Metric Uncapacitated Facility Location

I: set of facilities

J: set of clients

f_i: opening cost of facility i

d_{ij}: cost of assigning client j to facility i

Goal: Open a subset of facilities and assign each client to an open facility while incurring in a minimum cost

Primal-Dual Approximation Algorithm (Jain & Vazirani, 2001)

$$(LP): \quad \min \sum_{i \in I} f_i y_i + \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{i \in F} x_{ij} \geq 1 \quad \forall j \in J \longrightarrow \alpha_j$$

$$x_{ij} \leq y_i \quad \forall i \in I, j \in J \longrightarrow \beta_{ij}$$

$$x_{ij}, y_i \geq 0 \quad \forall i \in I, j \in J$$

$$(D): \quad \max \sum_{j \in J} \alpha_j$$

$$\text{s.t.} \quad \sum_{j \in J} \beta_{ij} \leq f_i \quad \forall i \in I \longrightarrow y_i$$

$$\alpha_j - \beta_{ij} \leq d_{ij} \quad \forall i \in I, j \in J \rightarrow x_{ij}$$

$$\alpha_j, \beta_{ij} \geq 0 \quad \forall i \in I, j \in J$$

Interpretation of dual:
- α_j: Total cost paid by j to get assigned to an open facility.
- β_{ij}: Cost paid by j to open facility i.

Why? Suppose (x,y) is integral and optimal for (LP). Let (α, β) be a corresponding dual solution.

By complementary slackness:
- $y_i > 0 \Rightarrow \sum_{j \in J} \beta_{ij} = f_i$ (To open i, f_i needs to be paid ~~for~~ by all clients)
- $x_{ij} > 0 \Rightarrow \alpha_j = d_{ij} + \beta_{ij}$ (If j is connected to i, j pays the assignment cost and its share for opening i)
- $\alpha_j > 0 \Rightarrow \sum_{i \in F} x_{ij} = 1$ (Not interesting)
- $\beta_{ij} > 0 \Rightarrow x_{ij} = y_i$ (Client j only pays its share of f_i if it uses facility i)

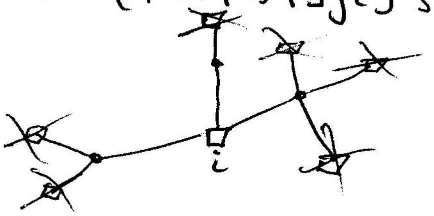
Algorithm: <u>Phase 1</u>

Definitions:
- j neighbors i if $\alpha_j \geq d_{ij}$
- j contributes to i if $\beta_{ij} > 0$

- $x, y, \alpha, \beta := 0$, $T := \{i \in I \mid \sum_{j \in J} \beta_{ij} = f_i\}$ (temporarily open facilities)
- while $\exists$ a client j not neighboring any facility in T.
  Raise α_j uniformly for all such j.
  - if j becomes neighbor of some $i \notin T \Rightarrow$ Increase β_{ij} at the same rate to maintain feasibility.
  - if j becomes neighbor of some $i \in T \Rightarrow$ Freeze variable α_j.

<u>Phase 2</u> (Cleanup)

- while $\exists$ a client contributing to at least two facilities in T:
  - Pick an arbitrary such i
  - $T := T - \{h \in T \setminus \{i\} \mid \exists j \in J \text{ s.t. } \beta_{ij} > 0 \text{ and } \beta_{hj} > 0\}$



<u>Phase 3</u>

- Set $y_i = 1 \quad \forall i \in T$

For all $j \in J$:
- Case 1: j contributes to a unique $i \in T \Rightarrow x_{ij} = 1$
- Case 2: j contributes to no $i \in T$ but neighbors some $i \in T$
  $\Rightarrow x_{ij} := 1$ for arbitrary such i.
- Case 3: j has no neighbors in $T \Rightarrow$ Let i be closest facility in T, $x_{ij} := 1$.

<u>Analysis:</u>

<u>Lemma</u>: If at the end of Phase 2 there is a client j with no neighbors in T, there exists $i \in T$ s.t. $d_{ij} \leq 3\alpha_j$.

<u>Proof</u>: Later

Theorem: The algorithm gives a 3-approximation.

Proof: For every i, let $A(i) = \{\text{neighbours of } i \text{ s.t. } x_{ij} = 1\}$

$$\sum_{i\in T}\Big(f_i + \sum_{j\in A(i)} d_{ij}\Big) = \sum_{i\in T}\sum_{j\in A(i)}(\beta_{ij} + d_{ij}) = \sum_{i\in T}\sum_{j\in A(i)} \alpha_j$$

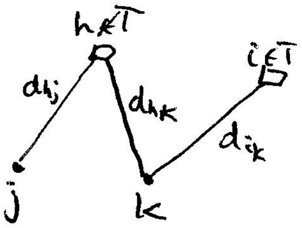
Let $Z := J \setminus \bigcup_{i\in T} A(i)$ (clients with no neighbors in T)

$$\sum_{i\in T}\sum_{j\in Z} d_{ij}x_{ij} = \sum_{j\in Z}\sum_{i\in T} d_{ij}x_{ij} \underset{\text{Lemma}}{\leq} \sum_{j\in Z} 3\alpha_j.$$

$$\Rightarrow \sum_{i\in T}\Big(f_i + \sum_{j\in J} d_{ij}x_{ij}\Big) = \sum_{i\in T}\sum_{j\in A(i)}\alpha_j + \sum_{j\in Z} 3\alpha_j \leq 3\sum_{j\in J}\alpha_j \leq 3\,OPT_D \leq 3\,OPT_{LP}. \quad \square$$

Proof of Lemma:

Claim 1: Given such j, there exists h, k, i s.t.:



Why?
- We stopped increasing α_j when j neighboured some $h \in T$.
- h was removed from T in Phase 2 because k was contributing both to it and to i (which remains in T).

Claim 2: $d_{ij} \leq 3\alpha_j$

Pf: By triangle inequality.

1. $d_{hj} \leq \alpha_j$ because h and j are neighbours.
2. k contributes to h $\Rightarrow \beta_{hk} > 0 \Rightarrow \alpha_k > d_{hk}$.
   When α_j stops growing, h is in T. k contributed to h $\Rightarrow$ At that time, k already neighbors h
   $\Rightarrow \alpha_k$ doesn't ~~keep~~ growing after α_j stops $\Rightarrow \alpha_j \geq \alpha_k \Rightarrow \alpha_j \geq d_{hk}$.
3. $\beta_{ik} > 0 \Rightarrow \alpha_k \geq d_{ik} \Rightarrow \alpha_j \geq d_{ik}$.

So, $d_{ij} \leq d_{hj} + d_{hk} + d_{ik} \leq 3\alpha_j$. $\square$

Best approximation known: 1.52-approx. by Mahdian, Ye, Zhang.

Hardness: No approximation better than 1.46 possible (unless $P = NP$)

References:
- Jain, Vazirani: Approx. algorithms for metric facility location and k-median problems (2001)
- Shmoys, Williamson: The Design of Approximation Algorithms (2010)
- Vazirani: Approximation Algorithms (2001)