Lecture 13: Facility Location Problems (Cont'd).

Last time: local search for k-medium, constant approx.

today: LP rounding algorithms for facility location / k-medium.

recall: same setup (almost) for both problems:

metric (V,i,d). clients C ∈ V

\[ \text{fac loc: open } F \subseteq V, \quad \text{cost } = \sum_{i \in F} f_i + \sum_{j \in C} d(j,F) \]

\[ \text{kmed: open } (F, k), F \subseteq V \quad \text{cost } = \sum_{j \in C} d(j,F). \]

local search for fac. loc can also give O(1) approx.

- Today see LP-based solutions that do better, also more versatile.
  
  and give better apx. (comparable to LP relaxation as opposed to optimal integer soln.).
Focus on Fac-Loc for now.

How to solve?

(1) Cast as set cover. Each set given by a center and some subset of clients in a set. See Exercise in HW1.

goto $O(\log n)$ approx!

Better? Yes.

Today: write an LP. round it (round and round again).

\[ \text{LP:} \quad \min \sum f_i y_i + \sum_{ij} d_{ij} x_{ij} \]
\[ \text{st:} \quad \sum_i x_{ij} = 1 \quad \forall i \in C \]
\[ x_{ij} \leq y_i \quad \forall i \in V, \forall j \in C \]
\[ x_{ij}, y_i \geq 0. \]

Thm: Can round and get solution $F \leq V$ st.

\[ \text{cost}(F) = \sum_i f_i + \sum_{j \in C} d(G, F) \]
\[ \leq \text{constant} \cdot \text{LP value}. \]

best known: $14.6$ or $80$.

Thm: Integrality gap is at least $1463$. 

Small gap here. But focus will be on the alg. ideas, not the actual numerical values.
In previous applications we've not considered the dual program, but that gives a lot of information as well.

\[
\begin{align*}
&\text{max} & \sum_{j \in C} \alpha_j \\
\text{subject to} & & \alpha_j - \beta_{ij} \leq d_{ij} & \forall ij \in E
\end{align*}
\]

dual variable per client

dual variable per client-facility pair

\[
\sum_{j \in C} \beta_{ij} \leq f_i & \forall i \in V.
\]

\[
\beta_{ij} \geq 0, \alpha_j \geq 0.
\]

but really set \( \alpha_j = \min_{i \in V} \left( \frac{d_{ij} + \beta_{ij}}{\beta_{ij}} \right) \) \( \forall i \in V \)

and \( \beta_{ij} \) so non-negative.

We will interpret this LP in greater detail soon, but for right now, just recall basic facts about LP duality.

\[\begin{align*}
\text{for:} & \quad \min c^T x \\
(\text{P}) & \quad Ax \leq b \quad x \geq 0
\end{align*}\]

\[\begin{align*}
\Rightarrow & \quad \max b^T y \\
(\text{D}) & \quad A^T y \leq c \\
& \quad y \geq 0
\end{align*}\]

Then (Weak duality): if \( x, y \) are feasible LP solutions to \( \text{P} \), \( \text{D} \) then \( c^T x \leq b^T y \).

\[\begin{align*}
\Rightarrow & \quad (A^T y)^T x = y^T A x \geq y^T b \quad \text{primal feasibility} \\
& \quad \text{and} \ x \geq 0 \quad \text{and} \ y \geq 0.
\end{align*}\]

Then (Strong duality). If \( x^* \) feasible primal & dual solution, then \( x^* \) optimal feasible, \( \alpha^* \geq 0 \).

Optimal feasible, \( \alpha^* \geq 0 \).

Not giving proof for now, see, e.g. Schrijver or Matoušek - Gärtner or...

Corollary: (Complementary slackness). For \( x^*, y^* \) optimal primal/dual solutions.

\[\begin{align*}
& x^T (A^* y^* - c) = 0 \quad \text{and} \ y^T (A^* x^* - b) = 0.
\end{align*}\]

Pf: both inequalities here must be tight if \( c^T x = b^T y \). ☺
In other words, if a dual variable is non-zero then the corresponding primal is constraint is tight.

As example: SPs $(x,y)$, $(x,\beta)$ optimal solutions

$$x_j > 0 \text{ then } d_j - \beta j = d_j = \beta j + d_j$$

$\text{BTW: other implications of strong duality.}$

Any dual $\preceq$ optimal dual LP soln $\preceq$ optimal primal LP soln $\preceq$ optimal primal IP soln before relaxing.

$\text{LP is relaxation of IP}$

$dual$ is like an accounting device.

So suffice to show: $\text{Solution to facility locating with cost} \leq c \cdot \text{dual value.}$

$(\text{and dual solution})$

* Solve the LP, and dual optimally. $(x,\beta)$ solutions.

* For each client $i \in C$,

  define its “neighbors” $N(i) = \{ i \in V / x_{ij} > 0 \}$.

  $\text{Fact: if } i \in N(j) \text{ then } d_j = \beta j + d_j.$

* What if we open the cheapest of these facilities and send $j$ there?
consider some disjoint clusters \( C \). A clique s.t. \( N(j) \cap N(j') = \emptyset \) for \( j, j' \in C \)

then opening these cheapest facilities in each cluster gives cost

\[
\sum_{j \in C} \left( \text{cost of cheapest facility in } N(j) + \text{dist of } j \text{ to this facility} \right).
\]

Hmm... how to account for this cost?

Indeed, s.t. \( i(j) \) is facility cheapest in \( N(j) \), then

\[
\sum_{i \in N(j)} x_{ij} \leq \sum_{i \in N(j)} f_{ij}x_{ij} \leq \sum_{i \in N(j)} f_{ij}y_i = \text{LP cost inside that cluster.}
\]

\( \text{min} \), \( \text{average} \), \( \text{subject to} \)

\( X_{ij} \; \text{s.t.} \; x_{ij} \leq y_i \text{ constraint} \)

\( \Rightarrow \text{bk clusters are disjoint,} \)

\[
\sum_{i \in N(j)} x_{ij} \leq \text{LP cost for opening facilities.}
\]

What about connecting costs?

\[
\text{if } i \in N(j) \text{ then } d_{ij} = B_{ij} + d_{ij} \geq d_{ij} \Rightarrow d_{ij} \leq d_{ij} \geq 0.
\]

\( B_{ij} \geq 0 \), \( \text{dual contribution of } i \)

\( \Rightarrow \text{for centers of these clusters, } \sum_{i \in N(j)} x_{ij} \leq \text{their dual contribution.} \)

But how to do this when clusters are not disjoint ???

- Pick an independent / disjoint set \( A \) of clusters of "small" radius.

\( \text{selected} \)

- Open cheapest facility within each cluster \( \text{(has low cost, already seen)} \)

- Show that routing clients of unselected clusters can be sent to these open facilities also with small cost.
Algorithm: Solve LP and dual.

Sort clients \( s \) so \( \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n \).

\( S \leftarrow \emptyset \).

\( \text{for } j = 1 \text{ to } n \)

\[
\begin{cases}
\text{if } N(j) \text{ is disjoint from all } N(j'), j' \in S \text{ }
\end{cases}
\]

\( S \leftarrow S \cup \{ j \}, \text{open cheapest facility in } N(j). \)

---

So open cheapest facilities in \( \bigcup_{j \in S} N(j) \).

- Facility cost: \( \sum_{j \in S} \min_{i \in N(j)} f_i \leq \sum_{j \in S} \sum_{i \in N(j)} f_i x_{ij} \leq \sum_{j \in S} f_j x_{ij} \leq \text{LP primal} \)

- connect arcs for \( j \in S \).

\( d_{ij} \leq \alpha_j \) because distance from \( j \) to \( F \)

\( \leq \text{distance from } j \text{ to open fac in } N(j) \)

\( \leq \alpha_{j'} \). (as above).

- What about \( j \notin S \).

Must be b/c \( N(j) \cap N(j') \neq \emptyset \) for \( j' \in S \)

say \( i \) liest

and cheapest fac in \( N(j') \) opened

\( d(j, i) \leq d(j, i') + d(j', i) + d(j', i') \)

\( \leq \alpha_i + \alpha_j + d_{j'} \leq 3 \alpha_j \)

by completeness.

by sorted order
Overall: Facility opening cost $\leq$ LP value

Connection cost $\leq 3$. Dual value $\leq 3 \cdot LP$ value

$\Rightarrow 4$-apx

---

Moral of Story: properties of dual solution / relationship to primal
allowed us to get quick apx-algo for Facility Location

- Deterministic rounding
- Clustering ensured that facility opening costs small.
  - Divis ensured that connection costs for non-cluster centers small.

---

What is the integrality gap? Showed upper bound $\leq 4$. Better? Yes.

Here is a randomized rounding idea.

- Having picked a set $S$ of cluster centers, open one facility randomly.

  - $\forall j \in S$, Pick $i \in N(j)$ w/ $X_{ij}$ (recall $\sum_{i \in N(j)} X_{ij} = 1$ so prob. dist).

  - For $i \notin N(j)$, pick indep $y_i = \sum_{j \in S} X_{ij}$ so pick one facility in each such set $N(i)$.

  - $E[\text{facility cost}] = \sum_{j \in S} E[\sum_{i \in N(j)} X_{ij}] = \sum_{j \in S} f_j X_{ij} \leq LP$ value.

  - Again, distance of each $j \in S$ to closest fac $\leq d_j$ is dual contribution.
What about \( j \neq 5 \),

since \( j \neq 5, \exists j \in S \iff \nu(j) \neq 0 \)

and \( \alpha_i \leq \alpha_j \)

\( \Rightarrow \) overall connect in cost (in expectation) \( \leq (1 + \frac{2}{\epsilon}) \sum_j \alpha_j = (1 + \frac{2}{\epsilon}) \text{dual value} \)

\( \Rightarrow \) integrality gap \( \leq 1 + (1 + \frac{2}{\epsilon}) \leq 2.7 \text{something} \)

Can save the "1 + .." loss as well to get \((1 + \frac{2}{\epsilon})\). See the [WS10] book.

And more ideas, but not today.

Let's see a different "primal-dual" way to use duality next time.