Approximating Connected Facility Location Problems via Random Facility Sampling and Core Detouring

Thomas Rothvoß

Institute of Mathematics
École Polytechnique Fédérale de Lausanne

This is joint work with Fritz Eisenbrand, Fabrizio Grandoni & Guido Schäfer

DISCRETE OPTIMIZATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Connected Facility Location

Given:

- graph $G = (V, E)$, with edge costs $c : E \to \mathbb{Q}^+$
- *facilities* $\mathcal{F} \subseteq V$, with *opening cost* $f : \mathcal{F} \to \mathbb{Q}^+$
- a set of *demands* $\mathcal{D} \subseteq V$
- a parameter $M \geq 1$,

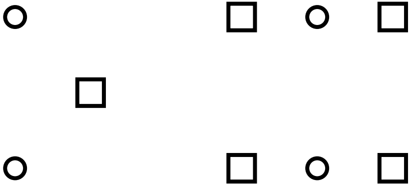
Goal:

- open facilities $F \subseteq \mathcal{F}$
- find Steiner tree T spanning opened facilities

minimizing

$$\underbrace{\sum_{i \in F} f(i)}_{\text{opening cost } O} \quad + \quad \underbrace{M \sum_{e \in T} c(e)}_{\text{Steiner cost } S} \quad + \quad \underbrace{\sum_{j \in \mathcal{D}} \ell(j, F)}_{\text{connection cost } C}$$

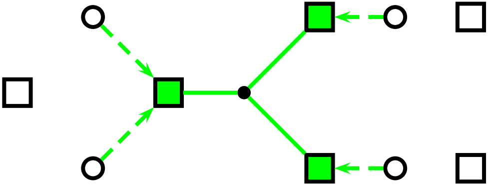
The Problem



□ = facility

○ = demand

The Problem



= facility

= demand

= open facility

= connection path

= Steiner tree edge

Previous Results

- APX-hard (reduction from Steiner tree)
- 10.66 worst-case approximation based on LP-rounding [Gupta et al STOC'01].
- primal-dual 8.55 worst-case approximation [Swamy, Kumar APPROX'02].

Previous Results

- APX-hard (reduction from Steiner tree)
- 10.66 worst-case approximation based on LP-rounding [Gupta et al STOC'01].
- primal-dual 8.55 worst-case approximation [Swamy, Kumar APPROX'02].

$\Rightarrow$ Here: 4-approximation

Single-Sink Rent-or-Buy

Def: CFL with $\mathcal{F} = V$ and $f(i) = 0$

Known Results:

- 9.01 worst-case approximation based on LP-rounding [Gupta et al STOC'01].
- primal-dual 4.55 worst-case approximation [Swamy, Kumar APPROX'02].
- 3.55 expected approximation based on random-sampling [Gupta, Kumar, Roughgarden STOC'03].
- 4.2 worst-case approximation based on the derandomized of GK&R [Gupta, Srinivasan, Tardos APPROX'04].
- 4.0 worst-case approximation based on the derandomized of GK&R [van Zuylen, Williamson ORL].

Single-Sink Rent-or-Buy

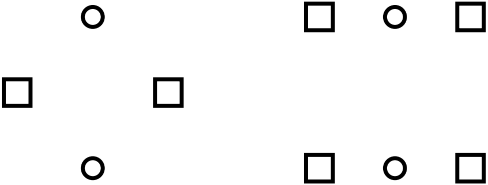
Def: CFL with $\mathcal{F} = V$ and $f(i) = 0$

Known Results:

- 9.01 worst-case approximation based on LP-rounding [Gupta et al STOC'01].
- primal-dual 4.55 worst-case approximation [Swamy, Kumar APPROX'02].
- 3.55 expected approximation based on random-sampling [Gupta, Kumar, Roughgarden STOC'03].
- 4.2 worst-case approximation based on the derandomized of GK&R [Gupta, Srinivasan, Tardos APPROX'04].
- 4.0 worst-case approximation based on the derandomized of GK&R [van Zuylen, Williamson ORL].

$\Rightarrow$ Here: 2.92-approximation

The Algorithm



- - = FL approximate solution

= sampled demand

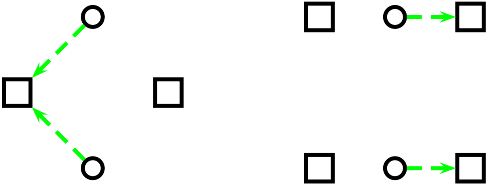
= sampled facility

— = Steiner tree in APX

- - = connections in APX

The Algorithm

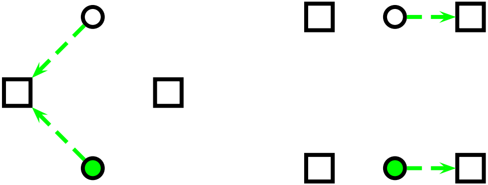
1. Run a (black-box) algo for (unconnected) facility location.



- - = FL approximate solution
○ = sampled demand
□ = sampled facility
— = Steiner tree in APX
- - = connections in APX

The Algorithm

1. Run a (black-box) algo for (unconnected) facility location.
2. Mark each demand independently with prob. $p \approx 1/M$
3. Mark a random demand



- - = FL approximate solution

= sampled demand

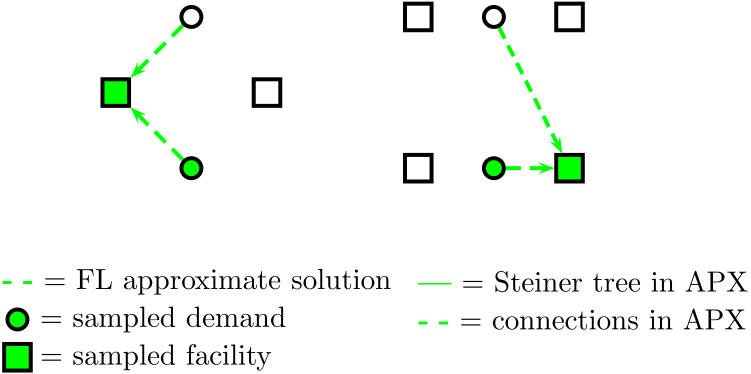
= sampled facility

— = Steiner tree in APX

- - = connections in APX

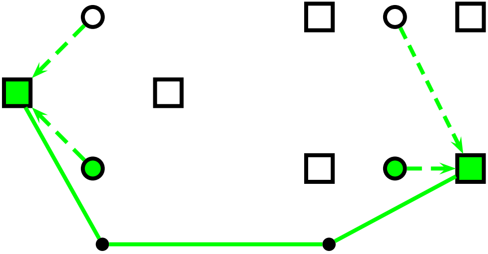
The Algorithm

1. Run a (black-box) algo for (unconnected) facility location.
2. Mark each demand independently with prob. $p \approx 1/M$
3. Mark a random demand
4. Open the facilities, serving sampled demand.



The Algorithm

1. Run a (black-box) algo for (unconnected) facility location.
2. Mark each demand independently with prob. $p \approx 1/M$
3. Mark a random demand
4. Open the facilities, serving sampled demand.
5. Compute 1.55-approx. Steiner Tree T spanning opened facilities



- - = FL approximate solution

= sampled demand

= sampled facility

— = Steiner tree in APX

- - = connections in APX

What to do...

We will bound separately

- Opening cost
- Steiner tree cost
- Connection cost

Each cost type will bounded by combination of

- O^* = opening cost
- S^* = Steiner tree cost
- C^* = connection cost

} in opt. solution

- O_{fl} = opening cost
- C_{fl} = connection cost

} in approx. facility location solution

We use

- $O^* + S^* + C^* = OPT$
- $O_{fl} + C_{fl} \leq 1.52 \cdot OPT_{fl} \leq 1.52 \cdot OPT$

Assume $|\mathcal{D}|/M \gg 1$ (otherwise we have a PTAS)

Analysis: Opening Cost

Lemma

The opening cost O_{apx} satisfies $O_{apx} \leq O_{fl}$.

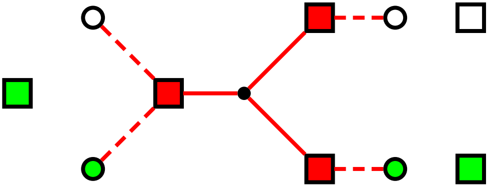
O_{fl} and C_{fl} denote the opening and connection costs in the approximate (unconnected) facility location solution computed in the first step of the algorithm.

Analysis: Steiner Cost

Lemma

The Steiner cost S_{apx} satisfies

$$E[S_{apx}] \leq \rho_{st}(S^* + pM(1 + o(1)) \cdot C^*) + pM(1 + o(1)) \cdot C_{fl})$$

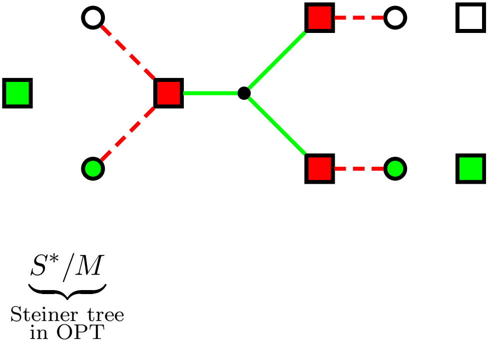


Analysis: Steiner Cost

Lemma

The Steiner cost S_{apx} satisfies

$$E[S_{apx}] \leq \rho_{st}(S^* + pM(1 + o(1)) \cdot C^*) + pM(1 + o(1)) \cdot C_{fl})$$

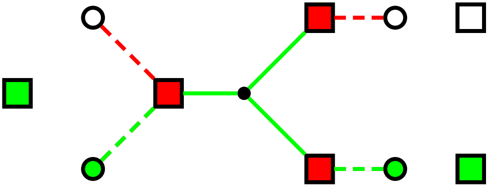


Analysis: Steiner Cost

Lemma

The Steiner cost S_{apx} satisfies

$$E[S_{apx}] \leq \rho_{st}(S^* + pM(1+o(1)) \cdot C^*) + pM(1+o(1)) \cdot C_{fl})$$



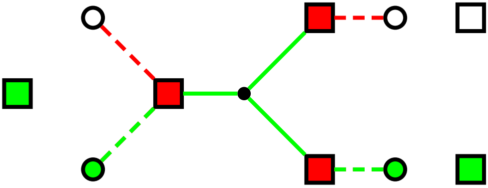
$$\underbrace{S^*/M}_{\text{Steiner tree in OPT}} + \underbrace{\left(p + \frac{1}{|\mathcal{D}|}\right) C^*}_{\text{paths sampled demands - Steiner tree}}$$

Analysis: Steiner Cost

Lemma

The Steiner cost S_{apx} satisfies

$$E[S_{apx}] \leq \rho_{st}(S^* + pM(1 + o(1)) \cdot C^*) + pM(1 + o(1)) \cdot C_{fl})$$



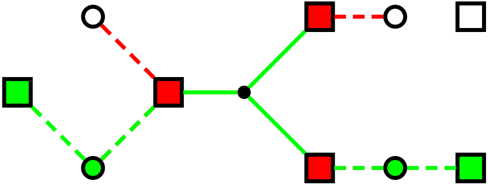
$$\underbrace{\rho_{St}}_{<1.55} \cdot \Big(\underbrace{S^*/M}_{\text{Steiner tree in OPT}} + \underbrace{(p + \frac{1}{|\mathcal{D}|})\, C^*}_{\text{paths sampled demands - Steiner tree}} \Big)$$

Analysis: Steiner Cost

Lemma

The Steiner cost S_{apx} satisfies

$$E[S_{apx}] \leq \rho_{st}(S^* + pM(1+o(1)) \cdot C^*) + pM(1+o(1)) \cdot C_{fl})$$



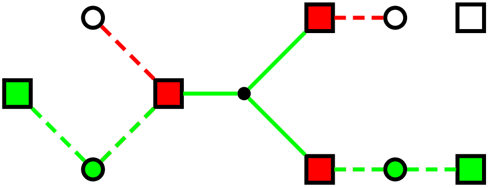
$$\underbrace{\rho_{St}}_{<1.55} \cdot \Big(\underbrace{S^*/M}_{\text{Steiner tree in OPT}} + \underbrace{\left(p + \frac{1}{|\mathcal{D}|}\right) C^*}_{\text{paths sampled demands - Steiner tree}} \Big) + \underbrace{\left(p + \frac{1}{|\mathcal{D}|}\right) C_{fl}}_{\text{paths sampled demands - sampled facilities}}$$

Analysis: Steiner Cost

Lemma

The Steiner cost S_{apx} satisfies

$$E[S_{apx}] \leq \rho_{st}(S^* + pM(1+o(1)) \cdot C^*) + pM(1+o(1)) \cdot C_{fl})$$



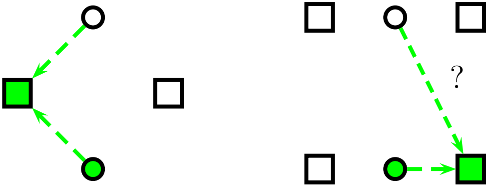
$$M \cdot (\underbrace{\rho_{St}}_{<1.55} \cdot (\underbrace{S^*/M}_{\text{Steiner tree in OPT}} + \underbrace{(p + \frac{1}{|\mathcal{D}|})\, C^*}_{\text{paths sampled demands - Steiner tree}}) + \underbrace{(p + \frac{1}{|\mathcal{D}|})\, C_{fl}}_{\text{paths sampled demands - sampled facilities}})$$

Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$

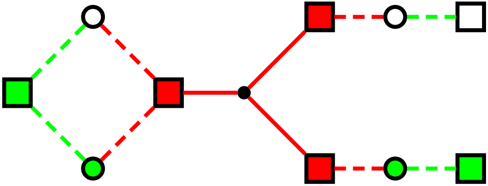


Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$



■ = facility in OPT

- - = connections in OPT

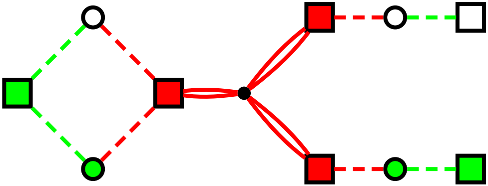
— = Steiner tree in OPT

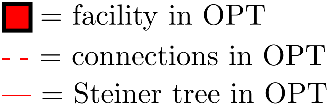
Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$



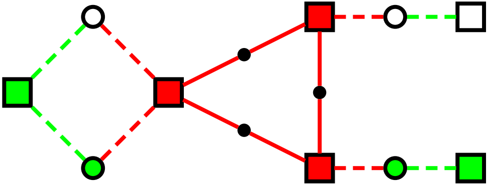


Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$

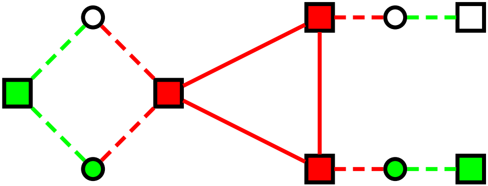


Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$

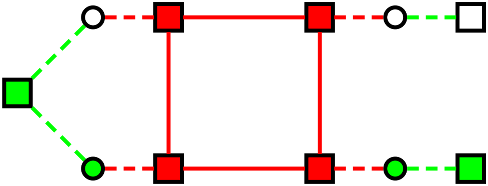


Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$

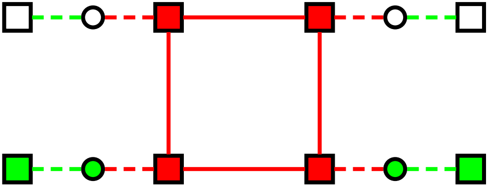


Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$

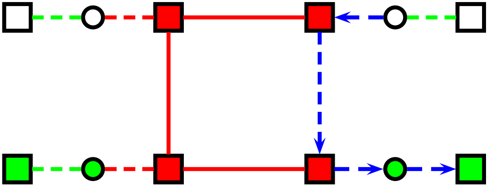


Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$



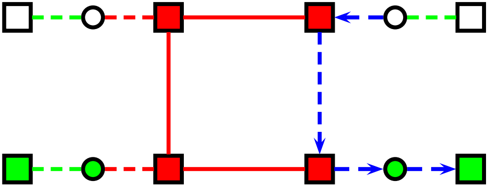
- Connect each demand to the closest open facility w.r.t. number of edges in the auxiliary graph above

Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$



By symmetry:

$E[$flow on  $] \leq 2$

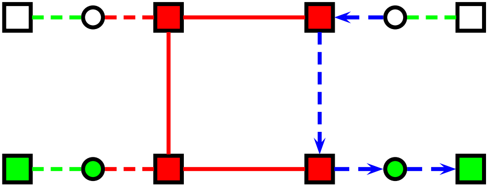
$E[$flow on  $] \leq 1$

Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$



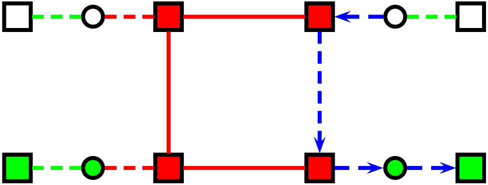
- $X_i :=$ # of cycle edges used by $i \in D$
- #demands = #cycle edges
- By symmetry: $E[\text{flow on cycle edge}] = E[X_i]$

Analysis: Connection Cost

Lemma

The connection cost C_{apx} satisfies

$$E[C_{apx}] \leq 2C^* + C_{fl} + S^*/(p\,M).$$



$$\Pr[X_i > k] \leq (1-p)^{2k+1}$$

$$E[X_i] = \sum_{k \geq 1} \Pr[X_i \geq k] = \sum_{k \geq 1} (1-p)^{2k-1} \leq \frac{1}{2} \cdot \frac{1}{p}$$

$$E[\text{cost of used cycle edges}] \leq \frac{1}{2} \cdot \frac{1}{p} \cdot \frac{2S^*}{M} = S^*/(pM)$$

Total Cost

Theorem

There is an expected 4.55 *approximation algorithm for CFL.*

Proof:

$$
\begin{aligned}
& E[APX] \\
\leq\ & O_{fl} + \rho_{st}(S^* + pMC^*) + pMC_{fl} + 2C^* + C_{fl} + \frac{S^*}{pM} \\
\leq\ & (1+pM)(O_{fl} + C_{fl}) + \rho_{st}(S^* + pMC^*) + 2C^* + \frac{S^*}{pM} \\
\overset{O^*_{fl}+C^*_{fl}\leq O^*+C^*}{\leq}\ & (1+pM)\rho_{fl}(O^* + C^*) + \rho_{st}(S^* + pMC^*) + 2C^* + \frac{S^*}{pM} \\
\overset{p=0.334/M}{\leq}\ & 2.03\,O^* + 4.55\,C^* + 4.55\,S^* \\
\leq\ & 4.55\,OPT
\end{aligned}
$$

Refinements

Theorem

There is an expected 4 approximation algorithm for CFL.

- Using bi-factor facility location, for a parameter $\delta \geq 1$, we obtain:

$$C_{fl} + O_{fl} \leq (1.11 + \ln \delta) O^* + (1 + 0.78/\delta) C^*.$$

- Using flow cancelling over the Euler tour, we can show that (for $|\mathcal{D}|/M \gg 1$)

$$C_{apx} \leq 2C^* + C_{fl} + 0.807 \frac{S^*}{pM}.$$

Refinements

Theorem

There is a 4.23 worst-case approximation algorithm for CFL.

Proof: Use idea of van Zuylen and Williamson to estimate expected Steiner cost

Corollary

There is an expected 2.92 and worst-case 3.28 approximation algorithm for SROB.

Proof: Same analysis with $C_{fl} = O_{fl} = 0$.

Summarizing

More results

Problem	**Now**	**Previous best**	
CFL	4.00*	8.55	Swamy, Kumar '02
	4.23		
SROB	2.92*	3.55*	Gupta, Kumar, Roughgarden '03
	3.28	4	van Zuylen, Williamson '07
k-CFL	6.85*	15.55*	Swamy and Kumar '02
	6.98		
tour-CFL	4.12*	5.83*	Ravi, Salman '99 (special case)
soft-CFL	6.33*	–	

The * indicates randomized results

A CFL PTAS for $|\mathcal{D}|/M = O(1)$

Algorithm:

1. Guess the best choice of $\leq k$ facilities F
2. Compute an optimum Steiner tree spanning F

A CFL PTAS for $|\mathcal{D}|/M = O(1)$

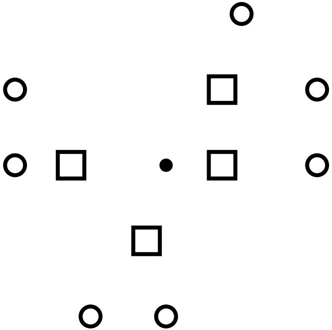
Algorithm:

1. Guess the best choice of $\leq k$ facilities F
2. Compute an optimum Steiner tree spanning F

Lemma

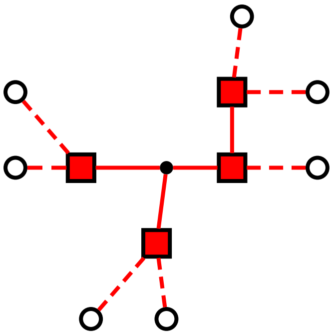
For $k := \frac{2|\mathcal{D}|}{\varepsilon M}$ the algorithm yields a solution of cost $\leq (1+\varepsilon)OPT$.

Proof



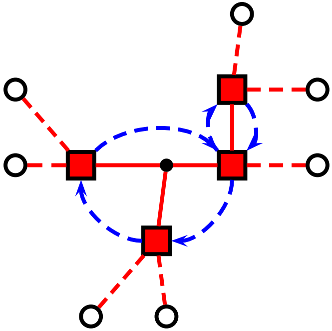
Proof

- T^* := Steiner tree in opt solution



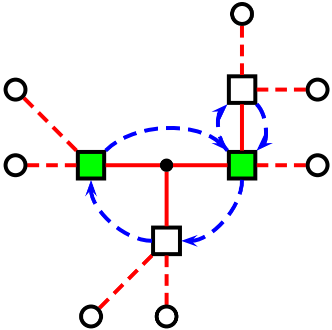
Proof

- T^* := Steiner tree in opt solution
- C := Euler tour (on opt. facilities) of cost $\leq 2c(T^*)$



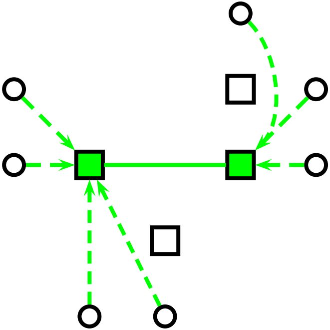
Proof

- $T^* :=$ Steiner tree in opt solution
- $C :=$ Euler tour (on opt. facilities) of cost $\leq 2c(T^*)$
- Mark facility "each" distance of $\frac{2c(T^*)}{k} \to F$
- k marked facilities F: $\forall i \in F^* : \ell(i, F) \leq \frac{2c(T^*)}{k}$



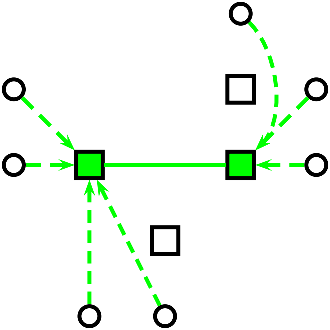
Proof

- $T^* :=$ Steiner tree in opt solution
- $C :=$ Euler tour (on opt. facilities) of cost $\leq 2c(T^*)$
- Mark facility "each" distance of $\frac{2c(T^*)}{k} \to F$
- k marked facilities F: $\forall i \in F^* : \ell(i, F) \leq \frac{2c(T^*)}{k}$



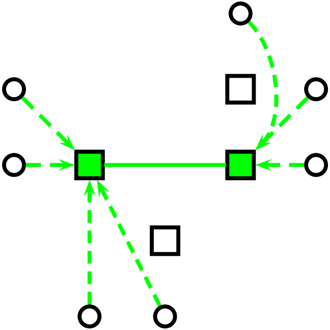
Proof

- $T^* :=$ Steiner tree in opt solution
- $C :=$ Euler tour (on opt. facilities) of cost $\leq 2c(T^*)$
- Mark facility "each" distance of $\frac{2c(T^*)}{k} \rightarrow F$
- k marked facilities F: $\forall i \in F^* : \ell(i, F) \leq \frac{2c(T^*)}{k}$
  - Opening costs: $\leq O^*$



Proof

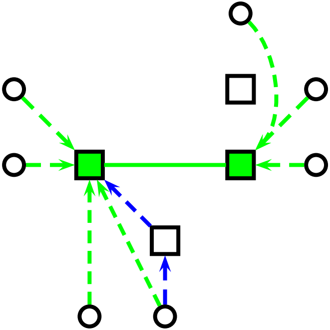
- $T^* :=$ Steiner tree in opt solution
- $C :=$ Euler tour (on opt. facilities) of cost $\leq 2c(T^*)$
- Mark facility "each" distance of $\frac{2c(T^*)}{k} \rightarrow F$
- k marked facilities F: $\forall i \in F^* : \ell(i, F) \leq \frac{2c(T^*)}{k}$



- Opening costs: $\leq O^*$
- Steiner costs: $\leq S^*$

Proof

- $T^* :=$ Steiner tree in opt solution
- $C :=$ Euler tour (on opt. facilities) of cost $\leq 2c(T^*)$
- Mark facility "each" distance of $\frac{2c(T^*)}{k} \to F$
- k marked facilities F: $\forall i \in F^* : \ell(i, F) \leq \frac{2c(T^*)}{k}$



- Opening costs: $\leq O^*$
- Steiner costs: $\leq S^*$
- Connection costs:

$$\begin{aligned} &\leq \quad C^* + \frac{2c(T^*)}{k} \cdot |\mathcal{D}| \\ &\leq \quad C^* + \frac{2|\mathcal{D}|}{M \cdot k} \cdot S^* \\ &\leq \quad C^* + \varepsilon S^* \end{aligned}$$

Open Problems

- Some of the best known approximation algorithms for network design are based on random sampling:
  - Single-Sink Buy-at-Bulk
  - Multi-Commodity Rent-or-Buy
  - Virtual Private Network Design
  - ...
- Can the improved bound on the connection cost help for these problems?

Open Problems

- Some of the best known approximation algorithms for network design are based on random sampling:
  - Single-Sink Buy-at-Bulk
  - Multi-Commodity Rent-or-Buy
  - Virtual Private Network Design
  - ...
- Can the improved bound on the connection cost help for these problems?

Thanks for your attention