Approximation Algorithms for Continuous Clustering and Facility Location Problems

Deeparnab Chakrabarty, Maryam Negahbani, **Ankita Sarkar**

Talk at TTI Chicago
May 12, 2023

Continuous center-based k-clustering

Continuous center-based k-clustering

Input.

- Metric space (X, d). $|X|$ possibly infinite.
- Clients $C \subseteq X$, $|C| = n$
- $k \in \mathbb{N}$

Continuous center-based k-clustering

Input.

- Metric space (X, d). $|X|$ possibly infinite.
- Clients $C \subseteq X$, $|C| = n$
- $k \in \mathbb{N}$

Output.

- Facilities $S \subseteq X$, $|S| = k$
- Minimize objective function

Continuous center-based k-clustering

Input.

- Metric space (X, d). $|X|$ possibly infinite.
- Clients $C \subseteq X$, $|C| = n$
- $k \in \mathbb{N}$

Output.

- Facilities $S \subseteq X$, $|S| = k$
- Minimize objective function
  - k-median: $\sum_{v \in C} d(v, S)$
  - k-means: $\sum_{v \in C} d(v, S)^2$

Continuous vs. discrete

- $S \subseteq X$ vs $S \subseteq C$

Continuous vs. discrete

- $S \subseteq X$ vs $S \subseteq C$
- Discrete:

  ...
  Hochbaum and Shmoys, 1985
  Shmoys, Tardos, and Aardal, 1997
  Guha and Khuller, 1999
  Charikar and Guha, 1999
  Charikar, Khuller, Mount, and Narasimhan, 2001
  Jain and Vazirani, 2001
  Arya, Garg, Khandekar, Meyerson, Munagala, and Pandit, 2001
  Charikar, Guha, Shmoys, and Tardos, 2002
  Jain, Mahdian, Markakis, Saberi, and Vazirani, 2003
  ...
  Li, 2013
  Li and Svensson, 2016
  Byrka, Pensyl, Rybicki, Srinivasan, and Trinh, 2017

  ...

Continuous vs. discrete

- $S \subseteq X$ vs $S \subseteq C$
- Discrete:
  ...
  Hochbaum and Shmoys, 1985
  Shmoys, Tardos, and Aardal, 1997
  Guha and Khuller, 1999
  Charikar and Guha, 1999
  Charikar, Khuller, Mount, and Narasimhan, 2001
  Jain and Vazirani, 2001
  Arya, Garg, Khandekar, Meyerson, Munagala, and Pandit, 2001
  Charikar, Guha, Shmoys, and Tardos, 2002
  Jain, Mahdian, Markakis, Saberi, and Vazirani, 2003
  ...
  Li, 2013
  Li and Svensson, 2016
  Byrka, Pensyl, Rybicki, Srinivasan, and Trinh, 2017

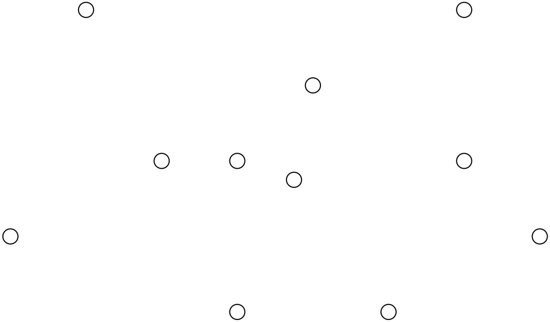
  ...
- Discrete best:
  - k-median: 2.613
    [Gowda, Pensyl, Srinivasan, and Trinh, SODA 2023]
  - k-means: 9
    [Kanungo, Mount, Netanyahu, Piatko, Silverman, and Wu, ComGeo 2004] ...

Continuous from discrete

- Reduction continuous $\rightarrow$ discrete

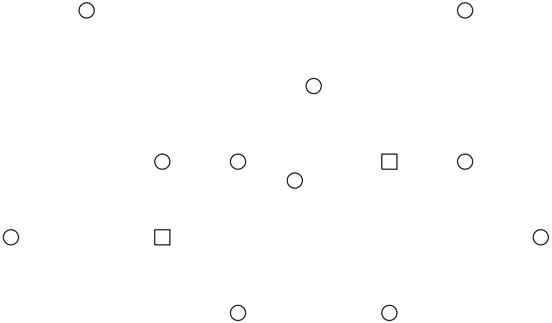
Continuous from discrete

- Reduction continuous $\rightarrow$ discrete



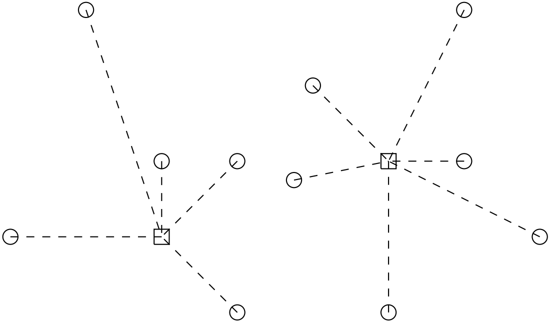
Continuous from discrete

- Reduction continuous → discrete



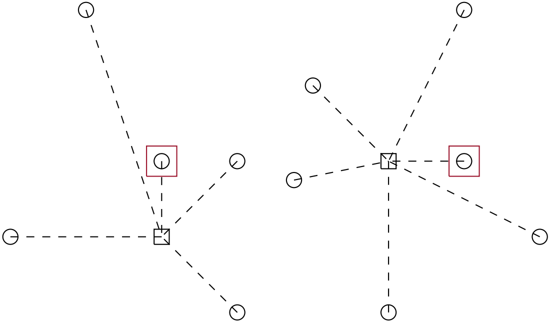
Continuous from discrete

- Reduction continuous $\to$ discrete



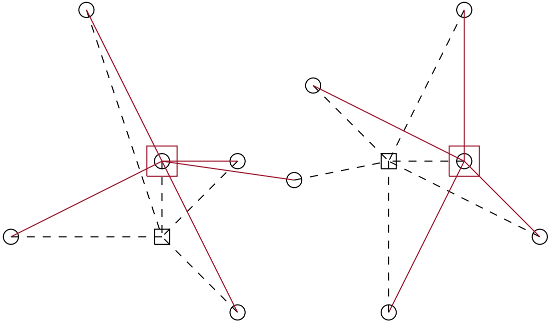
Continuous from discrete

- Reduction continuous $\rightarrow$ discrete



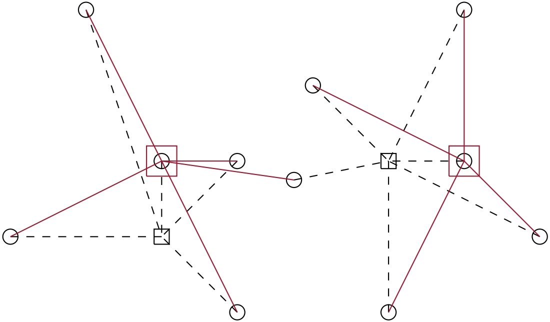
Continuous from discrete

- Reduction continuous → discrete



Continuous from discrete

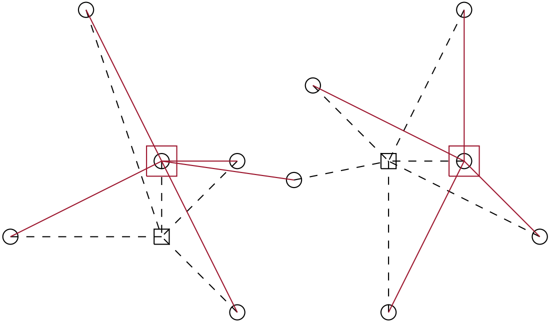
- Reduction continuous → discrete



k-median: $2\times$

Continuous from discrete

- Reduction continuous $\rightarrow$ discrete



k-median: $2\times$
k-means: $4\times$

Continuous vs discrete

- α-approx for discrete $\implies$ $\beta \cdot \alpha$-approx for continuous.
  - k-median: $\beta = 2$
  - k-means: $\beta = 4$

Continuous vs discrete

- α-approx for discrete $\implies$ $\beta \cdot \alpha$-approx for continuous.
  - k-median: $\beta = 2$
  - k-means: $\beta = 4$
- Constant k:
  - Continuous hardness $= \beta$
    [Cohen-Addad, Karthik, and Lee, 2021]
  - Discrete: polytime

Continuous vs discrete

- α-approx for discrete $\implies$ $\beta \cdot \alpha$-approx for continuous.
  - k-median: $\beta = 2$
  - k-means: $\beta = 4$
- Constant k:
  - Continuous hardness $= \beta$
    [Cohen-Addad, Karthik, and Lee, 2021]
  - Discrete: polytime
- Case $(\mathbb{R}^m, \ell_2)$, m small

Matǒusek, 2000
Har-Peled and Mazumdar, 2004
de la Vega, Karpinski, Kenyon, and Rabani, 2003
Kumar, Sabharwal, and Sen, 2004
Ke Chen, 2006
Feldman, Monemizadeh, and Sohler, 2007
...
Friggstad, Rezapour, and Salavatipour, 2019
Cohen-Addad, Klein, and Mathieu, 2019
Cohen-Addad, Saulpic, and Schwiegelshohn, 2021

...

Continuous vs discrete

- α-approx for discrete $\implies$ $\beta \cdot \alpha$-approx for continuous.
  - k-median: $\beta = 2$
  - k-means: $\beta = 4$
- Constant k:
  - Continuous hardness $= \beta$
    [Cohen-Addad, Karthik, and Lee, 2021]
  - Discrete: polytime
- Case $(\mathbb{R}^m, \ell_2)$, m small

  Matŏusek, 2000
  Har-Peled and Mazumdar, 2004
  de la Vega, Karpinski, Kenyon, and Rabani, 2003
  Kumar, Sabharwal, and Sen, 2004
  Ke Chen, 2006
  Feldman, Monemizadeh, and Sohler, 2007
  ...
  Friggstad, Rezapour, and Salavatipour, 2019
  Cohen-Addad, Klein, and Mathieu, 2019
  Cohen-Addad, Saulpic, and Schwiegelshohn, 2021

  ...
- General (X, d)?

Continuous vs discrete: gap of β?

Continuous vs discrete: gap of β?

Motivating question

In general (X, d), can we do better for continuous case than reducing to discrete?

Continuous vs discrete: gap of β?

Motivating question

In general (X, d), can we do better for continuous case than reducing to discrete?

Our results.

- λ-UFL. **Strong yes**.

$$\underbrace{2.32}_{\text{cont. ratio}} < \underbrace{2}_{\beta} \times \underbrace{1.27}_{\text{dis. hardness}}$$

Continuous vs discrete: gap of β?

Motivating question

In general (X, d), can we do better for continuous case than reducing to discrete?

Our results.

- λ-UFL. **Strong yes**.

$$\underbrace{2.32}_{\text{cont. ratio}} < \underbrace{2}_{\beta} \times \underbrace{1.27}_{\text{dis. hardness}}$$

- k-means. **Yes**.

$$\underbrace{32}_{\text{cont. ratio}} < \underbrace{4}_{\beta} \times \underbrace{9}_{\text{dis. ratio}}$$

Continuous vs discrete: gap of β?

Motivating question

In general (X, d), can we do better for continuous case than reducing to discrete?

Our results.

- λ-UFL. **Strong yes**.

$$\underbrace{2.32}_{\text{cont. ratio}} < \underbrace{2}_{\beta} \times \underbrace{1.27}_{\text{dis. hardness}}$$

- k-means. **Yes**.

$$\underbrace{32}_{\text{cont. ratio}} < \underbrace{4}_{\beta} \times \underbrace{9}_{\text{dis. ratio}}$$

- Fair-k-median, k-center with outliers : **Yes**.

Techniques

Techniques

A new LP.

Techniques

A new LP.

- Existing LP: $|X|$ variables, may be ∞

Techniques

A new LP.

- Existing LP: $|X|$ variables, may be ∞
- Our LP: $\mathrm{poly}(n)$ variables, $\exp(n)$ constraints

Techniques

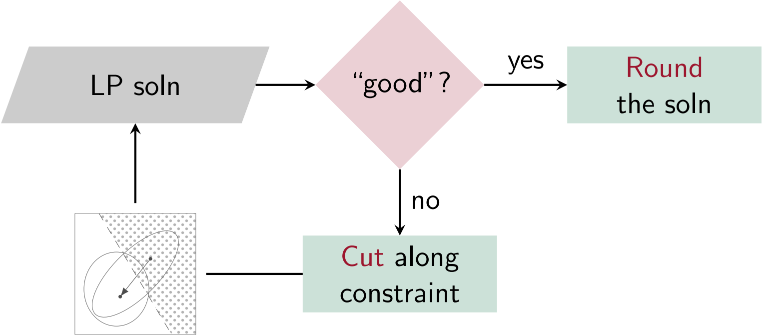
A new LP.

- Existing LP: $|X|$ variables, may be ∞
- Our LP: $\mathrm{poly}(n)$ variables, $\exp(n)$ constraints
- **Round-or-cut**

Techniques

A new LP.

- Existing LP: $|X|$ variables, may be ∞
- Our LP: $\mathrm{poly}(n)$ variables, $\exp(n)$ constraints
- **Round-or-cut**



New LP

Running example: 6.67-approx, cont. k-median

New LP

Running example: 6.67-approx, cont. k-median
Variables.

New LP

Running example: 6.67-approx, cont. k-median

Variables.

- C_v: cost share of $v \in C$

New LP

Running example: 6.67-approx, cont. k-median

Variables.

- C_v: cost share of $v \in C$
- $y(v, r)$: $v \in C$, $r \in \mathbb{R}$,

New LP

Running example: 6.67-approx, cont. k-median

Variables.

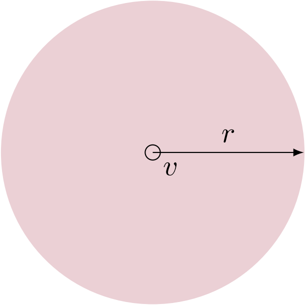
- C_v: cost share of $v \in C$
- $y(v, r)$: $v \in C$, $r \in \mathbb{R}$,

New LP

Running example: 6.67-approx, cont. k-median

Variables.

- C_v: cost share of $v \in C$
- $y(v, r)$: $v \in C$, $r \in \mathbb{R}$, $B(v, r) = \{x \in X \mid d(v, x) \leq r\}$

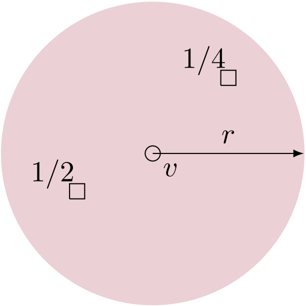


New LP

Running example: 6.67-approx, cont. k-median

Variables.

- C_v: cost share of $v \in C$
- $y(v, r)$: $v \in C$, $r \in \mathbb{R}$, $B(v, r) = \{x \in X \mid d(v, x) \leq r\}$

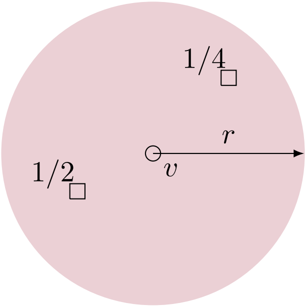


New LP

Running example: 6.67-approx, cont. k-median

Variables.

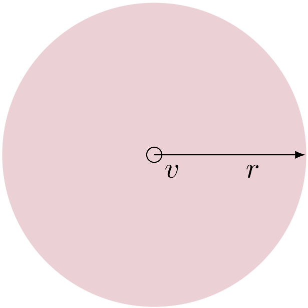
- C_v: cost share of $v \in C$
- $y(v, r)$: $v \in C$, $r \in \mathbb{R}$, $B(v, r) = \{x \in X \mid d(v, x) \leq r\}$



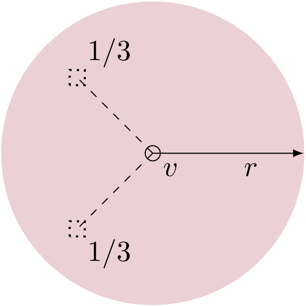
$y(v, r) = 3/4$

New LP: Consistency constraints

New LP: Consistency constraints

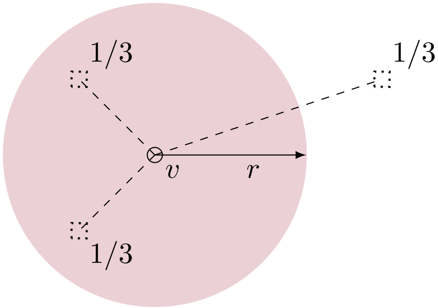


New LP: Consistency constraints



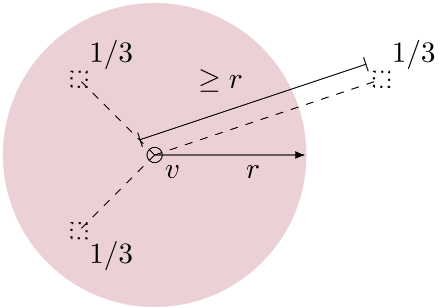
$$\underbrace{y(v, r)}_{2/3}$$

New LP: Consistency constraints



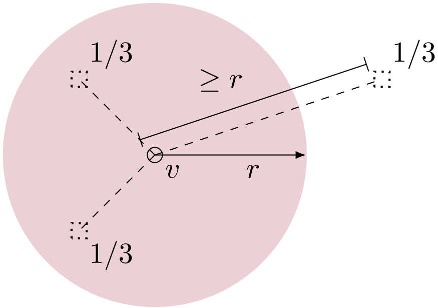
$$\underbrace{1 - y(v,r)}_{1/3}$$

New LP: Consistency constraints



$$\geq r \underbrace{(1 - y(v,r))}_{1/3}$$

New LP: Consistency constraints

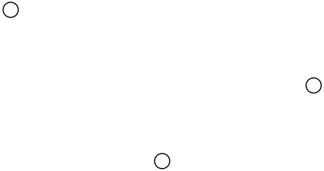


$$C_v \geq r \underbrace{(1 - y(v, r))}_{1/3}$$

New LP: Disjoint balls constraints

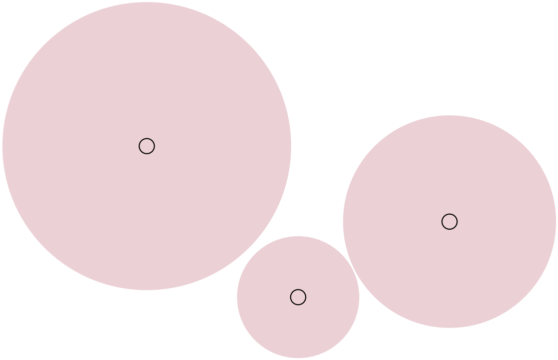
New LP: Disjoint balls constraints

$$k = 1$$

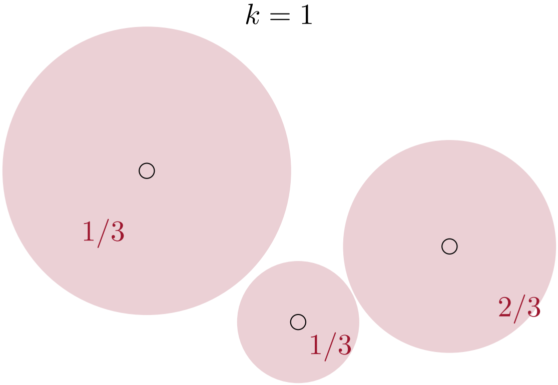


New LP: Disjoint balls constraints

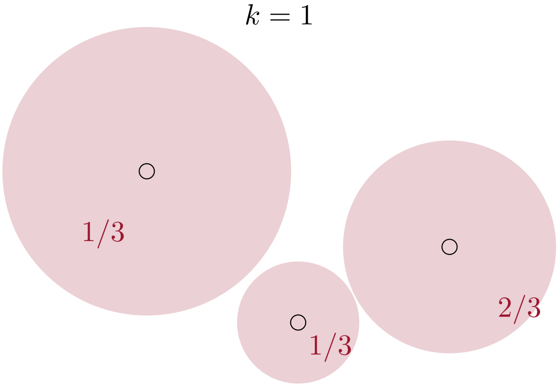
$k = 1$



New LP: Disjoint balls constraints

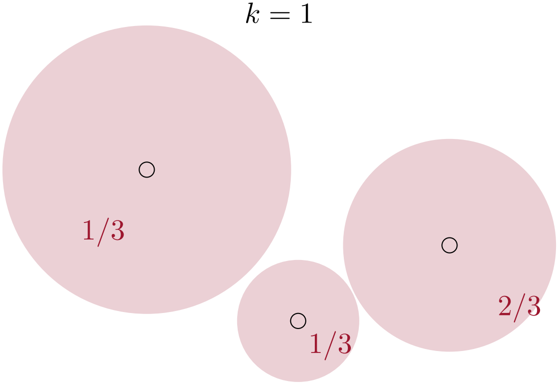


New LP: Disjoint balls constraints



For $\mathcal{B} \subseteq \{B(v,r)\}_{v \in C, r \in \mathbb{R}}$, pairwise disjoint

New LP: Disjoint balls constraints

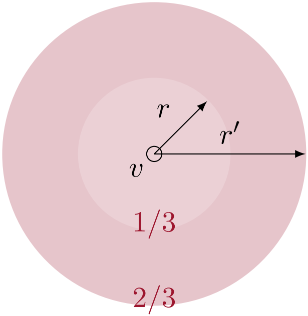


For $\mathcal{B} \subseteq \{B(v,r)\}_{v \in C, r \in \mathbb{R}}$, pairwise disjoint

$$\sum_{B \in \mathcal{B}} y(B) \leq k$$

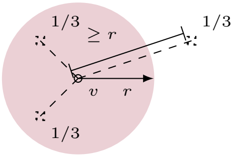
New LP: Monotonicity constraints

$y(v,r) \le y(v,r'),\ \forall v \in C,\ r \le r' \in \mathbb{R}$

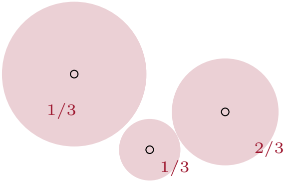


New LP: Constraints

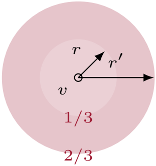
Consistency. $\forall v \in C, r \in \mathbb{R}$,
$$C_v \geq r(1 - y(v, r))$$



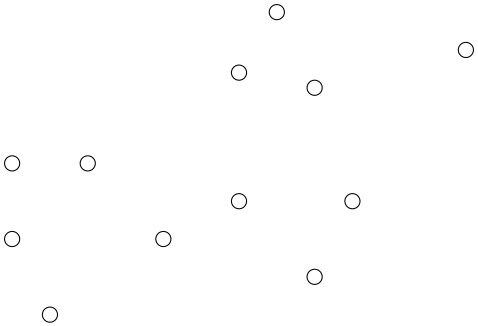
Disjoint balls. $\forall$ pairwise disj. $\mathcal{B}$,
$$\sum_{B \in \mathcal{B}} y(B) \leq k$$



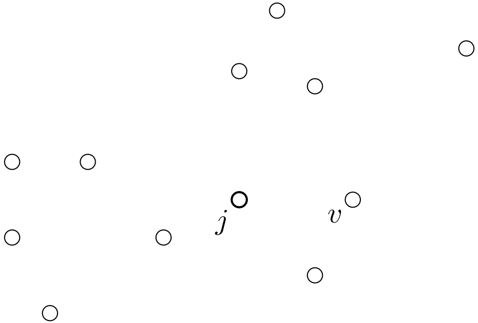
Monotonicity. $\forall v \in C, r \leq r' \in \mathbb{R}$,
$$y(v, r) \leq y(v, r')$$



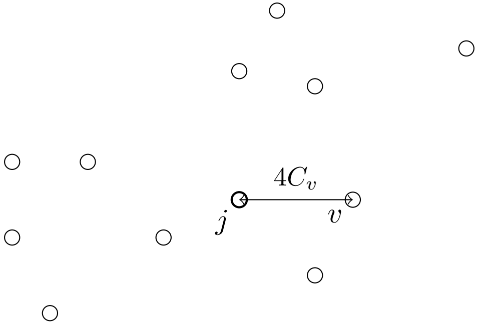
Satisfying disjoint balls constraints: filtering



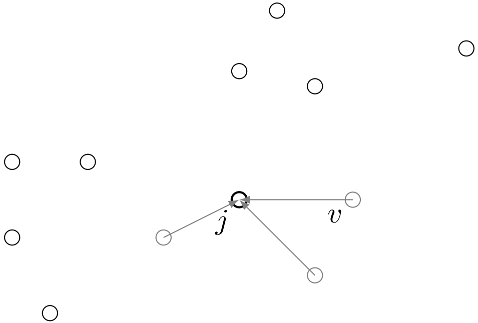
Satisfying disjoint balls constraints: filtering



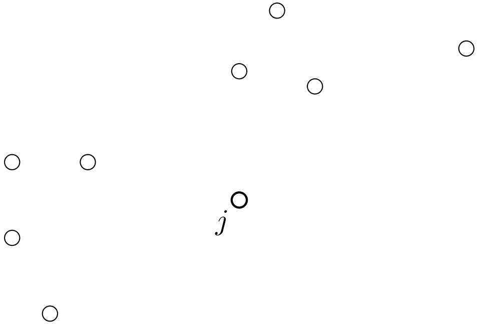
Satisfying disjoint balls constraints: filtering



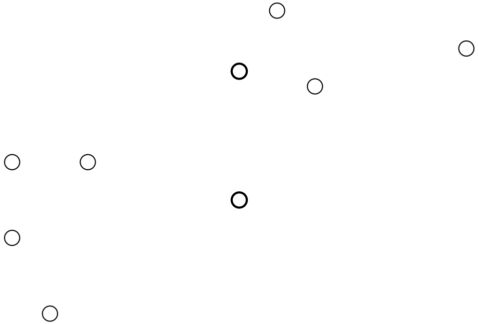
Satisfying disjoint balls constraints: filtering



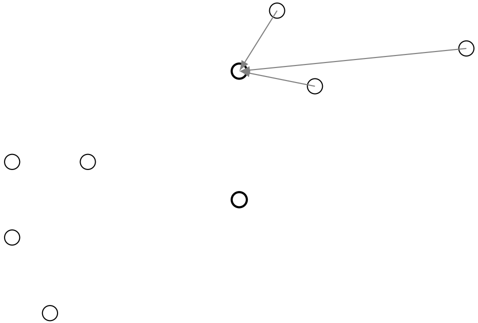
Satisfying disjoint balls constraints: filtering



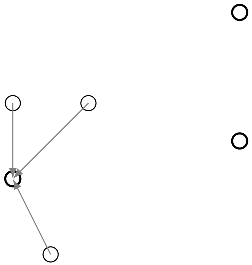
Satisfying disjoint balls constraints: filtering



Satisfying disjoint balls constraints: filtering

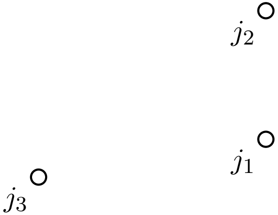


Satisfying disjoint balls constraints: filtering

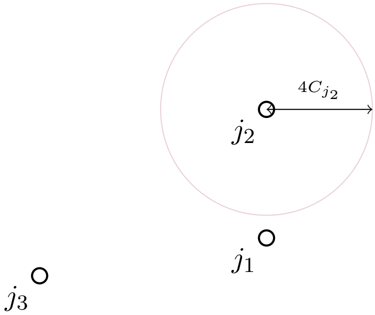


Satisfying disjoint balls constraints: filtering

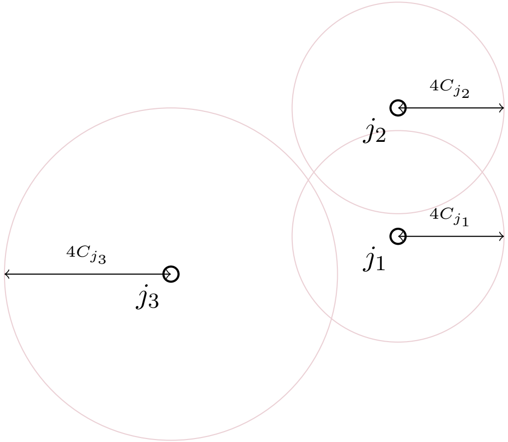
Satisfying disjoint balls constraints: filtering



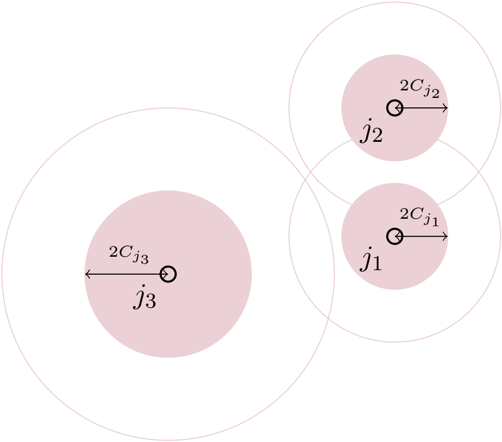
Satisfying disjoint balls constraints: filtering



Satisfying disjoint balls constraints: filtering



Satisfying disjoint balls constraints: filtering

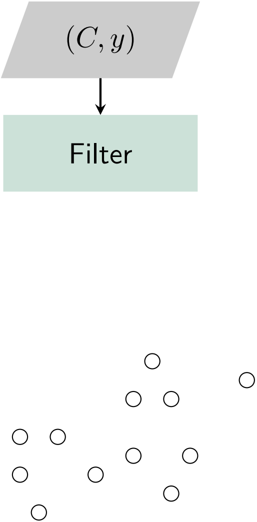


Satisfying disjoint balls constraints: round-or-cut

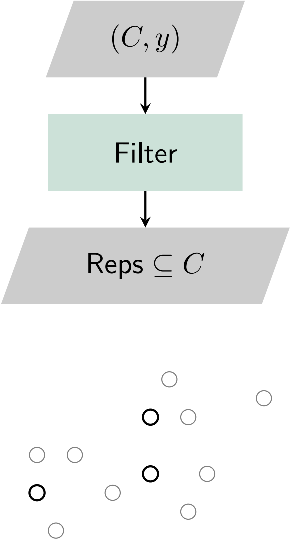
Satisfying disjoint balls constraints: round-or-cut

(C, y)

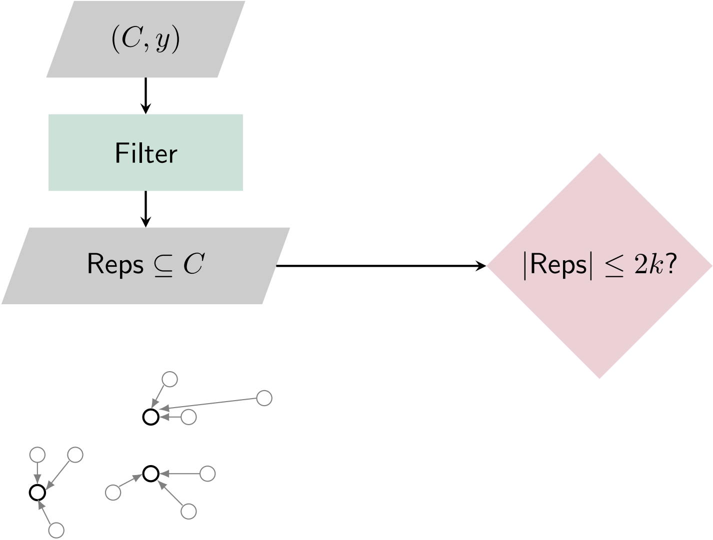
Satisfying disjoint balls constraints: round-or-cut



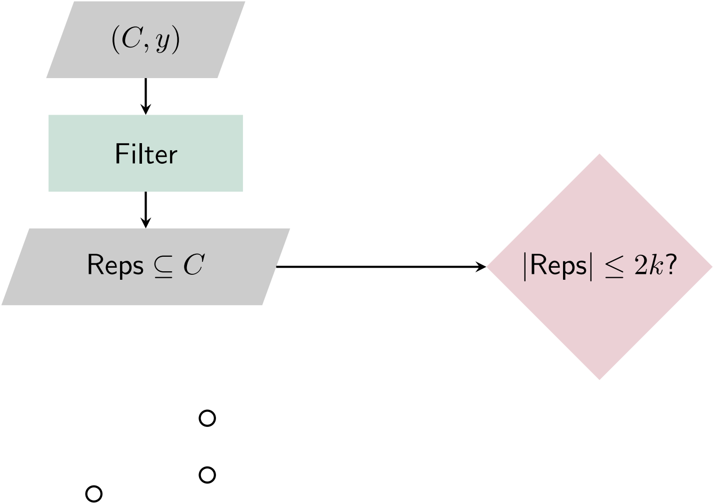
Satisfying disjoint balls constraints: round-or-cut



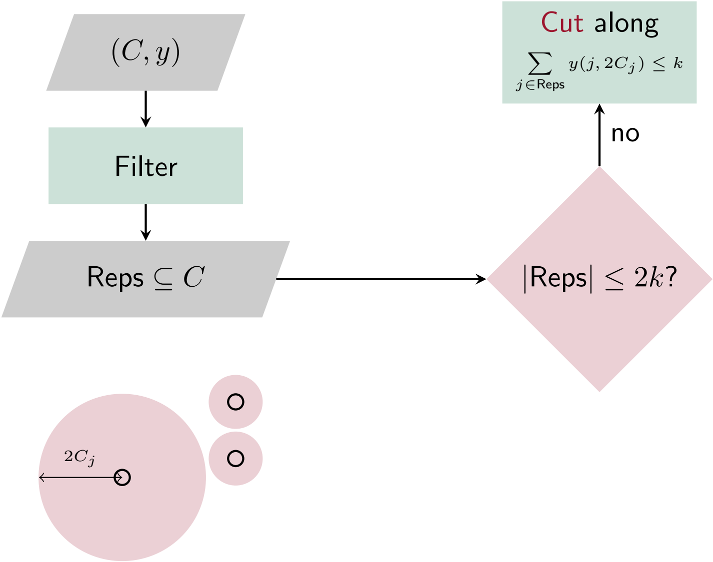
Satisfying disjoint balls constraints: round-or-cut



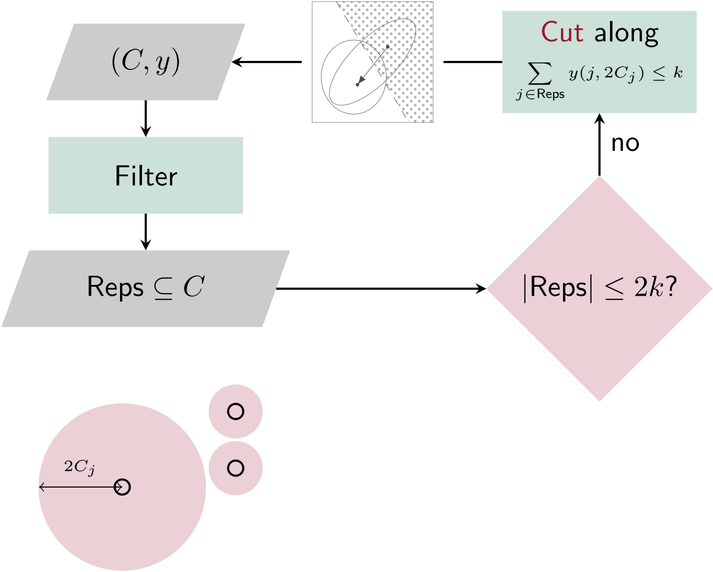
Satisfying disjoint balls constraints: round-or-cut



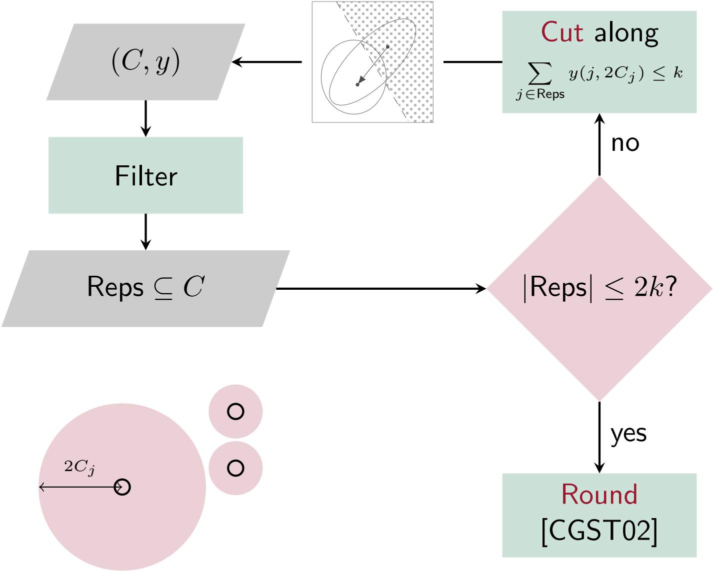
Satisfying disjoint balls constraints: round-or-cut



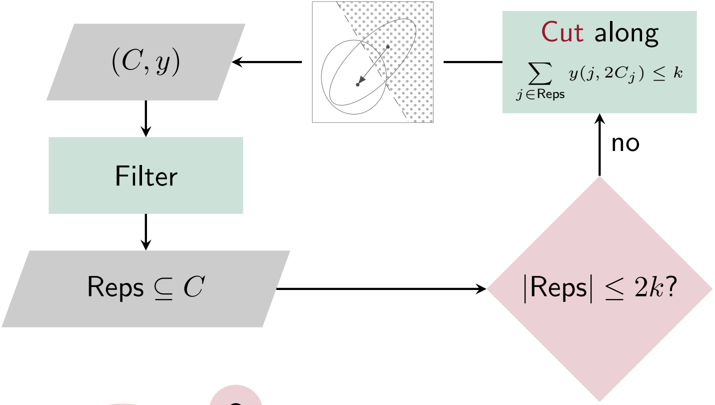
Satisfying disjoint balls constraints: round-or-cut

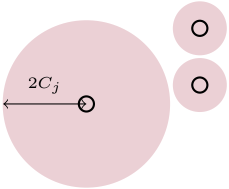


Satisfying disjoint balls constraints: round-or-cut



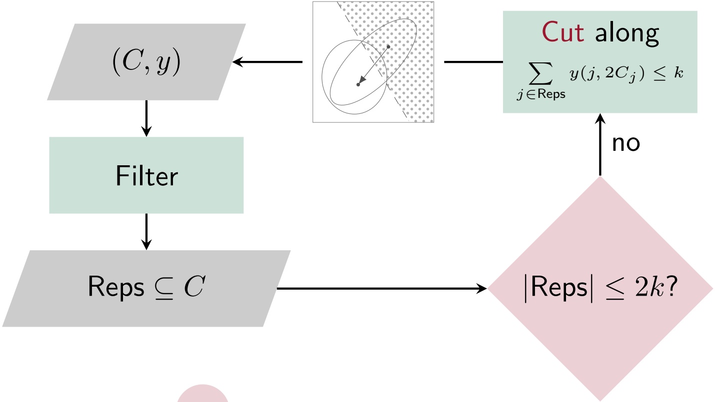
Satisfying disjoint balls constraints: round-or-cut

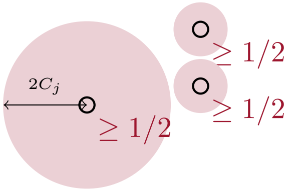




$$C_j \geq 2C_j(1 - y(j, 2C_j))$$

Satisfying disjoint balls constraints: round-or-cut





$$C_j \geq 2C_j(1 - y(j, 2C_j))$$
$$\implies \quad y(j, 2C_j) \geq 1/2$$

Open: continuous vs discrete for k-median

Open: continuous vs discrete for k-median

- Best known discrete: 2.613
  [Gowda, Pensyl, Srinivasan, and Trinh, SODA 2023]

Open: continuous vs discrete for k-median

- Best known discrete: 2.613
  [Gowda, Pensyl, Srinivasan, and Trinh, SODA 2023]
- $2 \times 2.613 = 5.226$

Open: continuous vs discrete for k-median

- Best known discrete: 2.613
  [Gowda, Pensyl, Srinivasan, and Trinh, SODA 2023]
- $2 \times 2.613 = 5.226$
- Ours: 6.67

Open: continuous vs discrete for k-median

- Best known discrete: 2.613
  [Gowda, Pensyl, Srinivasan, and Trinh, SODA 2023]
- $2 \times 2.613 = 5.226$
- Ours: 6.67

Question

Better than 5.226-approx for continuous k-median?

Next: Incorporating fairness

Continuous fair k-Median

Input.

- Metric space (X, d). $|X|$ possibly infinite.
- Clients $C \subseteq X$, $|C| = n$
- $k \in \mathbb{N}$
- Fairness radii $\{r_v\}_{v \in C}$

Continuous fair k-Median

Input.

- Metric space (X, d). $|X|$ possibly infinite.
- Clients $C \subseteq X$, $|C| = n$
- $k \in \mathbb{N}$
- Fairness radii $\{r_v\}_{v \in C}$

Output.

- Facilities $S \subseteq X$, $|S| = k$
- $\forall v \in C$, $d(v, S) \leq r_v$
- Minimize $\sum_{v \in C} d(v, S)$

Continuous fair k-Median

- α-approximation: S s.t. $\sum_{v \in C} d(v, S) \leq \alpha \cdot \texttt{opt}$

Continuous fair k-Median

- α-approximation: S s.t. $\sum_{v \in C} d(v, S) \leq \alpha \cdot \texttt{opt}$
- r_v's feasible? NP-hard.

Continuous fair k-Median

- α-approximation: S s.t. $\sum_{v \in C} d(v, S) \leq \alpha \cdot \texttt{opt}$
- r_v's feasible? NP-hard.
- S s.t. $\forall v \in C$, $d(v, S) \leq \gamma \cdot r_v$

Bicriterion (α, γ)-approximation

- $S \subseteq X$, $|S| = k$
- $\sum_{v \in C} d(v, S) \leq \alpha \cdot \texttt{opt}$

Bicriterion (α, γ)-approximation

- $S \subseteq X$, $|S| = k$
- $\sum_{v \in C} d(v, S) \leq \alpha \cdot \texttt{opt}$

Bicriterion (α, γ)-approximation

- $S \subseteq X$, $|S| = k$
- $\sum_{v \in C} d(v, S) \leq \alpha \cdot \texttt{opt}$
- $\forall v \in C$, $d(v, S) \leq \gamma \cdot r_v$

Continuous vs discrete

(α, γ) for discrete $\implies$ $(2\alpha, 2\gamma)$ for continuous

Continuous vs discrete

(α, γ) for discrete $\implies$ $(2\alpha, 2\gamma)$ for continuous
$(7.081, 3)$ for discrete [VY22] $\implies$ $(14.162, 6)$ for continuous

Continuous vs discrete

(α, γ) for discrete $\implies$ $(2\alpha, 2\gamma)$ for continuous
$(7.081, 3)$ for discrete [VY22] $\implies$ $(14.162, 6)$ for continuous

Our result

There is an $(8, 8)$-approximation for continuous fair k-median.

Algorithm: the LP

- Consistency constraints
- Disjoint-balls constraints

Algorithm: the LP

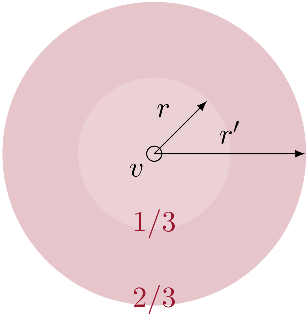
- Consistency constraints
- Disjoint-balls constraints
- Fairness constraints $y(v, r_v) \geq 1$

Algorithm: the LP

- Consistency constraints
- Disjoint-balls constraints
- Fairness constraints $y(v, r_v) \geq 1$
- Monotonicity constraints

Algorithm: the LP

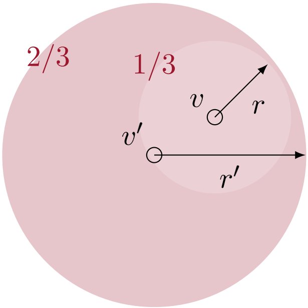
- Consistency constraints
- Disjoint-balls constraints
- Fairness constraints $y(v, r_v) \geq 1$
- Monotonicity constraints



$$y(v, r) \leq y(v, r')$$

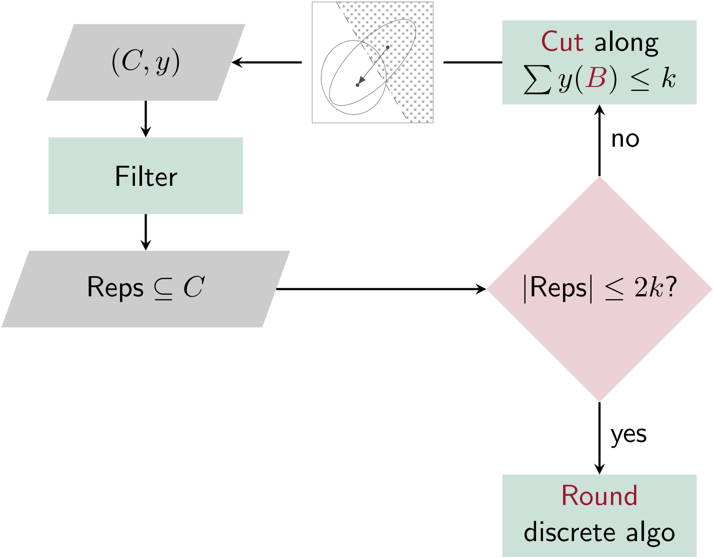
Algorithm: the LP

- Consistency constraints
- Disjoint-balls constraints
- Fairness constraints $y(v, r_v) \geq 1$
- Non-concentric monotonicity constraints



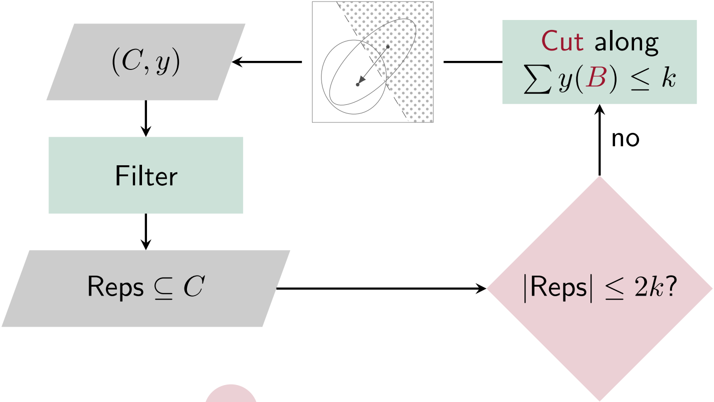
$$B(v, r) \subseteq B(v', r') \implies y(v, r) \leq y(v', r')$$

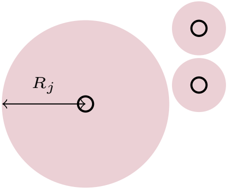
Round-or-cut



(C, y)
Filter
Reps $\subseteq C$
$|\text{Reps}| \leq 2k$?
no
Cut along
$\sum y(B) \leq k$
yes
Round
discrete algo

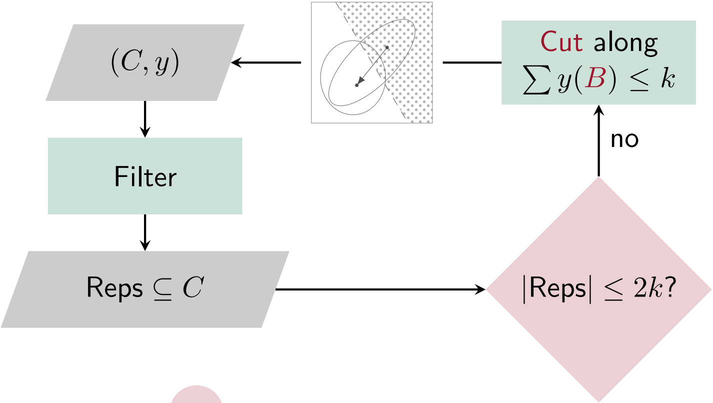
Round-or-cut

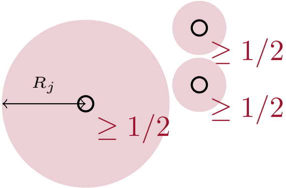




$$R_j := \min(r_j, 2C_j)$$

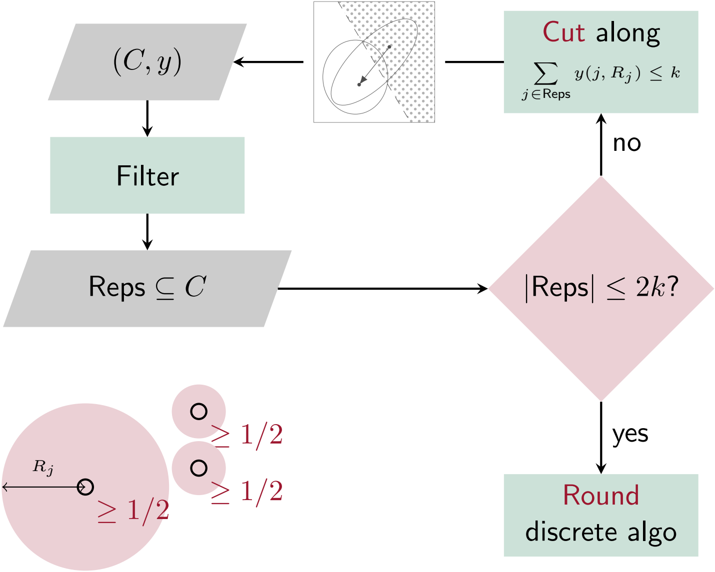
Round-or-cut





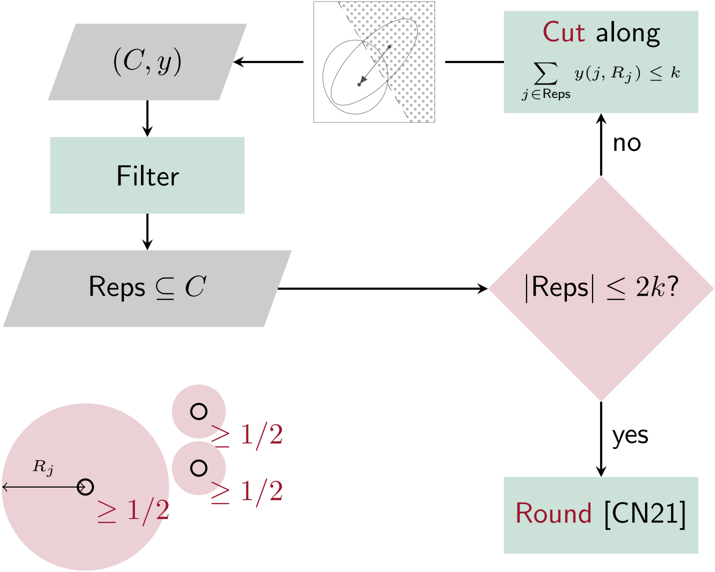
$$R_j := \min(r_j, 2C_j)$$
$$y(j, 2C_j) \geq 1/2,\ y(j, r_j) \geq 1$$

Round-or-cut



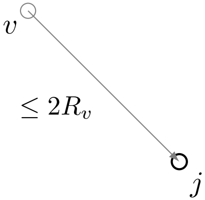
(C, y)
Cut along
$\sum_{j \in \text{Reps}} y(j, R_j) \leq k$
Filter
no
Reps $\subseteq C$
$|\text{Reps}| \leq 2k$?
yes
$\geq 1/2$
$\geq 1/2$
R_j
$\geq 1/2$
Round
discrete algo

Round-or-cut

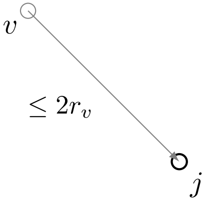


(C, y)
Cut along
$\sum_{j \in \text{Reps}} y(j, R_j) \le k$
Filter
no
Reps $\subseteq C$
$|\text{Reps}| \le 2k$?
yes
Round [CN21]
R_j
$\ge 1/2$
$\ge 1/2$
$\ge 1/2$

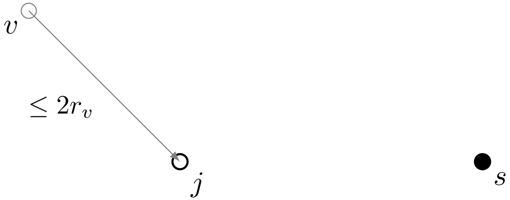
Fairness guarantee [CN21]



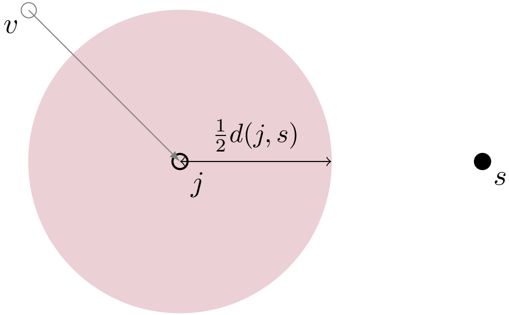
Fairness guarantee [CN21]



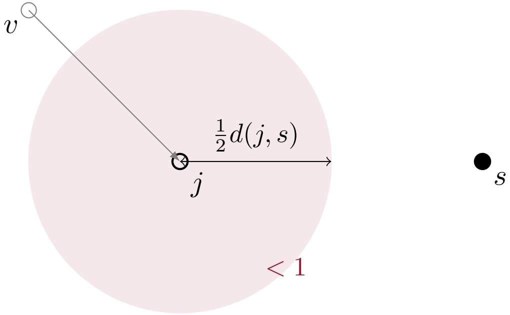
Fairness guarantee [CN21]



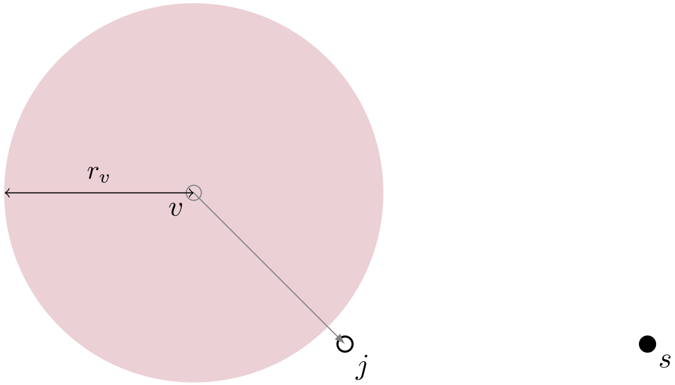
Fairness guarantee [CN21]



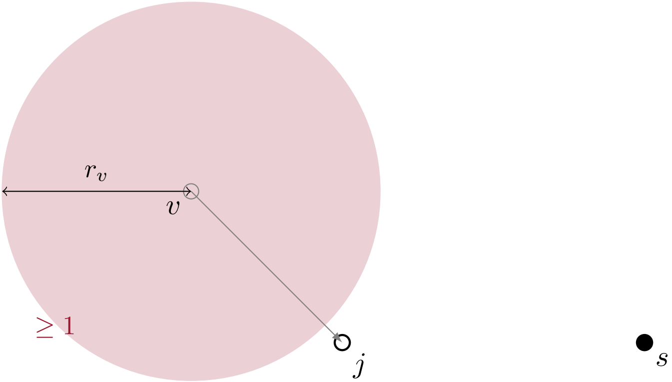
Fairness guarantee [CN21]



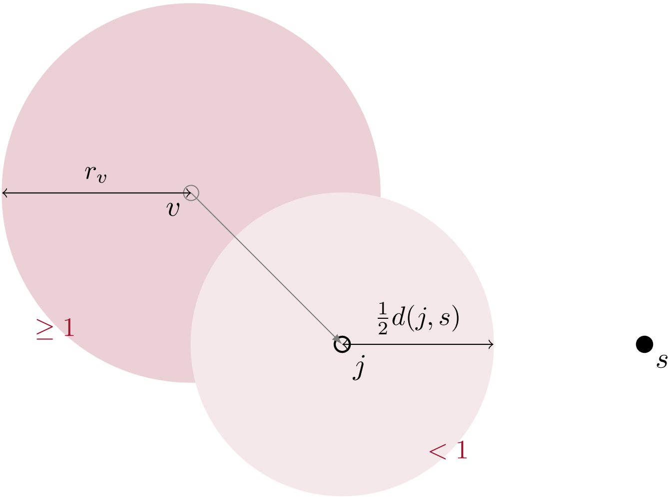
Fairness guarantee [CN21]



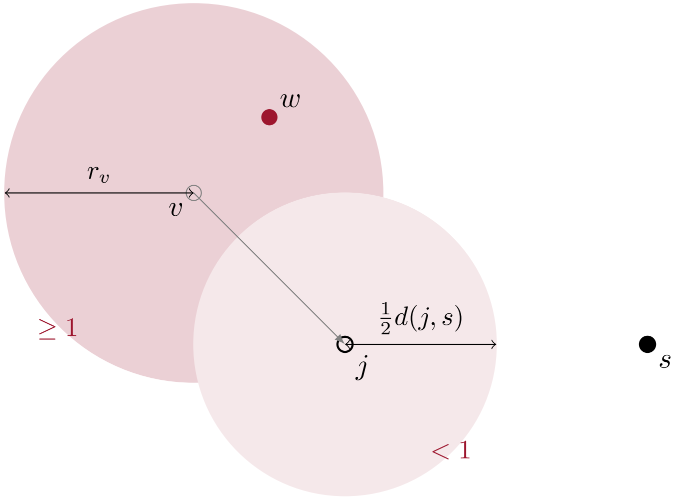
Fairness guarantee [CN21]



Fairness guarantee [CN21]

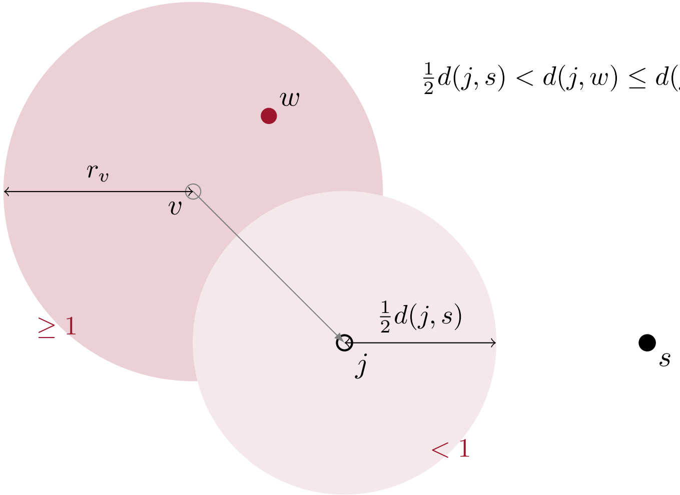


Fairness guarantee [CN21]



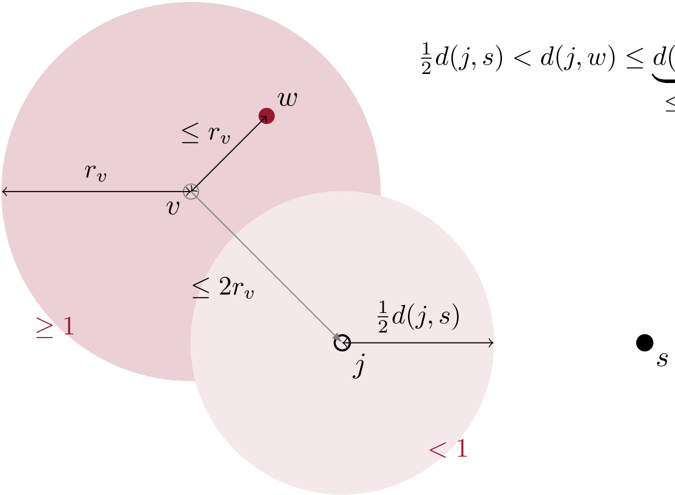
Fairness guarantee [CN21]

$$\frac{1}{2}d(j,s) < d(j,w) \leq d(j,v) + d(v,w)$$



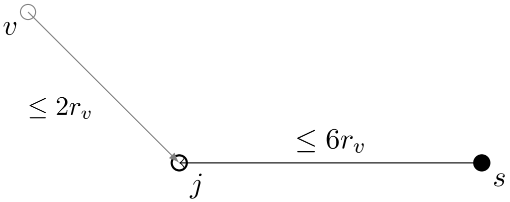
Fairness guarantee [CN21]

$$\tfrac{1}{2}d(j,s) < d(j,w) \leq \underbrace{d(j,v)}_{\leq 2r_v} + \underbrace{d(v,w)}_{\leq r_v}$$



Fairness guarantee [CN21]

$$d(j, s) \leq 6r_v \implies d(v, s) \leq 8r_v$$



Summary

- Continuous vs discrete: β gap?
- New LP and round-or-cut
- Open: continuous k-median (6.67 vs 5.226)
- Continuous fair k-median: $(8, 8)$-approximation