Fair and Fast k-Center Clustering for Data Summarization

Haris Angelidakis [1,2] Adam Kurpisz [2] Leon Sering [2] Rico Zenklusen [2]

[1]CoW Protocol

[2]ETH Zürich

SUMMARIZATION PROBLEMS

https://www.craneco.com › home

Crane Co. - Home

We serve some of the most demanding end markets, from aerospace & defense, bill validation and counterfeit detection, and oil & gas, to chemical, pharmaceutical ...

https://en.wikipedia.org › wiki › Crane_(bird)

Crane (bird) - Wikipedia

Cranes are a family, the Gruidae, of large, long-legged, and long-necked birds in the group Gruiformes. The 15 species of **cranes** are placed in three genera, ...

Family: Gruidae; Vigors, 1825
Kingdom: Animalia
Class: Aves
Order: Gruiformes

Common crane · Whooping crane · Sarus crane · Red-crowned crane

https://en.wikipedia.org › wiki › Crane_(machine)

Crane (machine) - Wikipedia

A **crane** is a type of machine, generally equipped with a hoist rope, wire ropes or chains, and sheaves, that can be used both to lift and lower materials and ...

https://dictionary.cambridge.org › dictionary › crane

CRANE | meaning in the Cambridge English Dictionary

7 days ago — a tall metal structure with a long horizontal part, used for lifting and moving heavy objects: The **crane** lifted the container off the ship.

https://www.konecranes.com

Konecranes: Overhead Cranes | Port Cranes | Crane Parts ...

Leading the global overhead **crane** industry with the largest service network and a complete range of lifting equipment.

SUMMARIZATION PROBLEMS

https://www.craneco.com › home

Crane Co. - Home

We serve some of the most demanding end markets, from aerospace & defense, bill validation and counterfeit detection, and oil & gas, to chemical, pharmaceutical ...

https://en.wikipedia.org › wiki › Crane_(bird)

Crane (bird) - Wikipedia

Cranes are a family, the Gruidae, of large, long-legged, and long-necked birds in the group Gruiformes. The 15 species of **cranes** are placed in three genera, ...

Family: Gruidae; Vigors, 1825 — Kingdom: Animalia
Class: Aves — Order: Gruiformes

Common crane · Whooping crane · Sarus crane · Red-crowned crane

https://en.wikipedia.org › wiki › Crane_(machine)

Crane (machine) - Wikipedia

A **crane** is a type of machine, generally equipped with a hoist rope, wire ropes or chains, and sheaves, that can be used both to lift and lower materials and ...

https://dictionary.cambridge.org › dictionary › crane

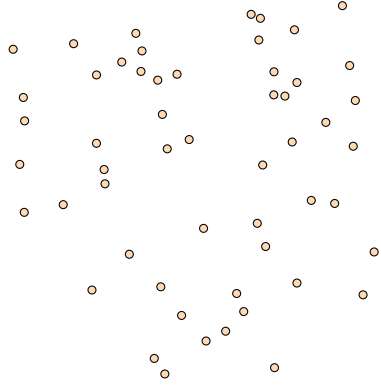
CRANE | meaning in the Cambridge English Dictionary

7 days ago — a tall metal structure with a long horizontal part, used for lifting and moving heavy objects: The **crane** lifted the container off the ship.

https://www.konecranes.com

Konecranes: Overhead Cranes | Port Cranes | Crane Parts ...

Leading the global overhead **crane** industry with the largest service network and a complete range of lifting equipment.



SUMMARIZATION PROBLEMS

https://www.craneco.com › home

Crane Co. - Home

We serve some of the most demanding end markets, from aerospace & defense, bill validation and counterfeit detection, and oil & gas, to chemical, pharmaceutical ...

https://en.wikipedia.org › wiki › Crane_(bird)

Crane (bird) - Wikipedia

Cranes are a family, the Gruidae, of large, long-legged, and long-necked birds in the group Gruiformes. The 15 species of **cranes** are placed in three genera, ...

Family: Gruidae; Vigors, 1825 — Kingdom: Animalia
Class: Aves — Order: Gruiformes

Common crane · Whooping crane · Sarus crane · Red-crowned crane

https://en.wikipedia.org › wiki › Crane_(machine)

Crane (machine) - Wikipedia

A **crane** is a type of machine, generally equipped with a hoist rope, wire ropes or chains, and sheaves, that can be used both to lift and lower materials and ...

https://dictionary.cambridge.org › dictionary › crane

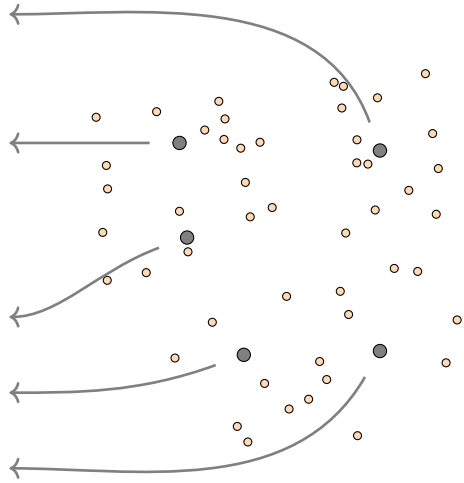
CRANE | meaning in the Cambridge English Dictionary

7 days ago — a tall metal structure with a long horizontal part, used for lifting and moving heavy objects: The **crane** lifted the container off the ship.

https://www.konecranes.com

Konecranes: Overhead Cranes | Port Cranes | Crane Parts ...

Leading the global overhead **crane** industry with the largest service network and a complete range of lifting equipment.



k-CENTER PROBLEM

k-Center problem

Input:

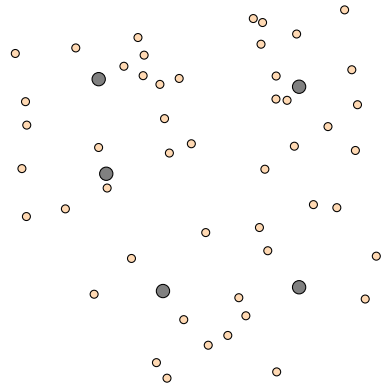
- metric space (X, d)
- integer k

Output: set of centers $C \subseteq X$ s.t

- $|C| \leqslant k$

Goal:

- minimize the distance: $r := \max_{x \in X} d(x, C)$



k-CENTER PROBLEM

k-Center problem

Input:

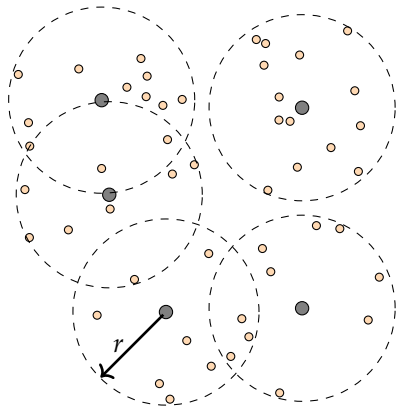
- metric space (X, d)
- integer k

Output: set of centers $C \subseteq X$ s.t

- $|C| \leqslant k$

Goal:

- minimize the distance: $r := \max_{x \in X} d(x, C)$



k-CENTER PROBLEM

k-Center problem

Input:

- metric space (X, d)
- integer k

Output: set of centers $C \subseteq X$ s.t

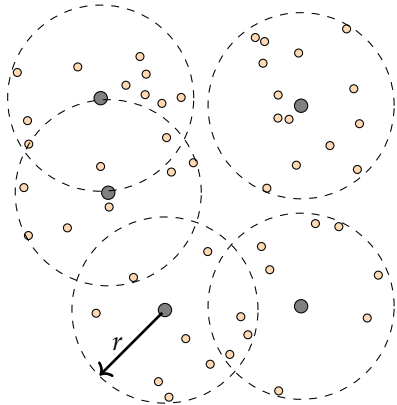
- $|C| \leqslant k$

Goal:

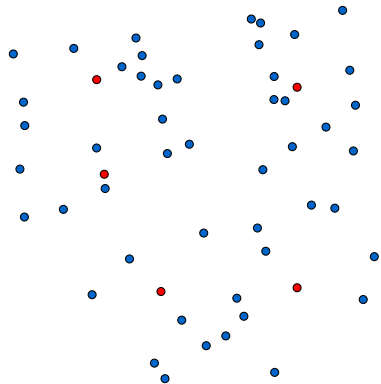
- minimize the distance: $r := \max_{x \in X} d(x, C)$

Known results:

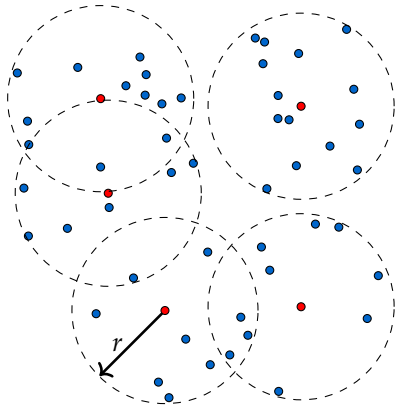
- $(2 - \epsilon)$-hard (APX-hard)
- 2-approx. (Gonzalez algorithm)



REPRESENTATIVE k-CENTER PROBLEM



REPRESENTATIVE k-CENTER PROBLEM



REPRESENTATIVE k-CENTER PROBLEM

Representative k-Center problem

Input:

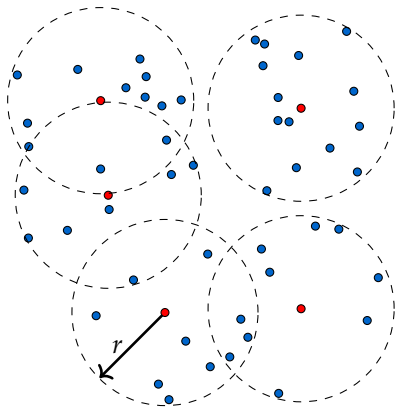
- metric space (X, d)
- integer $k \in \mathbb{Z}_{\geqslant 0}$
- partition $\{X_1, \ldots, X_\gamma\}$
- integers $a_i \leqslant b_i$ for $i \in [\gamma]$

Output: set of centers $C \subseteq X$ s.t

- $|C| \leqslant k$
- $a_i \leqslant |C \cap X_i| \leqslant b_i$ for $i \in [\gamma]$

Goal:

- minimize the distance: $r := \max_{x \in X} d(x, C)$



REPRESENTATIVE k-CENTER PROBLEM

Representative k-Center problem

Input:

- metric space (X, d)
- integer $k \in \mathbb{Z}_{\geqslant 0}$
- partition $\{X_1, \ldots, X_\gamma\}$
- integers $a_i \leqslant b_i$ for $i \in [\gamma]$

Output: set of centers $C \subseteq X$ s.t

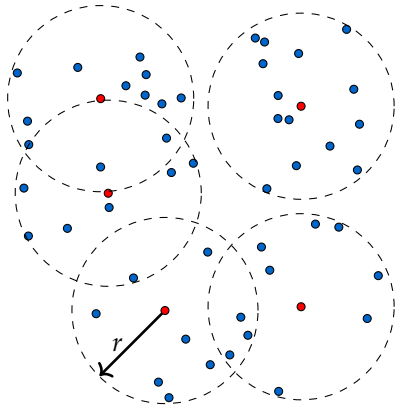
- $|C| \leqslant k$
- $a_i \leqslant |C \cap X_i| \leqslant b_i$ for $i \in [\gamma]$

Goal:

- minimize the distance: $r := \max_{x \in X} d(x, C)$

Known results:

- $(3 - \epsilon)$- hard (APX-hard)
- Chen et.al, Kleindessner et al., Jones et.al $\rightarrow$ 3 **approx. in** $O(nk)$ **time**

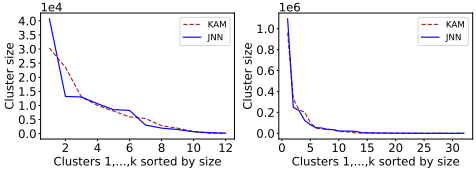


PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

We want every point in the summary to represent subsets of comparable size

PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

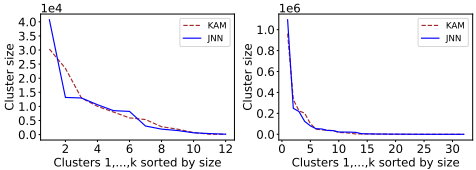
We want every point in the summary to represent subsets of comparable size



(a) Diabetes: $k = 12$ (b) Electric: $k = 32$

PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

We want every point in the summary to represent subsets of comparable size



(a) Diabetes: $k = 12$ (b) Electric: $k = 32$

Priv-Rep k-Center problem

Input:

- metric space (X, d)
- integer $k, L \in \mathbb{Z}_{\geqslant 0}$
- partition $\{X_1, \ldots, X_\gamma\}$
- integers $a_i \leqslant b_i$ for $i \in [\gamma]$

Output: set of centers $C \subseteq X$, assignment $\phi : X \to C$ s.t

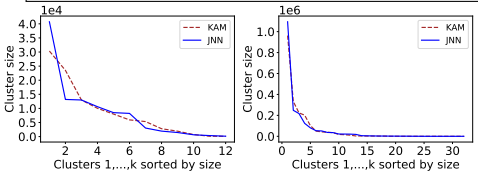
- $|C| \leqslant k$
- $a_i \leqslant |C \cap X_i| \leqslant b_i$ for $i \in [\gamma]$
- $|\phi^{-1}(c)| \geqslant L$ for $c \in C$

Goal:

- minimize the distance: $r := \max_{x \in X} d(x, \phi(x))$

PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

We want every point in the summary to represent subsets of comparable size



(a) Diabetes: $k = 12$ (b) Electric: $k = 32$

NAIVE IDEA: Expand Rep-k-Center solution

Open: 2 blue, 1 red. **What if $L = 2$?**

Priv-Rep k-Center problem

Input:

- metric space (X, d)
- integer $k, L \in \mathbb{Z}_{\geqslant 0}$
- partition $\{X_1, \ldots, X_\gamma\}$
- integers $a_i \leqslant b_i$ for $i \in [\gamma]$

Output: set of centers $C \subseteq X$, assignment $\phi : X \to C$ s.t

- $|C| \leqslant k$
- $a_i \leqslant |C \cap X_i| \leqslant b_i$ for $i \in [\gamma]$
- $|\phi^{-1}(c)| \geqslant L$ for $c \in C$

Goal:

- minimize the distance: $r := \max_{x \in X} d(x, \phi(x))$

PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

Main theorem:

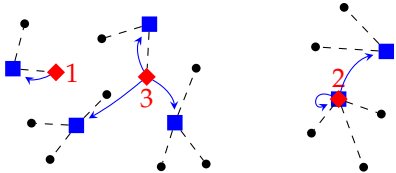
There is a 15 approximation algorithm for PRIV-REP-KC that runs in time $O(nk^2 + k^5)$.

PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

Main theorem:

There is a 15 approximation algorithm for PRIV-REP-KC that runs in time $O(nk^2 + k^5)$.

Backbone: $(\Pi, (1, 3, 2)), L = 2)$

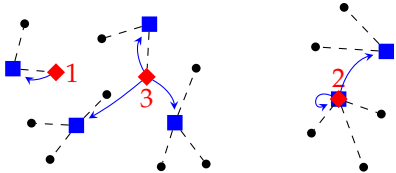


PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

Main theorem:

There is a 15 approximation algorithm for PRIV-REP-KC that runs in time $O(nk^2 + k^5)$.

Backbone: $(\Pi, (1, 3, 2)), L = 2)$



Algorithm sketch

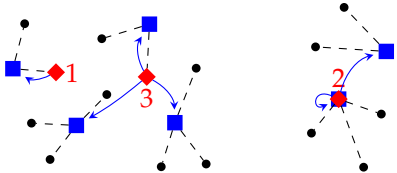
1. Create $q \leqslant k^2$ backbones $(\Pi_1, \eta_1), \ldots, (\Pi_q, \eta_q)$.
2. **for** $i = 1, \ldots, q$ **do**:
   - Solve MIN-REP-REALIZATION for backbone (Π_i, η_i).
   - Compute private clustering with centers C_i of smallest possible radius r_i.
3. **return** the clustering with smallest r_i.

PRIVATE-REPRESENTATIVE k-CENTER PROBLEM

Main theorem:

There is a 15 approximation algorithm for PRIV-REP-KC that runs in time $O(nk^2 + k^5)$.

Backbone: $(\Pi, (1, 3, 2)), L = 2)$



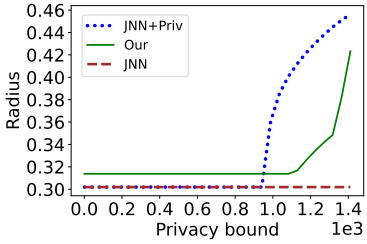
REMARKS:

- Beating nk^2 time algorithm is hard, even for computing private assignment for given set of centers.
- Achieving overall $O(nk^2 + k^5)$ running time is highly nontrivial.
- Backbones can be extended to combine privacy with other constraints.

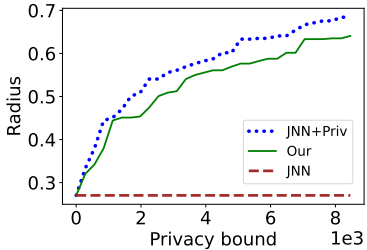
Algorithm sketch

1. Create $q \leqslant k^2$ backbones $(\Pi_1, \eta_1), \ldots, (\Pi_q, \eta_q)$.
2. **for** $i = 1, \ldots, q$ **do**:
   - Solve MIN-REP-REALIZATION for backbone (Π_i, η_i).
   - Compute private clustering with centers C_i of smallest possible radius r_i.
3. **return** the clustering with smallest r_i.

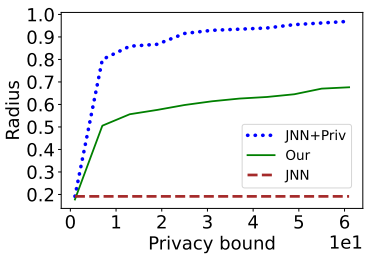
EXPERIMENTS

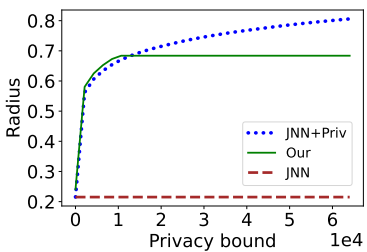


(a) Query: $k = 7$



(b) Diabetes: $k = 12$





Thank you