Lecture 5

Facility Location Algorithm

October 16, 2004
Lecturer: Kamal Jain
Notes: Ioannis Giotis

We are given a bipartite graph which represents a set of n facilities F and a set of m locations. An edge from facility i to location j represents the cost c_{ij} of connecting facility i to location j. Associated with each facility is the cost of opening that facility, f_i. A feasible solution is a subset of facilities that are open and connection edges such that every location is connected to at least one open facility. The total cost of such a solution is given by

$$\text{Total cost: } \sum_{i:i\ is\ open} f_i + \sum_{i,j:j\ connected\ to\ i} c_{ij}$$

We are obviously interested in finding the minimum cost solution. We will present an approximation algorithm for the case where the connection costs satisfy the triangle inequality.

$$\forall i, i', j, j'',\ c_{ij} \leq c_{ij'} + c_{i'j'} + c_{i'j}$$

Without this assumption there doesn't exist a constant factor approximation algorithm.

Consider the variable $y_i \in \{0,1\}$ which represents whether facility i is open, and the variable $x_{ij} \in \{0,1\}$ that represents whether facility i is connected to location j.

$$\forall i \in F,\ y_i = \begin{cases} 1, & \text{when facility is open;} \\ 0, & \text{when facility is closed.} \end{cases}$$

$$\forall i,j,\ x_{ij} = \begin{cases} 1, & \text{when facility } i \text{ is connected to location } j; \\ 0, & \text{otherwise.} \end{cases}$$

Rewriting our goal as a LP

$$\begin{aligned} & min \sum_{i,j} c_{ij}x_{ij} + \sum_i f_i y_i \\ s.t. \qquad & \forall j \sum_i x_{ij} \geq 1 \\ & \forall i,j\ y_i \geq x_{ij} \end{aligned}$$

We will relax the constraints such that

$$\forall i,j\ y_i \geq 0, x_{ij} \geq 0$$

Notice that there's no reason for these variables to take values greater than 1.

Rewriting the constraints we also introduce the dual variables.

$$\forall i,j\ y_i - x_{ij} \geq 0 \quad \longleftarrow \beta_{ij}$$
$$\forall i,j\ \sum_i x_{ij} \geq 1 \quad \longleftarrow \alpha_{ij}$$

Our dual LP now is

$$\begin{aligned} & max \sum_j \alpha_j \\ s.t.\ \forall i,j \qquad & \alpha_j \geq 0 \\ & \beta_{ij} \geq 0 \\ & \alpha_j - \beta_{ij} \leq c_{ij} \\ & \sum_j \beta_{ij} \leq f_i \end{aligned}$$

Let's look at the complementary slackness conditions

Primal C.S.		
1.	$x_{ij} > 0$	$\alpha_j - \beta_{ij} - c_{ij}$
2.	$y_i > 0$	$\sum_j \beta_{ij} = f_i$
Dual C.S.		
1.	$\alpha_j > 0$	$\sum_i x_{ij} = 1$
2.	$\beta_{ij} > 0$	$y_i - x_{ij} = 0$

Satisfying all the conditions is possible only with an optimal solution. We will relax our constraints by keeping only the last 3 constraints and be relaxing the last one to be

$$3(\alpha_j - \beta_{ij}) \geq c_{ij}$$

this will give us a 3-approximation algorithm. Our algorithm consists of two phases.

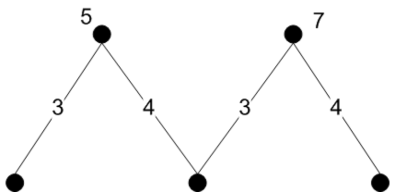
Phase 1

1: Start with zero duals at time=0
2: Increase duals in some uniform way
3: Pick some feasible primal solution

Phase 2

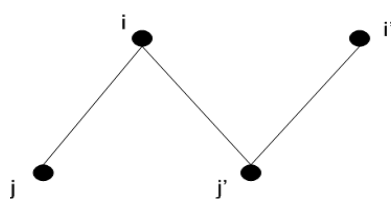
1: Throw away redundancy

Some crucial intuition is to understand that a bid form a location might be seen by more than one facility. There are two ways to throw away redundancy, either by trying to bound the number of facilities that see the bids, or by unpurchasing some of the facilities. Let's look at an example. Facilities are placed on top and locations at the bottom.



Time	Left bid	Center bid	Right bid	Comments
0	0	0	0	All α values are 0
1	1	1	1	
2	2	2	2	
3	3	3	3	
4	4	4	4	Left facility sees left location, right facility center location
5	5	5	5	
6	6	6	6	Left facility reaches bids of value 5 and opens
		7	7	
		8	8	right facility sees bids of value 7 and opens

At the end of the first phase all locations are connected to an open facility. However, we may have more than required open facilities as multiple facilities see the same bids from some location. Trying to reduce the number of open facilities we construct a graph where vertices represent facilities and edges connect facilities that see the same bid from some location. We pick a maximal independent set. At this point however, we may have locations that are not assigned to an open facility. For these locations, we enforce them to pay 3 times their bids. As we will see, this will be enough for them to be connected. In other words there exists a facility i such that $c_{ij} \leq 3\alpha_j$. Suppose that facilities i, i' had a conflict after the first phase because both were



seeing the bid of location j. Suppose another location j' that was connected to facility i is now unconnected because of we chose i' in our maximal independent set. Suppose facility i' opened at time t_1 and facility i at time t_2. The following must hold

$$\begin{aligned} c_{ij} &\leq \alpha_j \\ c_{ij'} &\leq \alpha_{j'} \\ c_{i'j'} &\leq \alpha_{j'} \end{aligned}$$

Since j' contributes to both i and i' it must be the case that those edges are tight, and

$$\alpha_{j'} \leq min(t_1, t_2)$$

On the other hand, facility i' opened exclusively by location j in which case $t_2 = \alpha_j$, or the facility opened using more contributions in which case $t_2 \leq \alpha_j$. Combining the last inequalities, we get

$$\alpha_{j'} \leq t_2 \leq \alpha_j$$

Therefore we reach the conclusion that α_j is greater or equal to $c_{ij}, c_{ij'}, c_{i'j'}$ and with the assumption of the triangle inequality, $3\alpha_j$ must be sufficient to connect location j to facility i which remains open.

This concludes the proof of our 3-approximation algorithm.