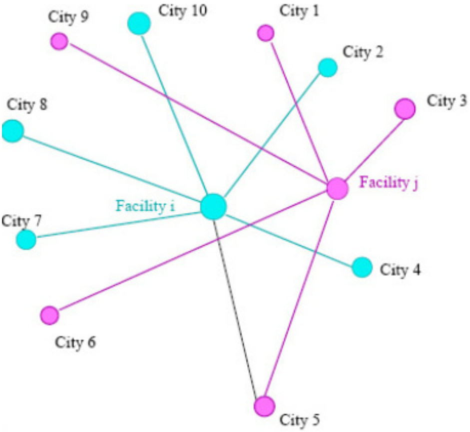
- *STAR* is a facility and several cities.



- *COST of a STAR* is the sum of the opening cost of the facility and the connection costs between the facility and the cities.



- Then the *Facility Location Problem* is a picking a minimum cost set of stars such that each city is in at least one star.



LP-formulation

- Let $\mathcal{S}$ be the set of all stars
- Let x_S be 0 or 1 depending on picking or not the star $S \in \mathcal{S}$
- Let c_S denote the cost of the star $S \in \mathcal{S}$

then the problem is:

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c_S x_S$$

$$\text{subject to} \quad \forall j \in \mathcal{C}: \quad \sum_{S : j \in S} x_S \geq 1$$

$$\text{where} \quad \forall S \in \mathcal{S}: \quad x_S \in \{0, 1\}$$

LP-formulation

- Let $\mathcal{S}$ be the set of all stars
- Let x_S be 0 or 1 depending on picking or not the star $S \in \mathcal{S}$
- Let c_S denote the cost of the star $S \in \mathcal{S}$

then the problem is:

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c_S x_S$$

$$\text{subject to} \quad \forall j \in \mathcal{C}: \quad \sum_{S : j \in S} x_S \geq 1$$

$$\text{where} \quad \forall S \in \mathcal{S}: \quad x_S \in \{0,1\}$$

LP-formulation

- Let $\mathcal{S}$ be the set of all stars
- Let x_S be 0 or 1 depending on picking or not the star $S \in \mathcal{S}$
- Let c_S denote the cost of the star $S \in \mathcal{S}$

then the problem is:

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c_S x_S$$

$$\text{subject to} \quad \forall j \in \mathcal{C}: \quad \sum_{S : j \in S} x_S \geq 1$$

$$\text{where} \quad \forall S \in \mathcal{S}: \quad x_S \in \{0, 1\}$$

LP-formulation

- Let $\mathcal{S}$ be the set of all stars
- Let x_S be 0 or 1 depending on picking or not the star $S \in \mathcal{S}$
- Let c_S denote the cost of the star $S \in \mathcal{S}$

then the problem is:

$$\text{minimize} \quad \sum_{S \in \mathcal{S}} c_S x_S$$

$$\text{subject to} \qquad \forall j \in \mathcal{C} : \quad \sum_{S : j \in S} x_S \geq 1$$

$$\text{where} \quad \forall S \in \mathcal{S} : \quad x_S \in \{0, 1\}$$

LP-relaxation

The LP-relaxation of the problem is:

$$\begin{aligned} \text{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ \text{subject to} \quad & \forall j \in \mathcal{C}: \quad \sum_{S : j \in S} x_S \geq 1 \\ \text{where} \quad & \forall S \in \mathcal{S}: \quad x_S \geq 0 \end{aligned}$$

And the dual program is:

$$\begin{aligned} \text{maximize} \quad & \sum_{j \in \mathcal{C}} \alpha_j \\ \text{subject to} \quad & \forall S \in \mathcal{S}: \quad \sum_{j \in S \cap \mathcal{C}} \alpha_j \leq c_S \\ \text{where} \quad & \forall j \in \mathcal{C}: \quad \alpha_j \geq 0 \end{aligned}$$

LP-relaxation

The LP-relaxation of the problem is:

$$\begin{aligned} \text{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ \text{subject to} \quad \forall j \in \mathcal{C} : \quad & \sum_{S : j \in S} x_S \geq 1 \\ \text{where} \quad \forall S \in \mathcal{S} : \quad & x_S \geq 0 \end{aligned}$$

And the dual program is:

$$\begin{aligned} \text{maximize} \quad & \sum_{j \in \mathcal{C}} \alpha_j \\ \text{subject to} \quad \forall S \in \mathcal{S} : \quad & \sum_{j \in S \cap \mathcal{C}} \alpha_j \leq c_S \\ \text{where} \quad \forall j \in \mathcal{C} : \quad & \alpha_j \geq 0 \end{aligned}$$

Approximation factor γ

Suppose that we have an algorithm that

- Finds solution of cost T and
- The values $\alpha_j, \quad j \in \mathcal{C}$ so that
  - $\sum_{j\in S} \alpha_j = T$ and
  - $\sum_{j\in S\cap\mathcal{C}} \alpha_j \leq \gamma c_S, \quad \gamma \geq 1$ fixed number
- Then the approximation ration is at most γ

Algorithm

- Denote by B_j the budget of the city j.
- Start at moment $t = 0$ and at every next moment increment the time and renew the unconnected cities budget setting $B_j := t$.
- At each moment every city offers some money from its budget to each unopened facility i:
  - If j is unconnected then its offer is $max(B_j - c_{ij}, 0)$
  - If j is already connected to some facility i' then its offer to the facility i is $max(c_{i'j} - c_{ij}, 0)$

Algorithm

- While there is an unconnected city increase the time and renew the unconnected cities budget, until one of the following occurs.
  - For some unopened facility i the total offer it receives from cities is equal to the cost of opening. Then we **open the facility** i and for all cities j that have non-zero offer to i we **connect** j **to** i.
  - For some opened facility i and unconnected city j, the budget of j is equal to the connection cost c_{ij} then we **connect** j **to** i

Algorithm

- At the end, when all cities are connected, set $\alpha_j = B_j, \forall j \in \mathcal{C}$
- This value is the time when j was connected for the first time.
- At any time: The budget of each connected city is equal to the total connection cost plus its total contribution toward open facilities.

Lemma

The total cost of the solution found by the algorithm is equal to the sum of α's.

Analysis of the Algorithm

- Consider a star S
  - of a facility f(with opening cost f)
  - and k cities: $1 \cdots k$
  - d_i-connection cost between f and these cities
  - α_i-the share of the expenses for these cities
  - Suppose $\alpha_1 \leq \alpha_2 \leq \alpha_3 \leq \cdots \leq \alpha_k$
- We imagine the algorithm at time $t = \alpha_i - \epsilon$ - just a moment before city i connects for the first time.
  - At this time all the cities $j \in 1 \cdots i-1$ are connected to some facility.
  - Let $r_{j,i}$ be the connection cost between this facility and city j
  - If some city $j \in 1 \cdots i-1$ is not connected then for this $r_{j,i} = \alpha_j$ and $\alpha_j = \alpha_i$
- We'll use this variables to bound the ratio $\frac{\sum_{i \in S \cup C} \alpha_i}{c_S}$

Analysis of the Algorithm

- From the moment a city gets connected it can not cancel its contribution to a facility so it can never connect to a city with a higher connection cost: $r_{j,j+1} \geq r_{j,j+2} \geq \cdots \geq r_{j,k}$
- At the moment $t = \alpha_i - \epsilon$ the amount city j offers to facility f is

$$max(r_{j,i} - d_j, 0) \quad if \quad j < i, \quad and$$

$$max(t - d_j, 0) \quad if \quad j \geq i$$

- By the algorithm the total offer to a facility can never be grater than its opening cost, so

$$\sum_{j=1}^{i-1} max(r_{j,i} - d_j, 0) + \sum_{j=i}^{k} max(t - d_j, 0) \leq f$$

Analysis of the Algorithm

- Consider cities i, j with $j < i$ at the moment $t = \alpha_i - \epsilon$
- Let $f\prime$ be the facility the city j is connected at the moment t
- From triangle inequality, $c_{f\prime i}$ is the connection cost between $f\prime$ and i then $c_{f\prime i} \leq r_{j,i} + d_i + d_j$
- $c_{f\prime i} \geq t$ because otherwise the algorithm could have connected the city i to the facility $f\prime$ at the time earlier than t-contradiction.
- So: $\alpha_i \leq r_{j,i} + d_i + d_j$

Analysis of the Algorithm

- Taking all the above inequalities together we have the following LP problem for each star S of k cities:

$$\text{maximize} \quad \frac{\sum_{i \in S \cup C} \alpha_i}{f + \sum_{i=1}^{k} d_i}$$

$$\begin{aligned}
\text{subject to} \quad & \forall 1 \leq i < k : \alpha_i \leq \alpha_{i+1} \\
& \forall 1 \leq j < i < k : r_{j,i} \geq r_{j,i+1} \\
& \forall 1 \leq j < i < k : \alpha_i \leq r_{j,i} + d_i + d_j \\
& \forall 1 \leq i < k : \\
& \sum_{j=1}^{i-1} max(r_{j,i} - d_j, 0) + \sum_{j=i}^{k} max(t - d_j, 0) \leq f \\
& \forall 1 \leq j < i < k : \alpha_i, d_i, f, r_{j,i} \geq 0
\end{aligned}$$

Analysis of the Algorithm

Lemma

If z_k denotes the solution of the factor-revealing LP, then for every star S consisting of a facility and k cities: $\sum_{i \in S \cap \mathcal{C}} \alpha_i \leq z_k c_S$

Lemma

Let z_k be the solution of the factor-revealing LP, and $\gamma := \sup_k \{z_k\}$. Then the described Algorithm solves the metric facility problem with an approximation factor of γ.

Comments on Solving the factor-revealing LP

- Solve factor-revealing LP for small values of k
- It seams that $\gamma \approx 1.6$
- In order to prove an upper bound on γ, the solution to the dual of the factor-revealing problem have to be presented.
- Doing this for the small values of k the upper bound 1.61 was found

k	$\max_{i<k} z_i$
10	1.54147
20	1.57084
50	1.58839
100	1.59425
200	1.59721
300	1.59819
400	1.59868
500	1.59898

Comments on Solving the factor-revealing LP

Theorem

The described Algorithm solves the facility location problem in time $O(n^3)$, where $n = max(n_f, n_c)$.

k-median problem

Differs from the facility location in two respects

- There is no cost for opening facility
- There is an upper bound k on the number of facilities that can be opened

Jain and Vazirani: Reduction the k-median to the facility location.

Definition: Suppose $\mathcal{A}$ is an approx. alg. for the FLP. Consider an instance $\mathcal{I}$ of the problem with optimum cost OPT, and let F and C be the facility and connection costs of the solution found by $\mathcal{A}$. Alg. $\mathcal{A}$ is called LMP α-approximation if for every $\mathcal{I}$, $C \leq \alpha(OPT - F)$.

Theorem

An LMP α-approximation algorithm for the facility location problem gives a 2α-approximation algorithm for the metric k-median problem.

k-median problem

Differs from the facility location in two respects

- There is no cost for opening facility
- There is an upper bound k on the number of facilities that can be opened

Jain and Vazirani: Reduction the k-median to the facility location.

Definition: Suppose $\mathcal{A}$ is an approx. alg. for the FLP. Consider an instance $\mathcal{I}$ of the problem with optimum cost OPT, and let F and C be the facility and connection costs of the solution found by $\mathcal{A}$. Alg. $\mathcal{A}$ is called LMP α-approximation if for every $\mathcal{I}$, $C \leq \alpha(OPT - F)$.

Theorem

An LMP α-approximation algorithm for the facility location problem gives a 2α-approximation algorithm for the metric k-median problem.

k-facility location problem

Is a common generalization of k-median and facility location. Here we have upper bound k on the number of facilities that can be opened, as well as costs for opening facilities.
Jain and Vazirani noted: An LMP α-approximation algorithm for the facility location problem gives a 2α-approximation algorithm for the metric k-facility problem.

Capacitated facility location problem

For every facility there is **one more parameter**, which indicates the **capacity of this facility**, i.e., the number of cities it can serve. If we are allowed to open each facility more than once then the problem is called *Capacitated facility location problem with soft capacities.*

Jain and Vazirani: Their facility location algorithm give rise to a 4-approximation algorithm for the metric capacitated facility location problem with soft capacities

Capacitated facility location problem

For every facility there is **one more parameter**, which indicates the **capacity of this facility**, i.e., the number of cities it can serve. If we are allowed to open each facility more than once then the problem is called *Capacitated facility location problem with soft capacities.*

Lemma

An LMP α-approximation algorithm for the facility location problem leads to an α-approximation algorithm for the capacitated facility location problem with soft capacities.

Capacitated facility location problem

- Described algorithm with some corrections gives an LMP 2-approximation algorithm for the *facility location problem*
- So, from the previous lemma it gives 3-approximation algorithm for the *metric capacitated facility location problem with soft capacities*

Capacitated facility location problem

- Described algorithm with some corrections gives an LMP 2-approximation algorithm for the *facility location problem*
- So, from the previous lemma it gives 3-approximation algorithm for the *metric capacitated facility location problem with soft capacities*

Thank you...