A Lottery Model for Center-type Problems With Outliers

David G. Harris Thomas Pensyl Aravind Srinivasan
Khoa Trinh

Department of Computer Science, University of Maryland, College Park, USA

APPROX 2017

The k-center problem

- Input:
  - set V of vertices (or *clients*)
  - symmetric distance metric d on V
  - a parameter $k \in \mathbb{Z}$
- Goal: pick a set $\mathcal{S} \in V$ of size $\leq k$ which *minimizes* the maximum connection cost R of any vertex:

$$R := \max_{j \in V} \min_{i \in \mathcal{S}} d(i,j)$$

- Known-results: 2-approximable; best possible unless P=NP [Hsu & Nemhauser '79, Hochbaum & Shmoys '86]

The matroid center problem

- The cardinality constraint $|\mathcal{S}| \leq k$ is replaced by a matroid constraint: $\mathcal{S}$ must be an independent set of some given matroid $\mathcal{M} = (V, \mathcal{I})$.
- Known-results: 3-approximable; best possible unless P=NP [CLLW '13]

The robust model

- Introduced by [Charikar, Khuller, Mount, Narasimhan '01]
- Given $t \leq n$, only need to cover $\geq t$ clients (i.e., we can ignore $\leq n - t$ "outliers" from the instance)
- The robust k-center problem:
  - Upper-bound: 3 [CKMN '01], 2 [CGK '16]
  - Lower-bound: 2
- The robust matroid center problem:
  - Upper-bound: 7 [CLLW '13]
  - Lower-bound: 3 [CLLW '13]

The lottery model

Motivations:

- We can think of the robust k-center as a facility-location problem: “unfairness” arises naturally as some clients might always be considered as outliers,
- In many contexts (e.g., unsupervised machine learning), each vertex j is “autonomous” – we do not want individuals excluded consistently.

The lottery model

The Lottery Model for robust center problems: each client $j \in \mathcal{C}$ submits a target probability p_j and we want a minimum radius R such that $\exists$ distribution $\mathcal{D}$ on subsets of V: for any $\mathcal{S} \sim \mathcal{D}$ (such an S should be efficiently sampleable),

- Feasibility constraint: $\mathcal{S}$ is feasible with prob. 1
- Coverage constraint: $\#$covered clients $\geq t$ with prob. 1
- *Fairness constraint*: Pr[j is covered] $\geq p_j$ for all j

Our main results

- A simple tight 3-approximation algorithm for the robust matroid center problem,
- Our results for the lottery model:

Problem	Radius	Feasibility	Coverage	Fairness
k-center	$2R$	$\leq k$	$\geq (1-\epsilon)t$	$\geq (1-\epsilon)p_j$ for all clients
MatCenter	$3R$ $3R$	Basis + "1" Basis	$\geq t$ $\geq t - \epsilon^2 n$	$\geq p_j$ for all clients $\geq p_j - \epsilon$ for $\geq (1-\epsilon)n$ clients

LP relaxation

- Notation: let $n := |V|$. For any $z \in \mathbb{R}_+^V$ and $S \subseteq V$, let $z(S) := \sum_{i \in S} z_i$ denote the *volume* or *mass* of S.

LP relaxation

- Notation: let $n := |V|$. For any $z \in \mathbb{R}_+^V$ and $S \subseteq V$, let $z(S) := \sum_{i \in S} z_i$ denote the *volume* or *mass* of S.
- Can guess the optimal radius R in $O(n^2)$ time.

LP relaxation

- Notation: let $n := |V|$. For any $z \in \mathbb{R}_+^V$ and $S \subseteq V$, let $z(S) := \sum_{i \in S} z_i$ denote the *volume* or *mass* of S.
- Can guess the optimal radius R in $O(n^2)$ time.
- Let y_i be the indicator for "center i is opened"
- Let x_{ij} be the indicator for "vertex j is connected to center i"

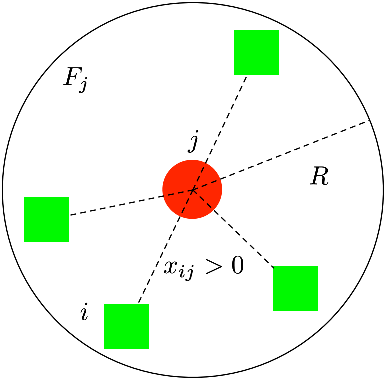
LP relaxation

- Notation: let $n := |V|$. For any $z \in \mathbb{R}_+^V$ and $S \subseteq V$, let $z(S) := \sum_{i \in S} z_i$ denote the *volume* or *mass* of S.
- Can guess the optimal radius R in $O(n^2)$ time.
- Let y_i be the indicator for "center i is opened"
- Let x_{ij} be the indicator for "vertex j is connected to center i"
- Find a fractional solution (x, y) such that

$$\begin{cases} \sum_{i:d_{ij} \leq R} x_{ij} \leq 1, \ \forall j \in \mathcal{C} & j \text{ connects to } \leq 1 \text{ center} \\ \sum_{j \in V} \sum_{i:d_{ij} \leq R} x_{ij} \geq t, & \geq t \text{ vertices are covered} \\ x_{ij} \leq y_i, \ \forall i \in V & j \text{ can only connect to } i \text{ if it's open} \\ y(U) \leq r_{\mathcal{M}}(U), \forall U \subseteq V & \text{matroid rank constraints} \\ 0 \leq x_{i,j}, y_i \leq 1 & \end{cases}$$

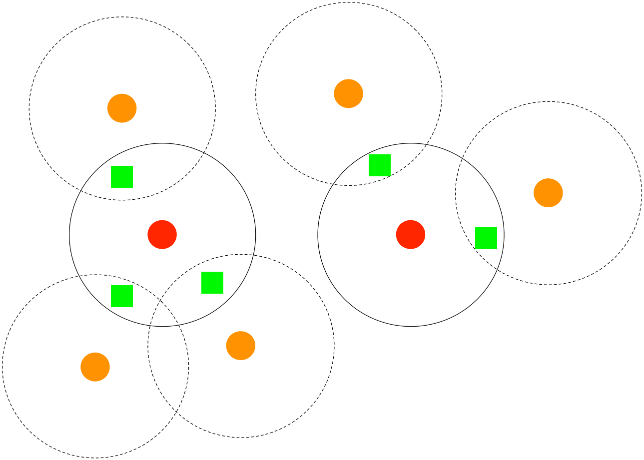
Filtering step

- For each $j \in V$, define *cluster* $F_j := \{i \in V : x_{ij} > 0\}$
- Let $s_j := \sum_{i \in F_j} x_{ij}$ denote the "mass" of F_j. Think of s_j as the *extent* that j gets connected.

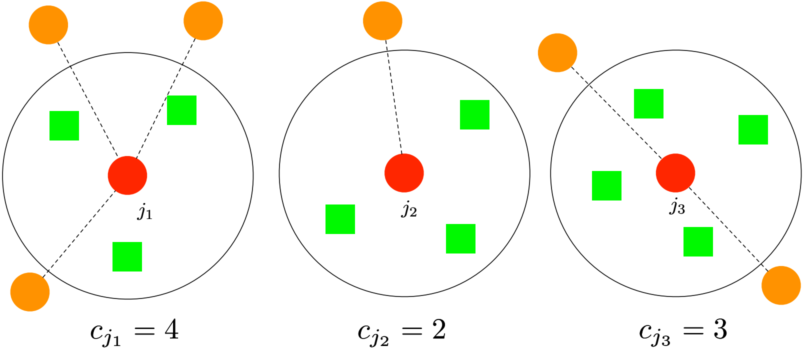


Filtering step

- Similar to the method by [CGTS '02] for k-median
- **Sort** all clusters in **decreasing order** of their mass
- Create a **maximal** set of pairwise **disjoint** clusters:
  - for each F_j: take F_j and remove all F_k that $F_j \cap F_j \neq \emptyset$



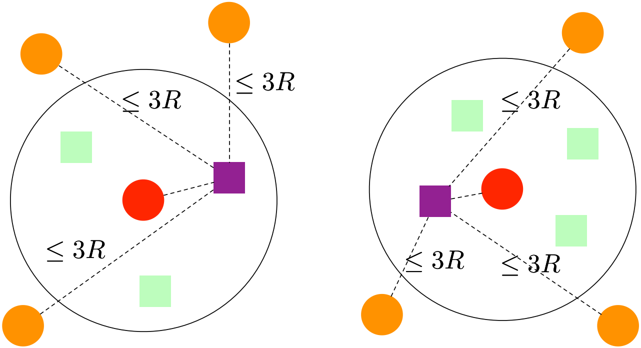
Filtering step



Let V' be the set of 

For each $j \in V'$, let $c_j := 1 + \#$  removed by j

Algorithm



Idea: for each $j \in V'$, opening 1 center in F_j will cover c_j clients within radius $3R$.

Algorithm

1. Define the rounding polytope:

$$\mathcal{P} := \left\{ z \in [0,1]^V : \underbrace{z(U) \leq r_{\mathcal{M}}(U)\ \forall U \subseteq V}_{\text{original matroid constraint}} \wedge \underbrace{z(F_i) \leq 1\ \forall i \in V'}_{\text{partition matroid constraint}} \right\}$$

$\longrightarrow$ $\mathcal{P}$ has integral extreme points.

Algorithm

1. Define the rounding polytope:

$$\mathcal{P} := \left\{ z \in [0,1]^V : \underbrace{z(U) \leq r_{\mathcal{M}}(U)\ \forall U \subseteq V}_{\text{original matroid constraint}} \wedge \underbrace{z(F_i) \leq 1\ \forall i \in V'}_{\text{partition matroid constraint}} \right\}$$

$\longrightarrow \mathcal{P}$ has integral extreme points.

2. Find an extreme point $Z \in \mathcal{P}$ which maximizes f defined as

$$f(z) := \underbrace{\sum_{j \in V'} c_j \cdot z(F_j)}_{\#\text{ clients covered by } z}$$

3. Output $\mathcal{S} := \{i \in V : Z_i = 1\}$

Analysis (sketch)

- Claim: $\mathcal{P} \neq \emptyset$. Define $z'_i := \begin{cases} x_{ij} & \text{if } i \in F_j \text{ for some } j \in V' \\ 0 & \text{otherwise} \end{cases}$

  $\implies z' \in \mathcal{P}$.

Analysis (sketch)

- Claim: $\mathcal{P} \neq \emptyset$. Define $z'_i := \begin{cases} x_{ij} & \text{if } i \in F_j \text{ for some } j \in V' \\ 0 & \text{otherwise} \end{cases}$

  $\implies z' \in \mathcal{P}$.

- At least t vertices are covered:

$$\begin{aligned} f(Z) &\geq f(z') && \text{by choice of } Y \\ &= \sum_{j \in V'} c_j \cdot z'(F_j) && \text{by def. of } f \\ &\geq \sum_{j \in V} s_j && \text{by greedy choice of } V' \\ &\geq t && \text{by LP constraints} \end{aligned}$$

The lottery model

Each client $j \in \mathcal{C}$ submits a target probability p_j.
WANT: a minimum radius R such that $\exists$ distribution $\mathcal{D}$ on subsets of V: for any $\mathcal{S} \sim \mathcal{D}$, we have

- Feasibility constraint: $\mathcal{S}$ is a basis of $\mathcal{M}$ with prob. 1
- Coverage constraint: $\#$covered clients $\geq t$ with prob. 1
- *Fairness constraint*: Pr[j is covered] $\geq p_j$ for all j

Our goal: a randomized algorithm which can sample from $\mathcal{D}$.

Modifying LP relaxation

- Again, assume WLOG that R is the optimal radius.
- Re-define

$$y_i := \Pr[\text{“center } i \text{ is opened”}]$$
$$x_{ij} := \Pr[\text{“vertex } j \text{ is connected to center } i\text{”}]$$

- Find a fractional solution (x, y) such that

$$\begin{cases} \sum_{i:d_{ij} \le R} x_{ij} \ge p_j, \ \forall j \in \mathcal{C} & \text{fairness constraint} \\ \sum_{i:d_{ij} \le R} x_{ij} \le 1, \ \forall j \in \mathcal{C} & j \text{ connects to} \le 1 \text{ center} \\ \sum_{j \in V} \sum_{i:d_{ij} \le R} x_{ij} \ge t, & \ge t \text{ clients are covered} \\ x_{ij} \le y_i, \ \forall i \in V & j \text{ can only connect to } i \text{ if it's open} \\ y(U) \le r_{\mathcal{M}}(U), \forall U \subseteq V & \text{matroid rank constraints} \\ 0 \le x_{i,j}, y_i \le 1 & \end{cases}$$

First attempt

Recall that $z'_i := \begin{cases} x_{ij} & \text{if } i \in F_j \text{ for some } j \in V' \\ 0 & \text{otherwise} \end{cases}$, and

$$\mathcal{P} := \left\{ z \in [0,1]^V : \underbrace{z(U) \le r_{\mathcal{M}}(U)\ \forall U \subseteq V}_{\text{original matroid constraint}} \wedge \underbrace{z(F_i) \le 1\ \forall i \in V'}_{\text{partition matroid constraint}} \right\}$$

A natural approach: decompose $z' \in \mathcal{P}$ into a convex combination of vertices of $\mathcal{P}$ and randomly choose a basis Y according to this distribution

- Y is always a basis of $\mathcal{M}$
- preserving marginals $\Longrightarrow$ fairness constraint is satisfied
- coverage constr. is only satisfied in expectation

Matroid polytopes

Let $\mathcal{M} = (\Omega, \mathcal{I})$ be any matroid. The matroid base polytope of $\mathcal{M}$ is

$$\mathcal{P}_{\mathcal{M}} := \left\{ x \in [0,1]^{\Omega} : x(S) \leq r_{\mathcal{M}}(S) \; \forall S \subseteq \Omega; \;\; x(\Omega) = r_{\mathcal{M}}(\Omega) \right\}.$$

Matroid polytopes

Let $\mathcal{M} = (\Omega, \mathcal{I})$ be any matroid. The matroid base polytope of $\mathcal{M}$ is

$$\mathcal{P}_{\mathcal{M}} := \left\{ x \in [0,1]^{\Omega} : x(S) \leq r_{\mathcal{M}}(S)\ \forall S \subseteq \Omega; \ \ x(\Omega) = r_{\mathcal{M}}(\Omega) \right\}.$$

Definition

Suppose $Ax \leq b$ is a valid inequality of $\mathcal{P}_{\mathcal{M}}$. A face D of $\mathcal{P}_{\mathcal{M}}$ (w.r.t this inequality) is the set $D := \{x \in \mathcal{P}_{\mathcal{M}} : Ax = b\}$.

Matroid polytopes

Let $\mathcal{M} = (\Omega, \mathcal{I})$ be any matroid. The matroid base polytope of $\mathcal{M}$ is

$$\mathcal{P}_{\mathcal{M}} := \left\{ x \in [0,1]^{\Omega} : x(S) \leq r_{\mathcal{M}}(S) \ \forall S \subseteq \Omega; \ \ x(\Omega) = r_{\mathcal{M}}(\Omega) \right\}.$$

Definition

Suppose $Ax \leq b$ is a valid inequality of $\mathcal{P}_{\mathcal{M}}$. A face D of $\mathcal{P}_{\mathcal{M}}$ (w.r.t this inequality) is the set $D := \{x \in \mathcal{P}_{\mathcal{M}} : Ax = b\}$.

Theorem (Edmonds)

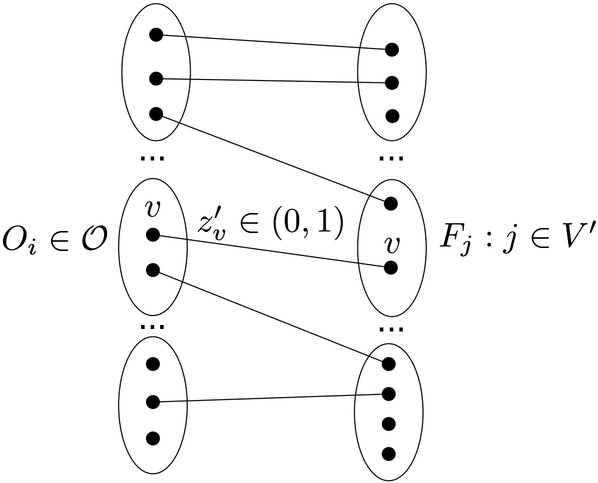
Any face D of $\mathcal{P}_{\mathcal{M}}$ can be characterized by

$$D = \{x \in \mathcal{P}_{\mathcal{M}} : x(T) = b_T \ \ \forall T \in \mathcal{O}; \quad x_i = 0 \ \ \forall i \in J\},$$

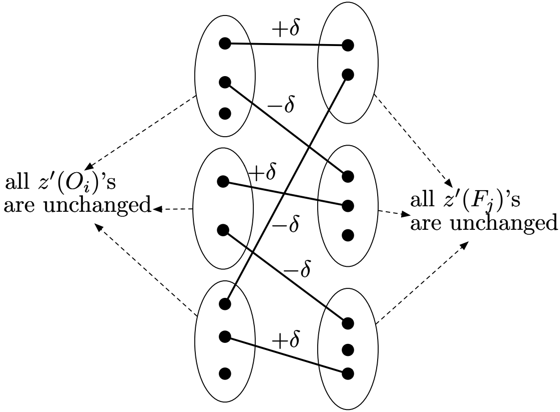
where $J \subseteq \Omega$ and $\mathcal{O}$ is a family of pairwise disjoint sets: $O_1, O_2, \ldots, O_t \subseteq \Omega$, and $b_{O_1}, \ldots, b_{O_t}$ are some integers.

Visualization of z' and $\mathcal{P}$

Compute disjoint sets $O_1, O_2, \ldots, O_t \subseteq V$ and integers $b_{O_1}, \ldots, b_{O_t}$ that characterize the face D of $\mathcal{P}$ w.r.t all tight constraints on z'. Let $O_0 := V \setminus \bigcup_{i=1}^{t} O_i$.

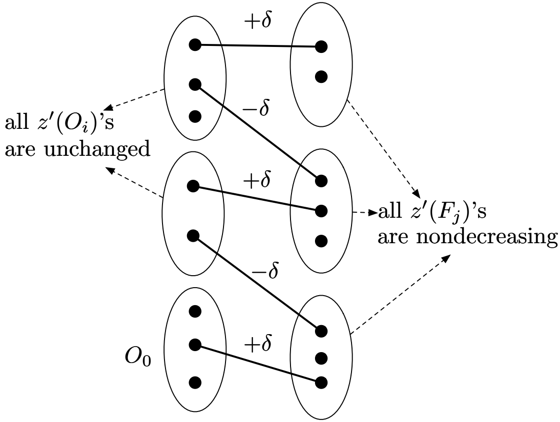


Case 1: $\exists$ a cycle



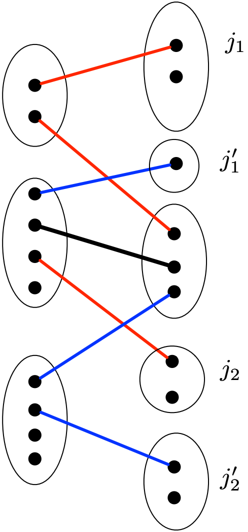
- δ is maximal so that either some z_v hits $\{0,1\}$ or some rank constraint of $\mathcal{P}_{\mathcal{M}}$ becomes tight.
- $f(z') = \sum_{j \in V'} c_j \cdot z'(F_j)$ is **unchanged**

Case 2: $\exists$ a maximal path O_0



- Again, δ is maximal so that either some z_v hits $\{0, 1\}$ or some rank constraint of $\mathcal{P}_{\mathcal{M}}$ becomes tight.
- $f(z') = \sum_{j \in V'} c_j \cdot z'(F_j)$ is **nondecreasing**

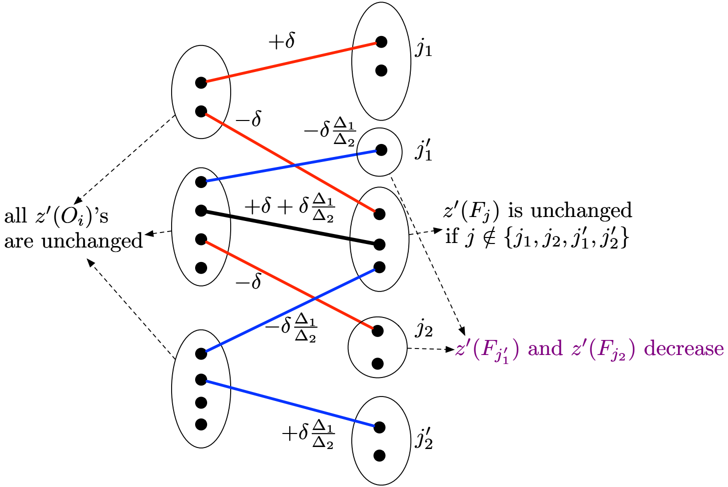
Case 3: $\exists \geq 2$ maximal paths



WLOG, assume $c_{j_1} \geq c_{j_2}$
and $c_{j_1'} \geq c_{j_2'}$

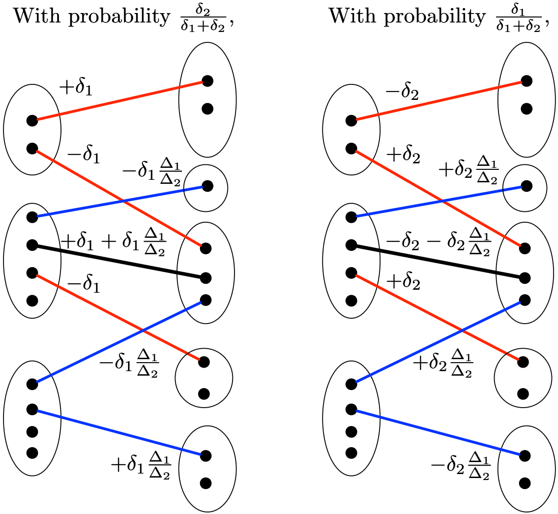
Define $\Delta_1 := c_{j_1} - c_{j_2}$
and $\Delta_2 := c_{j_1'} - c_{j_2'}$

Case 3: $\exists \geq 2$ maximal paths



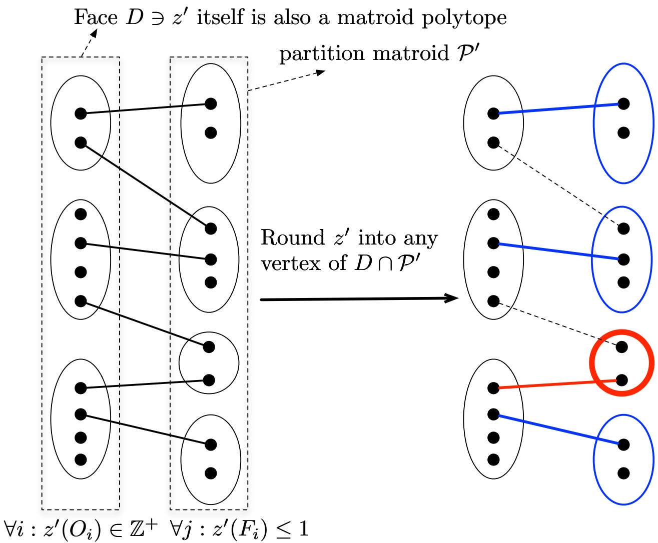
- $\Delta f = \delta c_{j_1} - \delta\frac{\Delta_1}{\Delta_2} c_{j'_1} - \delta c_{j_2} + \delta\frac{\Delta_1}{\Delta_2} c_{j'_2} = 0$ by choice of Δ_1, Δ_2

Case 3: $\exists \geq 2$ maximal paths



- E.g., $\mathbf{E}[\Delta z'(F_{j_1})] = \delta_1 \cdot \frac{\delta_2}{\delta_1+\delta_2} - \delta_2 \cdot \frac{\delta_1}{\delta_1+\delta_2} = 0$.
  Similarly, $\mathbf{E}[\Delta z'(F_{j_2})] = \mathbf{E}[\Delta z'(F_{j_1'})] = \mathbf{E}[\Delta z'(F_{j_2'})] = 0$.

Remaining case



- $f(z') = \sum_{j \in V'} c_j \cdot z'(F_j)$ is **nondecreasing** if we open one "extra" center

Analysis (sketch)

- **Feasibility:** Basis + 1 extra center

Analysis (sketch)

- **Feasibility:** Basis + 1 extra center
- **Coverage:** $\geq t$ clients are covered with prob. 1 since f is nondecreasing during the process

Analysis (sketch)

- **Feasibility:** Basis + 1 extra center
- **Coverage:** $\geq t$ clients are covered with prob. 1 since f is nondecreasing during the process
- **Fairness:**
  - For any $j \in V'$:

$$\Pr[\text{j is covered}] = \mathbf{E}\left[z'(F_j)\right] \geq \text{original mass of } F_j \geq p_j$$

  - For any $j \notin V'$: let $k \in V'$ be the center that removed j

$$\begin{aligned}\Pr[\text{j is covered}] &= \mathbf{E}\left[z'(F_k)\right] \\ &\geq \text{original mass of } F_k \\ &\geq \text{original mass of } F_j \quad \text{by greedy choice} \\ &\geq p_j\end{aligned}$$

Thank you for listening!
Questions?