Constant Factor Approximation for Capacitated k-Center with Outliers

Marek Cygan and **Tomasz Kociumaka**

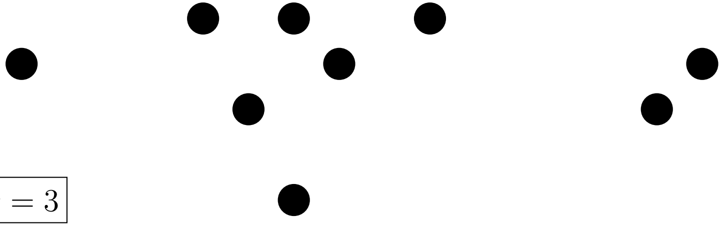
Institute of Informatics
University of Warsaw

STACS 2014
Lyon, France
March 7, 2013

k-Center

k-Center

Input: A finite set V, a metric function d on V, an integer k.

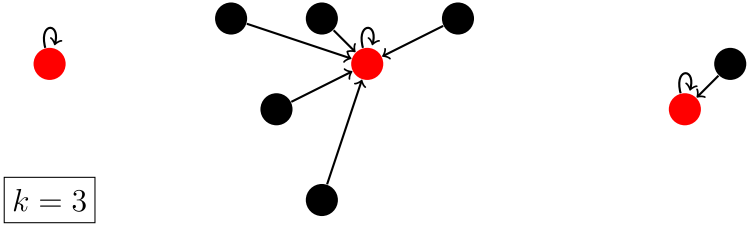


$k = 3$

k-Center

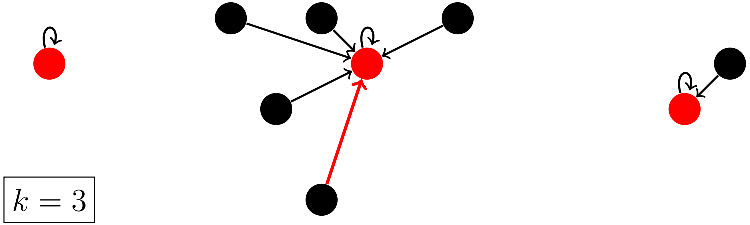
k-Center

Input: A finite set V, a metric function d on V, an integer k.
Output: A set $F \subseteq V$ of size k and a function $\phi : V \to F$.



k-CENTER

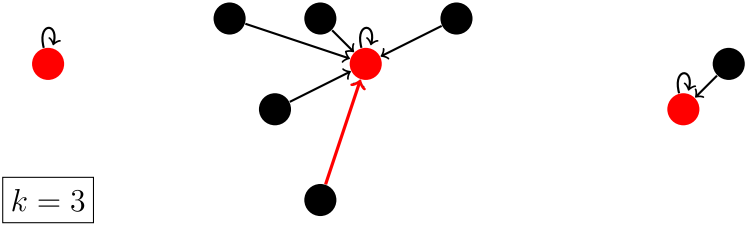
Input: A finite set V, a metric function d on V, an integer k.
Output: A set $F \subseteq V$ of size k and a function $\phi : V \to F$.
Minimize: $\max_{v \in V} d(v, \phi(v))$.



k-Center

k-Center

Input: A finite set V, a metric function d on V, an integer k.
Output: A set $F \subseteq V$ of size k and a function $\phi : V \to F$.
Minimize: $\max_{v \in V} d(v, \phi(v))$.

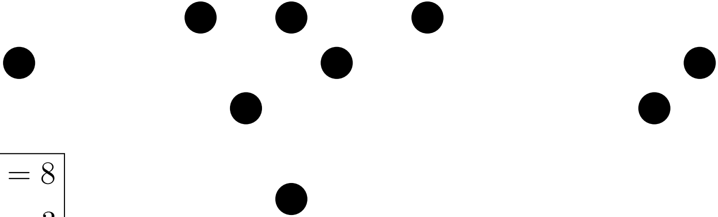


- simple 2-approximation known to be tight under $P \neq NP$ (Hochbaum & Shmoys, 1985; Gonzalez 1985).

k-Center With Outliers

k-Center with Outliers

Input: A finite set V, a metric function d on V, integers k, p.

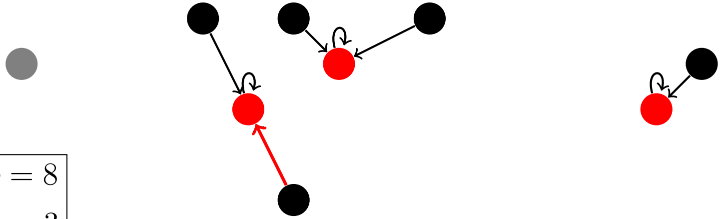


$p = 8$

$k = 3$

k-Center with Outliers

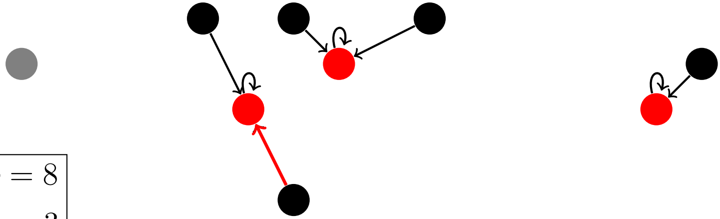
Input: A finite set V, a metric function d on V, integers k, p.
Output: Set $F \subseteq V$ of size k, $C \subseteq V$ of size p and a function $\phi : C \to F$.
Minimize: $\max_{v \in C} d(v, \phi(v))$.



$p = 8$
$k = 3$

k-Center with Outliers

Input: A finite set V, a metric function d on V, integers k, p.
Output: Set $F \subseteq V$ of size k, $C \subseteq V$ of size p and a function $\phi : C \to F$.
Minimize: $\max_{v \in C} d(v, \phi(v))$.



$p = 8$

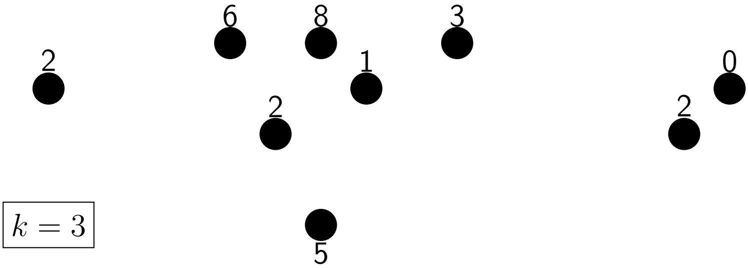
$k = 3$

- 3-approximation algorithm by Charikar, Khuller, Mound & Narasimhan (SODA 2001).

Capacitated k-Center

Capacitated k-Center

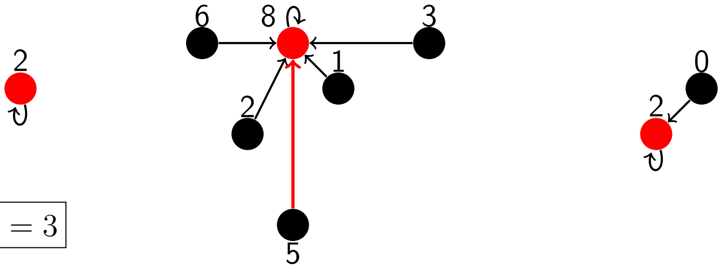
Input: A finite set V, a metric function d on V, an integer k, and a capacity function $L : V \to \mathbb{Z}_{\geq 0}$.



Capacitated k-Center

Capacitated k-Center

Input: A finite set V, a metric function d on V, an integer k, and a capacity function $L : V \to \mathbb{Z}_{\geq 0}$.
Output: A set $F \subseteq V$ of size k and a function $\phi : V \to F$ satisfying $|\phi^{-1}(v)| \leq L(v)$ for each $v \in F$.
Minimize: $\max_{v \in V} d(v, \phi(v))$.



$k = 3$

CAPACITATED k-CENTER: previous results

- an $\mathcal{O}(1)$-approximation bound by Cygan, Hajiaghayi & Khuller (FOCS'2012),
- improved to a 9-approximation by An, Bhaskara & Svensson (arXiv'2013) and independently by Chekuri, Gupta & Madan (joint paper accepted to IPCO'2014),
- $3 - \varepsilon$ lower bound by reduction from COST k-CENTER (Chuzhoy et al.; STOC'2004)
- a 6-approximation for *uniform* capacities and a 5-approximation for *uniform soft* capacities by Khuller & Sussmann (ESA'1996),

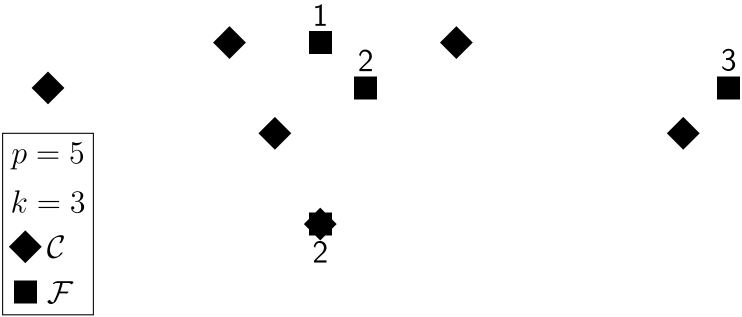
CAPACITATED k-CENTER: previous results

- an $\mathcal{O}(1)$-approximation bound by Cygan, Hajiaghayi & Khuller (FOCS'2012),
- improved to a 9-approximation by An, Bhaskara & Svensson (arXiv'2013) and independently by Chekuri, Gupta & Madan (joint paper accepted to IPCO'2014),
- $3 - \varepsilon$ lower bound by reduction from COST k-CENTER (Chuzhoy et al.; STOC'2004)
- a 6-approximation for *uniform* capacities and a 5-approximation for *uniform soft* capacities by Khuller & Sussmann (ESA'1996),
- soft capacities: multiple facilities can be opened in a single location

Capacitated k-Supplier with Outliers

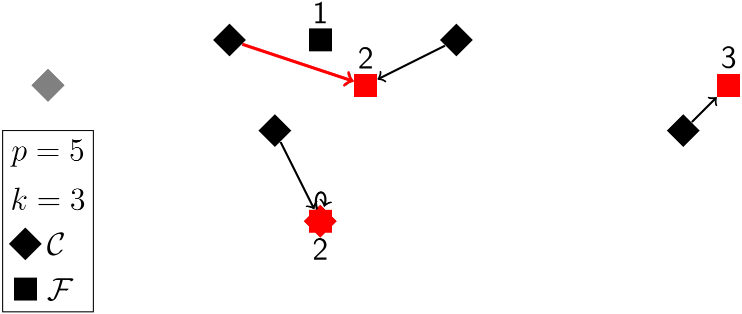
Capacitated k-Supplier with Outliers

Input: Finite sets $\mathcal{F}$ and $\mathcal{C}$, a (pseudo)metric function d on $\mathcal{F} \cup \mathcal{C}$, integers k, p, and a capacity function $L : \mathcal{F} \to \mathbb{Z}_{\geq 0}$.



Capacitated k-Supplier with Outliers

Input: Finite sets $\mathcal{F}$ and $\mathcal{C}$, a (pseudo)metric function d on $\mathcal{F} \cup \mathcal{C}$, integers k, p, and a capacity function $L : \mathcal{F} \to \mathbb{Z}_{\geq 0}$.
Output: Sets $F \subseteq \mathcal{F}$ of size k, $C \subseteq \mathcal{C}$ of size p and a function $\phi : C \to F$ satisfying $|\phi^{-1}(v)| \leq L(v)$ for each $v \in F$.
Minimize: $\max_{v \in C} d(v, \phi(v))$.



Capacitated k-Supplier with Outliers

Capacitated k-Supplier with Outliers

Input: Finite sets $\mathcal{F}$ and $\mathcal{C}$, a (pseudo)metric function d on $\mathcal{F} \cup \mathcal{C}$, integers k, p, and a capacity function $L : \mathcal{F} \to \mathbb{Z}_{\geq 0}$.
Output: Sets $F \subseteq \mathcal{F}$ of size k, $C \subseteq \mathcal{C}$ of size p and a function $\phi : C \to F$ satisfying $|\phi^{-1}(v)| \leq L(v)$ for each $v \in F$.
Minimize: $\max_{v \in C} d(v, \phi(v))$.

- Natural generalization of Capacitated k-Center with Outliers,
- Can be reduced to Capacitated k-Center with Outliers preserving the approximation factor.

Our results

Theorem (main result)

Capacitated k-Supplier with Outliers *admits a 25-approximation algorithm.*

Fact

The approximation ratio can be reduced to 23 for uniform capacities and to 13 for soft uniform capacities.

Corollary

Capacitated k-Center with Outliers *admits*
a 25-approximation algorithm in the general case,
a 23-approximation for uniform capacities and
a 13-approximation for uniform soft capacities.

Thresholding

Definition

A distance-r solution is a triple (C, F, ϕ) such that $|C| = p$ and $|F| = k$ and $\phi : C \to F$ obeys the capacities and satisfies $d(v, \phi(v)) \leq r$ for each $v \in C$.

Binary search (already used in most algorithms for k-CENTER) makes it sufficient to solve the following problem.

Graphic instances: r-approximation

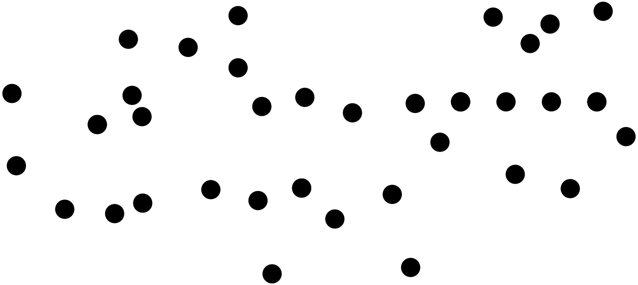
Input: An unweighted, undirected bipartite graph $G = (\mathcal{C}, \mathcal{F}, E)$, integers k and p, and a capacity function $L : \mathcal{F} \to \mathbb{Z}_{\geq 0}$.
Output: A distance r-solution or NO, if there is no distance-1 solution (with respect to metric d_G).

Skeletons

Definition

A set $S \subseteq \mathcal{F}$ is called a *skeleton* if

- **(separation)** $d(u, u') \geq 6$ for any $u, u' \in S$, $u \neq u'$,
- there exists a distance-1 solution (C_ϕ, F_ϕ, ϕ) such that:
  - **(covering)** $d(u, S) \leq 4$ for each $u \in F_\phi$,
  - **(injection)** there exists an injection $f : S \hookrightarrow F_\phi$ satisfying $d(u, f(u)) \leq 2$ for each $u \in S$.

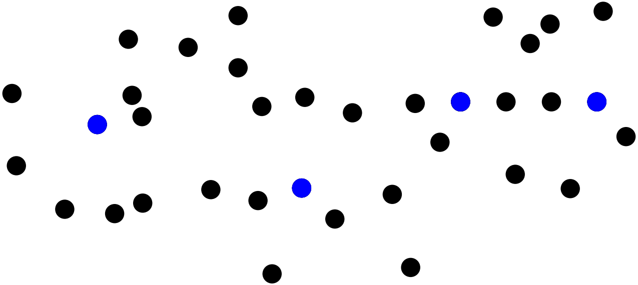


Skeletons

Definition

A set $S \subseteq \mathcal{F}$ is called a *skeleton* if

- **(separation)** $d(u, u') \geq 6$ for any $u, u' \in S$, $u \neq u'$,
- there exists a distance-1 solution (C_ϕ, F_ϕ, ϕ) such that:
  - **(covering)** $d(u, S) \leq 4$ for each $u \in F_\phi$,
  - **(injection)** there exists an injection $f : S \hookrightarrow F_\phi$ satisfying $d(u, f(u)) \leq 2$ for each $u \in S$.

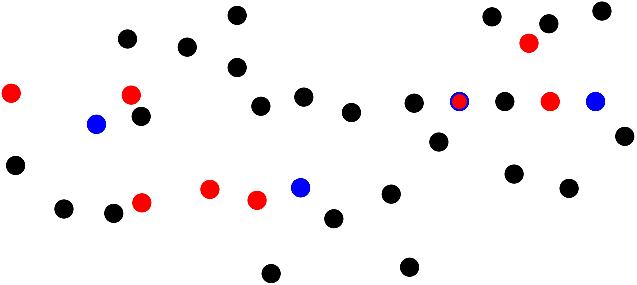


Skeletons

Definition

A set $S \subseteq \mathcal{F}$ is called a *skeleton* if

- **(separation)** $d(u, u') \geq 6$ for any $u, u' \in S$, $u \neq u'$,
- there exists a distance-1 solution (C_ϕ, F_ϕ, ϕ) such that:
  - **(covering)** $d(u, S) \leq 4$ for each $u \in F_\phi$,
  - **(injection)** there exists an injection $f : S \hookrightarrow F_\phi$ satisfying $d(u, f(u)) \leq 2$ for each $u \in S$.

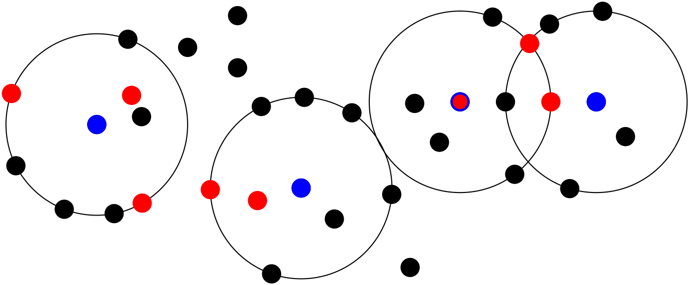


Skeletons

Definition

A set $S \subseteq \mathcal{F}$ is called a *skeleton* if

- (**separation**) $d(u, u') \geq 6$ for any $u, u' \in S$, $u \neq u'$,
- there exists a distance-1 solution (C_ϕ, F_ϕ, ϕ) such that:
  - (**covering**) $d(u, S) \leq 4$ for each $u \in F_\phi$,
  - (**injection**) there exists an injection $f : S \hookrightarrow F_\phi$ satisfying $d(u, f(u)) \leq 2$ for each $u \in S$.

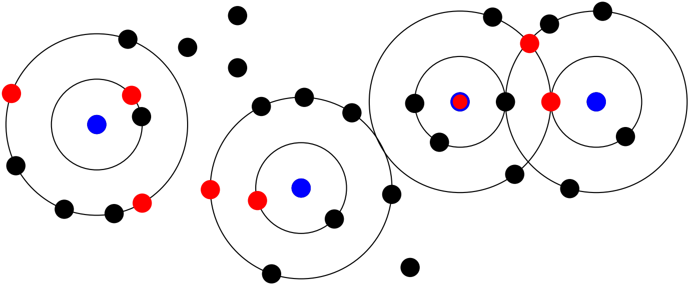


Skeletons

Definition

A set $S \subseteq \mathcal{F}$ is called a *skeleton* if

- (**separation**) $d(u, u') \geq 6$ for any $u, u' \in S$, $u \neq u'$,
- there exists a distance-1 solution (C_ϕ, F_ϕ, ϕ) such that:
  - (**covering**) $d(u, S) \leq 4$ for each $u \in F_\phi$,
  - (**injection**) there exists an injection $f : S \hookrightarrow F_\phi$ satisfying $d(u, f(u)) \leq 2$ for each $u \in S$.



Constructing a skeleton

Greedy algorithm:

1. $S_0 = \emptyset$.
2. $S_{i+1} = S_i + \arg\max\{\min(L(u), \deg(u)) : d_G(u, S_i) \geq 6\}$.

Choose a vertex which would not violate the separation property and can serve the largest number of clients in a distance-1 solution.

Lemma

If there exists a distance-1 solution, then at least one of the sets $S_0, \ldots, S_k$ is a skeleton.

Linear program

$$\sum_{u\in\mathcal{F}} y_u = k$$

$$\sum_{u\in\mathcal{F},v\in\mathcal{C}} x_{uv} = p$$

$$x_{uv} \le y_u \qquad \text{for } u \in \mathcal{F}, v \in \mathcal{C}$$

$$\sum_v x_{uv} \le L(u)\cdot y_u \qquad \text{for } u \in \mathcal{F}$$

$$\sum_u x_{uv} \le 1 \qquad \text{for } v \in \mathcal{C}$$

$$\sum_{u\in\mathcal{F}\cap N^2[s]} y_u \ge 1 \qquad \text{for } s \in S$$

$$x_{uv} = 0 \qquad \text{for } u \in \mathcal{F}, v \in \mathcal{C} \text{ such that } (v,u) \notin E$$

$$\mathbf{0} \le x, y \le \mathbf{1}$$

$LP_{k,p}(G, L, S)$. In the corresponding integer program $y_u = 1$ means $u \in F$ and $x_{uv} = 1$ means $\phi(v) = u$.

Linear program

$$\sum_{u \in \mathcal{F}} y_u = k$$

$$\sum_{u \in \mathcal{F}, v \in \mathcal{C}} x_{uv} = p$$

$$x_{uv} \le y_u \qquad \text{for } u \in \mathcal{F}, v \in \mathcal{C}$$

$$\sum_v x_{uv} \le L(u) \cdot y_u \qquad \text{for } u \in \mathcal{F}$$

$$\sum_u x_{uv} \le 1 \qquad \text{for } v \in \mathcal{C}$$

$$\sum_{u \in \mathcal{F} \cap N^2[s]} y_u \ge 1 \qquad \text{for } s \in S$$

$$x_{uv} = 0 \qquad \text{for } u \in \mathcal{F}, v \in \mathcal{C} \text{ such that } (v,u) \notin E$$

$$\mathbf{0} \le x, y \le \mathbf{1}$$

$LP_{k,p}(G, L, S)$. In the corresponding integer program $y_u = 1$ means $u \in F$ and $x_{uv} = 1$ means $\phi(v) = u$.

Integrality gap & clustering

Definition

The *integrality gap* of $LP_{k,p}(G, L, S)$ is the smallest r such that if the LP is feasible, then there is a distance-r solution.

Integrality gap & clustering

Definition

The *integrality gap* of $LP_{k,p}(G, L, S)$ is the smallest r such that if the LP is feasible, then there is a distance-r solution.

- The integrality gap of $LP_{k,p}(G, L, S)$ is in general unbounded, even if G is connected (unlike the analogous LP for the Capacitated k-Center).
- Use the covering property of a skeleton to remove vertices v such that $d_G(v, S) > 5$ and only then consider connected components separately.
- Apply dynamic programming to distribute k open facilities and p served clients among components, so that the LPs corresponding to the components are all feasible.

Integrality gap & clustering

Definition

The *integrality gap* of $LP_{k,p}(G, L, S)$ is the smallest r such that if the LP is feasible, then there is a distance-r solution.

- The integrality gap of $LP_{k,p}(G, L, S)$ is in general unbounded, even if G is connected (unlike the analogous LP for the Capacitated k-Center).
- Use the covering property of a skeleton to remove vertices v such that $d_G(v, S) > 5$ and only then consider connected components separately.
- Apply dynamic programming to distribute k open facilities and p served clients among components, so that the LPs corresponding to the components are all feasible.
- This may fail if S is not a skeleton
- If it does not, we will compute an approximate solution.

Rounding

Lemma

Let $I = (G = (\mathcal{C}, \mathcal{F}, E), L, k, p)$ *be an instance of* Capacitated k-supplier with Outliers *and let* $S \subseteq \mathcal{F}$*. If the following four conditions are satisfied:*

(i) G *is connected,*

(ii) *for any* $u, u' \in S$, $u \neq u'$ *we have* $d(u, u') \geq 6$*,*

(iii) $N^5[S] = \mathcal{F} \cup \mathcal{C}$*,*

(iv) $LP_{k,p}(G, L, S)$ *admits a feasible solution,*

then one can find a distance-25 solution for I *in polynomial time.*

Distance-r transfers

Distance-r transfers

- introduced by An, Bhaskara & Svensson (arXiv'2013) to round y preserving its sum
- each "portion" of y makes at most r hops, and lands in a vertex of capacity no smaller than the original one.

Distance-r transfers

Distance-r transfers

- introduced by An, Bhaskara & Svensson (arXiv'2013) to round y preserving its sum
- each "portion" of y makes at most r hops, and lands in a vertex of capacity no smaller than the original one.

Lemma

If $F \subseteq \mathcal{F}$ is a distance-r transfer of y in G with respect to L, then a distance-$(r+1)$ solution can be determined in polynomial time.

Rounding cnt'd

Lemma (An, Bhaskara & Svensson, arXiv 2013)

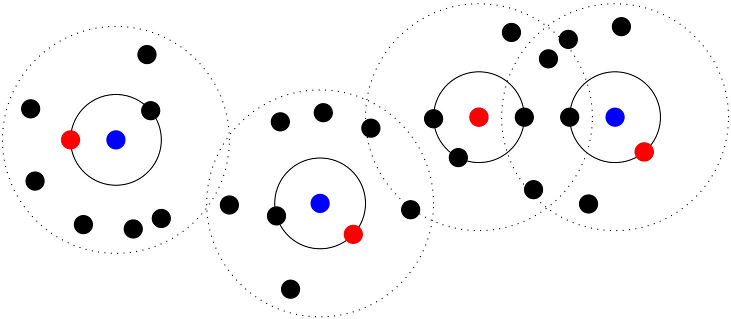
Integral distance-2 transfer can be found for a tree T *and* y *such that* $y_u = 1$ *for all non-leaves* u*.*

Rounding cnt'd

Lemma (An, Bhaskara & Svensson, arXiv 2013)

Integral distance-2 transfer can be found for a tree T and y such that $y_u = 1$ for all non-leaves u.

For simplicity we sketch a distance-32 transfer instead of a distance-24 transfer.

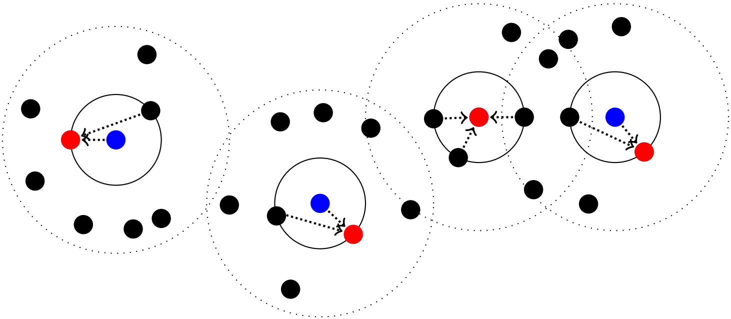


Rounding cnt'd

Lemma (An, Bhaskara & Svensson, arXiv 2013)

Integral distance-2 transfer can be found for a tree T and y such that $y_u = 1$ for all non-leaves u.

For simplicity we sketch a distance-32 transfer instead of a distance-24 transfer.

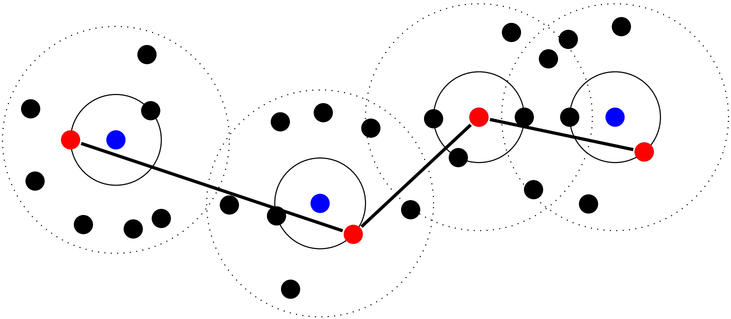


Rounding cnt'd

Lemma (An, Bhaskara & Svensson, arXiv 2013)

Integral distance-2 transfer can be found for a tree T and y such that $y_u = 1$ for all non-leaves u.

For simplicity we sketch a distance-32 transfer instead of a distance-24 transfer.

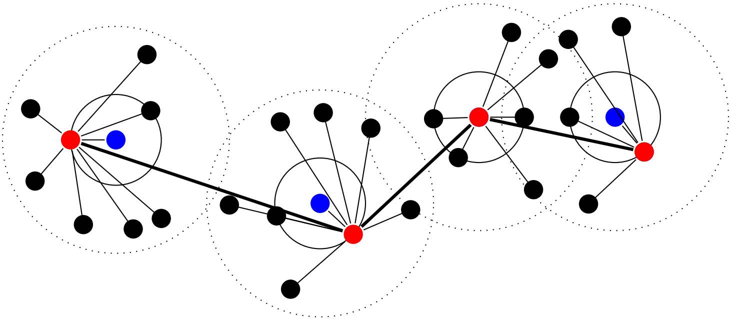


Rounding cnt'd

Lemma (An, Bhaskara & Svensson, arXiv 2013)

Integral distance-2 transfer can be found for a tree T and y such that $y_u = 1$ for all non-leaves u.

For simplicity we sketch a distance-32 transfer instead of a distance-24 transfer.



Conclusions & open problems

Theorem

CAPACITATED k-SUPPLIER WITH OUTLIERS *admits a 25-approximation algorithm.*

Conclusions & open problems

Theorem

CAPACITATED k-SUPPLIER WITH OUTLIERS *admits a 25-approximation algorithm.*

Open problems:

- Shrink the gap between the approximation ratio (25) and the lower bound $(3 - \varepsilon)$.

Conclusions & open problems

Theorem

CAPACITATED k-SUPPLIER WITH OUTLIERS *admits a 25-approximation algorithm.*

Open problems:

- Shrink the gap between the approximation ratio (25) and the lower bound $(3 - \varepsilon)$.
- $\mathcal{O}(1)$-approximation for CAPACITATED k-MEDIAN.

CAPACITATED k-MEDIAN

Input: A finite set V, a metric function d on V, an integer k and a capacity function $L : V \to \mathbb{Z}_{\geq 0}$.
Output: A set $F \subseteq V$ of size k and a function $\phi : V \to F$ satisfying $|\phi^{-1}(v)| \leq L(v)$ for each $v \in F$.
Minimize: $\sum_v d(v, \phi(v))$.

Thank you

Thank you for your attention!