6.891 Approximation Algorithms	April 7, 2000
Lecture 9	
Lecturer: David P. Williamson	*Scribe: Clarence Bruce Applegate*

1 Uncapacitated Facility Location

Today we will spend the lecture looking at applications of techniques we have seen previously to a single problem, which we give below.

Uncapacitated Facility Location Problem (UFL)

- **Input:**
  - A finite set of locations V.
  - Facilities $F \subset V$, cost $f_i \geq 0$, $\forall i \in F$.
  - Clients $D = V - F$, cost $c_{ij} \geq 0$ for assigning client $j \in D$ to facility $i \in F$.
  - Costs c_{ij} obey the trangle inequality.
- **Goal:** Find facilities $F' \subseteq F$ and assignment of clients to facilities in F' to minimize the total cost.

We begin with an integer program formulation. First, define the following variables:

$$y_i = \begin{cases} 1 & \text{if facility } i \text{ is opened} \\ 0 & \text{otherwise} \end{cases}$$

and

$$x_{ij} = \begin{cases} 1 & \text{if client } j \text{ is assigned to facility } i \\ 0 & \text{otherwise.} \end{cases}$$

We can then get the following integer programming formulation for the UFL:

$$\begin{array}{rcll} Z^*_{LP} & = & \text{Min} \ \sum_{i \in F} f_i y_i + \sum_{i \in F, j \in D} c_{ij} x_{ij} & \\ & & \text{subject to:} & \\ & & \sum_{i \in F} x_{ij} = 1 & \forall j \in D \\ & & x_{ij} \leq y_i & \forall i \in F, j \in D \\ & & x_{ij} \in \{0,1\} & \forall i \in F, j \in D \\ & & y_i \in \{0,1\} & \forall i \in F. \end{array}$$

The first set of equalities enforce that each client is assigned to some facility. The inequalities $x_{ij} \leq y_i$ enforce that if client j is assigned to facility i, then facility i should be open. We relax this IP to an LP by replacing $x_{ij} \in \{0,1\}$ with $x_{ij} \geq 0$ and $y_i \in \{0,1\}$ with $y_i \geq 0$.

The dual of this is

$$\begin{array}{rcll} Z_D^* & = & \text{Max} & \sum_{j\in D} v_j \\ & \text{subject to:} & & \\ & & \sum_{j\in D} w_{ij} \leq f_i & \forall i \in F \\ & & v_j - w_{ij} \leq c_{ij} & \forall i \in F, j \in D \\ & & w_{ij} \geq 0 & \forall i \in F, j \in D. \end{array}$$

The following lemma relates the optimal LP and dual solutions.

Lemma 1 *If (x^*, y^*) is an optimal solution to LP and (v^*, w^*) is an optimal solution to its dual, then $x_{ij}^* > 0 \Rightarrow c_{ij} \leq v_j^*$.*

Proof:

$$\begin{aligned} x_{ij}^* > 0 \quad & \Rightarrow \quad v_j^* - w_{ij}^* = c_{ij} & (1) \\ & \Rightarrow \quad c_{ij} \leq v_j^*, & (2) \end{aligned}$$

where (1) follows by complementary slackness, and (2) since $w_{ij}^* \geq 0$. □

We now make the following definition which we will use many times.

Definition 1 *Given an LP solution x^*, we say that i neighbors j if $x_{ij}^* > 0$.*

Given the definition, we let

$$N(j) = \{i : i \text{ neighbors } j\}$$

and

$$N^2(j) = \{k : k \text{ neighbors some } i \in N(j)\}.$$

Here's an intuition about why neigboring facilities are useful. If we open a set of facilities S such that $\forall j \in D$, $\exists i \in N(j)$ such that i is open, then the cost of assigning i to j $c_{ij} \leq v_j^*$ by Lemma 1. Thus the total assignment cost is no more than $\sum_{j\in D} v_j^* = Z_D^* \leq OPT$.

1.1 Deterministic rounding

We now propose the following skeleton of an algorithm:

UFLMain

> Solve (LP), get optimal solution (x^*, y^*), (v^*, w^*).
> $S \leftarrow D$.
> while $S \neq \emptyset$
>   Choose some $j \in S$.
>   Choose some $i \in N(j)$ and open i.
>   Assign j and all unassigned clients in $N^2(j)$ to i.
> $S \leftarrow S - \{j\} - N^2(j)$.

The details we have not stated in the algorithm are the rule with which we choose $j \in S$ (call this the Center Rule) and the rule with which we choose $i \in N(j)$ (call this the Facility Rule).

Rule Set 1

- Center Rule 1: Choose $j \in S$ to minimize v_j^*
- Facility Rule 1: Choose cheapest facility $i \in N(j)$

Theorem 2 *(Shmoys, Tardos, Aardal '97; Chudak '98) UFLMain with Rule Set 1 is a 4-approximation algorithm.*

Proof: Consider each iteration, and let j be the element of S picked in the iteration. Choose $i \in N(j)$ as given by the rules. Then

$$f_i = f_i \sum_{k \in N(j)} x_{kj}^* \leq \sum_{k \in N(j)} f_k x_{kj}^* \leq \sum_{k \in N(j)} f_k y_k^*,$$

since $\sum_{k \in N(j)} x_{kj}^* = 1$ and $x_{kj}^* \leq y_k^*$ by the LP constraints. We want to ensure that by summing over all opened facilities O,

$$\sum_{i \in O} f_i \leq \sum_{i \in F} f_i y_i^*.$$

This is true because once a facility $i \in N(j)$ is opened, all clients in $j + N^2(j)$ are removed; thus no facility in $N(j)$ will ever be considered again.

Obviously the cost of assigning j to i is $c_{ij} \leq v_j^*$. The cost of assigning $k \in N^2(j)$ to i, where k neighbors l neighbors j is

$$\begin{aligned} c_{ik} &\leq c_{ij} + c_{lj} + c_{lk} \\ &\leq v_j^* + v_j^* + v_k^* \\ &\leq 3v_k^*. \end{aligned}$$

Therefore the total cost is

$$\begin{aligned}\sum_{i \in F} f_i y_i^* + 3 \sum_{j \in D} v_j^* &\leq Z_{LP}^* + 3Z_D^* \\ &= 4Z_{LP}^* \leq 4OPT\end{aligned}$$

□

1.2 Deterministic rounding with randomization

To introduce randomizing rounding, we begin by defining:

$$C_j^* = \sum_{i \in F} c_{ij} x_{ij}^* \tag{3}$$

Now consider using the following rule set with the algorithm UFLMain. **Rule Set 2**

- Center Rule 2: Choose $j \in S$ to minimize $v_j^* + C_j^*$
- Facility Rule 2: Choose $i \in N(j)$ according to probability distribution x_{ij}^* (the sum equals 1)

We can now prove the following.

Theorem 3 *(Chudak '98) UFLMain with Rule Set 2 gives a 3-approximation algorithm.*

Proof: Consider each iteration, and let j be the element of S picked in the iteration. Let

$$\begin{aligned}F &\equiv \text{random variable denoting facility cost} \\ E[F] &= \sum_{i \in N(j)} f_i x_{ij}^* \leq \sum_{i \in N(j)} f_i y_i^* \\ A_k &\equiv \text{random variable denoting cost of assigning } k \text{ to opened facility } i\end{aligned}$$

Then the cost of assigning j to i is

$$E[A_j] = \sum_{i \in N(j)} c_{ij} x_{ij}^* = c_j^*.$$

The cost of assigning $k \in N^2(j)$ to i, where k neighbors l neighbors j is

$$\begin{aligned}E[A_k] &\leq c_{jk} + \sum_{i \in N(j)} c_{ij} x^* ij \\ &\leq c_{lk} + c_{lj} + C_j^* \\ &\leq v_k^* + v_j^* + C_j^* \\ &\leq v_k^* + v_k^* + C_k^* = 2v_k^* + C_k^*, \end{aligned} \tag{4}$$

where (4) follows by our choice that j minimizes $v_j^* + C_j^*$.

Thus the total cost of the solution given by the algorithm is no more than

$$\begin{aligned}\sum_{i\in F} f_i y_i^* + \sum_{j\in D}(2v_j^* + C_j^*) &= \sum_{i\in F} f_i y_i^* + \sum_{j\in D} C_j^* + 2\sum_{j\in D} v_j^* \\ &= Z_{LP}^* + 2z_D^* \leq 3OPT.\end{aligned}$$

□

1.3 Randomized rounding with determinism

The "natural" randomized rounding algorithm is to open each facility i with probability y_i^* independently, then connect each client to the nearest open facility. If we do this, then

$$E[\text{facility cost}] = \sum_{i\in F} f_i y_i^*,$$

and the expected facility cost is at most its contribution to the objective function. Then the assignment cost of j if $i \in N(j)$ is open is $c_{ij} \leq v_j^*$, and the probability that a neighbor of j is open is reasonably high:

$$\begin{aligned}Pr[\text{no neighbor of } j \text{ is open}] &= \prod_{i\in N(j)} (1 - y_i^*) \\ &\leq \prod_{i\in N(j)} e^{-y_i^*} \\ &= e^{-\sum_{i\in N(j)}(y_i^*)} \\ &\leq e^{-\sum_{i\in N(j)}(x_{ij}^*)} = e^{-1}\end{aligned}$$

However, it is not clear how to bound the assignment cost if some neighbor of j is not open. What can we do then?

In order to address this point, we consider an algorithm that is intermediate to this "pure" randomized rounding algorithm and the deterministic ones we consider before. First, we need the followign claim.

Claim 4 *Given any UFL instance, there is an equivalent polynomial-sized UFL instance that is* **complete***; that is, there exists optimal LP solution (x^*, y^*) to the instance with $x_{ij}^* = y_i^* \forall x_{ij}^* > 0$.*

We now consider the following modification to our central algorithm.

UFLMainMod

> Get optimal complete LP solution (x^*, y^*), (v^*, w^*).
> $S \leftarrow D$.
> $T \leftarrow F$.
> while $S \neq \emptyset$
>   Choose some $j \in S$.
>   Choose some $i \in N(j)$ and open i.
> $S \leftarrow S - \{j\} - N^2(j)$.
> $T \leftarrow T - N(j)$.
> For each $i \in T$, open i (independently) with probability y_i^*.
> Assign each $j \in D$ to the nearest open facility.

As before, two steps of the algorithm are not specified. Consider the following rules for choosing a center and a facility.

Rule Set 3

- Center Rule 1: Choose $j \in S$ to minimize v_j^*
- Facility Rule 2: Choose $i \in N(j)$ according to probability distribution $x_{ij}^* = y_i^*$

Note that the probabilities are no longer independent: you must pick exactly one facility for each center. However, we can still bound the probability that no neighboring facility of j is opened by $1/e$, but we have to be more careful. Pick an arbitrary $j \in D$. Let X_p be members of $N(j)$ removed from T in iteration p. Each $i \in N(j)$ opened independently in the very last step is put in its own set X_p.

Lemma 5 *Pr[no neighbor of j opened]* $\leq \frac{1}{e}$

Proof: Let $O_p \equiv$ event that some facility in X_p opened. Events O_p are independent. Let $Y_p^* = \Pr[O_p] = \sum_{i \in X_p} y_i^*$. Then

$$\begin{aligned}
\sum_p Y_p^* &= \sum_p \sum_{i \in X_p} y_i^* \\
&= \sum_{i \in N(j)} y_i^* \\
&= \sum_{i \in N(j)} x_{ij}^* = 1.
\end{aligned}$$

Thus

$$\Pr[\text{no neighbor of j is open}] = \prod_p (1 - Y_p^*) \leq \frac{1}{e}.$$

□

We can now prove the following theorem.

Theorem 6 *(Chudak '98, Chudak, Shmoys '98) UFLMainMod with Rule Set 3 is a (1 + $\frac{3}{e}$)-approximation algorithm* ($1 + \frac{3}{e} \approx 2.104$)

Proof: By the reasoning above,

$$E[\text{facility cost}] = \sum_{i \in F} f_i y_i^*$$

Note that if no neighbor of j is opened, one can assign j to some facility $N^3(j)$ for a cost at most $3v_j^*$ (same analysis as before, use rule 1 to get the v_j^*). Therefore, the expected assignment cost for j is

$$\sum_{i \in N(j)} c_{ij} y_i^* + Pr[\text{no neighbor of j is open}] \cdot 3v_j^* \quad \leq \quad \sum_{i \in N(j)} c_{ij} x_{ij}^* + \frac{3}{e} v_j^*.$$

Thus the overall expected cost is

$$\sum_{i \in F} f_i y_i^* + \sum_{i,j} c_{ij} c_{ij}^* + \frac{3}{e} \sum_{j \in D} v_j^* \quad = \quad Z_L^* P + \frac{3}{e} Z_D^* \leq (1 + \frac{3}{e}) OPT.$$

□

In fact, one can show a slightly better performance guarantee by using different rules.

Theorem 7 *(Chudak, Shmoys '98) UFLMainMod using Rule Set 2 gives a* ($1 + \frac{2}{e}$)*-approximation algorithm* ($1 + \frac{2}{e} \approx 1.736$)

The best known result is $\approx 1 + \frac{2}{e} - .001$, or something to that effect (result due to Charikar and Guha '99). There is a limit to how well the problem can be approximated.

Theorem 8 *(Guha, Khuller '98, Srividenko '98) There is no α-approximation algorithm for $\alpha < 1.427$ for UFL unless P=NP.*

1.4 The primal-dual method

As usual, our primal-dual method will construct a feasible dual solution, and use the information from it to infer a primal solution. To this end, we need to modify our definition of a neighbor.

Definition 2 *Given a feasible dual solution (v, w), we say that i neighbors j if $v_j \geq c_{ij}$.*

We then define $N(j)$ and $N^2(j)$ as before.

We consider the standard primal-dual technique of choosing the set of items such that the corresponding dual constraint is tight. In this case, let $A = \{i \in F : \sum_{j \in D} w_{ij} = f_i\}$. Then

$$\begin{aligned}\sum_{i \in A} f_i &= \sum_{i \in A} \sum_{j \in D} w_{ij} \\ &= \sum_{i \in A} \sum_{j \in D} \max(0, v_j - c_{ij}) \\ &= \sum_{j \in D} \sum_{i \in N(j) \cap A} (v_j - c_{ij}).\end{aligned}$$

Now suppose $|N(j) \cap A| = 1$ for all j, and $i(j)$ is the unique element in $N(j) \cap A$. Then by the above,

$$\sum_{i \in A} f_i = \sum_{j \in D} (v_j - c_{i(j),j}).$$

In this situation, if we open facilities in A, and assign each j to $i(j)$, the cost is

$$\sum_{i \in A} f_i = \sum_{j \in D} c_{i(j),j} = \sum_{j \in D} v_j \leq OPT.$$

Obviously, this can't happen in the general case, but maybe we can do something similar. The basic idea is that we will prune A such that $\forall j \in D$, $|N(j) \cap A| \leq 1$. Then the facility cost is paid for and assignment cost is paid for $\forall j$ such that $|N(j) \cap A| = 1$. Use ideas from previous algorithms to take care of the case when $N(j) \cap A = \emptyset$ by assigning j to a nearby facility.

Here is the overall algorithm.

UFLPrimalDual

$v \leftarrow 0$, $w \leftarrow 0$
$A \leftarrow \emptyset$
$l \leftarrow 0$
$S \leftarrow D$.
while $S \neq \emptyset$
    $l \leftarrow l + 1$
    Increase v_j for all $j \in S$ and w_{ij} for all $i \in N(j), j \in S$ uniformly until
        $\exists i_l \in F : \sum_{j \in D} w_{i_l j} = f_{i_l}$
    $A \leftarrow A \cup \{i_l\}$
    $S \leftarrow S - N(i_l)$.
$A' \leftarrow A$
for $k \leftarrow 1$ to l
    if $i_k \in A'$
        Open i_k, $A' \leftarrow A' - i_k - N^3(i_k)$
        Assign unassigned $j \in N(i_k) \cup N^3(i_k)$ to i_k

Before we begin the analysis, let A' be the final set of opened facilities and $i(j)$ be the facility j is assigned to.

The following lemma shows that we have achieved the goal we set out for.

Lemma 9 $\forall j \in D, |A' \cap N(j)| \leq 1$

Proof: By contradiction. Suppose $i_a, i_b \in A' \cap N(j)$, where $a < b$. Then i_b was removed from A' when i_a was opened. □

Theorem 10 *(Jain, Vazirani '99)*

$$\sum_{j \in D} c_{i(j),j} + 3 \sum_{i \in A'} f_i \leq 3 \sum_{j \in D} v_j \leq 3OPT.$$

Proof:

Carrying out our analysis of the facility cost as before, we have

$$\begin{aligned} 3 \sum_{i \in A'} f_i &= 3 \sum_{i \in A'} \sum_{j \in D} w_{ij} \\ &= 3 \sum_{i \in A'} \sum_{j \in D} \max(0, v_j - c_{ij}) \\ &= 3 \sum_{j \in D} \sum_{i \in N(j) \cap A'} (v_j - c_{ij}) \\ &= 3 \sum_{j \in D : |N(j) \cap A'| = 1} (v_j - c_{i(j),j}) \end{aligned}$$

What about j such that $N(j) \cap A' = \emptyset$? We claim that $c_{i(j),j} \leq 3v_j$. Note that for such j all facilities $i \in N(j) \cap A$ have indices higher than $i(j)$. Therefore, $v_j \geq v_k$ for all $k \in N(i(j))$. This follows because facilities were indexed as their dual constraints became tight, so the neighboring clients of lower indexed facilities stopped increasing before the clients of higher indexed facilities. Assume that j neighbors i' neighbors k neighbors $i(j)$, so that:

$$\begin{aligned} c_{i(j),j} &\leq c_{i',j} + c_{i',k} + c_{i(j),k} \\ &\leq v_j + v_k + v_k \\ &\leq 3v_j. \end{aligned}$$

From above, we have that

$$3 \sum_{i \in A'} f_i = 3 \sum_{j \in D : |N(j) \cap A'| = 1} (v_j - c_{i(j),j}),$$

which implies that

$$3 \sum_{i \in A'} f_i + \sum_{j \in D : |N(j) \cap A'| = 1} c_{i(j),j} \leq 3 \sum_{j \in D : |N(j) \cap A'| = 1} v_j.$$

Thus,

$$\begin{aligned}
3\sum_{i\in A'} f_i + \sum_{j\in D:|N(j)\cap A'|=1} c_{i(j),j} + \sum_{j\in D:N(j)\cap A=\emptyset} c_{i(j),j} &\leq 3\sum_{j\in D:|N(j)\cap A'|=1} v_j + 3\sum_{j\in D:N(j)\cap A'=\emptyset} v_j \\
&\leq 3\sum_{j\in D} v_j \leq 3OPT
\end{aligned}$$

□

We will use the statement of the proof for an approximation algorithm for the k-median problem, to be presented in the next lecture.