LP rounding algorithms for Facility Location

Jaroslaw Byrka

University of Wroclaw

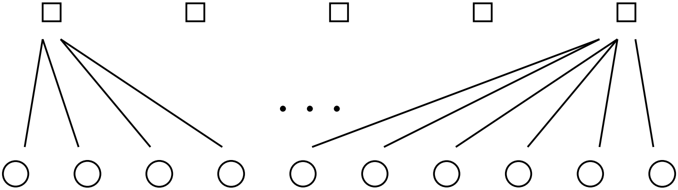
Algorithmic Workshop 10-12 October, 2014

Outline

1. Uncapacitated Facility Location Problem
   - Problem definition
   - Approximation results
   - Bifactor approximationn
2. LP-rounding algorithms
   - The algorithm of Chudak and Shmoys
   - Problem with irregular instances
   - Improvements: how we obtained 1.5-apx

UFLP

- set of clients C
- set of facilities F with opening costs f_i for $i \in F$
- connection costs c_{ij} for $i \in F$ and $j \in C$



IP formulation

minimize $\sum_{i\in\mathcal{F},j\in\mathcal{C}} c_{ij}x_{ij} + \sum_{i\in\mathcal{F}} f_i y_i$

subject to

$\sum_{i\in\mathcal{F}} x_{ij} = 1$	for all $j \in \mathcal{C}$	(i)
$x_{ij} - y_i \leq 0$	for all $i \in \mathcal{F}, j \in \mathcal{C}$	(ii)
$x_{ij}, y_i \in \{0, 1\}$	for all $i \in \mathcal{F}, j \in \mathcal{C}$	(iii)

Approximation for general connection costs

Definition

λ-approximation algorithm for a minimization problem $=$ polynomial time alg. producing solutions with cost $\leq \lambda \times OPT$

- Hochbaum '82 - $O(\log n)$-approximation algorithm
- An approximation preserving reduction from the Set Cover problem

Metric case

- c_{ij} symmetric
- satisfy the triangle inequality
  ($c_{ij} \leq c_{ik} + c_{kj}$ for all $i, j, k \in C \cup F$)

- Shmoys et al. '97 - 3.16-appr. by LP-rounding
- Chudak '98 - 1.73-appr. by randomized LP-rounding
- Jain et al. '02 - 1.61-appr. by a greedy algorithm
- Mahdian et al. '02 - 1.52-appr. by a greedy algorithm
- B., Aardal '07 - 1.5-appr. by a combination
- Li '11 - 1.488-appr. by a more randomized combination

Lower bound : 1.463

Metric case

- c_{ij} symmetric
- satisfy the triangle inequality ($c_{ij} \leq c_{ik} + c_{kj}$ for all $i, j, k \in C \cup F$)

- Shmoys et al. '97 - 3.16-appr. by LP-rounding
- Chudak '98 - 1.73-appr. by randomized LP-rounding
- Jain et al. '02 - 1.61-appr. by a greedy algorithm
- Mahdian et al. '02 - 1.52-appr. by a greedy algorithm
- B., Aardal '07 - 1.5-appr. by a combination
- Li '11 - 1.488-appr. by a more randomized combination

Lower bound : 1.463

Hardness of approximation

- Guha and Khuller '98:
  If there exists an 1.463-approximation algorithm for the metric UFLP, then $NP \subset DTIME(n^{loglogn})$
- Sviridenko: 1.463 lower bound, unless $P = NP$
- Jain et al. '02:
  Bifactor lower bound : $(\lambda_f, 1 + 2e^{-\lambda_f})$

Bifactor approximation

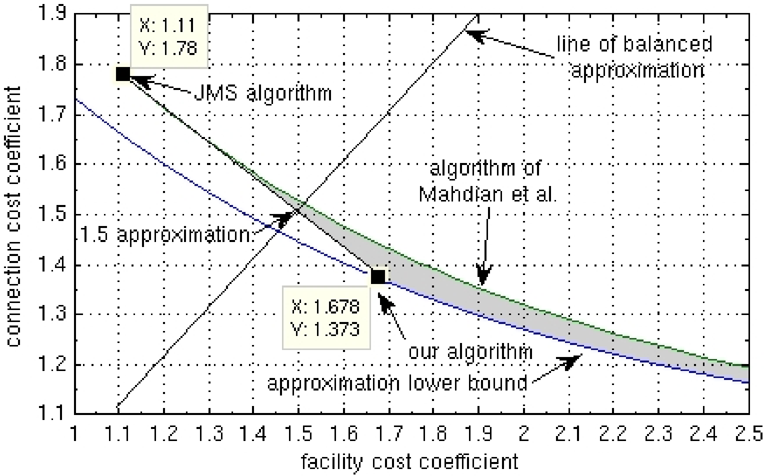
minimize $\sum_{i\in\mathcal{F},j\in\mathcal{C}} c_{ij}x_{ij} + \sum_{i\in\mathcal{F}} f_i y_i$

subject to

$\sum_{i\in\mathcal{F}} x_{ij} = 1$	for all $j \in \mathcal{C}$	(i)
$x_{ij} - y_i \leq 0$	for all $i \in \mathcal{F}, j \in \mathcal{C}$	(ii)
$x_{ij}, y_i \in \{0, 1\}$	for all $i \in \mathcal{F}, j \in \mathcal{C}$	(iii)

- Standard approximation:
  $ALG \leq \lambda \cdot OPT = \lambda \cdot \sum_{i\in\mathcal{F},j\in\mathcal{C}} c_{ij}x_{ij} + \lambda \cdot \sum_{i\in\mathcal{F}} f_i y_i$
- Bifactor approximation:
  $ALG \leq \lambda_c \cdot \sum_{i\in\mathcal{F},j\in\mathcal{C}} c_{ij}x_{ij} + \lambda_f \cdot \sum_{i\in\mathcal{F}} f_i y_i$

Bifactor approximation results



Outline

1. Uncapacitated Facility Location Problem
   - Problem definition
   - Approximation results
   - Bifactor approximationn
2. LP-rounding algorithms
   - The algorithm of Chudak and Shmoys
   - Problem with irregular instances
   - Improvements: how we obtained 1.5-apx

Sketch of the algorithm of Chudak and Shmoys

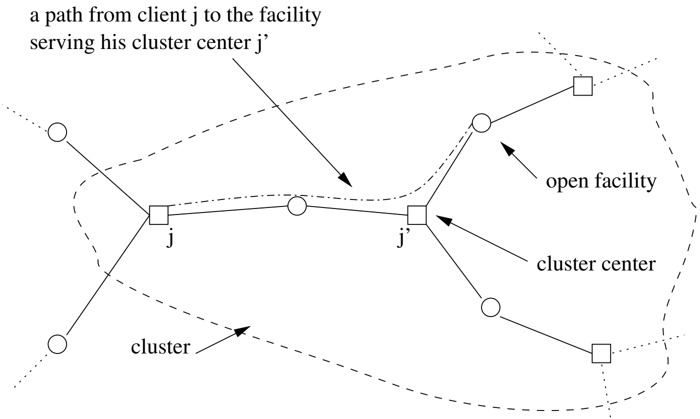
- Solve the LP-relaxation of the problem
- Divide clients into clusters
- Open facilities (by a randomized procedure)
- Connect every client to his nearest facility

Greedy clustering procedure

While not all the clients are clustered:

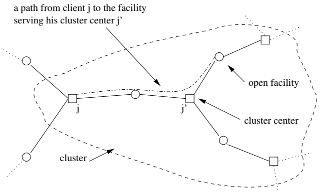
- Choose a not clustered client (greedily)
- Make him a center of the new cluster
- Take all his not clustered neighbors into the new cluster (two clients fractionally served by a facility are neighbors)

A cluster



Squares represent clients, circles - facilities

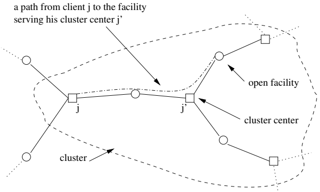
Facilities opening



- Open one facility around the cluster center, each facility j with pr. y_j
- Open other facilities independently with pr. y_j

Expected facility opening cost $=$ fractional fac. opening cost F^*

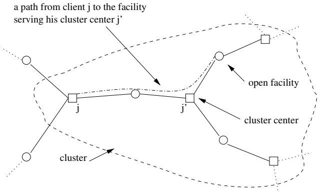
Expected connection cost



- Open one facility around the cluster center, each facility j with pr. y_j
- Open other facilities independently with pr. y_j

- dual LP -> budget $v_j^* = F_j^* + C_j^*$
- $E[C_j] \leq (1 - 1/e) \cdot C_j^* + (1/e) \cdot (2F_j^* + 3C_j^*)$
- $E[\sum_{j \in C} C_j] + E[F] = (1 + 2/e) \cdot (C^* + F^*)$

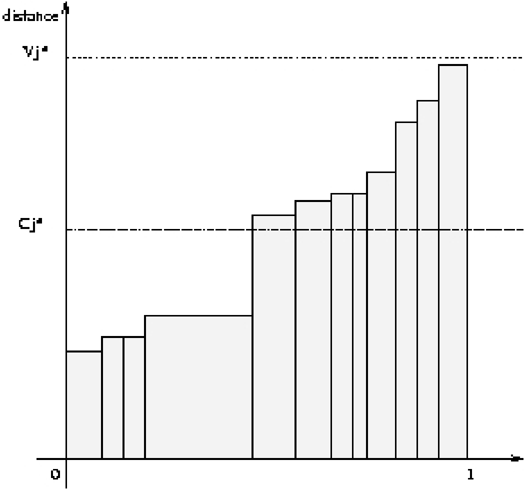
Expected connection cost



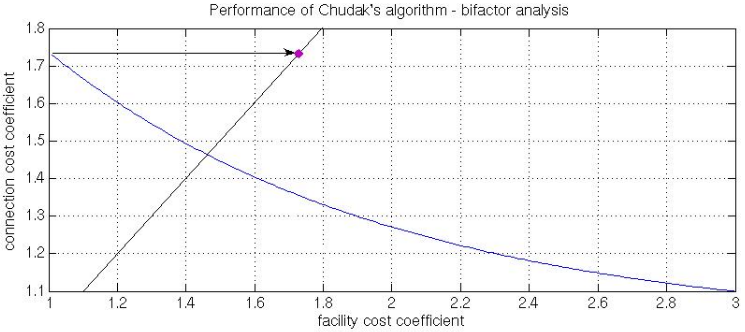
- Open one facility around the cluster center, each facility j with pr. y_j
- Open other facilities independently with pr. y_j

- dual LP -> budget $v_j^* = F_j^* + C_j^*$
- $E[C_j] \leq (1 - 1/e) \cdot C_j^* + (1/e) \cdot (2F_j^* + 3C_j^*)$
- $E[\sum_{j \in \mathcal{C}} C_j] + E[F] = (1 + 2/e) \cdot (C^* + F^*)$

Graph of distances in a fractional solution



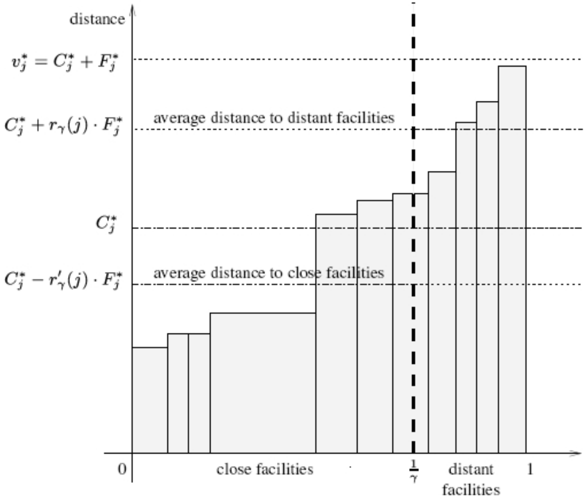
Optimality for regular instances



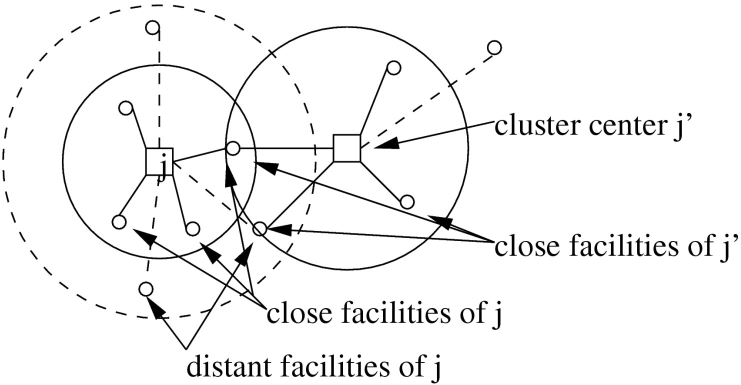
Our algorithm:

- solve LP relaxation
- scale facility opening variables *y* by a constant $\gamma > 1$
- adjust connection variables *x* to use closest open facilities
- compute clusters of clients in the modified solution
- randomly open facilities (like in the Chudak's algorithm)
- connect each client to his closest open facility

How sparsening looks on the graph of distances

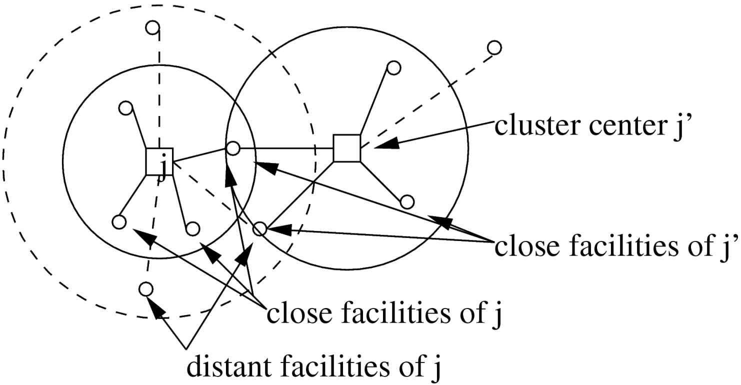


Close and distant facilities of a client



Expected support distance $\leq D^{D}_{av}(j) + D^{C}_{max}(j') + D^{C}_{av}(j')$.

Close and distant facilities of a client



Expected support distance $\leq D^D_{av}(j) + D^C_{max}(j') + D^C_{av}(j')$.

Calculating the expected connection cost

$$E[C_{SOL}] \leq$$
$$\sum_{j\in\mathcal{C}} (p_c \cdot D^C_{av}(j) + p_d \cdot D^D_{av}(j) + p_s \cdot (D^D_{av}(j) + D^C_{max}(j') + D^C_{av}(j')))$$
$$\leq \sum_{j\in\mathcal{C}} ((p_c + p_s) \cdot D^C_{av}(j) + (p_d + 2p_s) \cdot D^D_{av}(j)) =$$
$$\sum_{j\in\mathcal{C}} \Big((p_c + p_s) \cdot (C^*_j - r'_\gamma(j) \cdot F^*_j) + (p_d + 2p_s) \cdot (C^*_j + r_\gamma(j) \cdot F^*_j)\Big)$$
$$= ((p_c + p_d + p_s) + 2p_s) \cdot C^*$$
$$+ \sum_{j\in\mathcal{C}} \Big((p_c + p_s) \cdot (-r_\gamma(j) \cdot (\gamma - 1) \cdot F^*_j) + (p_d + 2p_s) \cdot (r_\gamma(j) \cdot F^*_j)\Big)$$
$$=$$
$$(1+2p_s)\cdot C^* + \sum_{j\in\mathcal{C}} \Big(F^*_j \cdot r_\gamma(j) \cdot (p_d + 2p_s - (\gamma - 1) \cdot (p_c + p_s))\Big)$$
$$\leq$$
$$(1+\tfrac{2}{e^\gamma})\cdot C^* + \sum_{j\in\mathcal{C}} \Big(F^*_j \cdot r_\gamma(j) \cdot (\tfrac{1}{e} + \tfrac{1}{e^\gamma} - (\gamma - 1) \cdot (1 - \tfrac{1}{e} + \tfrac{1}{e^\gamma}))\Big)$$

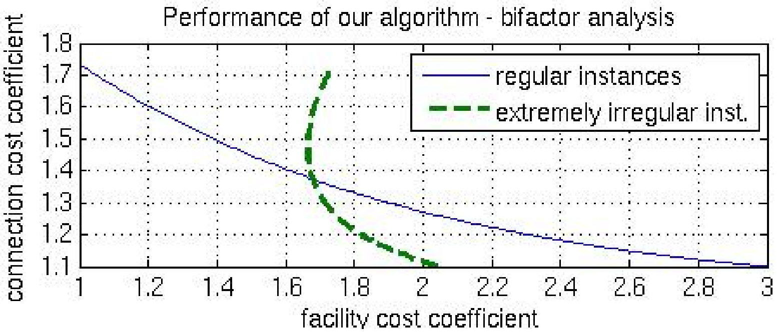
$\gamma = \gamma_0 \approx 1.67736$;
$E[C_{SOL} + F_{SOL}] \leq \gamma_0 \cdot F^* + (1 + \frac{2}{e^{\gamma_0}}) \cdot C^*$

Calculating the expected connection cost

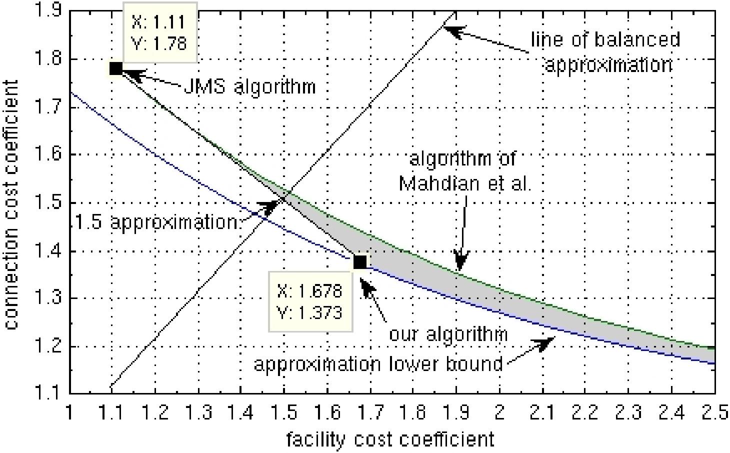
$$E[C_{SOL}] \leq$$
$$\sum_{j\in\mathcal{C}} (p_c \cdot D^C_{av}(j) + p_d \cdot D^D_{av}(j) + p_s \cdot (D^D_{av}(j) + D^C_{max}(j') + D^C_{av}(j')))$$
$$\leq \sum_{j\in\mathcal{C}} ((p_c + p_s) \cdot D^C_{av}(j) + (p_d + 2p_s) \cdot D^D_{av}(j)) =$$
$$\sum_{j\in\mathcal{C}} \Big((p_c + p_s) \cdot (C^*_j - r'_\gamma(j) \cdot F^*_j) + (p_d + 2p_s) \cdot (C^*_j + r_\gamma(j) \cdot F^*_j)\Big)$$
$$= ((p_c + p_d + p_s) + 2p_s) \cdot C^*$$
$$+ \sum_{j\in\mathcal{C}} \Big((p_c + p_s) \cdot (-r_\gamma(j) \cdot (\gamma - 1) \cdot F^*_j) + (p_d + 2p_s) \cdot (r_\gamma(j) \cdot F^*_j)\Big)$$
$$=$$
$$(1+2p_s)\cdot C^* + \sum_{j\in\mathcal{C}} \Big(F^*_j \cdot r_\gamma(j) \cdot (p_d + 2p_s - (\gamma - 1) \cdot (p_c + p_s))\Big)$$
$$\leq$$
$$(1+\tfrac{2}{e^\gamma})\cdot C^* + \sum_{j\in\mathcal{C}} \Big(F^*_j \cdot r_\gamma(j) \cdot (\tfrac{1}{e} + \tfrac{1}{e^\gamma} - (\gamma - 1) \cdot (1 - \tfrac{1}{e} + \tfrac{1}{e^\gamma}))\Big)$$

$$\gamma = \gamma_0 \approx 1.67736;$$
$$E[C_{SOL} + F_{SOL}] \leq \gamma_0 \cdot F^* + (1 + \tfrac{2}{e^{\gamma}_0}) \cdot C^*$$

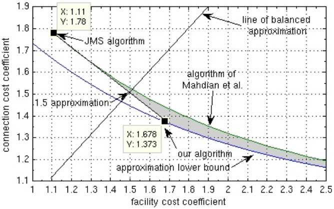
Graph of performance of our algorithm



Bifactor picture once more



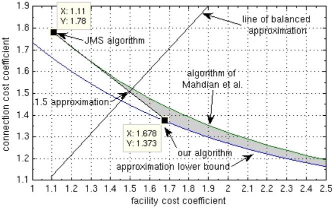
Combining two algorithms



Our (1.6774, 1.3738)-appr. with (1.11, 1.7764)-appr. of JMS:

- Randomly choose the algorithm:
  run JMS with pr. $p = 0.313$ and our alg. with pr. $1 - p$.
- If $\frac{F^*}{C^*} > \frac{2764}{3900}$ take JMS, else take our algorithm.

The main open problem



- Traversing the curve rightwards is easy (scaling + greedy augmentation).
- But how to go leftwards?

Random clustering would also work

Our algorithm

- solve LP relaxation
- scale facility opening variables *y* by a constant $\gamma > 1$
- adjust connection variables *x* to use closest open facilities
- compute clusters of clients in the modified solution
  (we could choose new cluster centers randomly)
- randomly open facilities (like in the Chudak's algorithm)
- connect each client to his closest open facility