Approximation Algorithms for Single and Multi-Commodity Connected Facility Location

Fabrizio Grandoni & Thomas Rothvoß

Department of Mathematics, M.I.T.

IPCO 2011

Massachusetts Institute of Technology

Alexander von Humboldt Stiftung / Foundation

Outline

1. Better approximation algorithm
for CONNECTED FACILITY LOCATION

Outline

1. Better approximation algorithm for Connected Facility Location
2. First $O(1)$-apx for Multi-Commodity Connected Facility Location

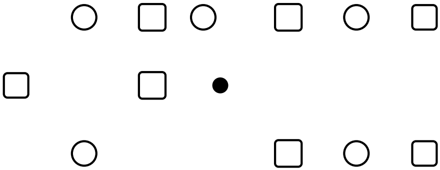
Outline

1. Better approximation algorithm for Connected Facility Location
2. First $O(1)$-apx for Multi-Commodity Connected Facility Location
3. Improved hardness results for several problems

Part 1: Connected Facility Location

Given:

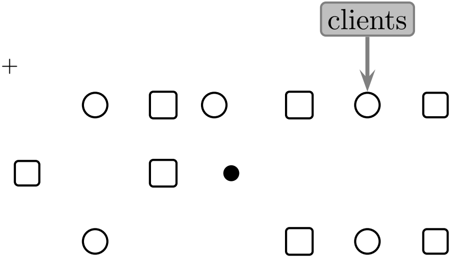
- graph $G = (V, E)$, with metric distances $c : E \to \mathbb{Q}^+$



Part 1: Connected Facility Location

Given:

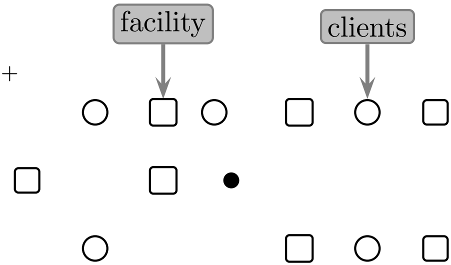
- graph $G = (V, E)$, with metric distances $c : E \to \mathbb{Q}^+$
- a set of **clients** $C \subseteq V$



Part 1: Connected Facility Location

Given:

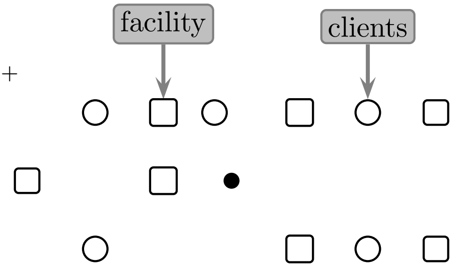
- graph $G = (V, E)$, with metric distances $c : E \to \mathbb{Q}^+$
- a set of **clients** $C \subseteq V$
- **facilities** $F \subseteq V$, with **opening cost** $o : F \to \mathbb{Q}_+$



Part 1: Connected Facility Location

Given:

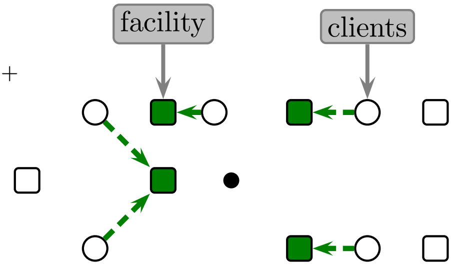
- graph $G = (V, E)$, with metric distances $c : E \to \mathbb{Q}^+$
- a set of **clients** $C \subseteq V$
- **facilities** $F \subseteq V$, with **opening cost** $o : F \to \mathbb{Q}_+$
- a parameter $M \geq 1$



Part 1: Connected Facility Location

Given:

- graph $G = (V, E)$, with metric distances $c : E \to \mathbb{Q}^+$
- a set of **clients** $C \subseteq V$
- **facilities** $F \subseteq V$, with **opening cost** $o : F \to \mathbb{Q}_+$
- a parameter $M \geq 1$



Goal:

- open facilities $F' \subseteq F$

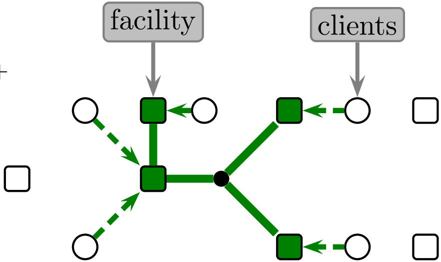
minimizing

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{\sum_{j \in C} c(j, F')}_{\text{connection cost}}$$

Part 1: Connected Facility Location

Given:

- graph $G = (V, E)$, with metric distances $c : E \to \mathbb{Q}^+$
- a set of **clients** $C \subseteq V$
- **facilities** $F \subseteq V$, with **opening cost** $o : F \to \mathbb{Q}_+$
- a parameter $M \geq 1$



Goal:

- open facilities $F' \subseteq F$
- find Steiner tree T spanning opened facilities

minimizing

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{\sum_{j \in C} c(j, F')}_{\text{connection cost}} \quad + \quad \underbrace{M \sum_{e \in T} c(e)}_{\text{Steiner cost}}$$

Previous Results for Connected Facility Location

- **APX**-hard (reduction from Steiner Tree)
- 10.66-apx based on LP-rounding [Gupta et al '01].
- 8.55-apx primal-dual algorithm [Swamy, Kumar '02].
- 3.92-apx based on random-sampling [Eisenbrand, Grandoni, R., Schäfer '08]

Previous Results for Connected Facility Location

- **APX**-hard (reduction from Steiner Tree)
- 10.66-apx based on LP-rounding [Gupta et al '01].
- 8.55-apx primal-dual algorithm [Swamy, Kumar '02].
- 3.92-apx based on random-sampling [Eisenbrand, Grandoni, R., Schäfer '08]

Our result

Simple 3.19-approximation algorithm.

Previous Results for Connected Facility Location

- **APX**-hard (reduction from Steiner Tree)
- 10.66-apx based on LP-rounding [Gupta et al '01].
- 8.55-apx primal-dual algorithm [Swamy, Kumar '02].
- 3.92-apx based on random-sampling [Eisenbrand, Grandoni, R., Schäfer '08]

Our result

Simple 3.19-approximation algorithm.

Use existing algorithms for subproblems:

- Facility Location: 1.5-apx [Byrka '07]
- Steiner Tree: 1.39-apx [Byrka,Grandoni,R.,Sanità 10]

Where is the difficulty?

Algorithm:

Where is the difficulty?

Algorithm:

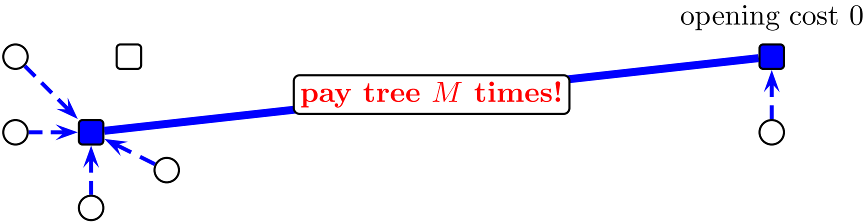
(1) Just solve Facility Location

Where is the difficulty?

Algorithm:

(1) Just solve FACILITY LOCATION

(2) Run STEINER TREE algorithm

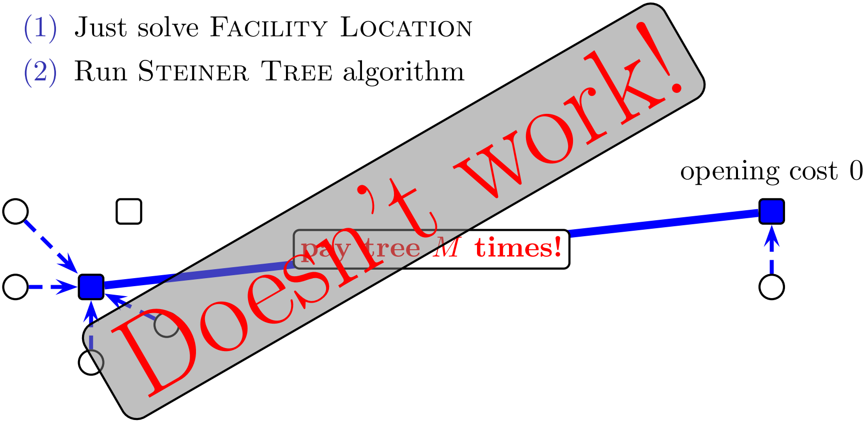


Where is the difficulty?

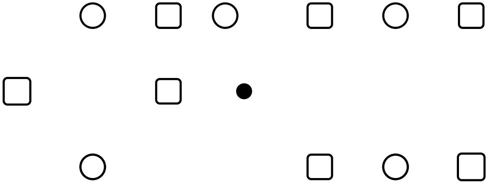
Algorithm:

(1) Just solve Facility Location

(2) Run Steiner Tree algorithm

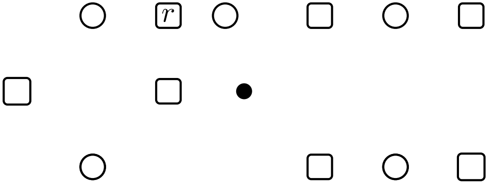


The CFL algorithm



The CFL algorithm

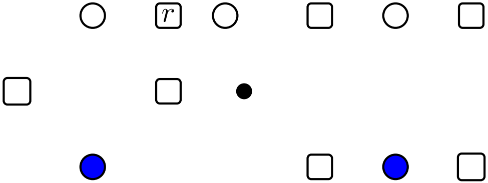
(1) Guess facility r from OPT.



The CFL algorithm

(1) Guess facility r from OPT.

(2) Sample each client with prob $\frac{1}{M} \to C'$

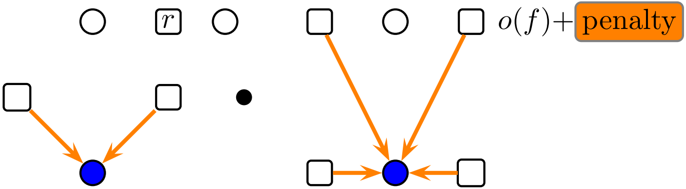


The CFL algorithm

(1) Guess facility r from OPT.

(2) Sample each client with prob $\frac{1}{M} \rightarrow C'$

(3) Compute 1.5-apx F' for Facility Location instance
$o'(f) := o(f) + M \cdot \boxed{c(f, \text{nearest node in } C' \cup \{r\})}$

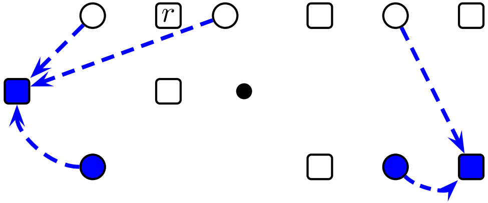


The CFL algorithm

(1) Guess facility r from OPT.

(2) Sample each client with prob $\frac{1}{M} \to C'$

(3) Compute 1.5-apx F' for Facility Location instance $o'(f) := o(f) + M \cdot \boxed{c(f, \text{nearest node in } C' \cup \{r\})}$



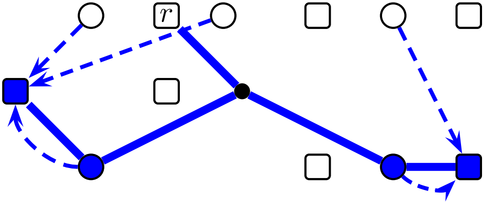
The CFL algorithm

(1) Guess facility r from OPT.

(2) Sample each client with prob $\frac{1}{M} \to C'$

(3) Compute 1.5-apx F' for Facility Location instance $o'(f) := o(f) + M \cdot \boxed{c(f, \text{nearest node in } C' \cup \{r\})}$

(4) Compute 1.39-apx Steiner Tree T' on $F' \cup \{r\}$



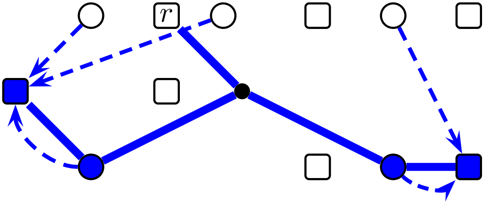
The CFL algorithm

(1) Guess facility r from OPT.

(2) Sample each client with prob $\frac{1}{M} \to C'$

(3) Compute 1.5-apx F' for Facility Location instance
$o'(f) := o(f) + M \cdot \boxed{c(f, \text{nearest node in } C' \cup \{r\})}$

(4) Compute 1.39-apx Steiner Tree T' on $F' \cup \{r\}$



- Observe:

$$E[APX] \leq E\begin{bmatrix} \text{cost of apx} \\ \text{Facility Location sol.} \\ +\text{penalties} \end{bmatrix} + E\begin{bmatrix} \text{cost of apx} \\ \text{Steiner Tree} \\ \end{bmatrix}$$

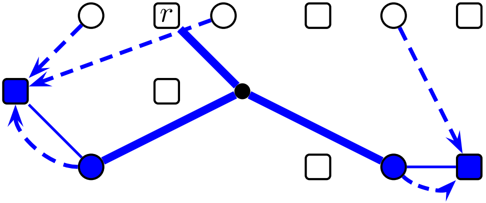
The CFL algorithm

(1) Guess facility r from OPT.

(2) Sample each client with prob $\frac{1}{M} \to C'$

(3) Compute 1.5-apx F' for Facility Location instance
$o'(f) := o(f) + M \cdot c(f, \text{nearest node in } C' \cup \{r\})$

(4) Compute 1.39-apx Steiner Tree T' on $F' \cup \{r\}$



- Observe:

$$E[APX] \leq E\begin{bmatrix} \text{cost of apx} \\ \text{Facility Location sol.} \\ +\text{penalties} \end{bmatrix} + E\begin{bmatrix} \text{cost of apx} \\ \text{Steiner Tree} \\ \text{on } C' \cup \{r\} \end{bmatrix}$$

Analysis: Steiner Cost

Notation:

$$OPT = \underbrace{O^*}_{\text{opening cost}} + \underbrace{S^*}_{\text{Steiner tree cost}} + \underbrace{C^*}_{\text{connection cost}}$$

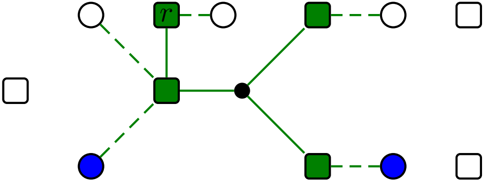
Analysis: Steiner Cost

Notation:

$$OPT = \underbrace{O^*}_{\text{opening cost}} + \underbrace{S^*}_{\text{Steiner tree cost}} + \underbrace{C^*}_{\text{connection cost}}$$

Lemma

$E[M \cdot apx$ Steiner Tree *on* $C' \cup \{r\}] \leq 1.39 \cdot (S^* + C^*)$



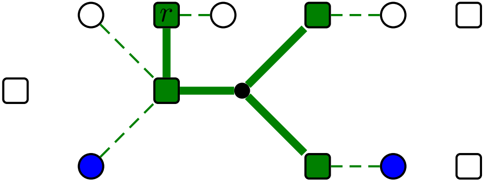
Analysis: Steiner Cost

Notation:

$$OPT = \underbrace{O^*}_{\text{opening cost}} + \underbrace{S^*}_{\text{Steiner tree cost}} + \underbrace{C^*}_{\text{connection cost}}$$

Lemma

$E[M \cdot apx$ Steiner Tree on $C' \cup \{r\}] \leq 1.39 \cdot (S^* + C^*)$



$\underbrace{S^*}_{\substack{\text{Steiner tree} \\ \text{in } OPT}}$

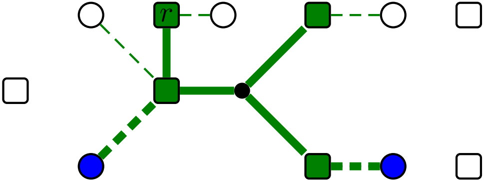
Analysis: Steiner Cost

Notation:

$$OPT = \underbrace{O^*}_{\text{opening cost}} + \underbrace{S^*}_{\text{Steiner tree cost}} + \underbrace{C^*}_{\text{connection cost}}$$

Lemma

$E[M \cdot apx$ Steiner Tree $on\ C' \cup \{r\}] \leq 1.39 \cdot (S^* + C^*)$



$$\underbrace{S^*}_{\substack{\text{Steiner tree}\\ \text{in } OPT}} + \underbrace{\frac{1}{M} \cdot M \cdot C^*}_{\substack{\text{need path in } C^*\\ \text{with prob. } 1/M}}$$

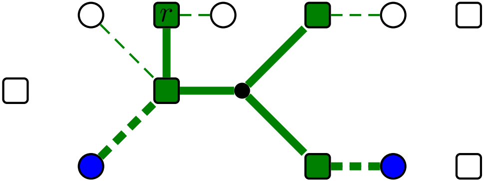
Analysis: Steiner Cost

Notation:

$$OPT = \underbrace{O^*}_{\text{opening cost}} + \underbrace{S^*}_{\text{Steiner tree cost}} + \underbrace{C^*}_{\text{connection cost}}$$

Lemma

$E[M \cdot apx$ Steiner Tree $on\ C' \cup \{r\}] \le 1.39 \cdot (S^* + C^*)$



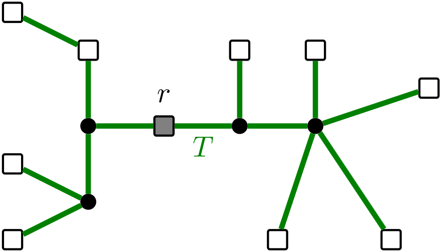
$$1.39 \cdot \Bigg(\underbrace{S^*}_{\substack{\text{Steiner tree} \\ \text{in } OPT}} + \underbrace{\frac{1}{M} \cdot M \cdot C^*}_{\substack{\text{need path in } C^* \\ \text{with prob. } 1/M}} \Bigg)$$

Core detouring theorem

Core Detouring Theorem [EGSR '08]

Given a spanning tree T with root $r \in T$ and distinguished terminals $D \subseteq V$. Sample any terminal in D with prob. $p \in]0,1]$

$$E\Big[\sum_{v \in D} \mathrm{dist}(v, \text{sampled node} \cup \{r\})\Big] \leq \frac{0.81}{p} \cdot c(T)$$

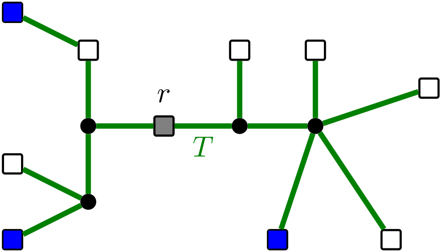


Core detouring theorem

Core Detouring Theorem [EGSR '08]

Given a spanning tree T with root $r \in T$ and distinguished terminals $D \subseteq V$. Sample any terminal in D with prob. $p \in]0,1]$

$$E\Big[\sum_{v \in D} \mathrm{dist}(v, \text{sampled node} \cup \{r\})\Big] \leq \frac{0.81}{p} \cdot c(T)$$

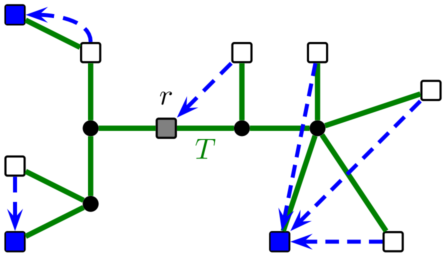


Core detouring theorem

Core Detouring Theorem [EGSR '08]

Given a spanning tree T with root $r \in T$ and distinguished terminals $D \subseteq V$. Sample any terminal in D with prob. $p \in]0, 1]$

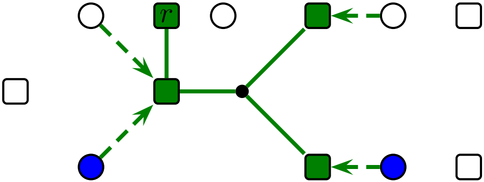
$$E\Big[\sum_{v \in D} \mathrm{dist}(v, \text{sampled node} \cup \{r\})\Big] \leq \frac{0.81}{p} \cdot c(T)$$



Analysis: Facility Location cost

Theorem

$E[apx \text{ Facility Location } cost] \leq 1.5 \cdot (O^* + 2C^* + 0.81S^*)$

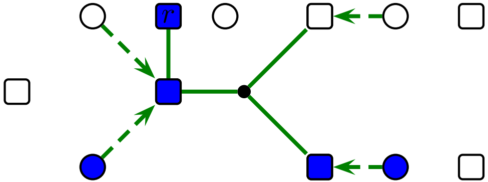


Analysis: Facility Location cost

Theorem

$E[apx \text{ Facility Location } cost] \leq 1.5 \cdot (O^* + 2C^* + 0.81S^*)$

- There is a good sol: $F :=$ facilities in OPT serving $C' \cup \{r\}$

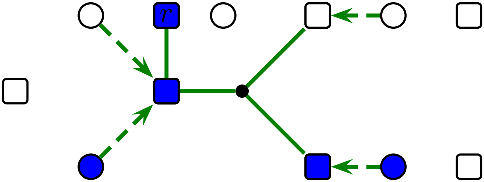


Analysis: Facility Location cost

Theorem

$E[apx \text{ Facility Location } cost] \leq 1.5 \cdot (O^* + 2C^* + 0.81S^*)$

- There is a good sol: $F :=$ facilities in OPT serving $C' \cup \{r\}$
- $E\left[\begin{array}{c}\text{opening cost}\\ +\text{ penalties}\end{array}\right] \leq \underbrace{O^*}_{\text{orig. cost}} + \underbrace{\frac{1}{M} \cdot M \cdot C^*}_{\text{expected penalty}}$

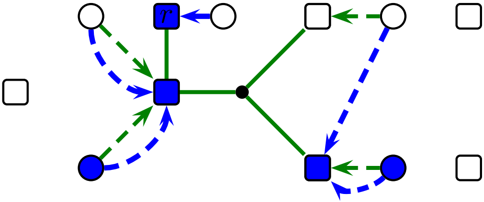


Analysis: Facility Location cost

Theorem

$E[apx \text{ Facility Location } cost] \leq 1.5 \cdot (O^* + 2C^* + 0.81S^*)$

- There is a good sol: $F :=$ facilities in OPT serving $C' \cup \{r\}$
- $E\left[\begin{array}{c}\text{opening cost}\\ + \text{penalties}\end{array}\right] \leq \underbrace{O^*}_{\text{orig. cost}} + \underbrace{\frac{1}{M} \cdot M \cdot C^*}_{\text{expected penalty}}$
- $E[\text{connection cost}] \leq C^* +$



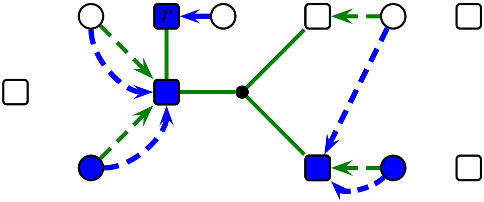
Analysis: Facility Location cost

Theorem

$E[apx \text{ FACILITY LOCATION } cost] \leq 1.5 \cdot (O^* + 2C^* + 0.81 S^*)$

- There is a good sol: $F :=$ facilities in OPT serving $C' \cup \{r\}$
- $E\left[\begin{array}{c}\text{opening cost}\\ +\text{ penalties}\end{array}\right] \leq \underbrace{O^*}_{\text{orig. cost}} + \underbrace{\frac{1}{M} \cdot M \cdot C^*}_{\text{expected penalty}}$
- $E[\text{connection cost}] \leq C^* + \frac{0.81}{1/M} \cdot \underbrace{\frac{S^*}{M}}_{\text{tree}}$

Use **Core Detouring Theorem** with $T = S^*$ and $p := \frac{1}{M}$.



Finishing the analysis for CFL

Conclusion:

$$E[APX] \quad \leq \quad 1.39 \cdot (S^* + C^*) + 1.5 \cdot (O^* + 2C^* + 0.81S^*)$$

Finishing the analysis for CFL

Conclusion:

$$\begin{aligned} E[APX] &\leq 1.39 \cdot (S^* + C^*) + 1.5 \cdot (O^* + 2C^* + 0.81S^*) \\ &\leq 1.5 \cdot O^* + 2.7 \cdot S^* + 4.39 \cdot C^* \end{aligned}$$

Finishing the analysis for CFL

Conclusion:

$$\begin{aligned} E[APX] &\leq 1.39 \cdot (S^* + C^*) + 1.5 \cdot (O^* + 2C^* + 0.81S^*) \\ &\leq 1.5 \cdot O^* + 2.7 \cdot S^* + 4.39 \cdot C^* \end{aligned}$$

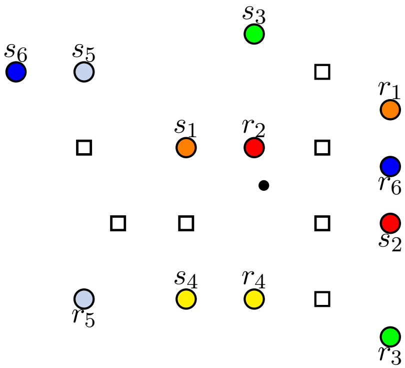
Improvement to 3.19:

- Adapting the sampling probability
- Using a bi-factor Facility Location algorithm

Part 2: Multi-Commodity Connected Facility Location

Input:

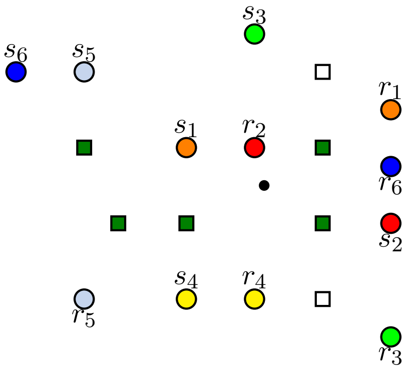
- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$



Part 2: Multi-Commodity Connected Facility Location

Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$



Goal: Find

- Facilities $F' \subseteq F$

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}}$$

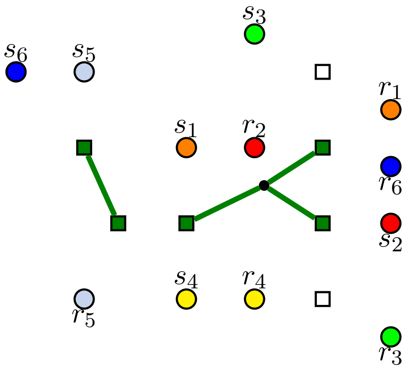
Part 2: Multi-Commodity Connected Facility Location

Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$

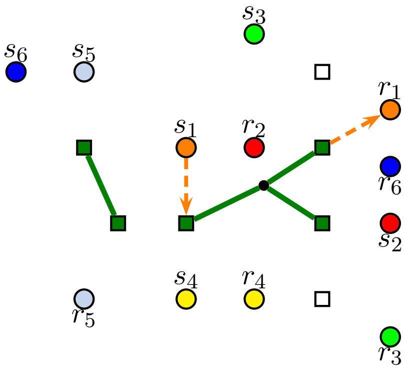
Goal: Find

- Facilities $F' \subseteq F$
- Forest T



$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}}$$

Part 2: Multi-Commodity Connected Facility Location



Input:

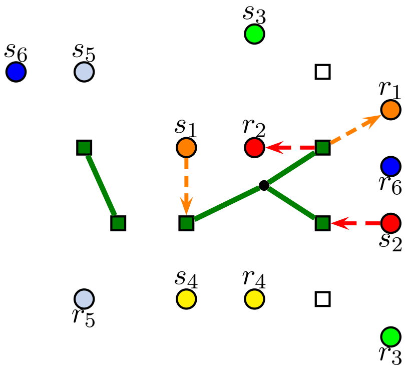
- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$

Goal: Find

- Facilities $F' \subseteq F$
- Forest T

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}} + \sum_{i=1}^{k} \underbrace{\mathrm{dist}_{F',T}(s_i, r_i)}_{\substack{\text{0 distance between}\\ \text{opened \& connected}\\ \text{facilities}}}$$

Part 2: Multi-Commodity Connected Facility Location



Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$

Goal: Find

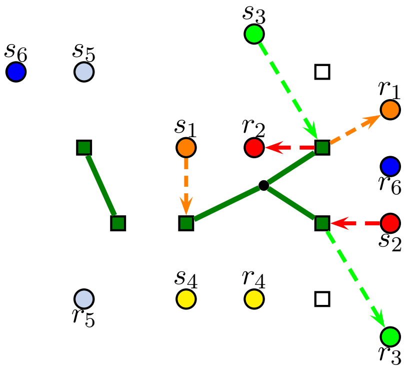
- Facilities $F' \subseteq F$
- Forest T

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}} + \sum_{i=1}^{k} \underbrace{\mathrm{dist}_{F',T}(s_i, r_i)}_{\substack{\text{0 distance between} \\ \text{opened \& connected} \\ \text{facilities}}}$$

Part 2: Multi-Commodity Connected Facility Location

Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$



Goal: Find

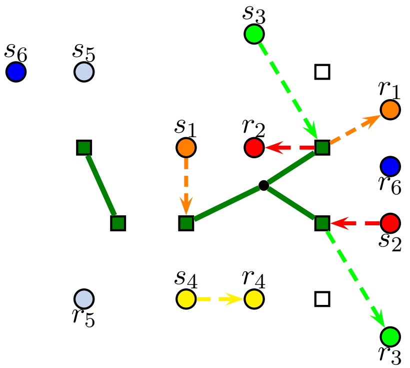
- Facilities $F' \subseteq F$
- Forest T

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}} + \sum_{i=1}^{k} \underbrace{\mathrm{dist}_{F',T}(s_i, r_i)}_{\substack{\text{0 distance between} \\ \text{opened \& connected} \\ \text{facilities}}}$$

Part 2: Multi-Commodity Connected Facility Location

Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$



Goal: Find

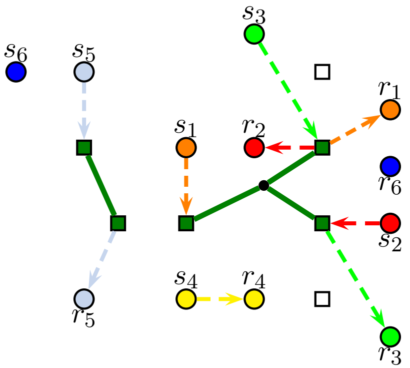
- Facilities $F' \subseteq F$
- Forest T

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}} + \sum_{i=1}^{k} \underbrace{\text{dist}_{F',T}(s_i, r_i)}_{\substack{\text{0 distance between} \\ \text{opened \& connected} \\ \text{facilities}}}$$

Part 2: Multi-Commodity Connected Facility Location

Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$



Goal: Find

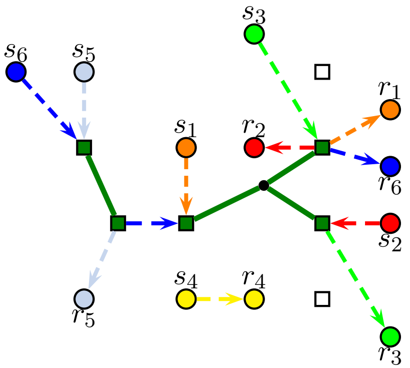
- Facilities $F' \subseteq F$
- Forest T

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}} + \sum_{i=1}^{k} \underbrace{\mathrm{dist}_{F',T}(s_i, r_i)}_{\text{0 distance between opened \& connected facilities}}$$

Part 2: Multi-Commodity Connected Facility Location

Input:

- Undirected graph $G = (V, E)$, metric distances $c : E \to \mathbb{Q}^+$
- source-sink pairs $(s_1, r_1), \ldots, (s_k, r_k)$
- a set of facilities $F \subseteq V$ with opening costs $o : F \to \mathbb{Q}^+$
- parameter $M \geq 1$



Goal: Find

- Facilities $F' \subseteq F$
- Forest T

$$\underbrace{\sum_{f \in F'} o(f)}_{\text{opening cost}} + \underbrace{M \cdot c(T)}_{\text{forest cost}} + \sum_{i=1}^{k} \underbrace{\mathrm{dist}_{F',T}(s_i, r_i)}_{\substack{\text{0 distance between} \\ \text{opened \& connected} \\ \text{facilities}}}$$

Our result

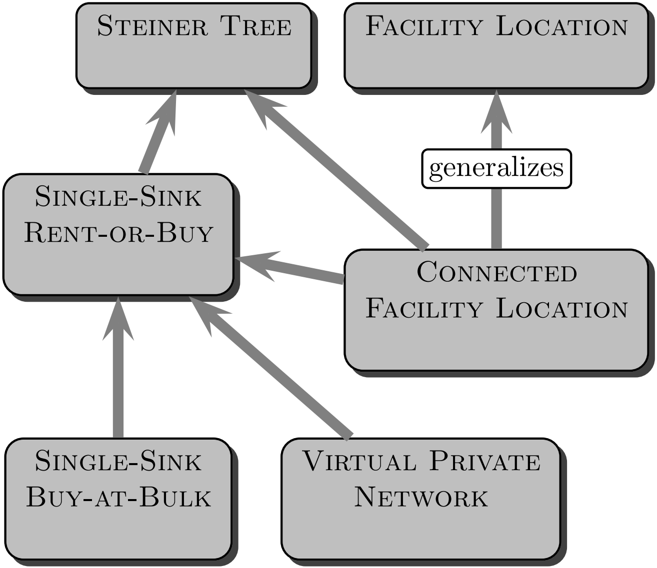
Our result

Simple 16.2-approximation algorithm for Multi-Commodity Connected Facility Location.

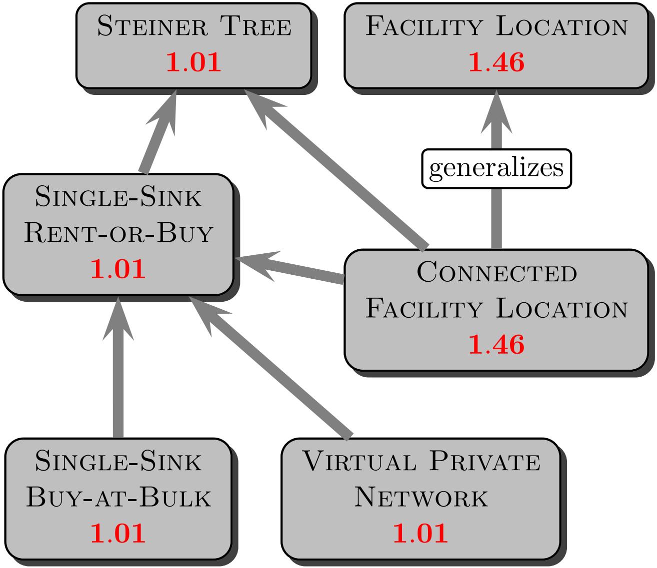
Ingredients:

- Random-sampling
- Use algorithms for
  - Price-Collecting Facility Location
  - Steiner Forest

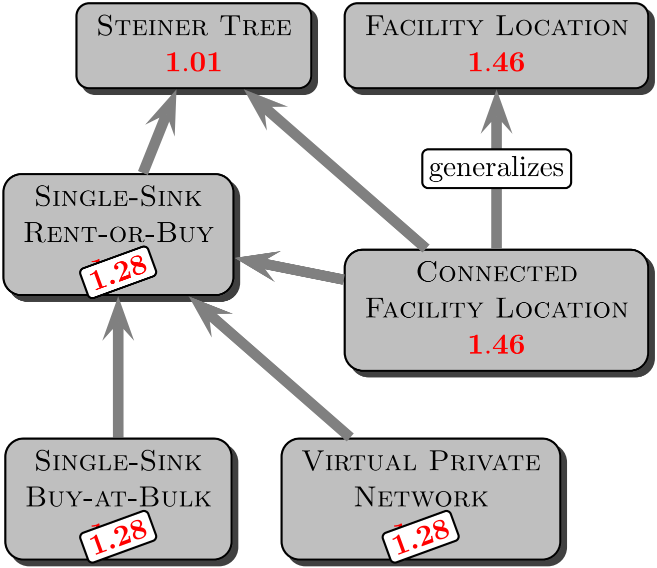
Part 3: Improved hardness results



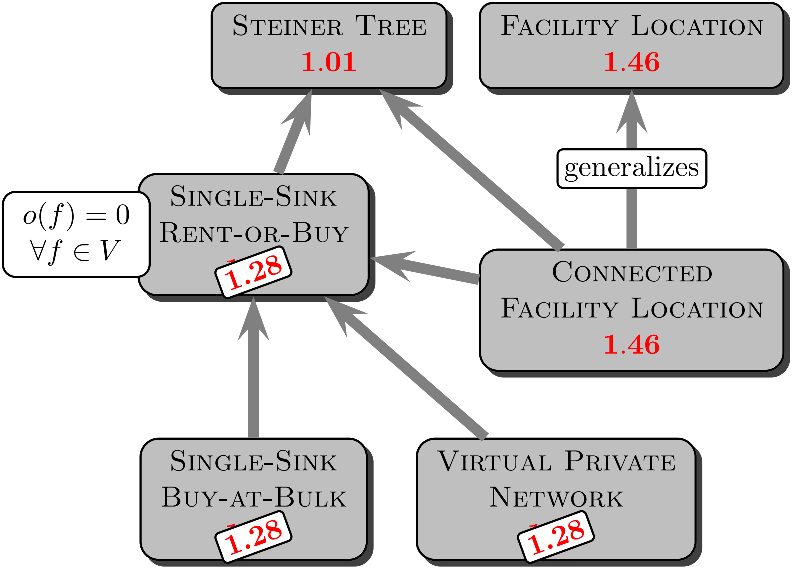
Part 3: Improved hardness results



Part 3: Improved hardness results

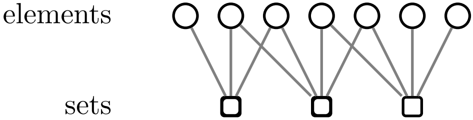


Part 3: Improved hardness results



The reduction (1)

- Reduce Set Cover to SROB

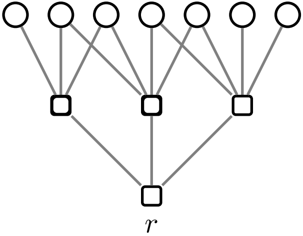


The reduction (1)

- Reduce Set Cover to SROB

clients/elements

facilities/sets

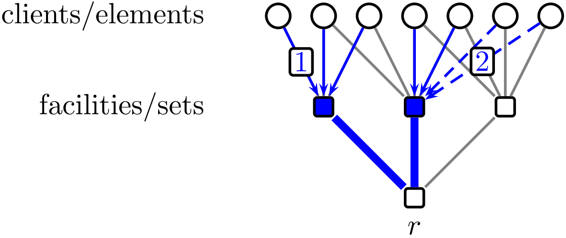


$c(e) \in \{1, 2\}$

$M = \dfrac{0.27n}{OPT_{SC}}$

The reduction (1)

- Reduce Set Cover to SROB



$c(e) \in \{1, 2\}$

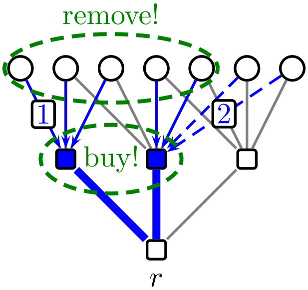
$$M = \frac{0.27n}{OPT_{SC}}$$

The reduction (1)

- Reduce SET COVER to SROB

clients/elements

facilities/sets



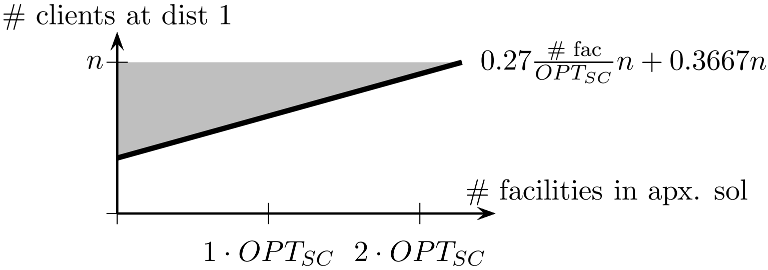
$c(e) \in \{1, 2\}$

$$M = \frac{0.27n}{OPT_{SC}}$$

(1) WHILE not all elements covered DO
   (2) Compute 1.27-apx SROB sol
   (3) Buy facilities/sets in sol. & remove covered elements

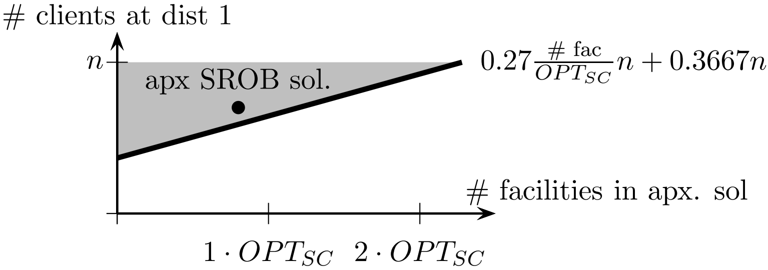
The reduction (2)

- Use idea from [Guha & Khuller '99]



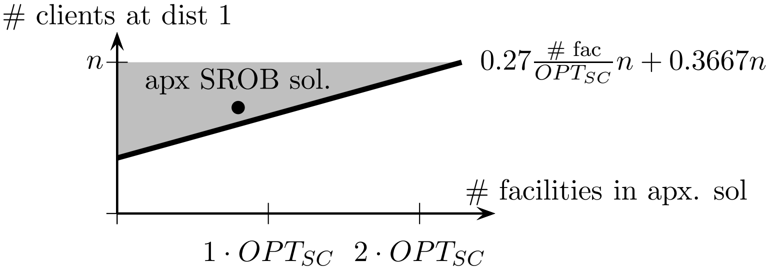
The reduction (2)

- Use idea from [Guha & Khuller '99]



The reduction (2)

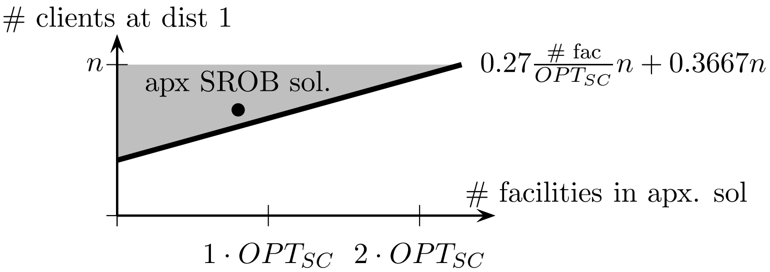
- Use idea from [Guha & Khuller '99]



$$\#\text{ needed sets} \quad \leq \quad [\ldots\text{some calc}\ldots] \leq 0.999\ln(n)\cdot OPT_{SC}$$

The reduction (2)

- Use idea from [Guha & Khuller '99]



$$\#\text{ needed sets} \quad \leq \quad [\ldots\text{some calc}\ldots] \leq 0.999 \ln(n) \cdot OPT_{SC}$$

- Contradiction!

Theorem (Feige '98)

Unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log\log n)})$, *there is no* $(1-\varepsilon)\cdot\ln(n)$*-apx for* Set Cover.

The end

Thanks for your attention