COM S 783 / OR&IE 634 Approximation Algorithms **Fall Semester, 2003**

Lecture 19: Nov 12, 2003

Lecturer: Yuval Rabani *Scribe: Sumit Kunnumkal*

1 Local search algorithm for facility location

We'll analyze a local search algorithm for the uncapacitated facility location problem. The local search procedure that we consider permits adding, dropping *and* swapping a facility. Korupolu et al. [2] showed that such a procedure gives a 5 approximation. The analysis presented here is due to Arya et al. [1] who improve the factor to 3.

2 Notation

Recall that in the uncapacitated facility location problem we are given a set F of facilities and a set C of clients. There is a specified cost of opening a facility and specified distance between every pair $i, j \in F \cup C$. We assume that the distances satisfy the triangle inequality. The goal is to identify a set of facilities in F to serve all the clients in C such that the total facility and connection cost is minimized. Observe that a given a set of open facilities, serving each client by the nearest open facility minimizes the connection cost.
Let
S^*: set of open facilities in the global optimum solution
S: set of open facilities in the local optimum solution
c_j^*: connection cost of client j in S^*
c_j: connection cost of client j in S
$N^*(i)$: set of clients assigned to facility i in S^*
$N(i)$: set of clients assigned to facility i in S
f_i: cost of opening facility i

3 Analysis

We will get bounds for the total connection and facility cost of the local optimum solution by analyzing them separately. The analysis will be based on the fact that since we are at a local minimum, a local move cannot decrease the objective value.

3.1 Connection cost

We will show that

$$\sum_j c_j \leq \sum_j c_j^* + \sum_{i \in S^*} f_i \tag{1}$$

Consider adding a facility $i \in S^*$ and assigning to it all clients in $N^*(i)$. The change in the objective value consists of the facility opening cost and the change in the connection costs of the clients assigned to this newly opened facility. From local optimality of S, we get:

$$f_i + \sum_{j \in N^*(i)} (c_j^* - c_j) \geq 0$$

Note that the above argument assumed that the facility i was not present in the local optimum solution. If i is in the local optimum solution, reassigning the clients in $N^*(i)$ to i results in a change in cost of $\sum_{j \in N^*(i)} (c_j^* - c_j)$ which is non-negative (from local optimality) and so the above inequality continues to hold.

Since each client is assigned to a facility in S^*, $\{N^*(i)\}$ is a partition of the set of clients and so summing over all facilities in S^*:

$$\sum_{i \in S^*} f_i + \sum_j (c_j^* - c_j) = \sum_{i \in S^*} f_i + \sum_{i \in S^*} \sum_{j \in N^*(i)} (c_j^* - c_j) \geq 0$$

which on re-arranging gives the desired result.

3.2 Facility cost

Next we try to bound the facility cost of the local optimum solution. We'll show that:

$$\sum_{i' \in S} f_{i'} \leq \sum_{i \in S^*} f_i + 2 \sum_j c_j^* \tag{2}$$

We'll differentiate facilities in S as "good" and "bad" (defined later) and obtain different bounds on the facility costs. To make this more precise, consider a facility $i \in S^*$. Define:
$P^i_{i'}$: set of clients j such that $j \in N^*(i)$ and $j \in N(i')$. So, $P^i_{i'}$ is the set of clients which are assigned to facility i in the global optimum solution, but which are assigned to facility i' in the local optimum solution.
Define a bijection π from $N^*(i)$ to itself, i.e., $\pi : N^*(i) \to N^*(i)$ with the following properties:

- If $|P^i_{i'}| \leq |N^*(i)|/2$, then $\pi(P^i_{i'}) \cap P^i_{i'} = \emptyset$
  Therefore π maps a "small" set to a different region in $N^*(i)$. So, if j belongs to a small set, then the client to which it is mapped is assigned a facility different from i' in the local optimum solution.

- If $|P^i_{i'}| > |N^*(i)|/2$, then $\forall j \in P^i_{i'}$ s.t. $\pi(j) \in P^i_{i'}$, $\pi(\pi(j)) = j$
  Now we cannot guarantee that a client in a "large" set is mapped outside the set. However, if client j_1 is mapped to client j_2 within the same set, it must be that j_2 is mapped to j_1. Observe that we can have at most one set of cardinality strictly greater $|N^*(i)|/2$.

Also, note that we can construct such a bijection by shifting cyclically the elements of $N^*(i)$ by $|N^*(i)|/2$.

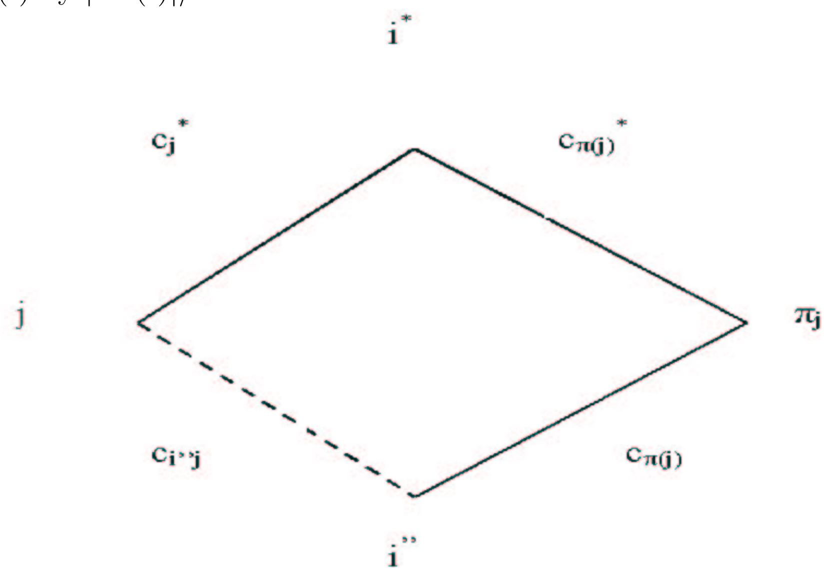

Figure 1

For the analysis of the facility cost, we distinguish two cases. Call a facility i' which is open in the local optimum solution *"good"* iff $\forall i \in S^*; \quad |P^i_{i'}| \leq |N^*(i)|/2$. Otherwise call i' *"bad"*. Therefore, for any facility in S^*, less than half its clients are assigned to a good facility; while a bad facility gets assigned more than half of some optimal facility's clients.

Case 1: $i' \in S$ is good

We will show that

$$-f_{i'} + \sum_{j \in N(i')} (c^*_j + c^*_{\pi(j)} + c_{\pi(j)} - c_j) \geq 0 \tag{3}$$

Consider removing i' from the local optimum solution. Re-assign all clients $j \in N(i')$ to the facility that $\pi(j)$ is attached to. So $\forall j \in N(i')$ assign j to i'', where $i'' : \pi(j) \in N(i'')$. Note that since i' is good, $i'' \neq i'$ and so this is a valid re-assignment.

Now, the change in the objective value on performing the above operations is

$$-f_{i'} + \sum_{j \in N(i')} (c_{i''j} - c_j)$$

where $c_{i''j}$ is the connection cost of j to i''. Using the triangle inequality (see Figure 1, recall that j and $\pi(j)$ are assigned to the same facility in S^*), we get the following bound on $c_{i''j}$:

$$c_{i''j} \leq c^*_j + c^*_{\pi(j)} + c_{\pi(j)}$$

And so,

$$-f_{i'} + \sum_{j \in N(i')} (c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j) \geq -f_{i'} + \sum_{j \in N(i')} (c_{i''j} - c_j) \geq 0$$

where the last inequality follows from local optimality

Case 2: If $i^{'} \in S$ is bad

This implies that there is at least one facility $i \in S^*$ s.t $|P_{i^{'}}^i| > |N^*(i)|/2$. We say that such a facility is captured by $i^{'}$. Let $\text{Cap}_{i'} = \{i_1, i_2, ..., i_k\}$ denote the facilities in S^* captured by $i^{'}$. Let $i \in \{i_1, i_2, ..., i_k\}$ be the closest facility to $i^{'}$. For i, the closest captured facility, we look at the cost change from swapping i and $i^{'}$. For the remaining captured facilities, we analyze the cost change on adding a captured facility.

Consider the operation of swapping $i^{'}$ with i. Now that $i^{'}$ is deleted, all the clients attached to it in the local optimum must be assigned to other facilities. For a client $j \in N(i^{'})$, consider 3 sub-cases

(a) If $\pi(j) \in N(i^{''})$ and $i^{''} \neq i^{'}$, then assign j to $i^{''}$. So if the bijection maps j to a client which is not in $N(i^{'})$, then simply assign j to $i^{''}$ — the facility to which $\pi(j)$ is assigned in the local optimum solution. We get a similar bound on the connection cost change as in Case(1):

$$c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j \geq c_{i''j} - c_j$$

(b) If $\pi(j) \in N(i^{'})$ and $j \in N^*(i)$, then assign j to i. This is the case when $\pi(j)$ is also assigned to $i^{'}$. So, there is no other facility in the local optimum solution that we can naturally assign j to. However, j is assigned to i in the global optimum (which is now added to the set of open facilities) and so we do the same. The change in connection cost is $c_{ij} - c_j = c_j^* - c_j$.

(c) Finally consider the situation when $\pi(j) \in N(i^{'})$ and $j \notin N^*(i)$. So j is assigned to some other facility, say t in S^*. Note that by the property of the bijective mapping, facility t is captured by $i^{'}$. In this case also, we assign j to i. The change in connection cost is $c_{ij} - c_j$. Now, we'll bound c_{ij} in terms of c_j and c_j^*:

$$c_{ij} \leq c_j + c_{i'i} \leq c_j + c_{i't} \leq c_j + c_j + c_j^* = 2c_j + c_j^*$$

The first and third inequalities follow from the triangle inequality (see Figure 2), while the second one results from i being the closest facility to $i^{'}$ in the set of captured facilities (which includes t). So we get:

$$2c_j + c_j^* - c_j \geq c_{ij} - c_j$$

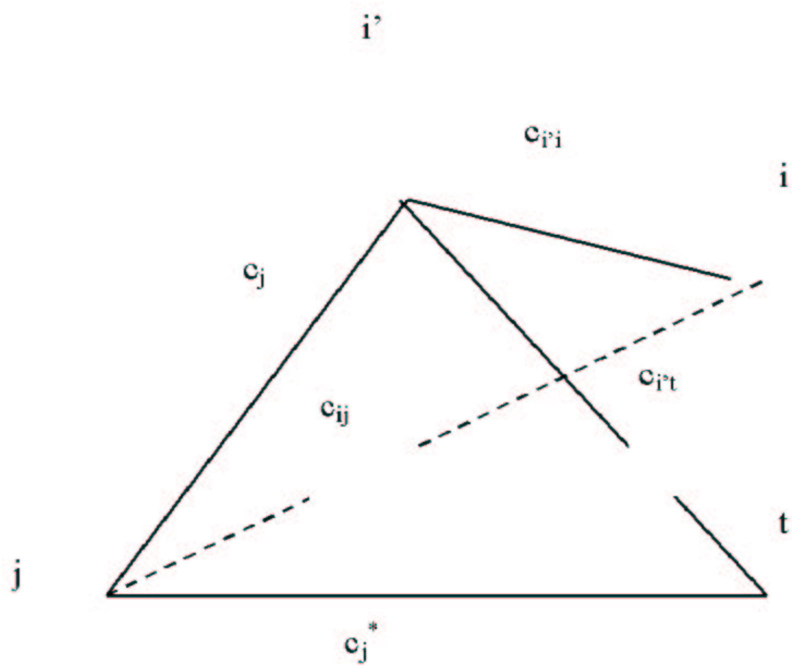

Figure 2

Note that the above 3 cases cover all the clients in $N(i')$. Now, the change in the objective value consists of the change in the facility cost $(f_i - f_{i'})$ and the change in the connection cost $(\sum_{j:(a)}(c_{i''j} - c_j) + \sum_{j:(b),(c)}(c_{ij} - c_j))$. Using the bounds obtained in (a), (b), (c):

$$f_i - f_{i'} + \sum_{j:(\mathrm{a})}(c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j) + \sum_{j:(\mathrm{b})}(c_j^* - c_j) + \sum_{j:(\mathrm{c})}(2c_j + c_j^* - c_j)$$

$$\geq f_i - f_{i'} + \sum_{j:(\mathrm{a})}(c_{i''j} - c_j) + \sum_{j:(\mathrm{b})}(c_{ij} - c_j) + \sum_{j:(\mathrm{c})}(c_{ij} - c_j) \geq 0 \qquad (4)$$

We also consider the move of adding $i'' \in \mathrm{Cap}_{i'} \backslash \{i\}$ to the set of open facilities. Therefore, i'' is not the closest captured facility. (Note that facility i' is not deleted now). Assign all $j \in N(i')$ s.t. $j \in N^*(i'')$ and $\pi(j) \in N(i')$ to i''. So, we assign all clients in $N(i')$ which are attached to i'' in S^* and which the bijection maps to within $N(i')$, to i''. Observe, this is a subset of the clients considered in Case 2(c). Considering the change in cost, we get the following inequality for each $i'' \in \mathrm{Cap}_{i'} \backslash \{i\}$:

$$f_{i''} + \sum_{j:\pi(j)\in N(i'), j \in N^*(i'')} (c_j^* - c_j) \geq 0 \qquad (5)$$

Summing over all facilities captured by i' (inequality(4) and inequalities(5))we get:

$$\sum_{i \in \mathrm{Cap}_{i'}} f_i - f_{i'} + \sum_{j:(\mathrm{a})} (c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j) + \sum_{j:(\mathrm{b})} (c_j^* - c_j) + \sum_{j:(\mathrm{c})} 2c_j^* \geq 0$$

which implies

$$\sum_{i \in \mathrm{Cap}_{i'}} f_i - f_{i'} + \sum_{j:(\mathrm{a})} (c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j) + \sum_{j:(\mathrm{b}),(\mathrm{c})} 2c_j^* \geq 0 \qquad (6)$$

As in the analysis of the connection cost, the above arguments assumed that facility $i' \notin S^*$. If facility $i' \in S^*$, then it clearly must be bad and the only facility it captures is itself. So inequality (6) continues to hold.

Finally we add up inequality (3) for all good facilities and inequality (6) for all bad facilities and noting that

- A facility in S^* is captured by at most one facility in S. So, $\sum_{i \in S^*} f_i \geq \sum_{i'} \sum_{i \in \mathrm{Cap}_{i'}} f_i$.
- $\{N(i')\}$ is a partition of the set of clients.
- $$\sum_{i':\mathrm{good}} \sum_{j \in N(i')} (c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j) + \sum_{i':\mathrm{bad}} \sum_{j:(\mathrm{a})} (c_j^* + c_{\pi(j)}^* + c_{\pi(j)} - c_j) = \sum_{i'} \sum_{j:\pi(j) \notin N(i')} 2c_j^*$$

  since $c_{\pi(j)}$ and c_j cancel out while summing over all such facilities and clients.

We get:

$$\sum_{i' \in S} f_{i'} \leq \sum_{i \in S^*} f_i + 2 \sum_j c_j^* \qquad (7)$$

Combining the bounds on the connection and facility costs (1,7) we get:

$$\sum_j c_j + \sum_{i' \in S} f_{i'} \leq 3(\sum_j c_j^* + \sum_{i \in S^*} f_i)$$

Therefore, the local search algorithm gives a 3 approximation.

Remark 1 We have not discussed the convergence of the local search algorithm. Arya et al. [1] argue that it is possible to replace the above local search algorithm with one that moves to a new solution only if there is some fixed improvement in the cost, with the approximation guarantee not "much worse". The latter algorithm clearly terminates in polynomial time.

Remark 2 Also see Arya et al. [1] for a tight example.

References

[1] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local Search Heuristics for k-median and Facility Location Problems. *STOC'01*, July 6-8, 2001.

[2] M. Korupolu, C. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1-10, January 1998.