On the Fixed-Parameter Tractability of Capacitated Clustering

Jason Li (CMU)

Joint work with

Vincent Cohen-Addad

(CNRS & Sorbonne Université)

ICALP 2019

(Capacitated) k-median problem

k-median: metric space (V, d)

clients $C \subseteq V$, facilities $\mathbb{F} \subseteq V$

Find set F of k facilities minimizing

$$\sum_{v \in C} \min_{f \in F} d(v, f)$$

Capacitated k-median: facilities have capacities

Find set F of k facilities and assignment of clients to facilities s.t.

- Every facility f is assigned $\leq cap(f)$ clients
- Minimize $\sum_{v \in C} d(v, \text{assignment}(v))$.

(Capacitated) k-median problem

k-median: metric space (V, d)

clients $C \subseteq V$, facilities $\mathbb{F} \subseteq V$

Find set F of k facilities minimizing

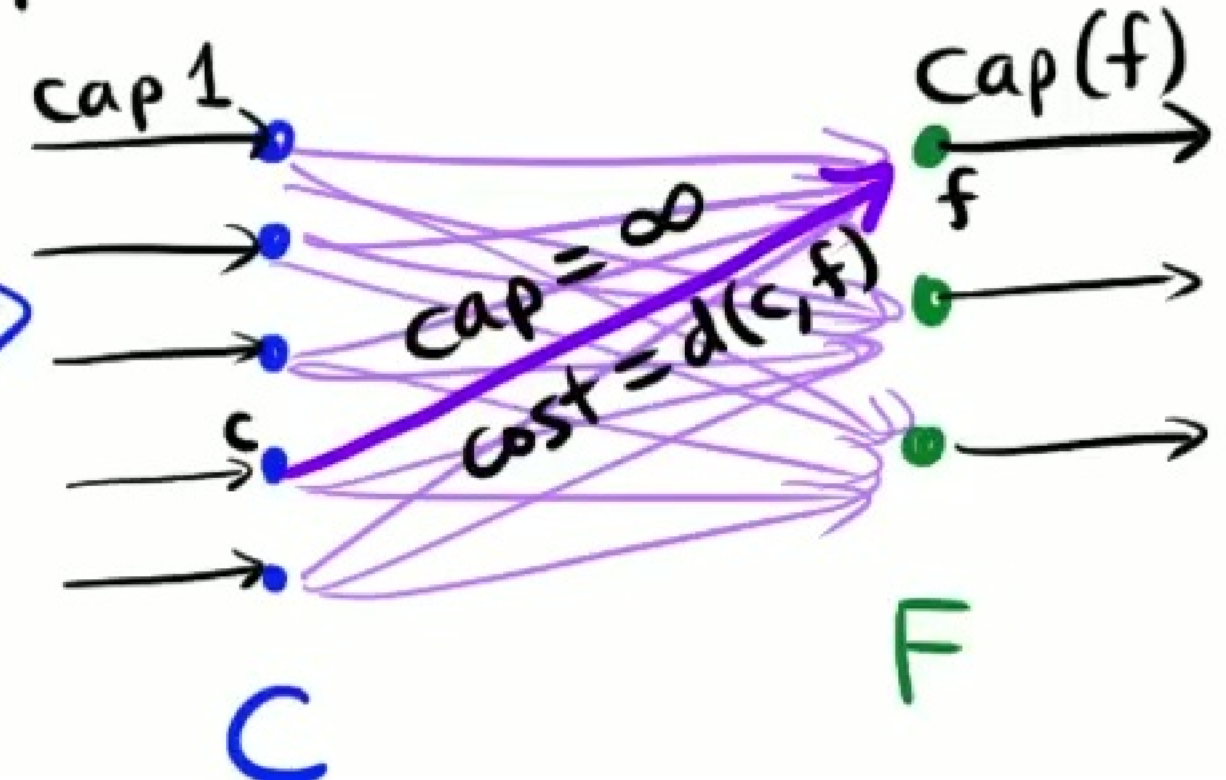
$$\sum_{v \in C} \min_{f \in F} d(v, f)$$

Capacitated k-median: facilities have capacities

Find set F of k facilities and assignment of clients to facilities s.t.

- Every facility f is assigned $\leq \operatorname{cap}(f)$ clients
- Minimize $\sum_{v \in C} d(v, \operatorname{assignment}(v))$.

$= \operatorname{MinCostFlow}(C, F)$



(Capacitated) k-median problem

k-median: metric space (V, d)

clients $C \subseteq V$, facilities $\mathbb{F} \subseteq V$

Find set F of k facilities minimizing

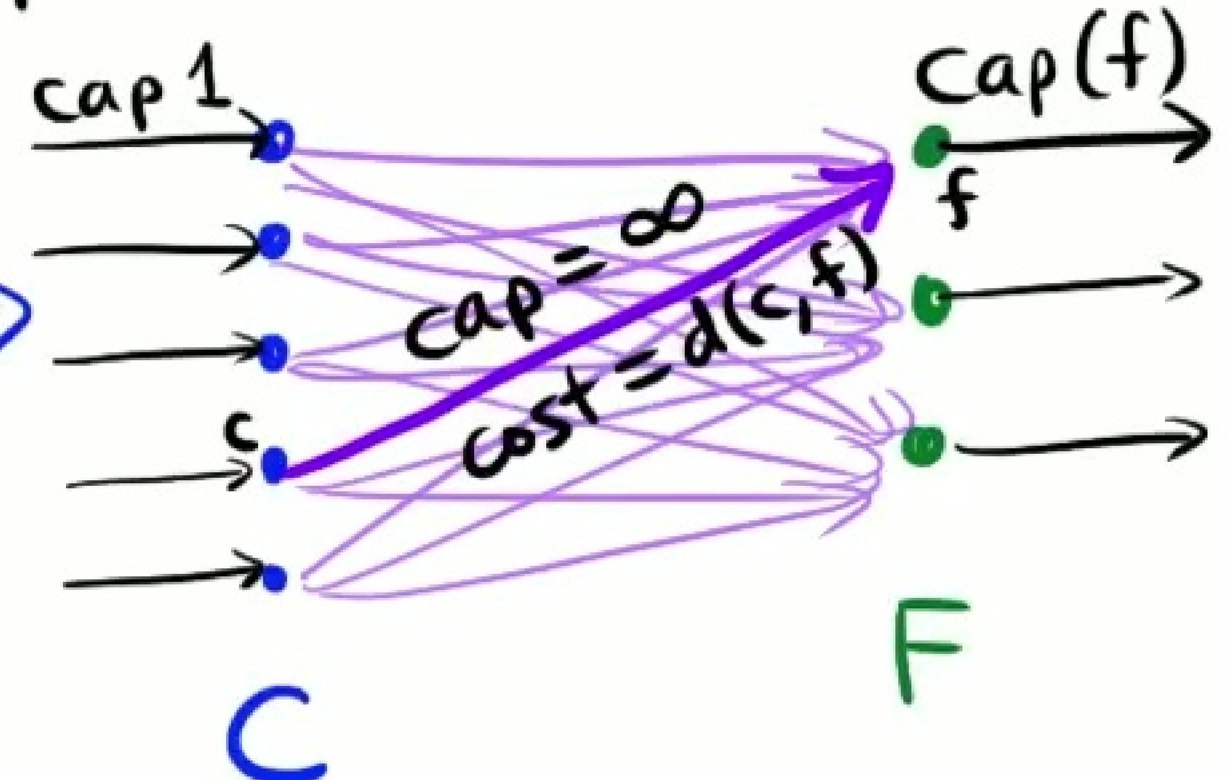
$$\sum_{v \in C} \min_{f \in F} d(v, f)$$

Capacitated k-median: facilities have capacities

Find set F of k facilities and assignment of clients to facilities s.t.

- Every facility f is assigned $\leq \mathrm{cap}(f)$ clients
- Minimize $\sum_{v \in C} d(v, \text{assignment}(v))$.

$= \text{MinCostFlow}(C, F)$ (integral)



Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time ($f(k) \cdot n^{O(1)}$ time)
[CGKLL'19]

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time ($f(k)\cdot n^{O(1)}$ time)
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard
Bicriteria: violate caps by $(1+\varepsilon)$, or
select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time $(f(k)\cdot n^{O(1)}$ time$)$
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard
Bicriteria: violate caps by $(1+\varepsilon)$, or
select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Open: $O(1)$-apx?

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time $(f(k)\cdot n^{O(1)}$ time$)$
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard
Bicriteria: violate caps by $(1+\varepsilon)$, or
select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Open: $O(1)$-apx?
[ABMMW'19] $(7+\varepsilon)$-apx in FPT(k) time

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time $(f(k)\cdot n^{O(1)}$ time$)$
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard
Bicriteria: violate caps by $(1+\varepsilon)$, or
select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Open: $O(1)$-apx?
[ABMMW'19] $(7+\varepsilon)$-apx in FPT(k) time
This talk: $(3+\varepsilon)$ " " " "

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time ($f(k)\cdot n^{O(1)}$ time)
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard
Bicriteria: violate caps by $(1+\varepsilon)$, or
select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Open: $O(1)$-apx?
[ABMMW'19] $(7+\varepsilon)$-apx in FPT(k) time
This talk: $(3+\varepsilon)$ " " " "
Also in work: $(9+\varepsilon)$ for k-means

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time ($f(k)\cdot n^{O(1)}$ time)
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard

Bicriteria: violate caps by $(1+\varepsilon)$, or select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Open: $O(1)$-apx?

[ABMMW'19] $(7+\varepsilon)$-apx in FPT(k) time

This talk: $(3+\varepsilon)$ " " " "

Also in work: $(9+\varepsilon)$ for k-means

Main Technical Contribution: coreset for capacitated k-median

Prior Work

k-median: ≈ 2.6 apx, $(1+2/e)$-apx hard,
tight $(1+2/e)$ in FPT(k) time ($f(k)\cdot n^{O(1)}$ time)
[CGKLL'19]

Cap-k-med: $O(\log k)$ apx, $(1+2/e)$-apx hard
Bicriteria: violate caps by $(1+\varepsilon)$, or
select $(1+\varepsilon)k$ facilities
$\Rightarrow O(1)$-apx [Li 15, 16, 17]

Open: $O(1)$-apx?
[ABMMW'19] $(7+\varepsilon)$-apx in FPT(k) time
This talk: $(3+\varepsilon)$ " " " "
Also in work: $(9+\varepsilon)$ for k-means

Main Technical Contribution: coreset for capacitated k-median
Then, use FPT algo to get $(3+\varepsilon)$-apx

Core-sets

Def[Coreset]: a subset $S \subseteq C$ with weights $w(v): v \in S$
(for cap-k-med) s.t.

$\forall F$ set of k facilities:

$$\mathrm{MinCostFlow}(S, F) \in (1 \pm \varepsilon)\, \mathrm{MinCostFlow}(C, F)$$

Core-sets

Def [Coreset]: a subset $S \subseteq C$ with weights $w(v) : v \in S$
(for cap-k-med) s.t.

$\forall F$ set of k facilities:

$$\text{MinCostFlow}(S, F) \in (1 \pm \varepsilon)\, \text{MinCostFlow}(C, F)$$

(flow can be fractional)

Core-sets

Def[Coreset]: a subset $S \subseteq C$ with weights $w(v): v \in S$
(for cap-k-med) s.t.

$\forall F$ set of k facilities:

$$\text{MinCostFlow}(S, F) \in (1 \pm \varepsilon)\,\text{MinCostFlow}(C, F)$$

(flow can be fractional)

$$\Longrightarrow \min_{|F|=k} \text{MinCostFlow}(S, F) \in \min_{|F|=k} (1 \pm \varepsilon)\,\text{MinCostFlow}(C, F$$

Core-sets

Def[Coreset]: a subset $S \subseteq C$ with weights $w(v): v \in S$
(for cap-k-med) s.t.

$\forall F$ set of k facilities:

$$\mathrm{MinCostFlow}(S, F) \in (1 \pm \varepsilon)\, \mathrm{MinCostFlow}(C, F)$$

(flow can be fractional)

$$\Longrightarrow \min_{|F|=k} \mathrm{MinCostFlow}(S, F) \in \min_{|F|=k} (1 \pm \varepsilon)\, \mathrm{MinCostFlow}(C, F$$

Suffices to solve : $|S|$ many weighted clients
(ideally $|S| \ll |C|$)

Core-sets

Def[Coreset]: a subset $S \subseteq C$ with weights $w(v): v \in S$
(for ~~cap-k-med~~ k-median) s.t.

$\forall F$ set of k facilities:

$\cancel{\mathrm{MinCostFlow}(S, F) \in (1 \pm \varepsilon)\,\mathrm{MinCostFlow}(C, F)}$

$$\sum_{v \in C} \min_{f \in F} d(v, f) \in (1 \pm \varepsilon) \sum_{v \in C} \min_{f \in F} d(v, f)$$

Thm[Chen; Feldman-Langberg]: $\exists$ coreset for k-median size $\mathrm{poly}(k \log n\, \varepsilon^{-1})$.

Core-sets

Def[Coreset] (for ~~cap-k-med~~ k-median): a subset $S \subseteq C$ with weights $w(v) : v \in S$ s.t.

$\forall F$ set of k facilities:

$\cancel{\text{MinCostFlow}(S, F) \in (1 \pm \varepsilon)\, \text{MinCostFlow}(C, F)}$

$$\sum_{v \in C} \min_{f \in F} d(v, f) \in (1 \pm \varepsilon) \sum_{v \in C} \min_{f \in F} d(v, f)$$

Thm[Chen; Feldman-Langberg]: $\exists$ coreset for k-median size $\text{poly}(k \log n\, \varepsilon^{-1})$.

This work: $\exists$ coreset for cap-k-med (and cap-k-means) size $\text{poly}(k \log n\, \varepsilon^{-1})$.

Chen's Coreset Algorithm

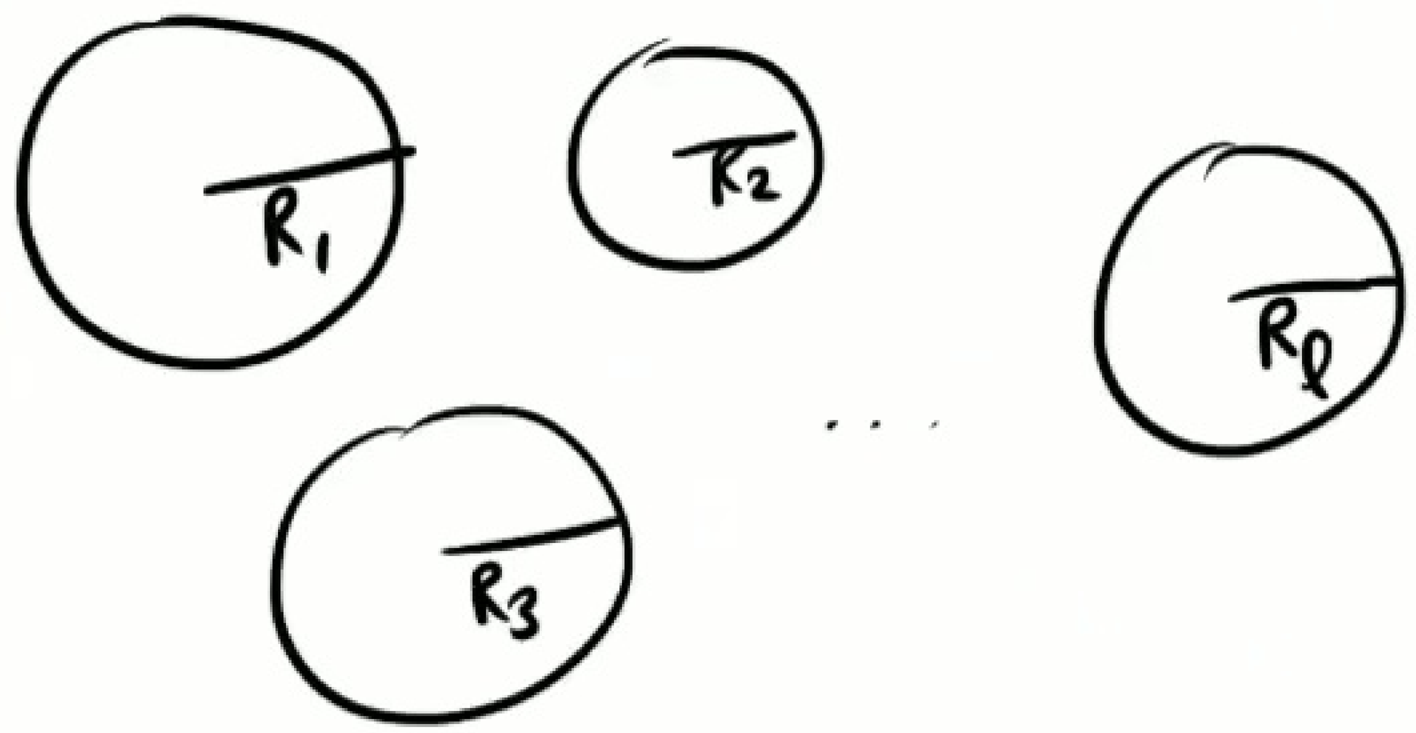
- Random sampling
- For each F $(|F|=k)$,

$$\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$$

Union bound over all F

Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $$\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$$
  Union bound over all F
- Suppose can $\underline{\text{cluster}}$ C into ℓ clusters

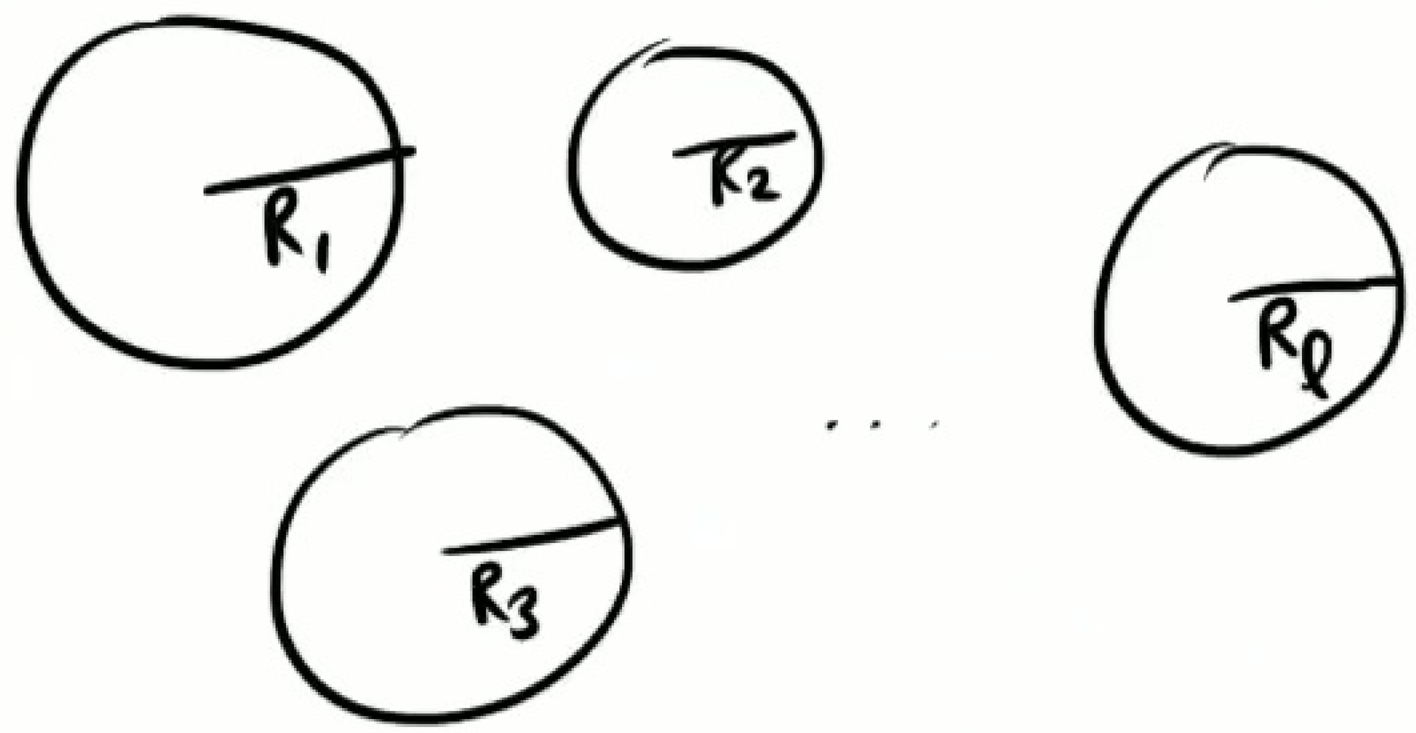


Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $$\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon) \sum_{v\in C} d(v,F) \right] \gg 1-n^{-k}$$

  Union bound over all F

- Suppose can <u>cluster</u> C into ℓ clusters



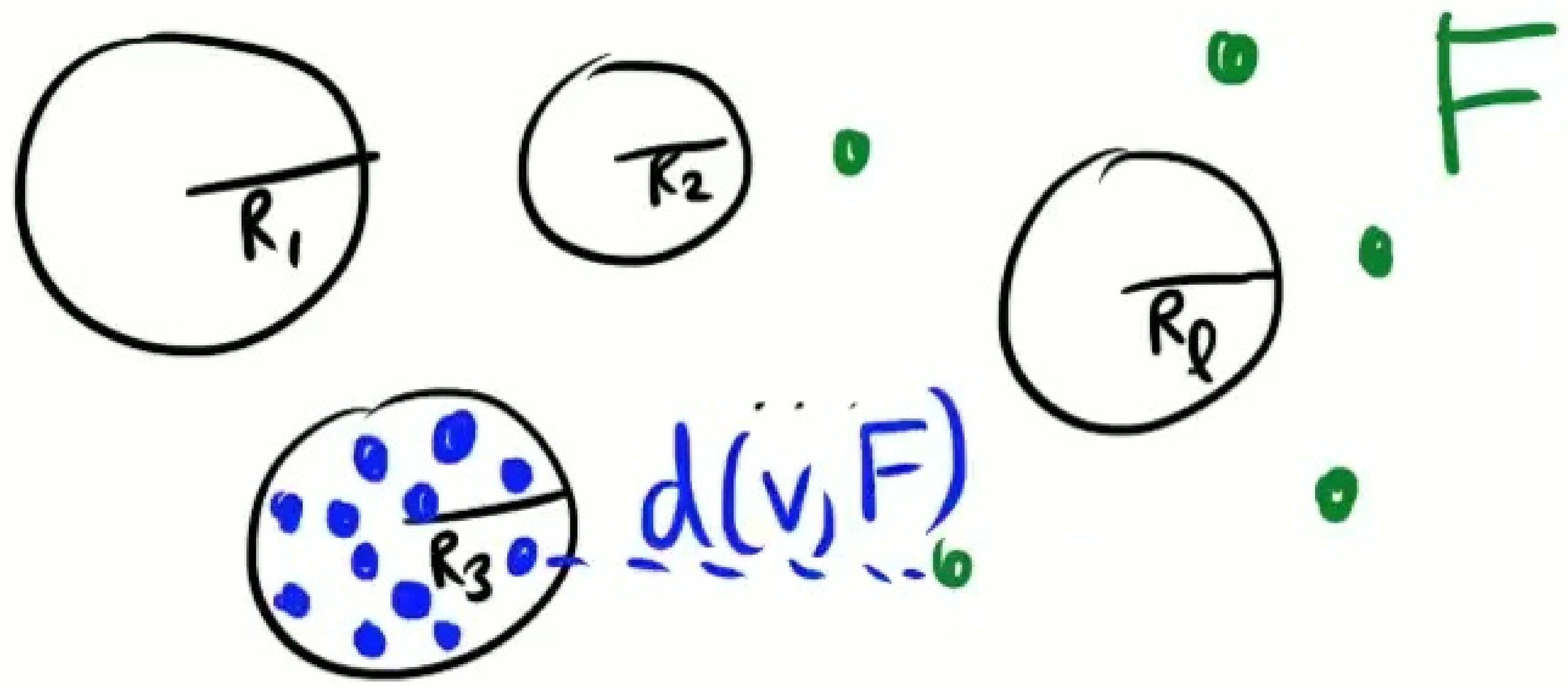
- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

  Union bound over all F

- Suppose can cluster C into ℓ clusters



- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

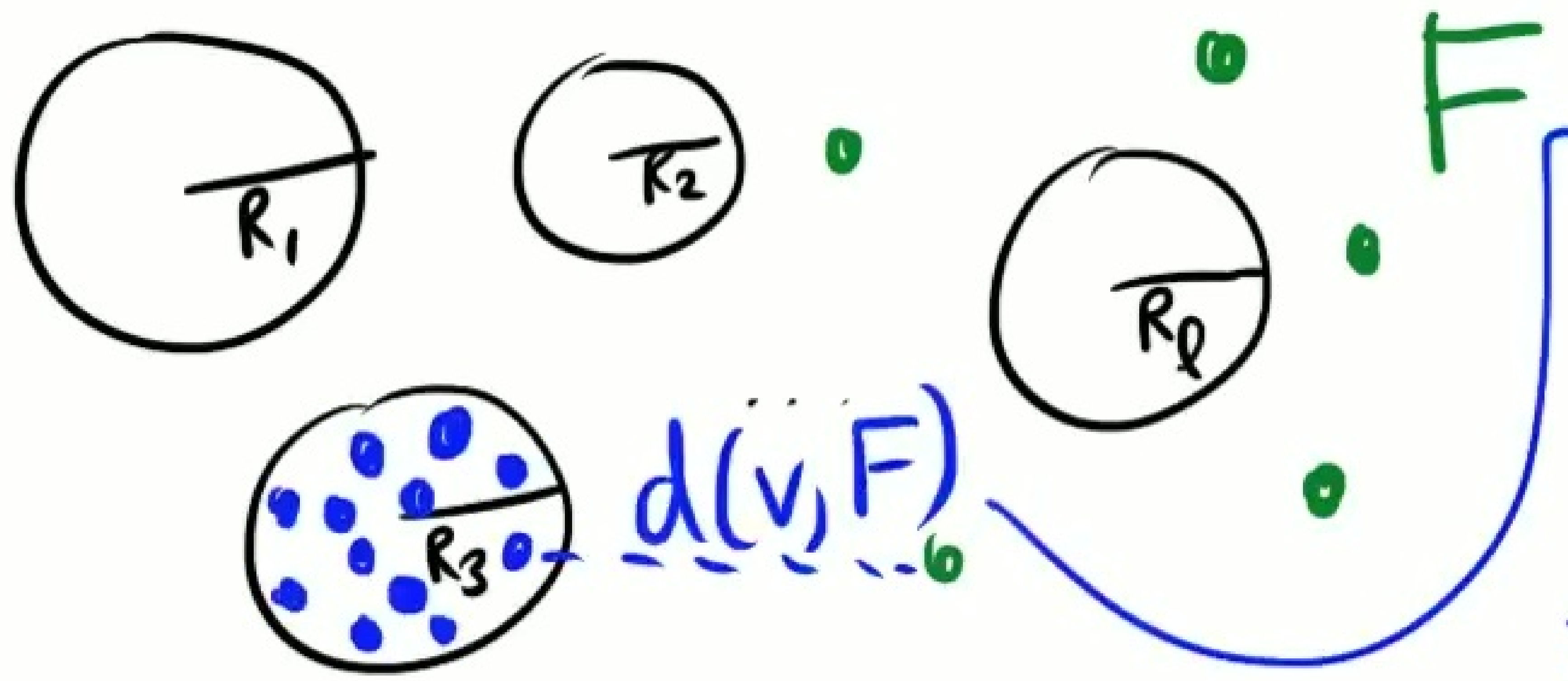
Chen's Coreset Algorithm

- Random sampling
- For each F ($|F| = k$),

$$\Pr\Big[\sum_{v \in S} w(v)\, d(v,F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v,F)\Big] \gg 1 - n^{-k}$$

Union bound over all F

- Suppose can <u>cluster</u> C into ℓ clusters



Claim:
$\{d(v,F) : v \in C_i\}$
contained in
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Chen's Coreset Algorithm

- Random sampling
- For each F ($|F| = k$),
  $\Pr\left[\sum_{v \in S} w(v)\, d(v,F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v,F) \right] \gg 1 - n^{-k}$

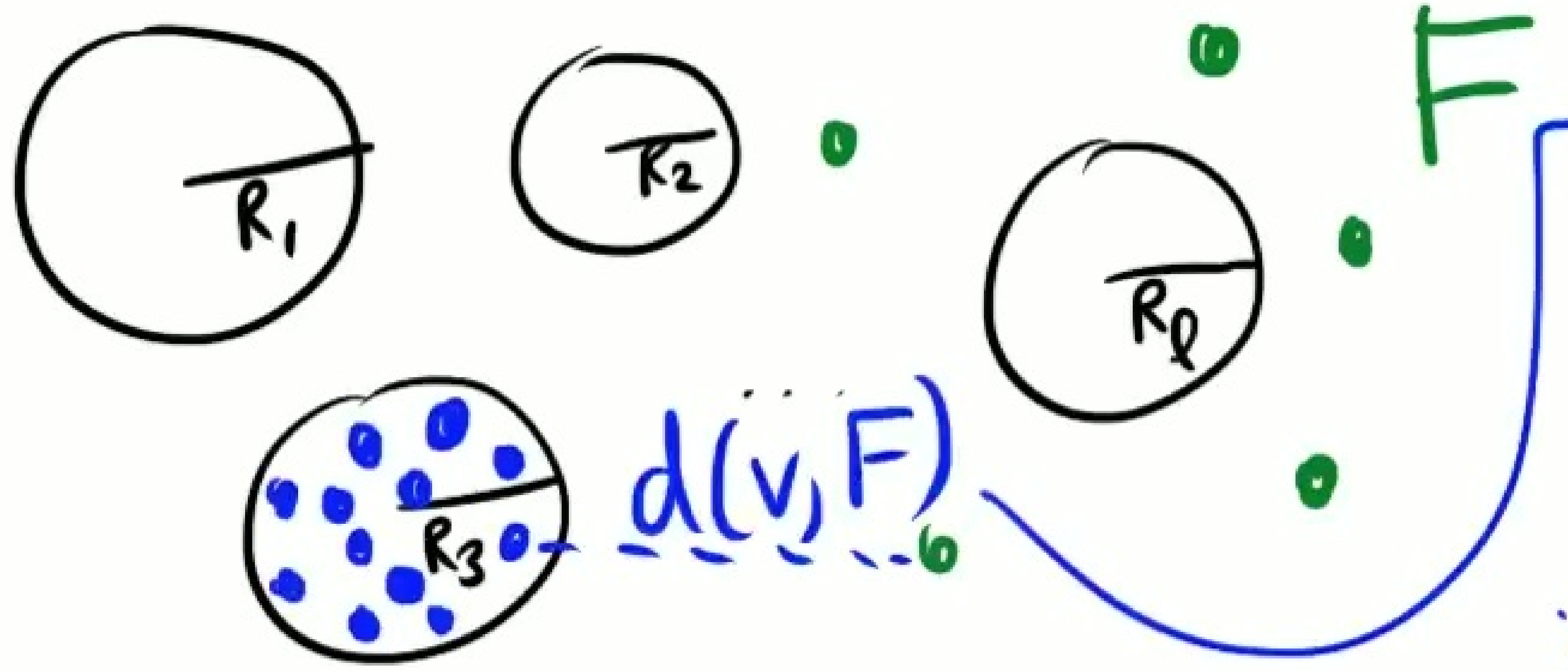
  Union bound over all F

Chernoff bound:

$\underset{v \in S}{\mathrm{avg}}(d(v,F)) \in \underset{v \in C_i}{\mathrm{avg}}(d(v,F)) + \varepsilon R_i$

w.p. $\geq 1 - n^{-k}$

- Suppose can <u>cluster</u> C into ℓ clusters



Claim:
$\{ d(v,F) : v \in C_i \}$
contained in interval size $O(R_i)$

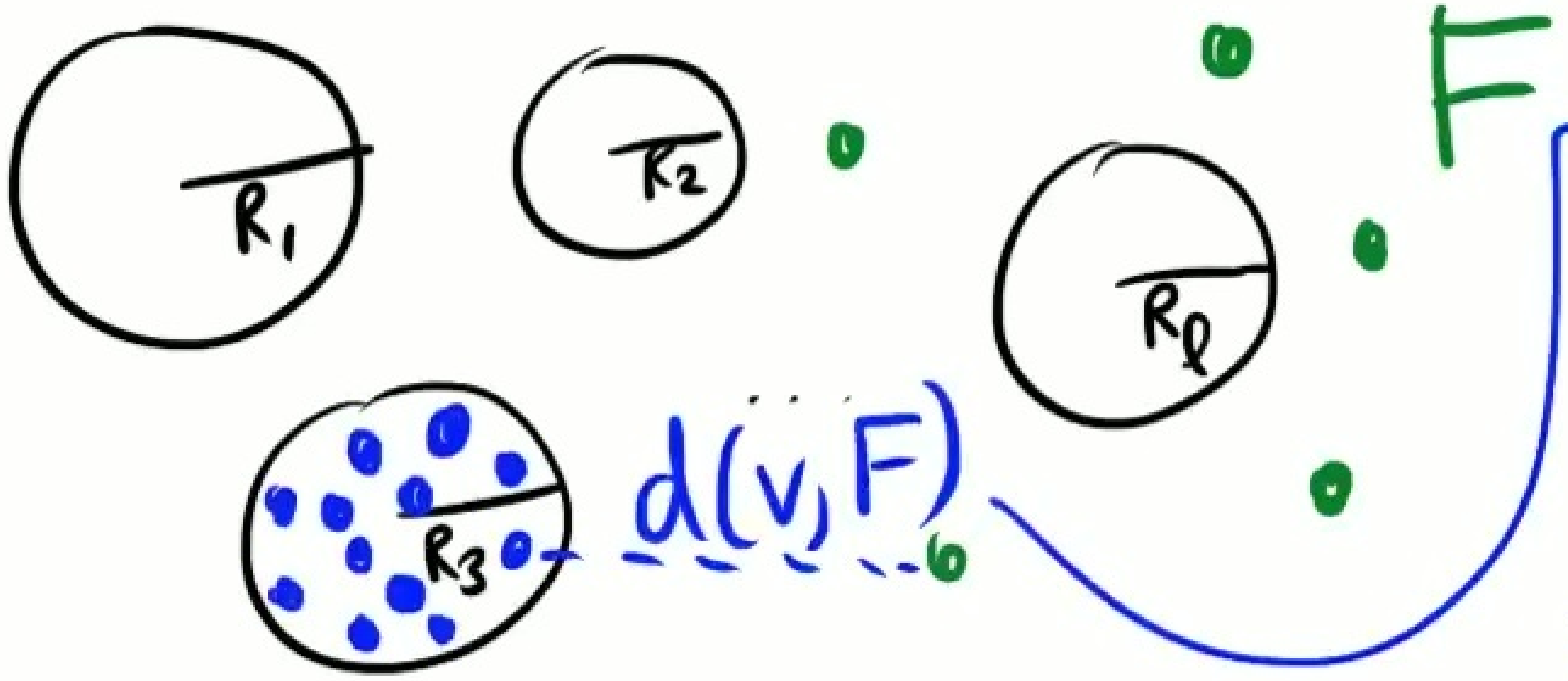
- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

  Union bound over all F

- Suppose can <u>cluster</u> C into ℓ clusters



→ Claim:
$\{d(v,F) : v\in C_i\}$
contained in
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

→ Chernoff bound:

$\operatorname{avg}_{v\in S}(d(v,F)) \in \operatorname{avg}_{v\in C_i}(d(v,F)) + \varepsilon R_i$

w.p. $\geq 1-n^{-k}$

$\Rightarrow \sum_{v\in S} \frac{|C_i|}{s} d(v,F)$

$= |C_i| \cdot \operatorname{avg}_{v\in S} d(v,F)$

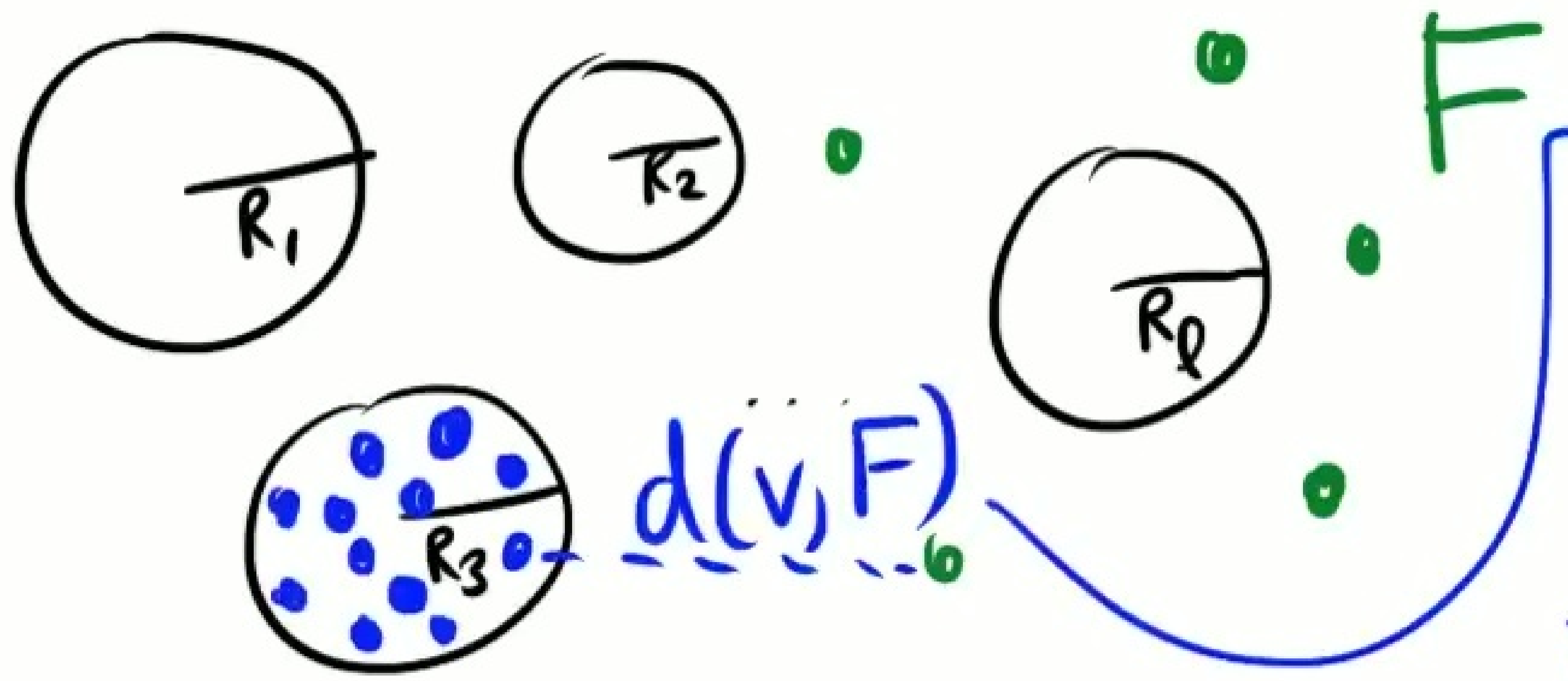
$\in |C_i| \cdot \operatorname{avg}_{v\in C_i}(d(v,F)) + |C_i|\cdot\varepsilon R_i$

$= \sum_{v\in C_i} d(v,F) \quad + |C_i|\cdot\varepsilon R_i$

Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $\Pr\left[\sum_{v \in S} w(v)\, d(v,F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v,F)\right] \gg 1 - n^{-k}$

  Union bound over all F
- Suppose can cluster C into ℓ clusters



- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Chernoff bound:

$\operatorname*{avg}_{v \in S}(d(v,F)) \in \operatorname*{avg}_{v \in C_i}(d(v,F)) + \varepsilon R_i$ w.p. $\geq 1 - n^{-k}$

$\Rightarrow \sum_{v \in S} \frac{|C_i|}{s} d(v,F)$

$= |C_i| \cdot \operatorname*{avg}_{v \in S} d(v,F)$

$\in |C_i| \cdot \operatorname*{avg}_{v \in C_i}(d(v,F)) + |C_i| \cdot \varepsilon R_i$

$= \sum_{v \in C_i} d(v,F) + |C_i| \cdot \varepsilon R_i$

Claim:
$\{d(v,F) : v \in C_i\}$
contained in
interval size $O(R_i)$

Total error:
$\sum_{C_i} |C_i| \cdot \varepsilon R_i = \varepsilon \sum_{C_i} |C_i| \cdot R_i$

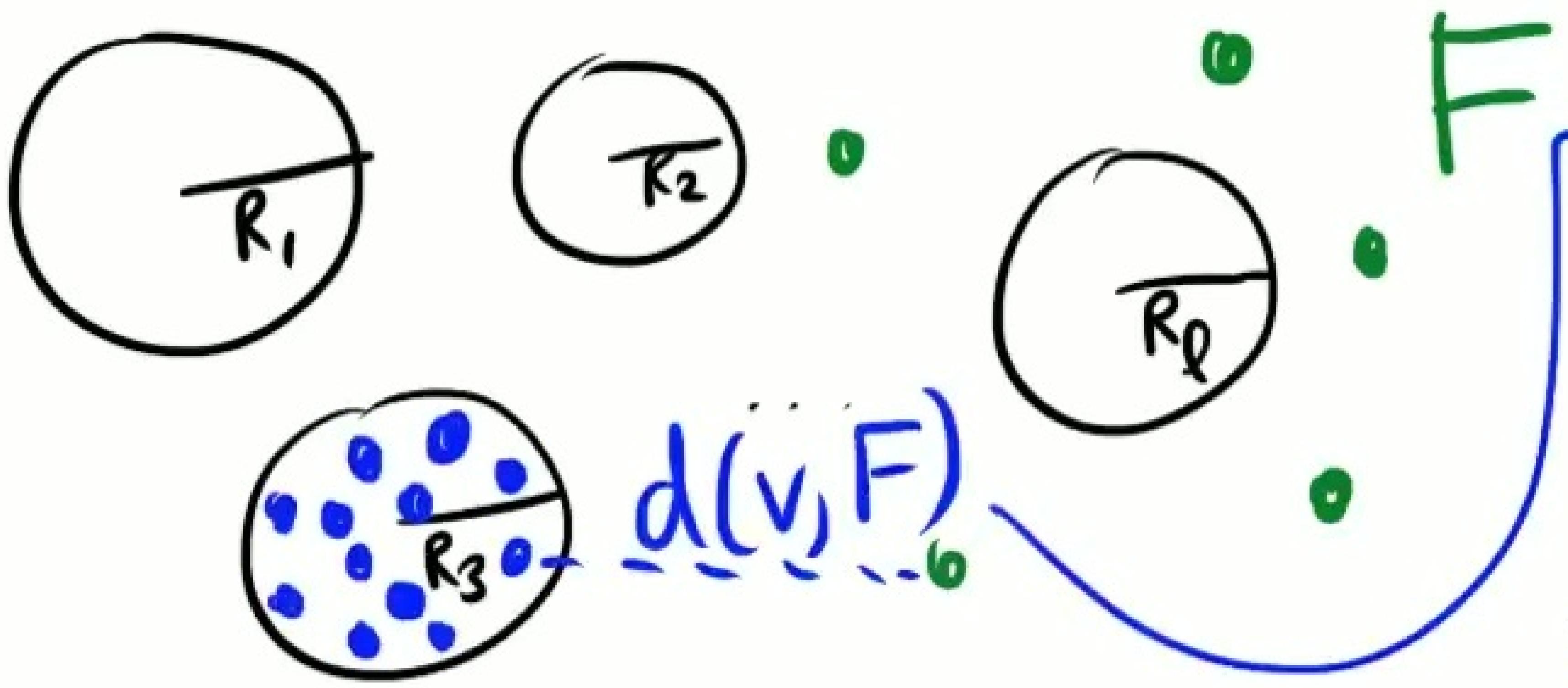
Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,

$$\Pr\Big[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\Big] \gg 1-n^{-k}$$

Union bound over all F

- Suppose can cluster C into ℓ clusters



→ Claim:
$\{d(v,F) : v \in C_i\}$ contained in interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \text{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

→ Chernoff bound:

$$\operatorname*{avg}_{v\in S}(d(v,F)) \in \operatorname*{avg}_{v\in C_i}(d(v,F)) + \varepsilon R_i$$

w.p. $\geq 1-n^{-k}$

$$\Rightarrow \sum_{v\in S} \frac{|C_i|}{s} d(v,F)$$

$$= |C_i| \cdot \operatorname*{avg}_{v\in S} d(v,F)$$

$$\in |C_i| \cdot \operatorname*{avg}_{v\in C_i}(d(v,F)) + |C_i|\cdot \varepsilon R_i$$

$$= \sum_{v\in C_i} d(v,F) \quad + |C_i|\cdot\varepsilon R_i$$

Total error:

$$\sum_{C_i} |C_i|\cdot \varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$$

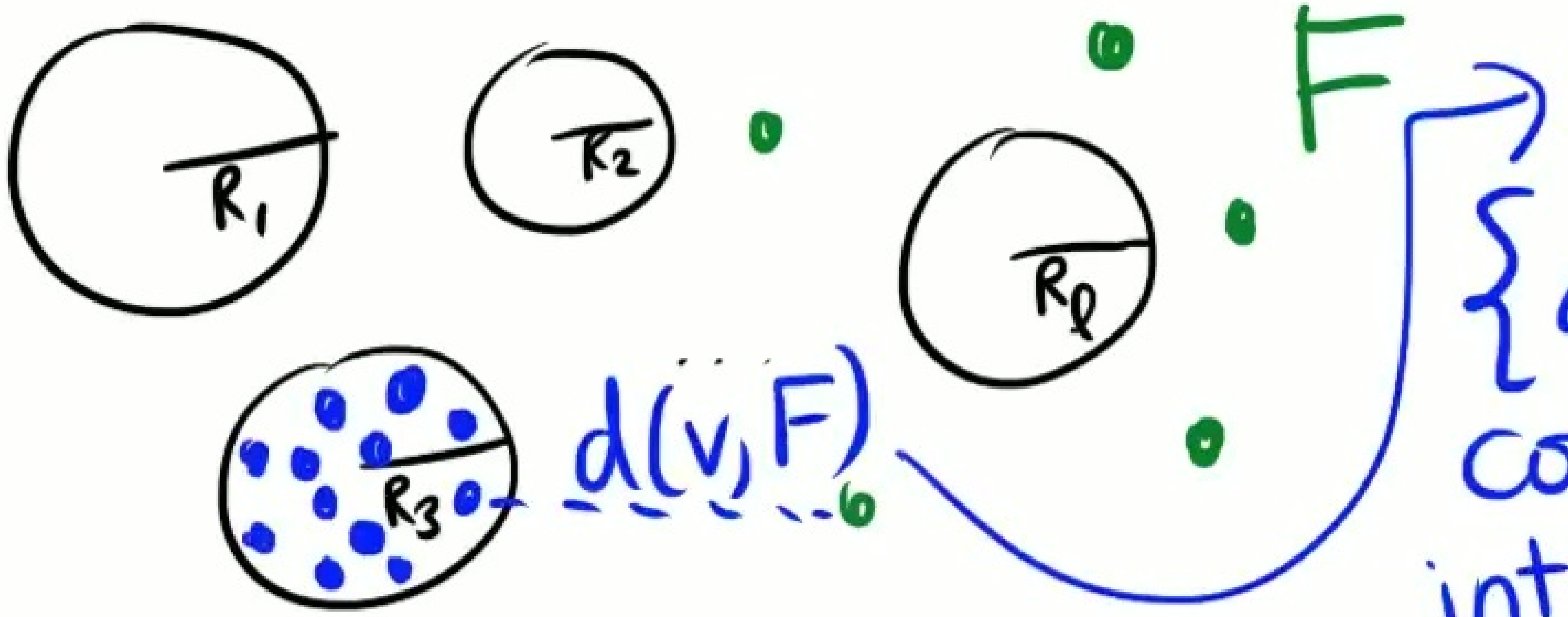
Chen's Coreset Algorithm

- Random sampling
- For each F ($|F|=k$),

$$\Pr\Big[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\Big] \gg 1-n^{-k}$$

Union bound over all F

- Suppose can cluster C into ℓ clusters



Claim:
$\{d(v,F) : v$
contained in
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Let F' be 3-approx k-median soln

Total error:

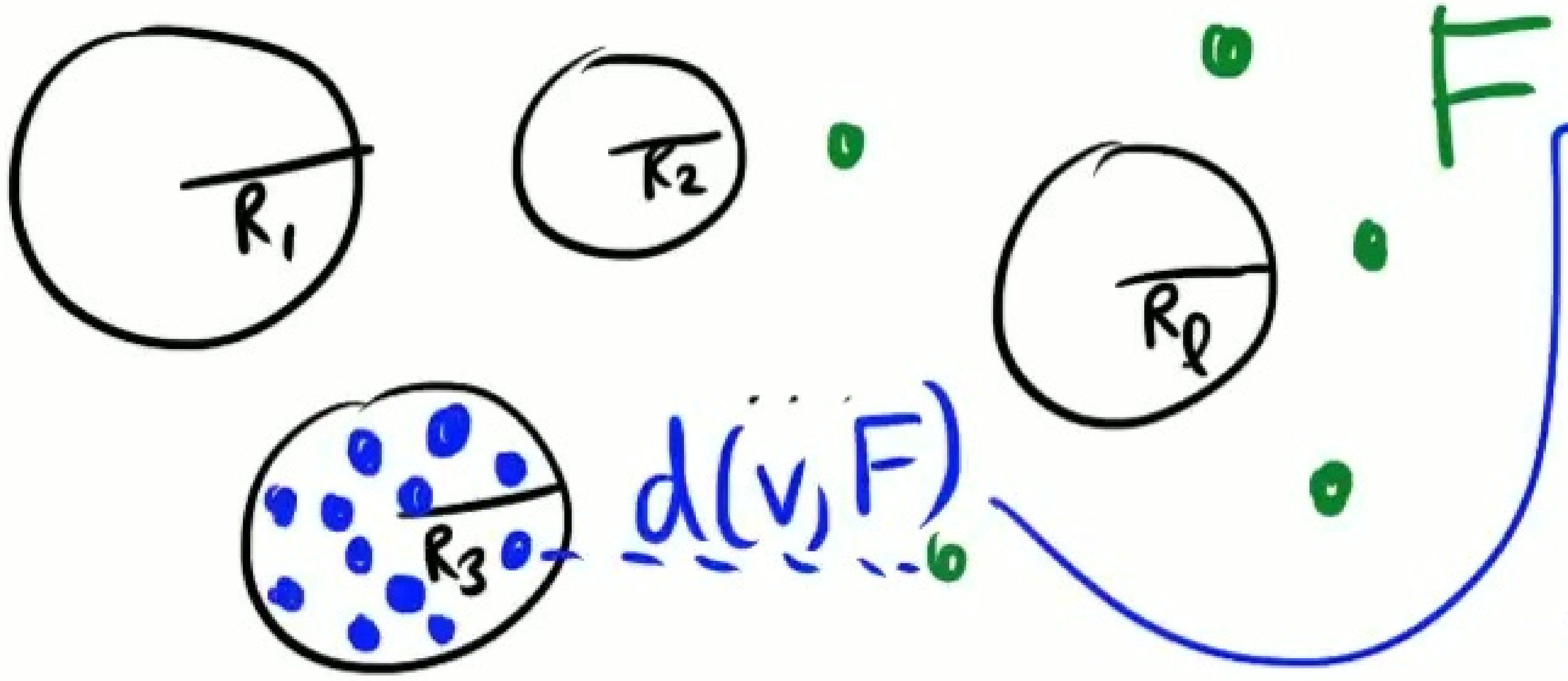
$$\sum_{C_i} |C_i|\cdot \varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$$

Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

  Union bound over all F

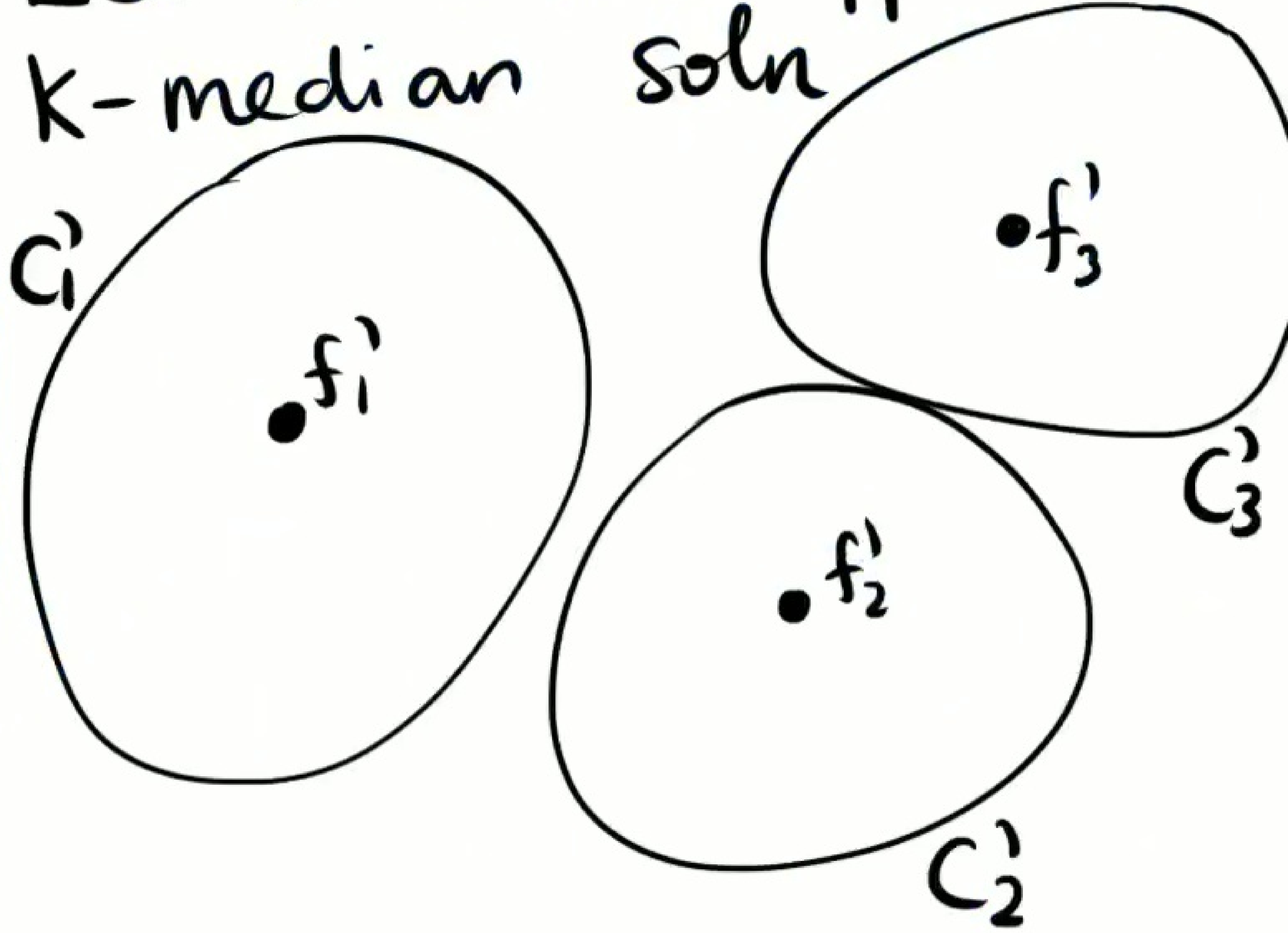
- Suppose can cluster C into ℓ clusters



Claim:
$\{d(v,F): v$
contained in
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Let F' be 3-approx k-median soln



Total error:
$\sum_{C_i} |C_i|\cdot \varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$

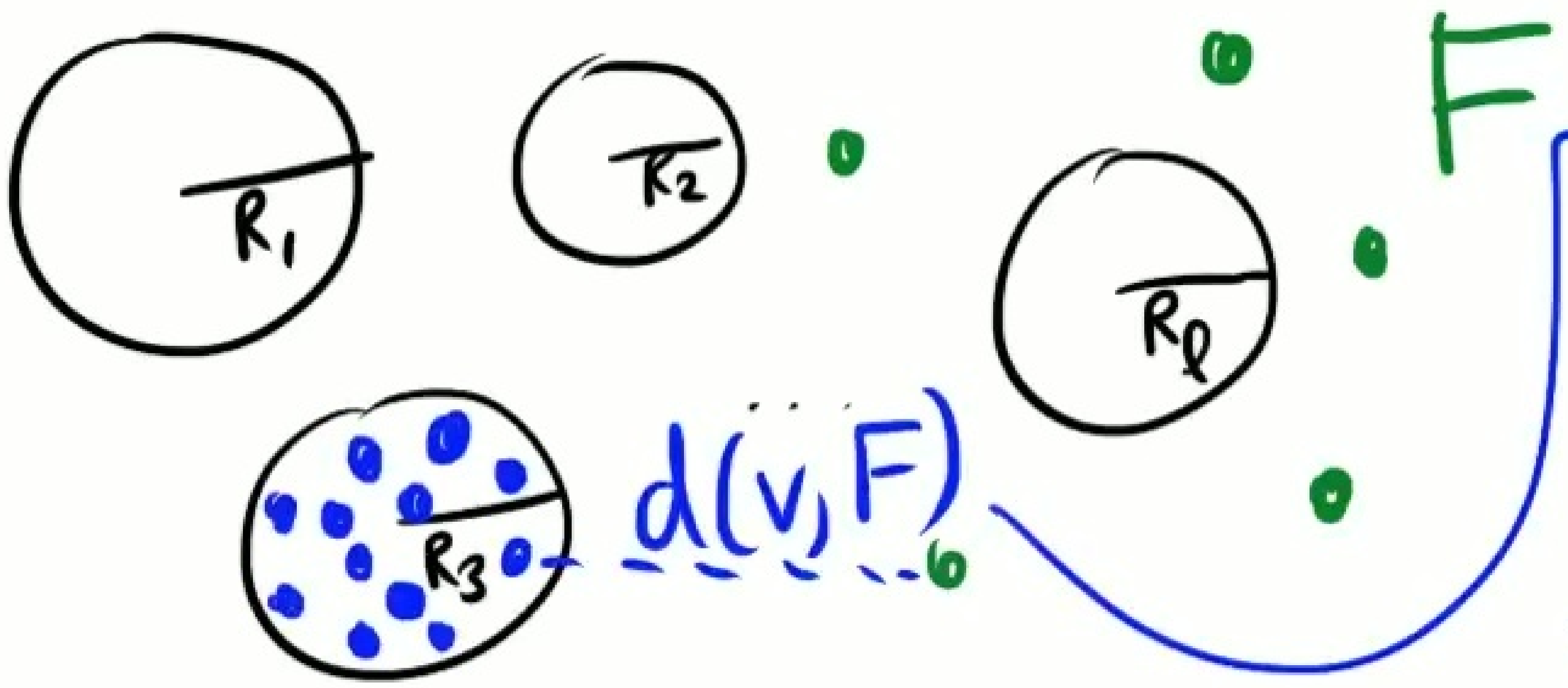
Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,

$$\Pr\Big[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\Big] \gg 1-n^{-k}$$

Union bound over all F

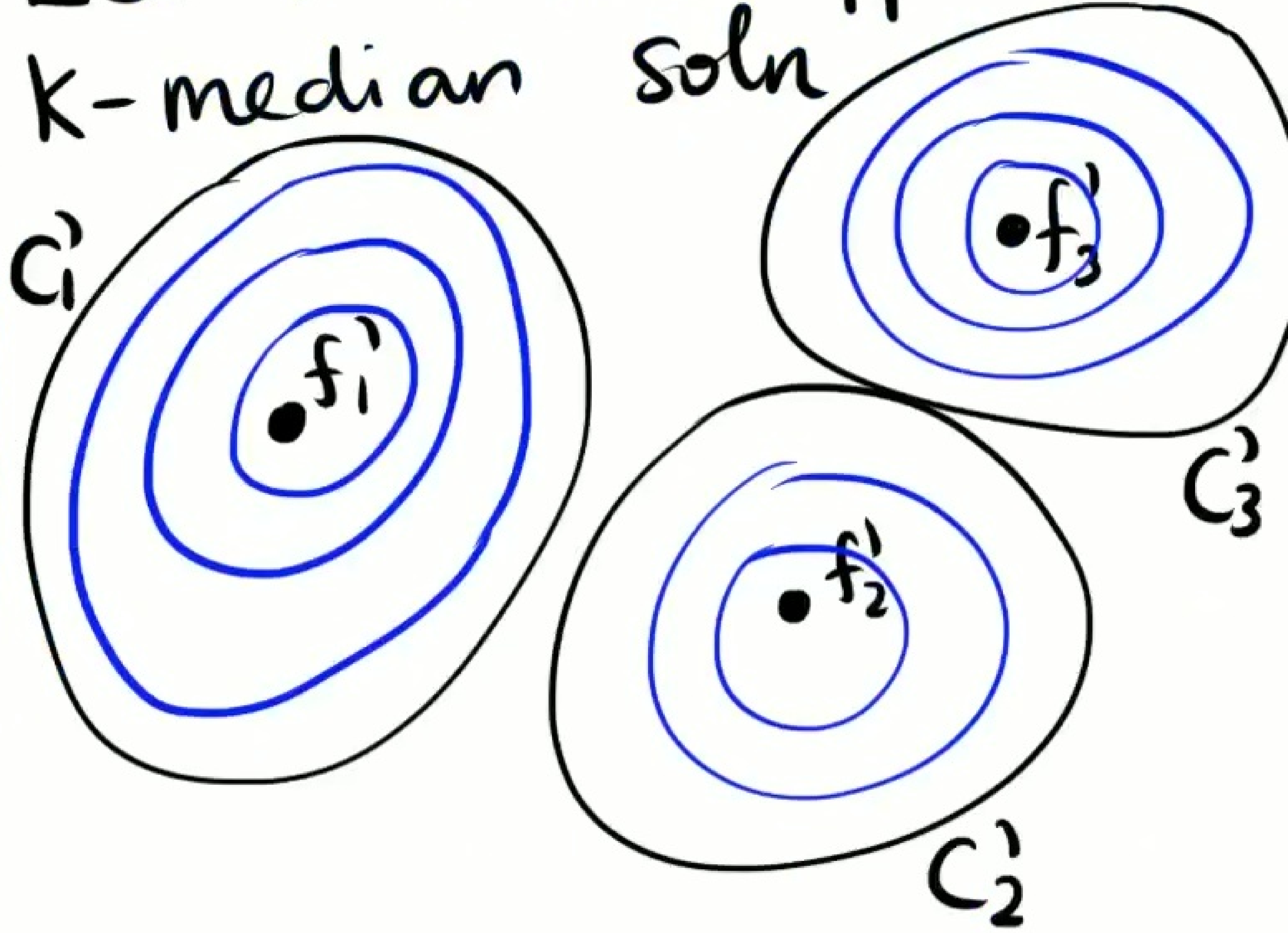
- Suppose can <u>cluster</u> C into ℓ clusters



Claim:
$\{d(v,F) : v$
contained i
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Let F' be 3-approx
k-median soln



Total error:

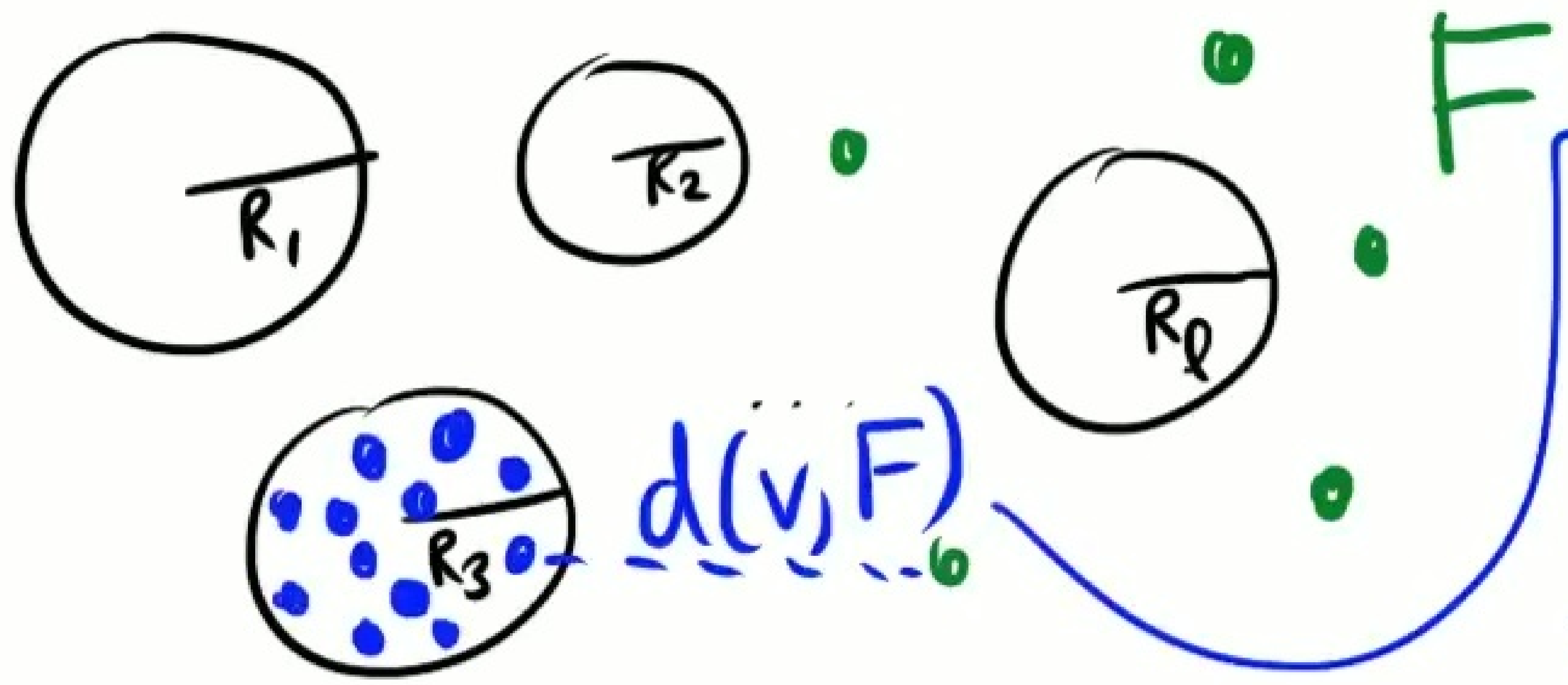
$$\sum_{C_i} |C_i|\cdot\varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$$

Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,
  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

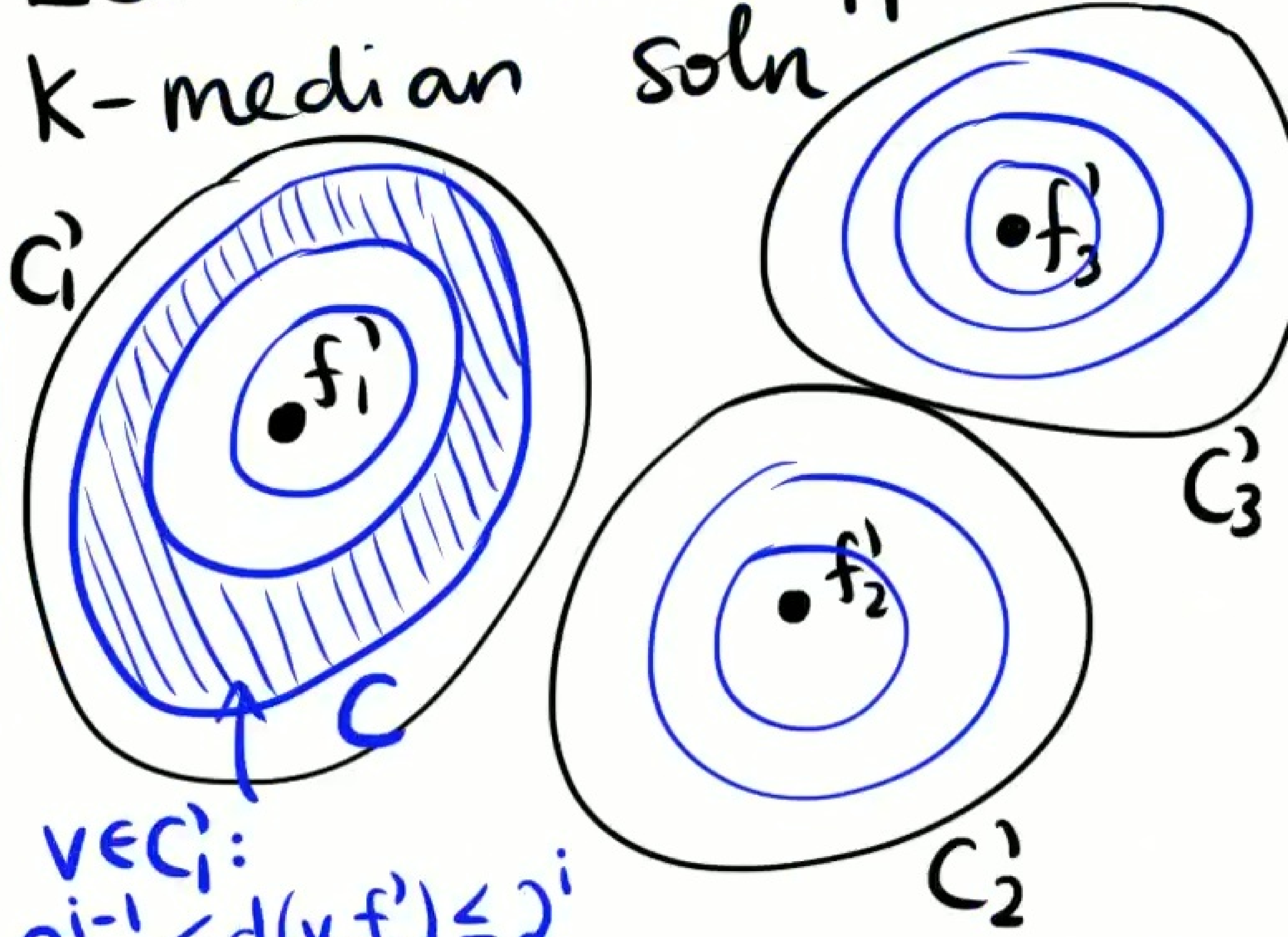
  Union bound over all F

- Suppose can cluster C into ℓ clusters



Claim:
$\{d(v,F) : v \in C_i\}$
contained in
interval size $O(R_i)$

Let F' be 3-approx
k-median soln



$v \in C_i$:
$2^{i-1} \le d(v,f_i') \le 2^i$

- Within each cluster C_i:
  - sample $s := \text{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Total error:
$\sum_{C_i} |C_i|\cdot\varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$

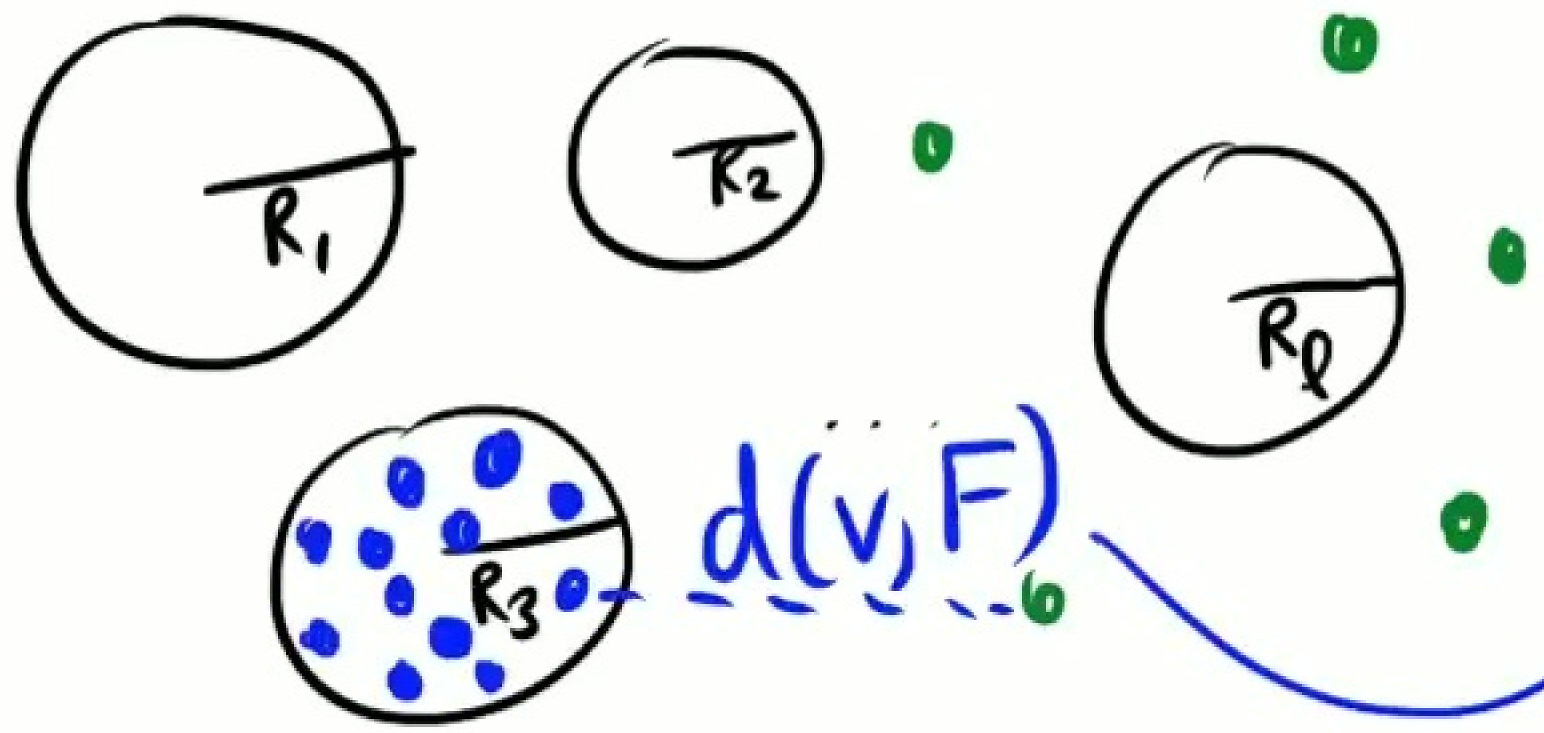
Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,

  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

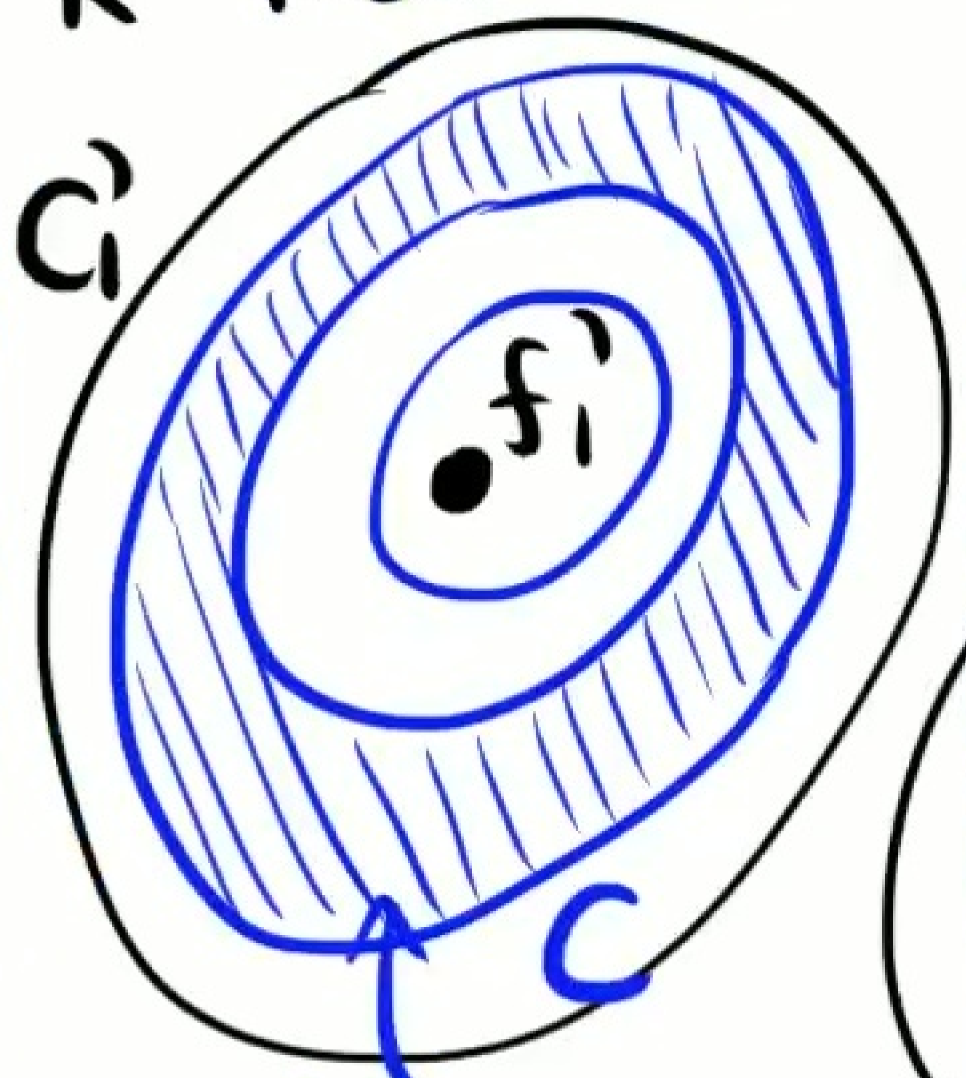
  Union bound over all F

- Suppose can cluster C into ℓ clusters



Claim:
$\{d(v,F) : v$ ...
contained in ...
interval size $O(R_i)$

Let F' be 3-approx k-median soln



$v\in C_i$:
$2^{i-1} \le d(v,f_i') \le 2^i$

$\to R \le 2\cdot 2^i$
error $\le |C|\cdot\varepsilon\cdot 2^{i+1}$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Total error:

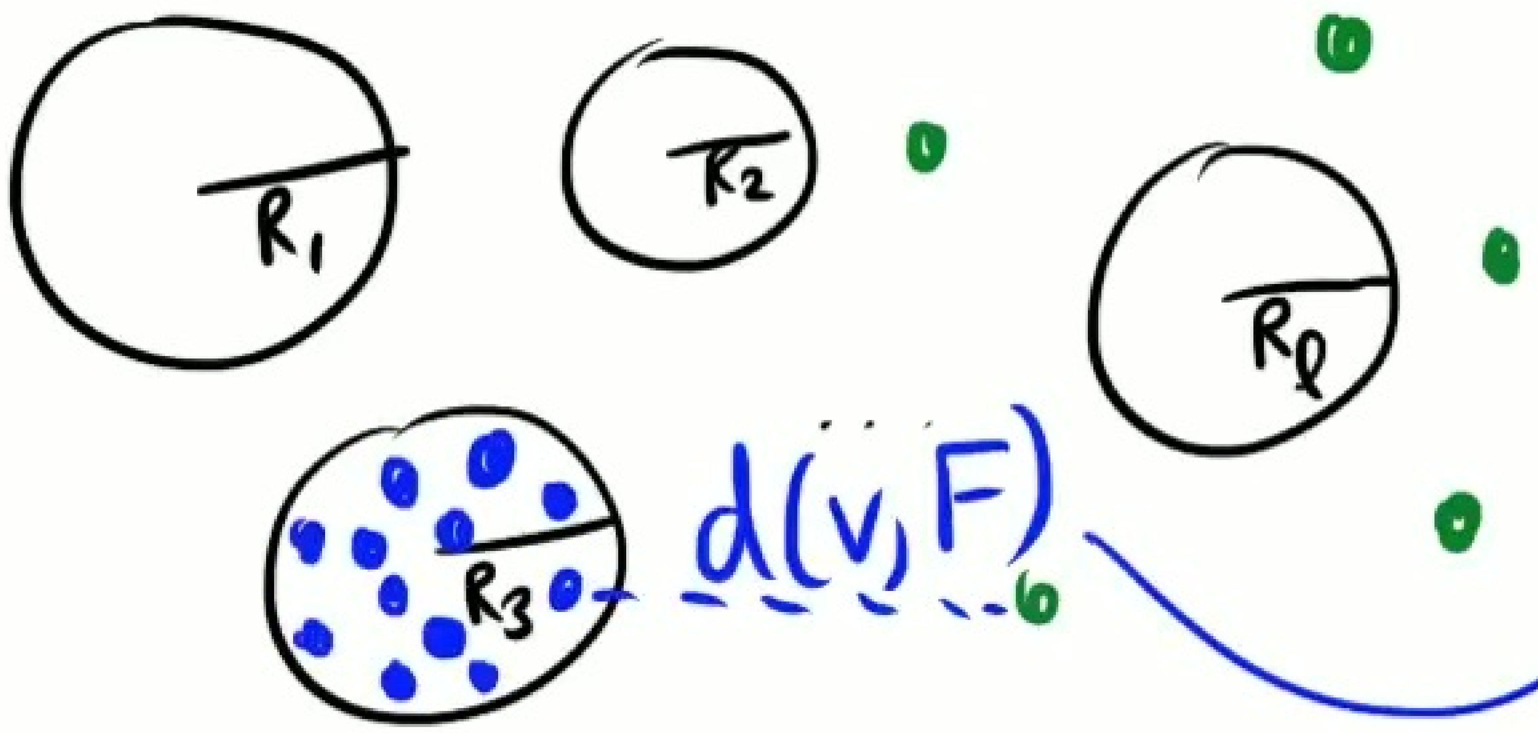
$\sum_{C_i} |C_i|\cdot\varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$

Chen's Coreset Algorithm

- Random sampling
- For each F ($|F|=k$),
  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

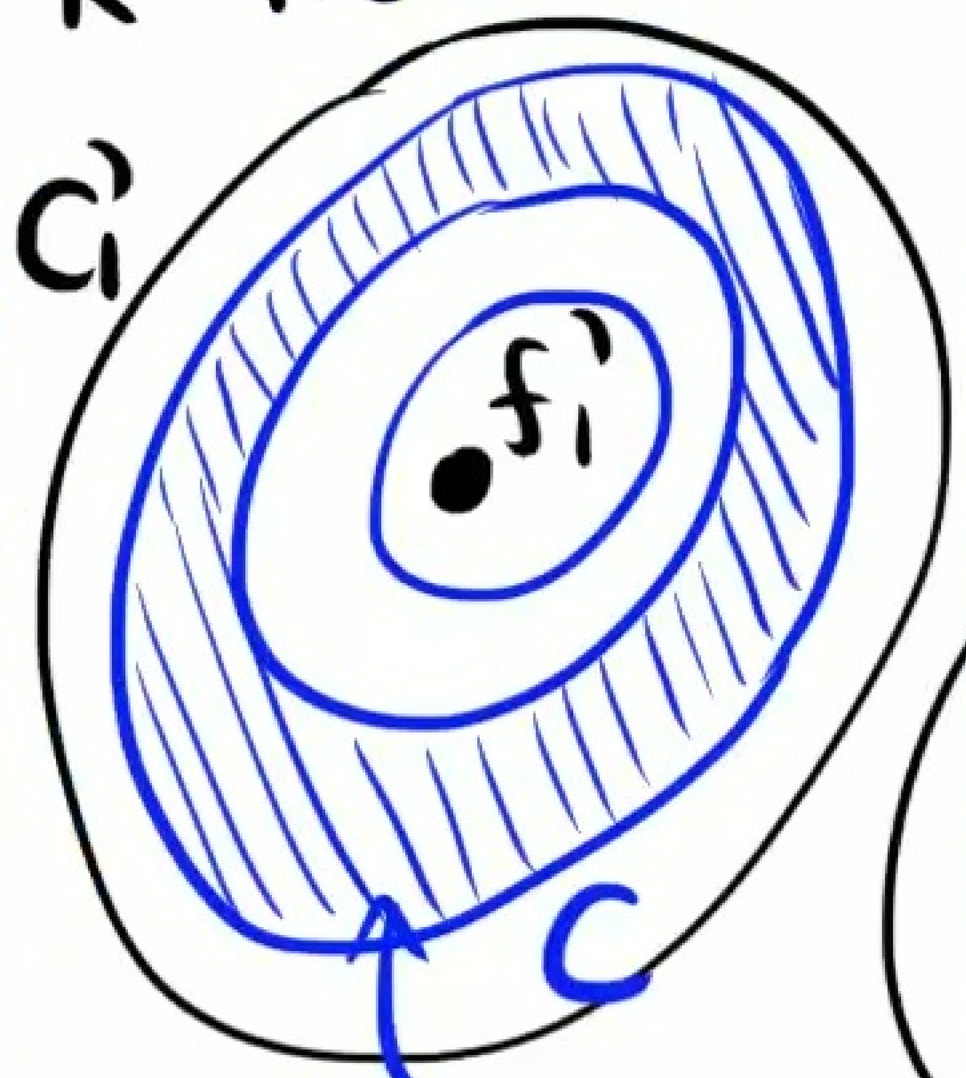
  Union bound over all F

- Suppose can cluster C into ℓ clusters



Claim:
$\{d(v,F) : v$
contained in
interval size $O(R_i)$

Let F' be 3-approx k-median soln



$v \in C_i$:
$2^{i-1} \le d(v,f_i') \le 2^i$

- $R \le 2\cdot 2^i$
- error $\le |C|\cdot\varepsilon\cdot 2^{i+1}$
- Contribution to soln $\ge |C|\cdot 2^{i-1}$

Total error:
$$\sum_{C_i} |C_i|\cdot\varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

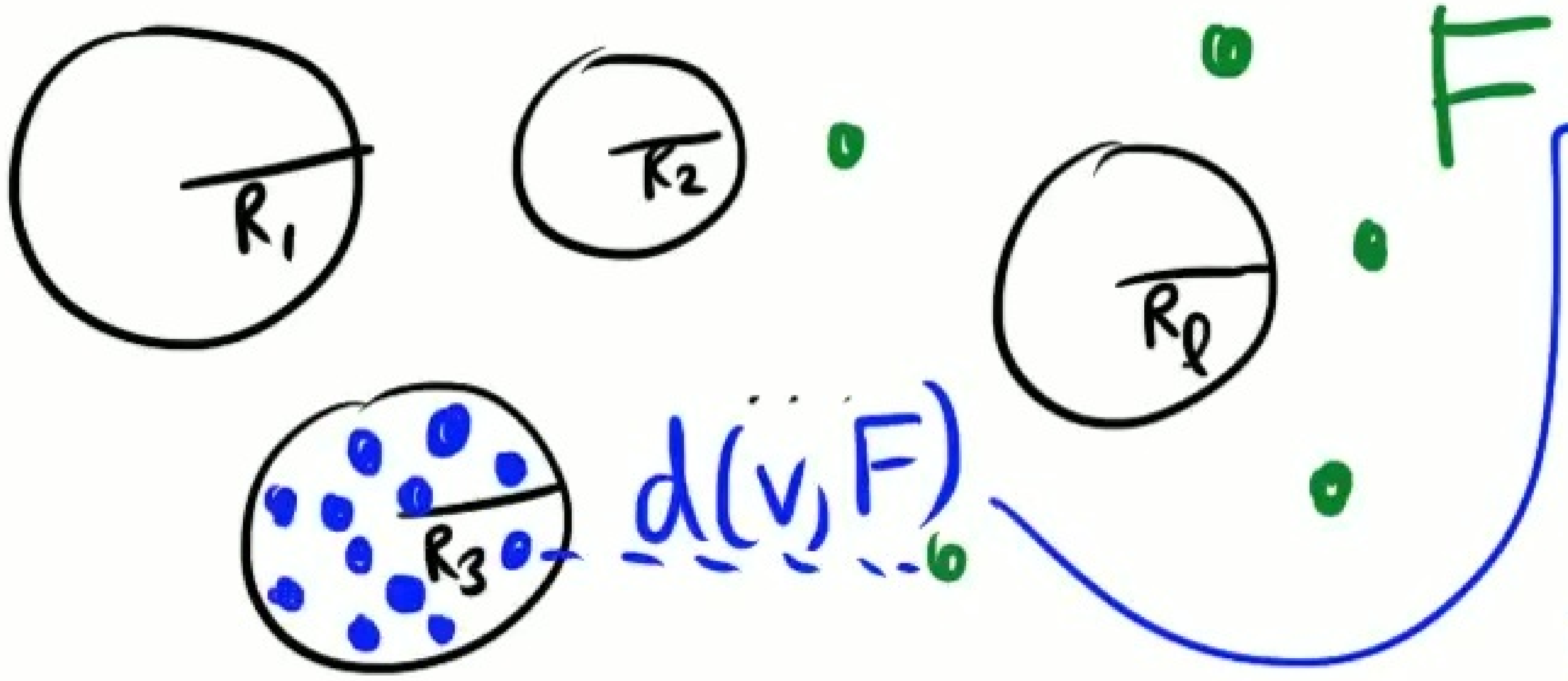
Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,

  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

  Union bound over all F

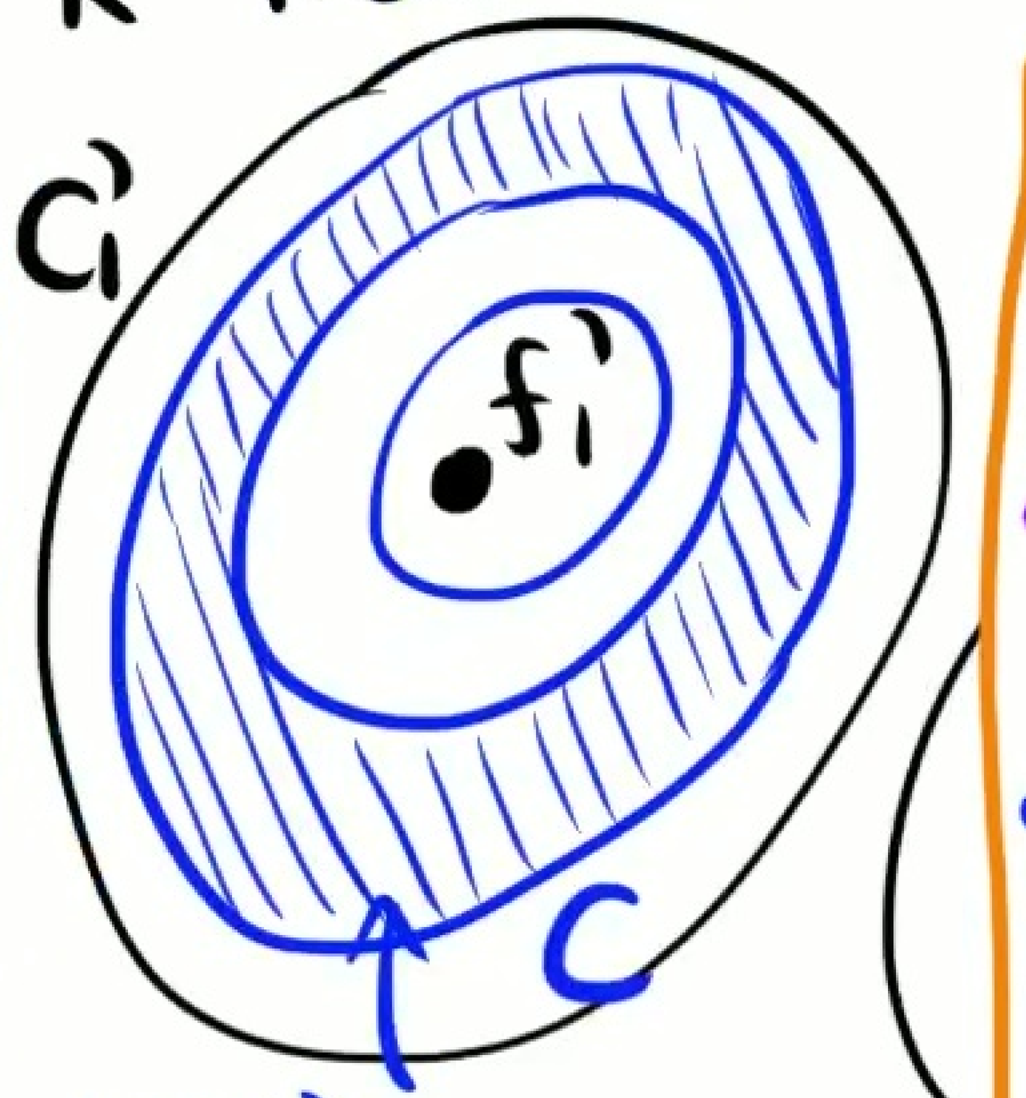
- Suppose can <u>cluster</u> C into ℓ clusters



→ Claim:
$\{d(v,F) : v$
contained in
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \text{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Let F' be 3-approx k-median soln



$v \in C_i$: $2^{i-1} \le d(v, f_i') \le 2^i$

→ $R \le 2\cdot 2^i$
error $\le |C|\cdot\varepsilon\cdot 2^{i+1}$

- Contribution to soln $\ge |C|\cdot 2^{i-1}$
- Total error $\le \sum_C |C|\,\varepsilon\, 2^{i+1} = 4\varepsilon \sum_C |C|\, 2^{i-1}$

Total error:
$$\sum_{C_i} |C_i|\cdot\varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{=O(OPT)?}$$

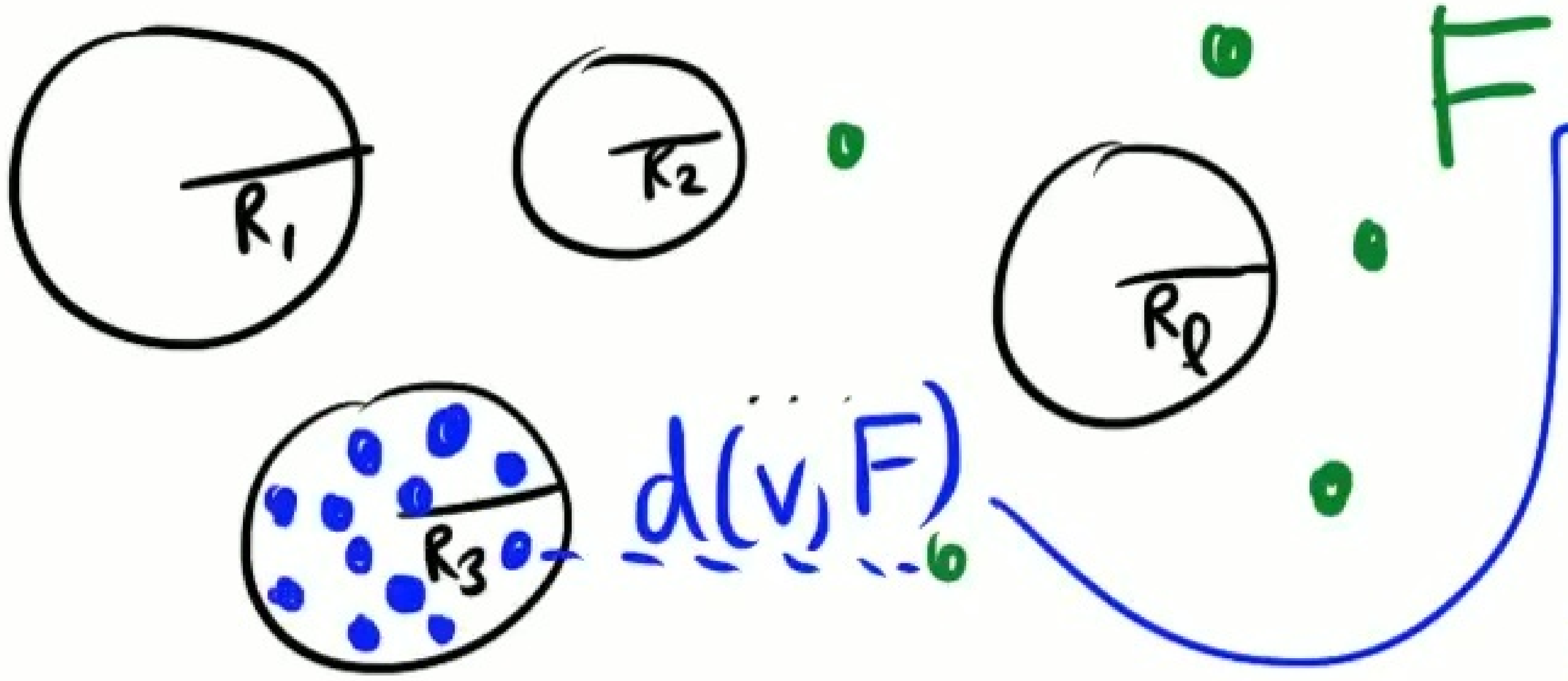
Chen's Coreset Algorithm

- Random sampling
- For each F $(|F|=k)$,

  $\Pr\left[\sum_{v\in S} w(v)\, d(v,F) \in (1\pm\varepsilon)\sum_{v\in C} d(v,F)\right] \gg 1-n^{-k}$

  Union bound over all F

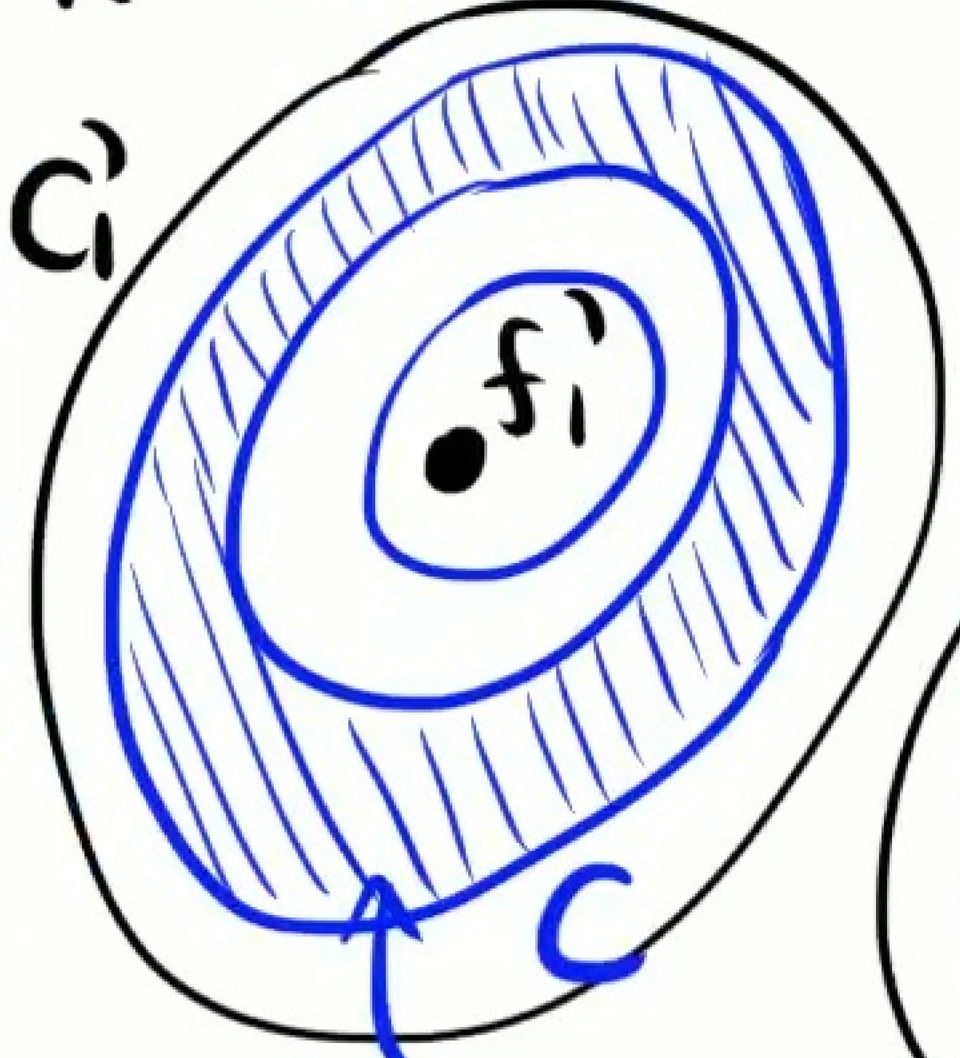
- Suppose can cluster C into ℓ clusters



Claim:
$\{d(v,F) : v$
contained in
interval size $O(R_i)$

- Within each cluster C_i:
  - sample $s := \mathrm{poly}(k \log n\, \varepsilon^{-1})$ vertices
  - each has weight $|C_i|/s$

Let F' be 3-approx k-median soln



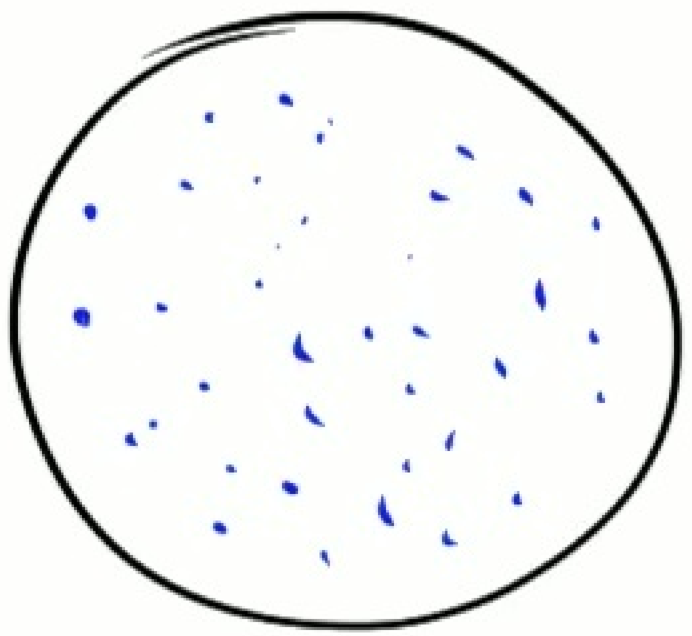
$v \in C_i'$:
$2^{i-1} \le d(v, f_i') \le 2^i$

- $R \le 2\cdot 2^i$
  error $\le |C|\cdot\varepsilon\cdot 2^{i+1}$
- Contribution to soln $\ge |C|\cdot 2^{i-1}$
- Total error
  $\le \sum_C |C|\,\varepsilon\, 2^{i+1}$
  $= 4\varepsilon \sum_C |C|\, 2^{i-1}$
  $\le 4\varepsilon(\text{soln}) \le 12\varepsilon\,\text{OPT}$

Total error:
$\sum_{C_i} |C_i|\cdot\varepsilon R_i = \varepsilon \underbrace{\sum_{C_i} |C_i|\cdot R_i}_{= O(\text{OPT})?}$

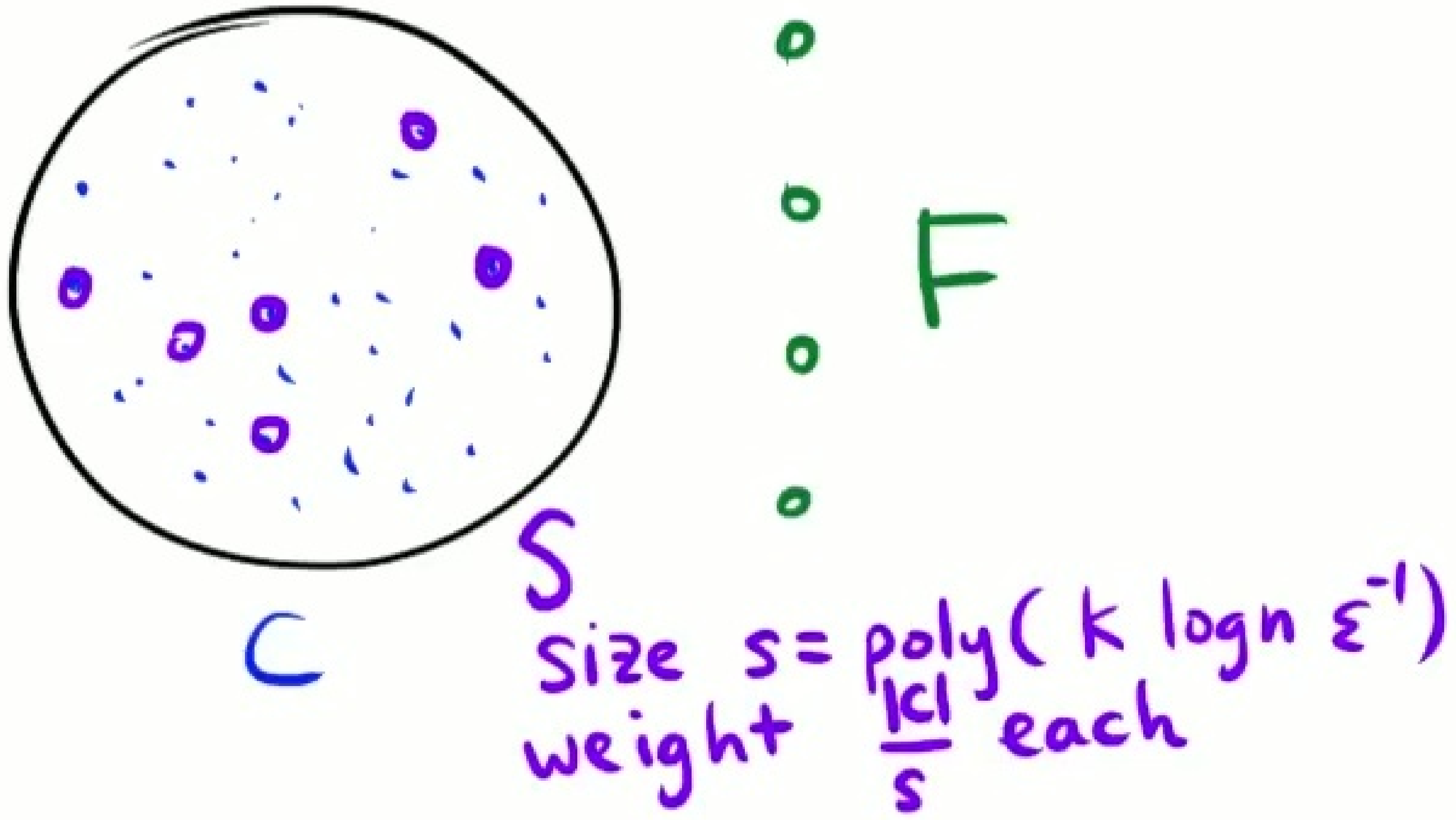
Extending to Capacitated

- This talk: single cluster
-

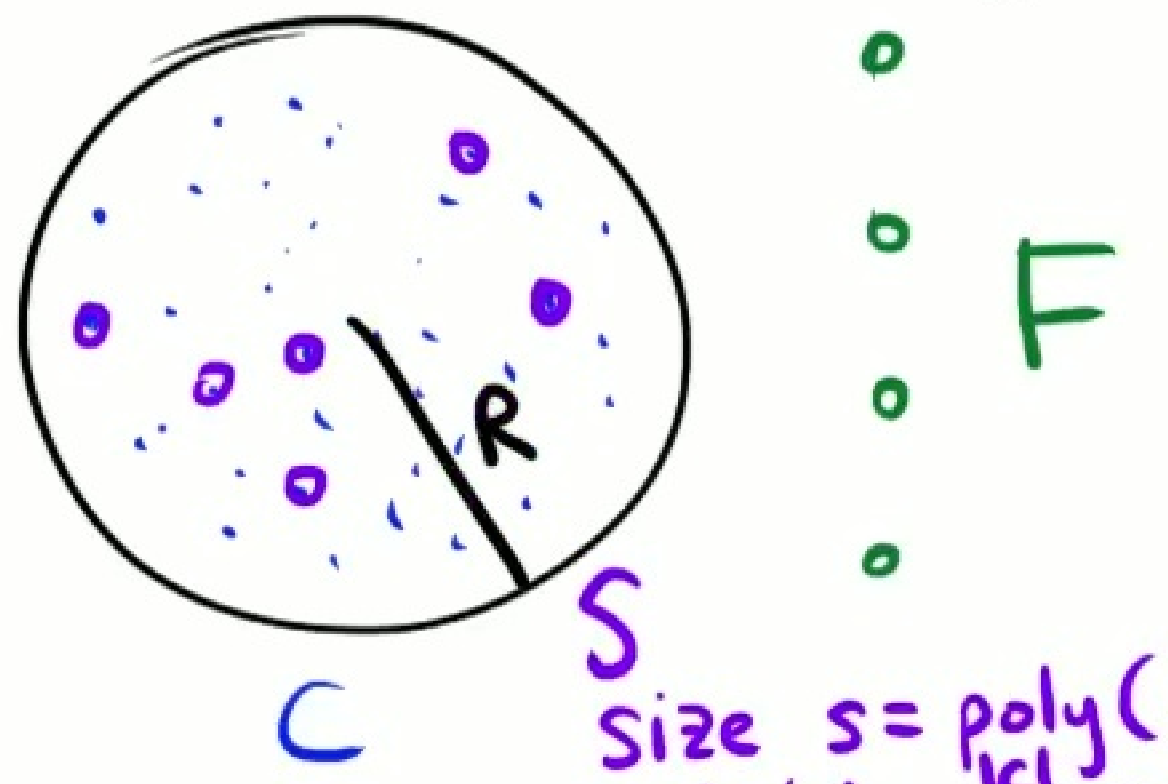


C

Extending to Capacitated

- This talk: single cluster
- 

Extending to Capacitated

- This talk: single cluster
- 

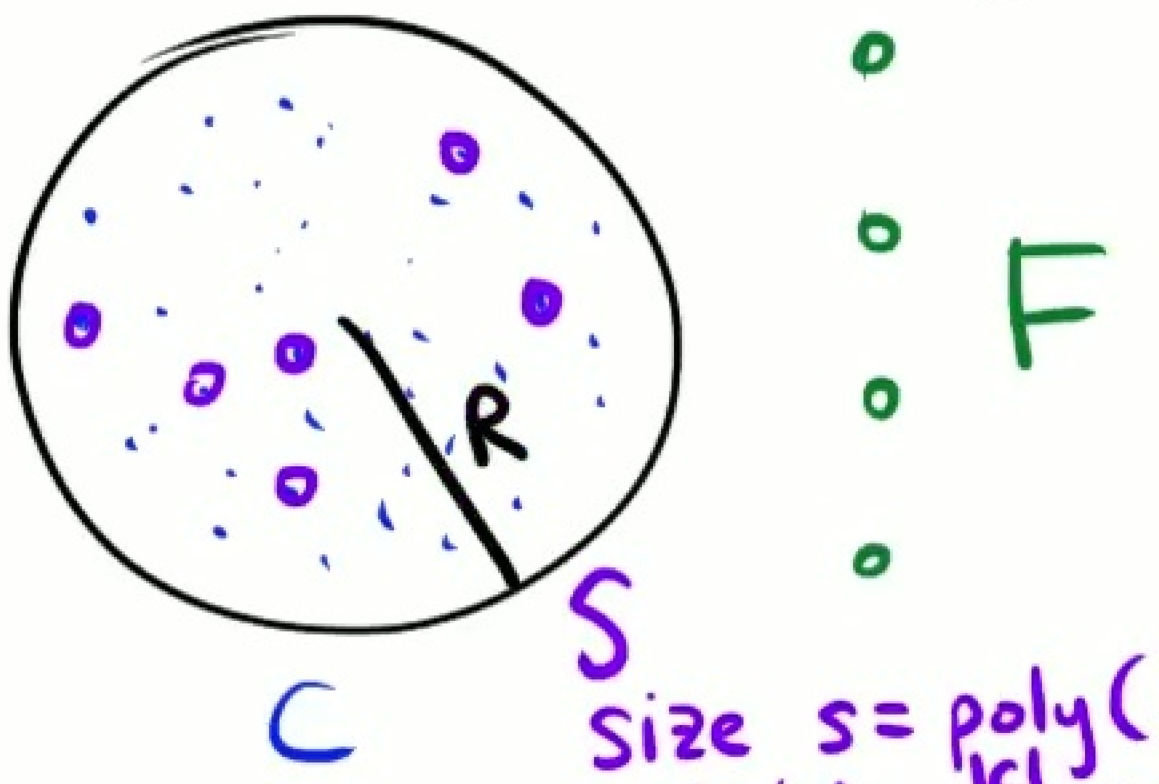
Size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C,F) \in \text{MinCostFlow}(S,F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

Extending to Capacitated

- This talk: single cluster
- 

Size $s = \mathrm{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

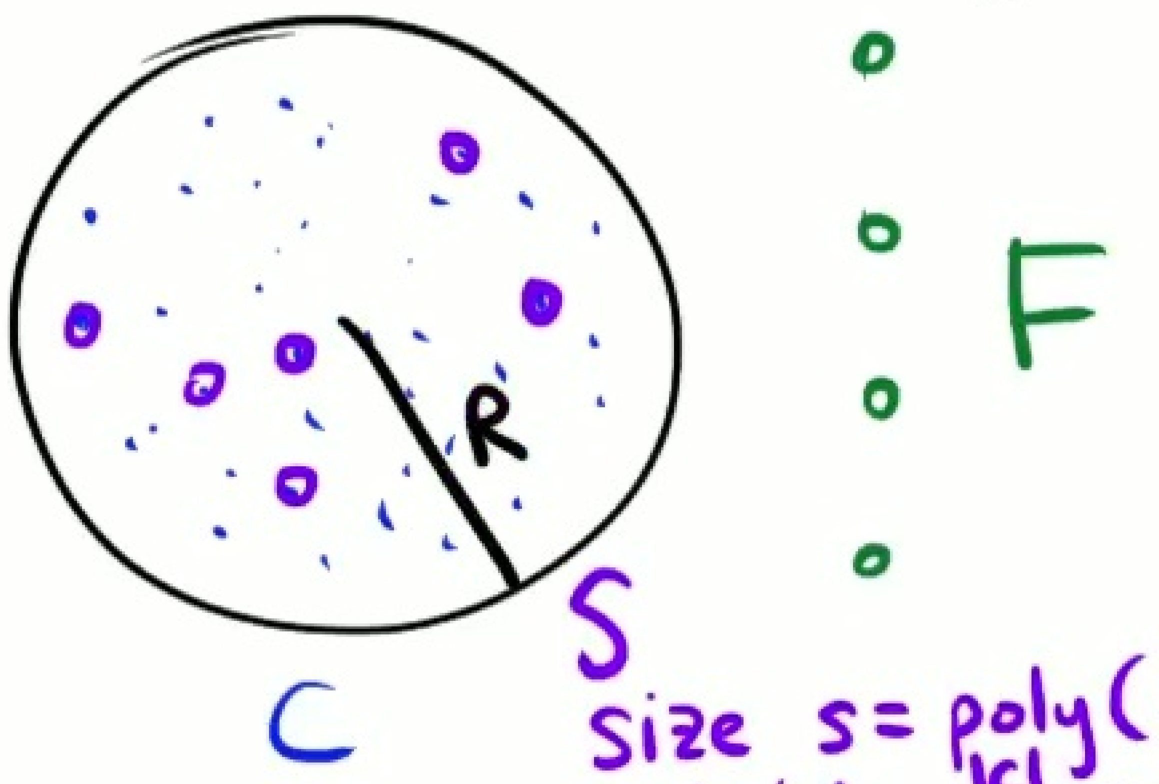
To show:

$\mathrm{MinCostFlow}(C, F) \in \mathrm{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.

Extending to Capacitated

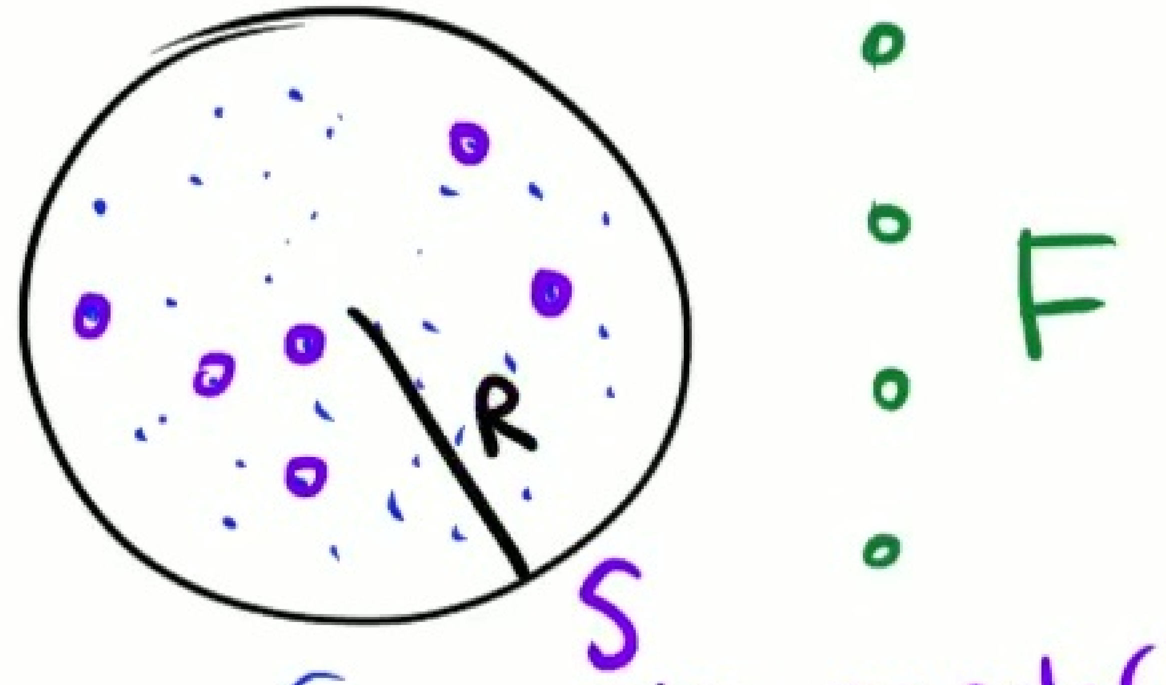
- This talk: single cluster
- 

Size $s = \mathrm{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:
$\mathrm{MinCostFlow}(C, F) \in \mathrm{MinCostFlow}(S, F) \pm \varepsilon |C| R$
w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Extending to Capacitated

- This talk: single cluster
- 

C

Size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

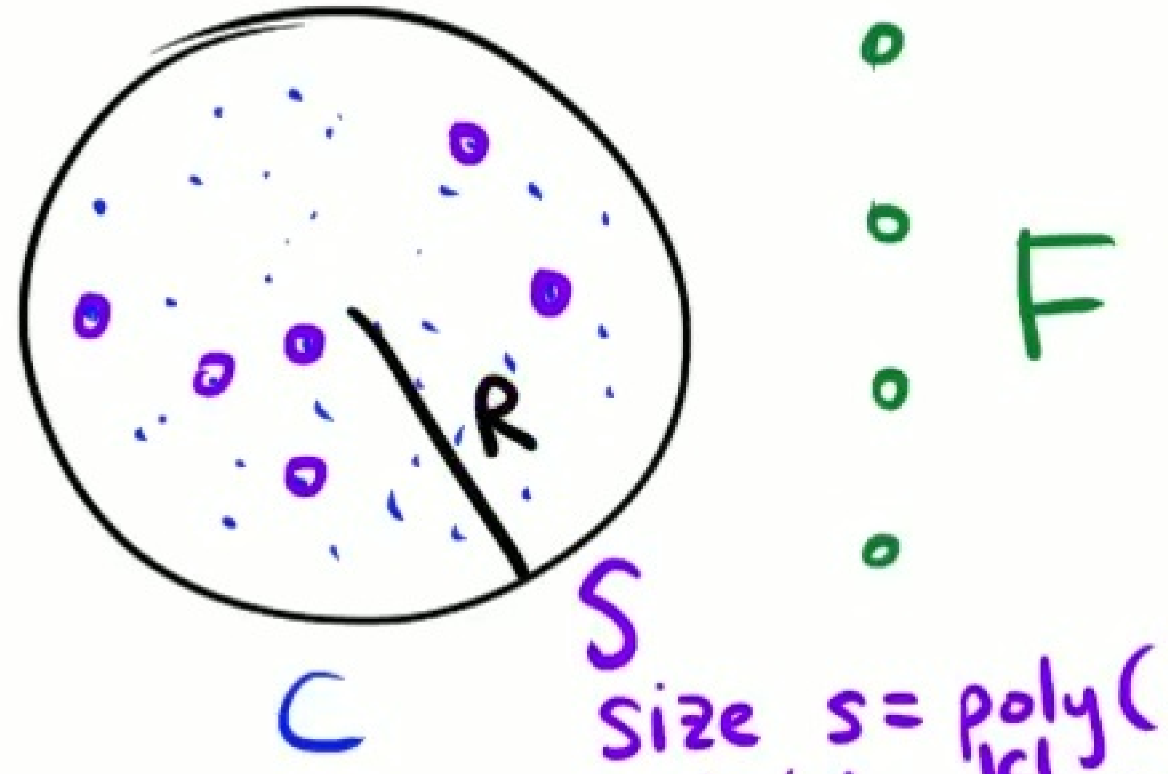
$\text{MinCostFlow}(C, F) \in \text{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C, F) \in \text{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

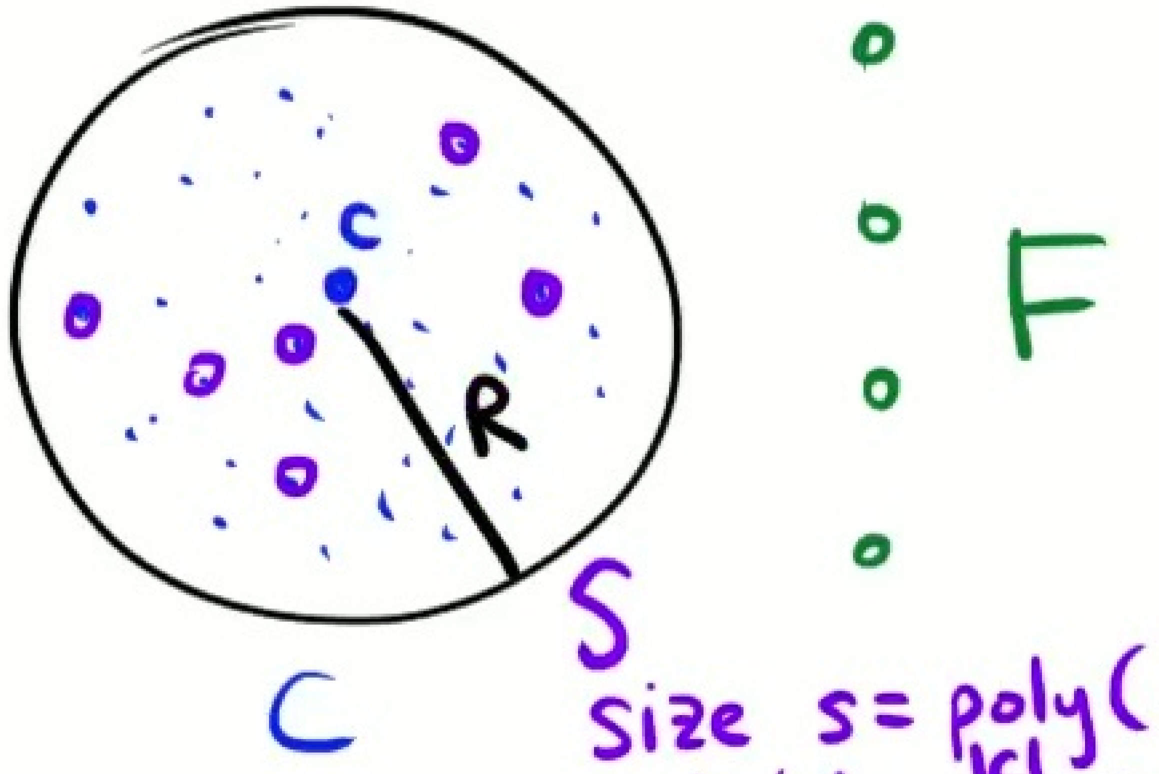
- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

Extending to Capacitated

- This talk: single cluster
- 

Size $s = \mathrm{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\mathrm{MinCostFlow}(C, F) \in \mathrm{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

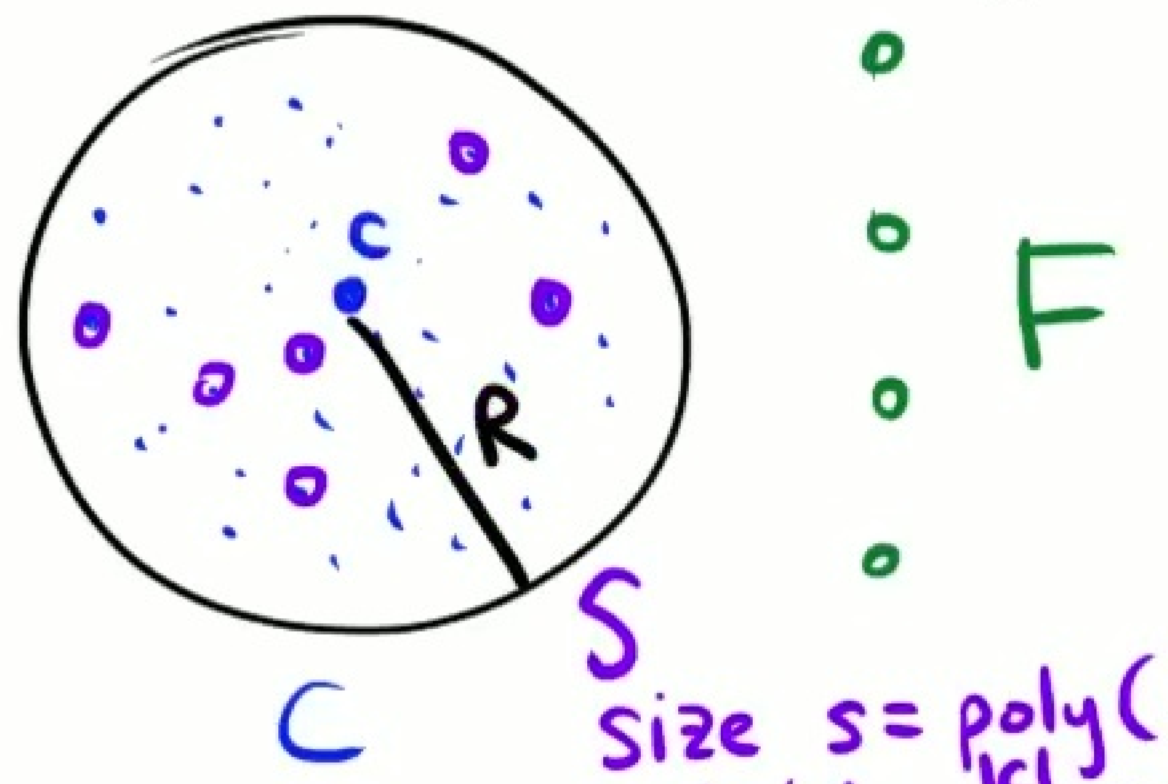
$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) := \mathrm{MinCostFlow}(\text{demands}, F)$

$-\mathrm{cap}(f)$ demand at each $f \in F$

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C, F) \in \text{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

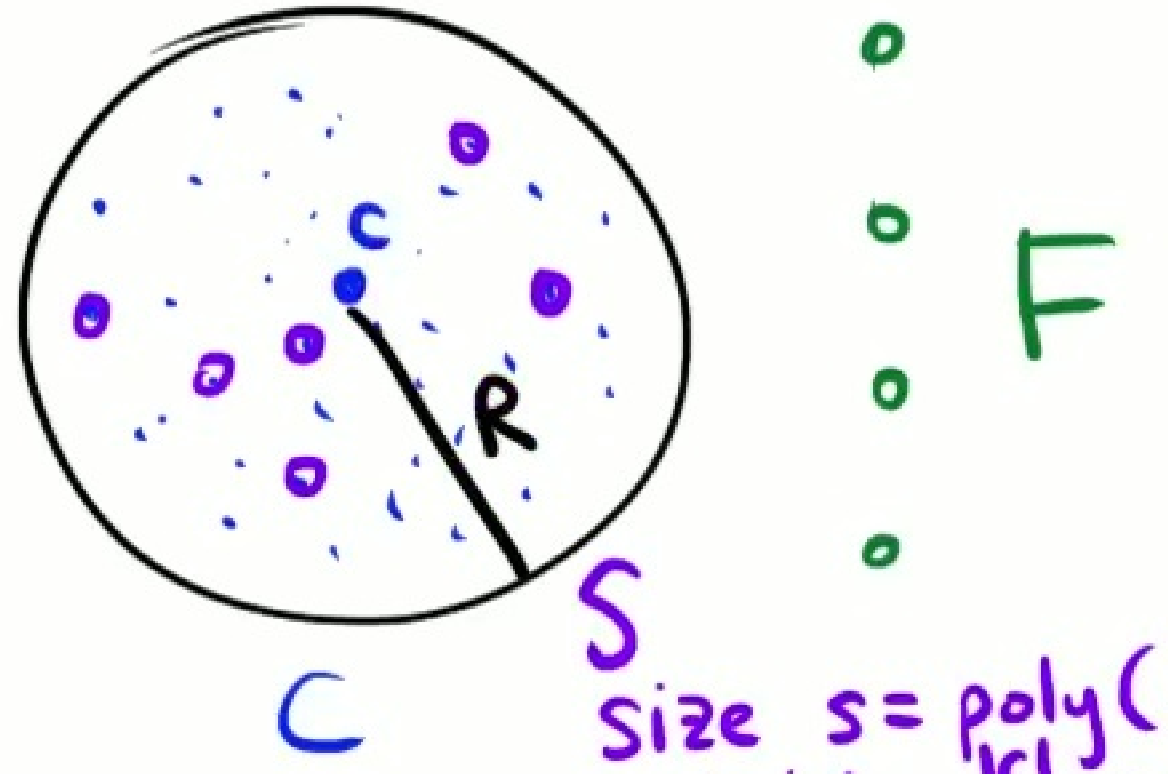
$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

Claim: $g(d)$ is R-Lipschitz:

$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C, F) \in \text{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) := \text{MinCostFlow}(\text{demands}, \underset{\uparrow}{F})$

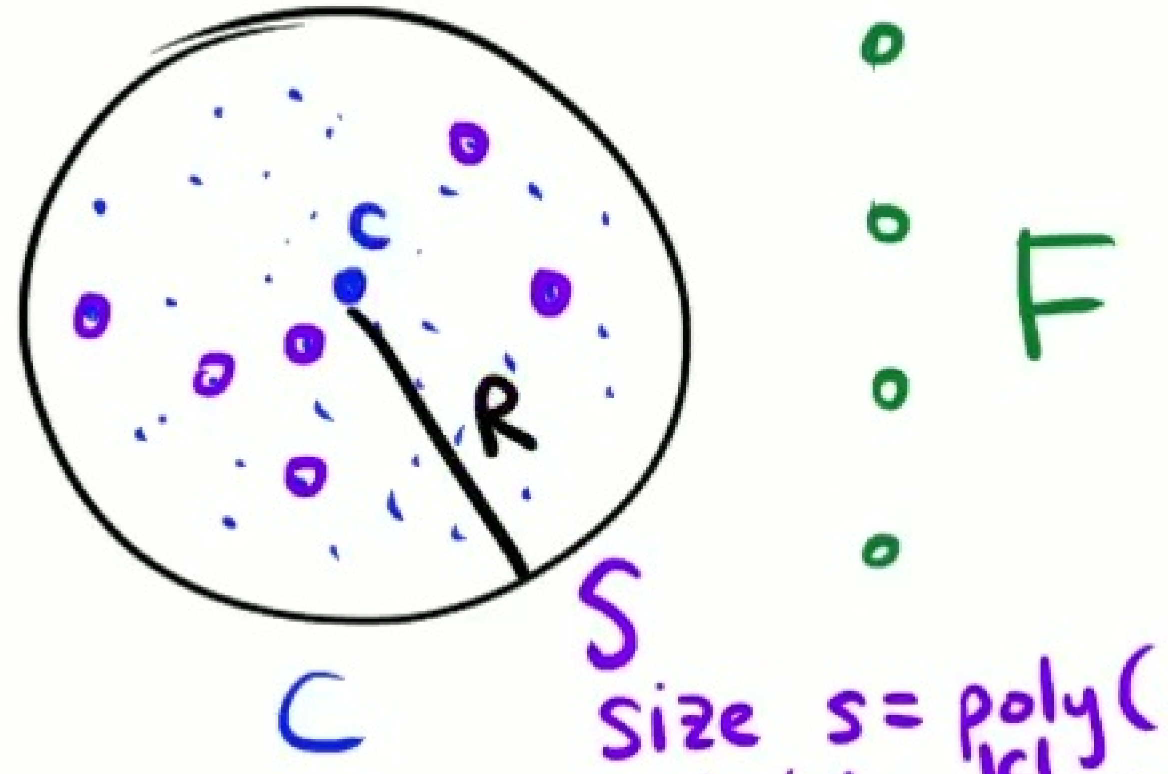
$-\text{cap}(f)$ demand at each $f \in F$

Claim: $g(d)$ is R-Lipschitz:

$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

MinCostFlow(C, F) $\in$ MinCostFlow(S, F) $\pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\text{cap}(f)$ demand at each $f \in F$

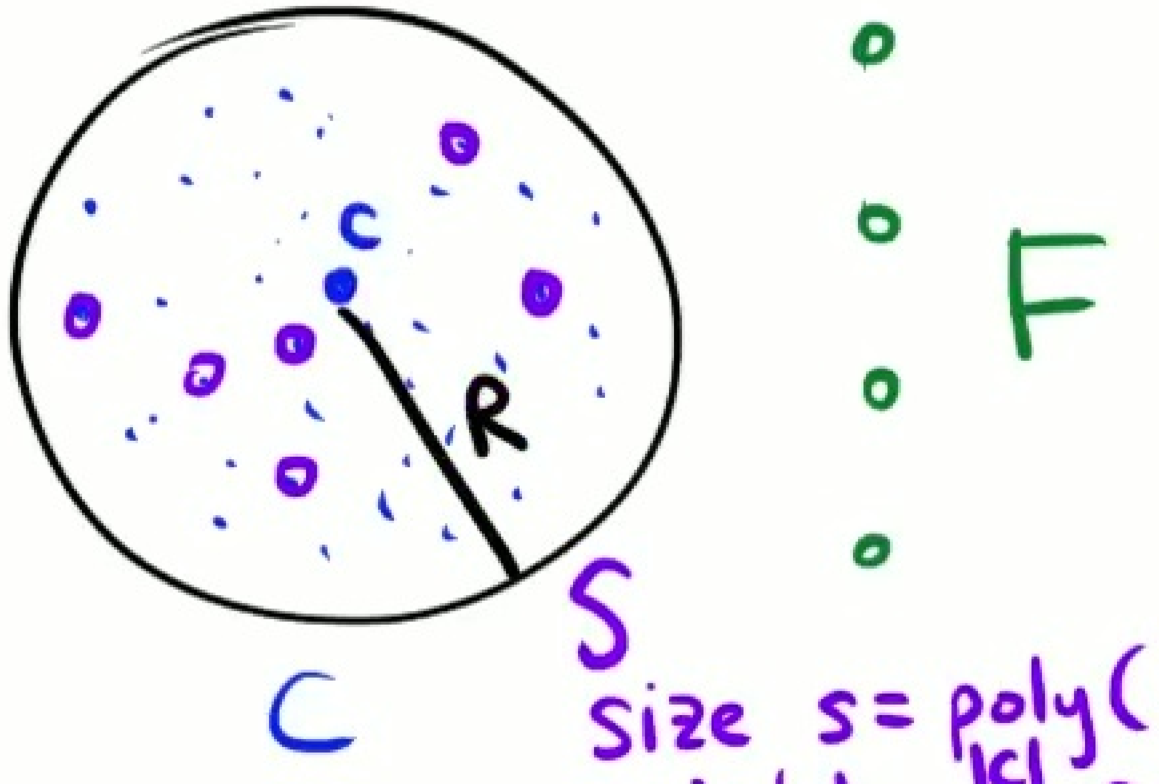
Claim: $g(d)$ is R-Lipschitz:

$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$

- Consider opt MinCostFlow for d

Extending to Capacitated

- This talk: single cluster
- 

Size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C, F) \in \text{MinCostFlow}(S, F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

$\Big\}$ total demand $|C|$

$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

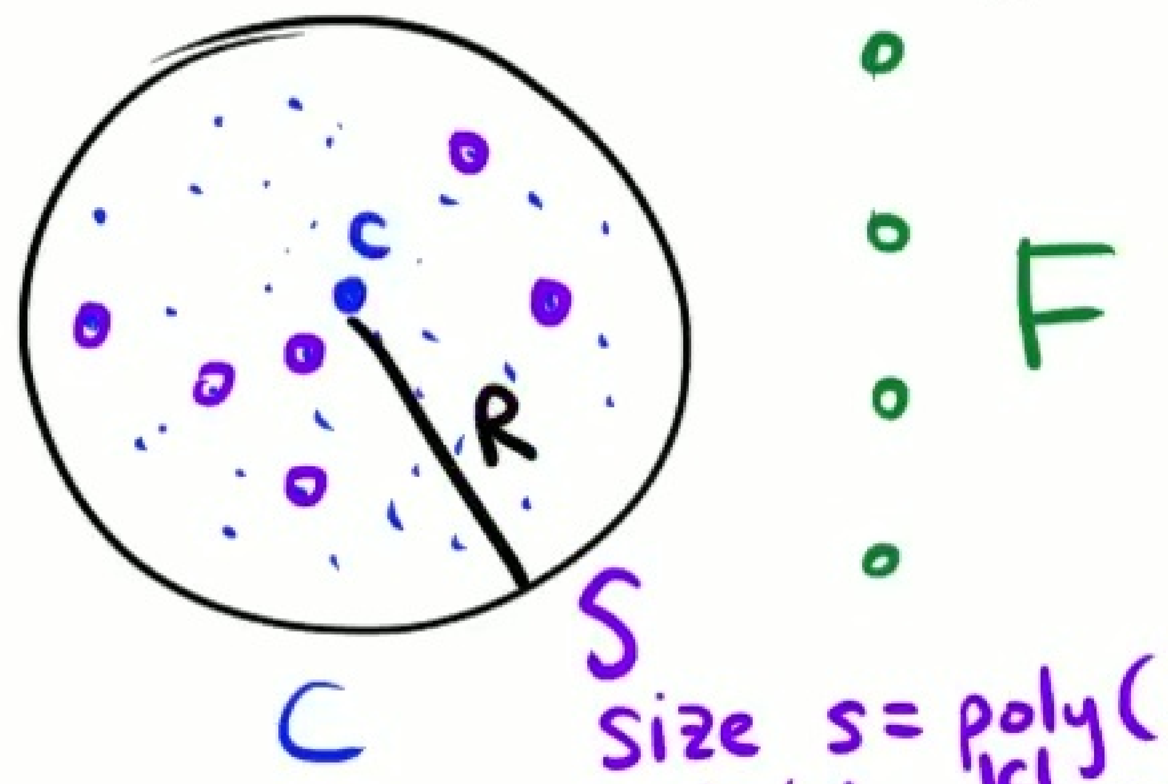
Claim: $g(d)$ is R-Lipschitz:

$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$

- Consider opt MinCostFlow for d
- Only difference between demands d and $d + \delta \mathbb{1}_v$: δ more demand at v, δ less demand at c

Extending to Capacitated

- This talk: single cluster
- 

S
size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C,F) \in \text{MinCostFlow}(S,F) \pm \varepsilon |C| R$

w.p. $\geq 1 - n^{-k}$.

- Want corresponding Chernoff bound for min-cost flow.
- Construct Lipschitz function for concentration

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

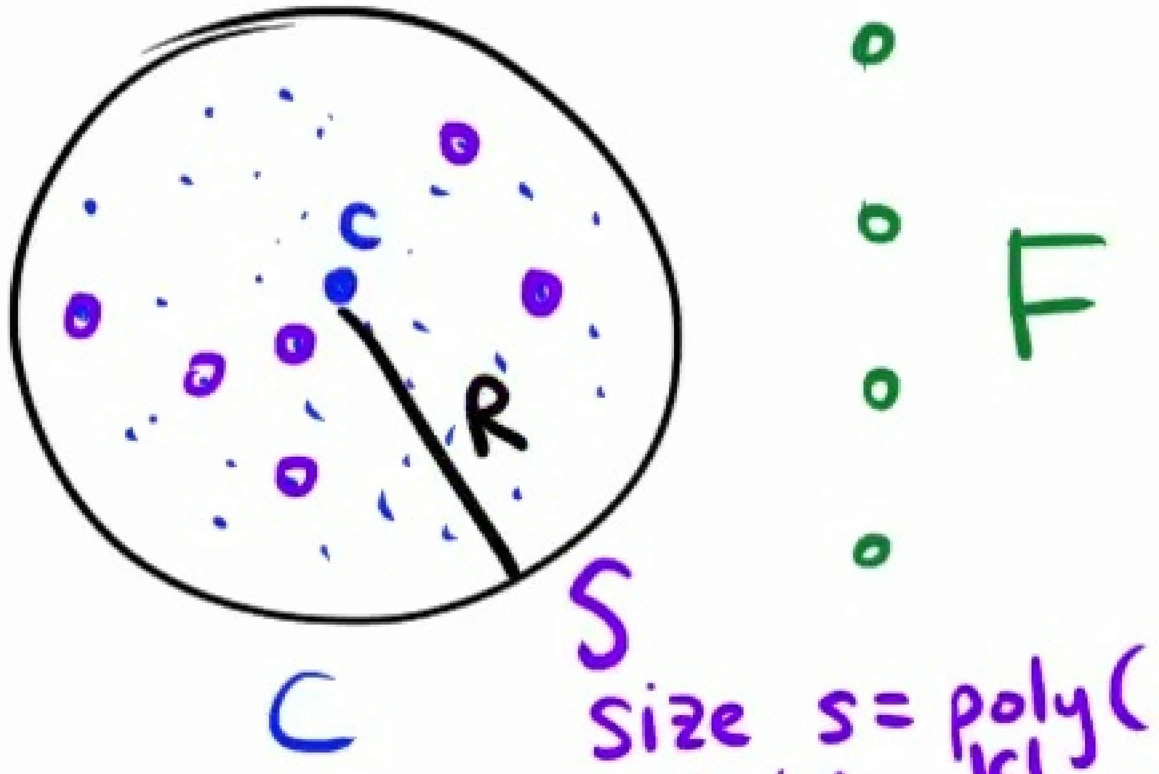
Claim: $g(d)$ is R-Lipschitz:

$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$

- Consider opt MinCostFlow for d
- Only difference between demands d and $d + \delta \mathbb{1}_v$: δ more demand at v, δ less demand at c
- Start from , reroute δ flow $v \leftrightarrow c$, costs $\leq \delta R$

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:
$\text{MinCostFlow}(C,F) \in \text{MinCostFlow}(S,F) \pm \varepsilon |C| R$
w.p. $\geq 1 - n^{-k}$.

Concentration:
$$\Pr\left[\,\left|g(d) - \mathbb{E}[g(d)]\right| > \varepsilon |C| R\right] \ll n^{-k}$$

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:
$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

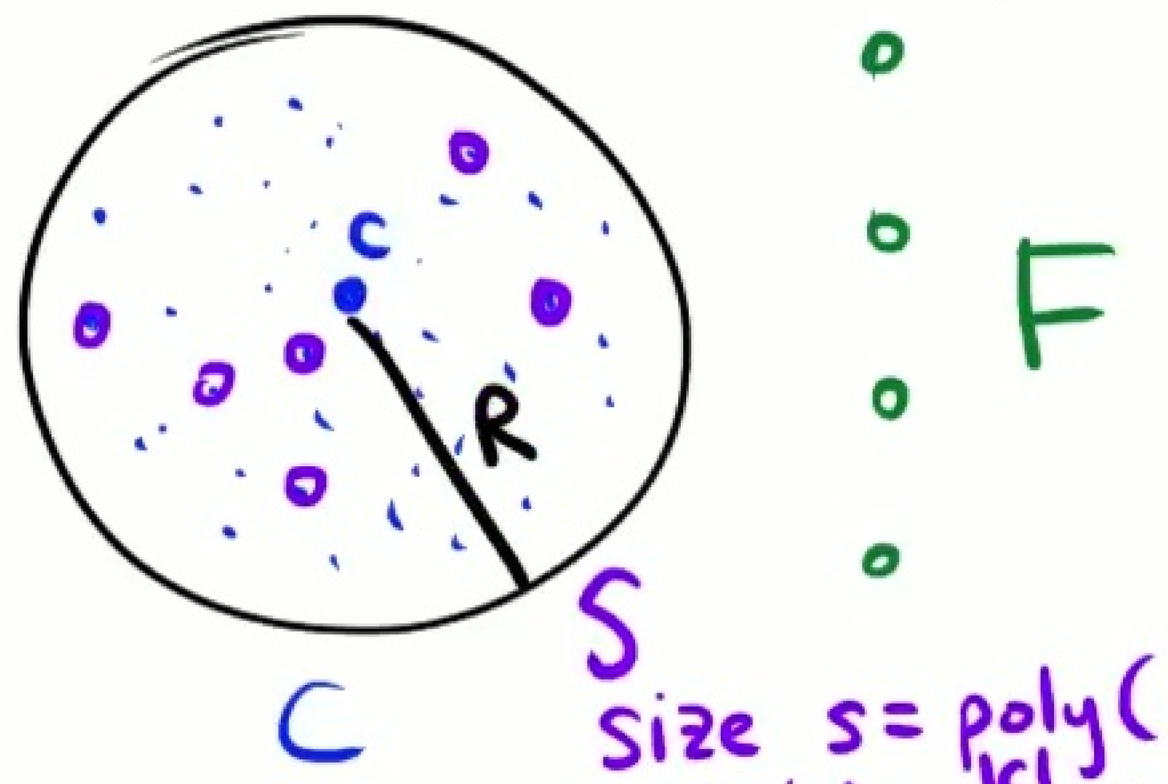
$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

Claim: $g(d)$ is R-Lipschitz:
$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$
- Consider opt MinCostFlow for d
- Only difference between demands d and $d + \delta \mathbb{1}_v$: δ more demand at v, δ less demand at c
- Start from , reroute δ flow $v \leftrightarrow c$, costs $\leq \delta R$

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:
$\text{MinCostFlow}(C, F) \in \text{MinCostFlow}(S, F) \pm \varepsilon |C| R$
w.p. $\geq 1 - n^{-k}$.

$= g(\mathbb{1}) = g(\mathbb{E}[d])$

Concentration:
$\Pr\left[\,|g(d) - \mathbb{E}[g(d)]| > \varepsilon |C| R\right] \ll n^{-k}$

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) := \text{MinCostFlow}(\text{demands}, F)$

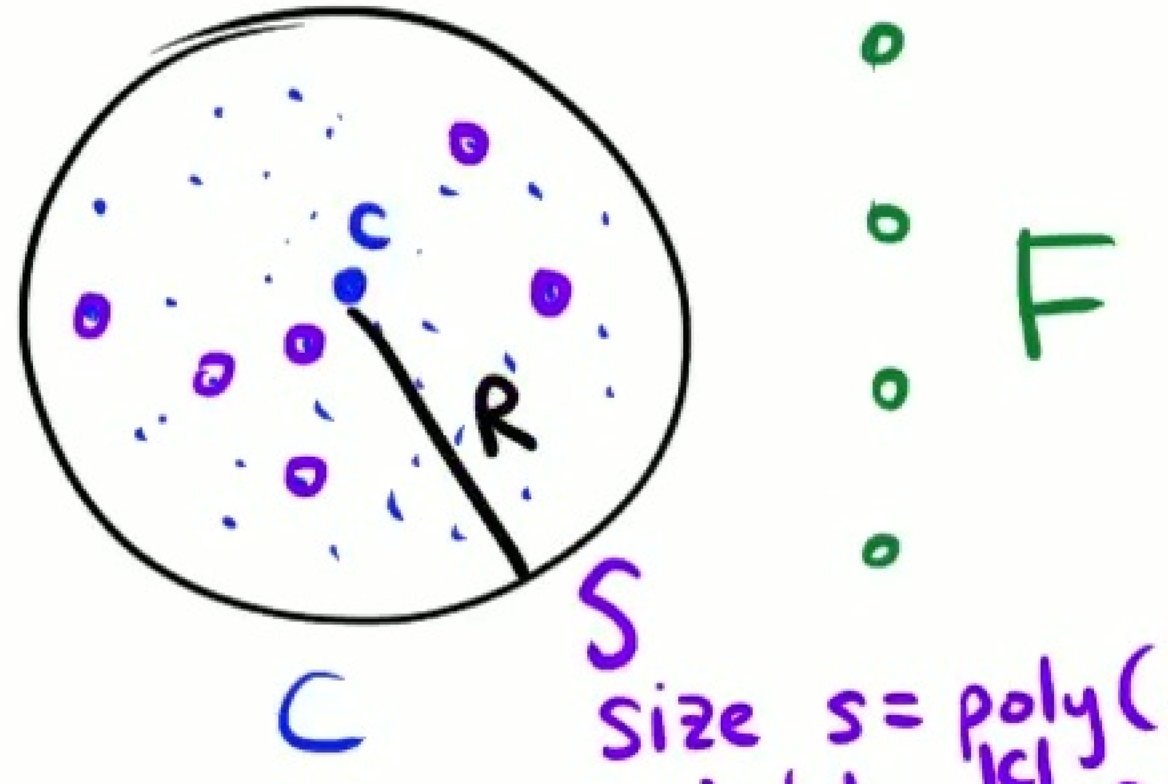
$-\text{cap}(f)$ demand at each $f \in F$

Claim: $g(d)$ is R-Lipschitz:
$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$

- Consider opt MinCostFlow for d
- Only difference between demands d and $d + \delta \mathbb{1}_v$: δ more demand at v, δ less demand at c
- Start from , reroute δ flow $v \leftrightarrow c$, costs $\leq \delta R$

Extending to Capacitated

- This talk: single cluster
- 

size $s = \text{poly}(k \log n\, \varepsilon^{-1})$
weight $\frac{|C|}{s}$ each

To show:

$\text{MinCostFlow}(C,F) \in \text{MinCostFlow}(S,F) \pm \varepsilon |C| R$
w.p. $\geq 1 - n^{-k}$.

$\boxed{= g(\mathbb{1}) = g(\mathbb{E}[d])}$

Concentration:

$\Pr\left[\,|g(d) - \mathbb{E}[g(d)]| > \varepsilon |C| R\,\right] \ll n^{-k}$

To show: $\mathbb{E}[g(d)] \approx g(\mathbb{E}[d])$

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

Claim: $g(d)$ is R-Lipschitz:

$$|g(d_1) - g(d_2)| \leq R \|d_1 - d_2\|_1$$

Pf: $g(d + \delta \mathbb{1}_v) \overset{?}{\leq} g(d) + \delta R$

- Consider opt MinCostFlow for d
- Only difference between demands d and $d + \delta \mathbb{1}_v$: δ more demand at v, δ less demand at c
- Start from , reroute δ flow $v \leftrightarrow c$, costs $\leq \delta R$

$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

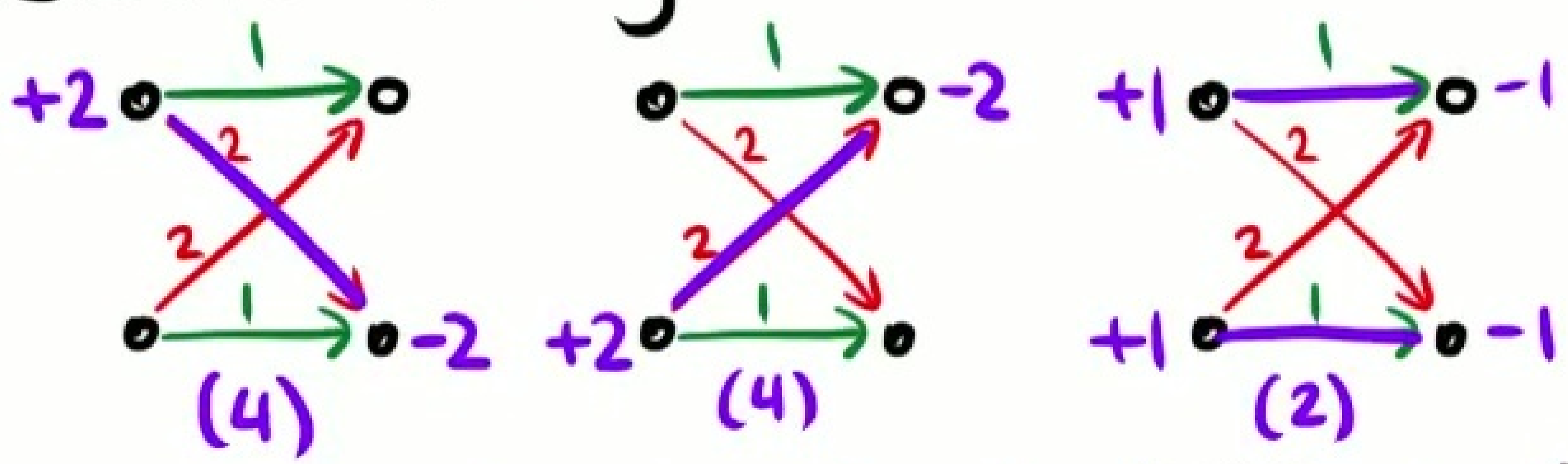
total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\mathrm{cap}(f)$ demand at each $f \in F$

$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

- Convexity of MinCostFlow:



$$\Rightarrow g(\mathbb{E}[d]) \leq \mathbb{E}[g(d)]$$

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

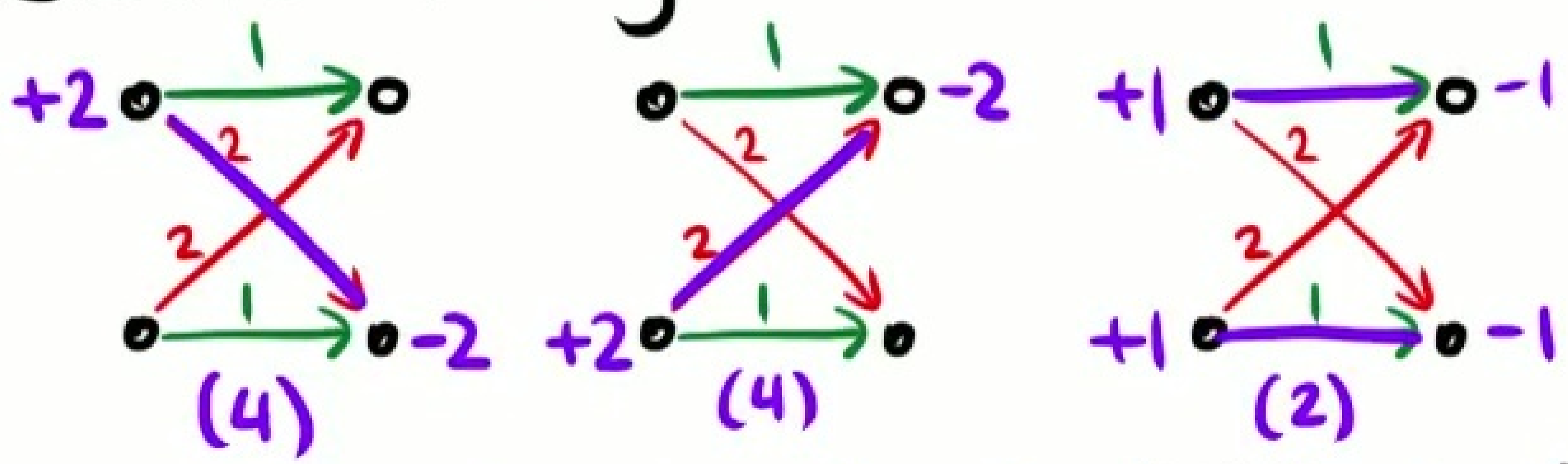
$g(d)$: $\left.\begin{array}{l}\bullet \text{ demand } +d_v \text{ at each } v \in C \\ \bullet \text{ demand } |C| - \sum_v d_v \text{ at center } c\end{array}\right\}$ total demand $|C|$

$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

- Convexity of MinCostFlow:



$$\Rightarrow g(\mathbb{E}[d]) \le \mathbb{E}[g(d)]$$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\le} g(\mathbb{E}[d]) + \varepsilon |C| R$

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

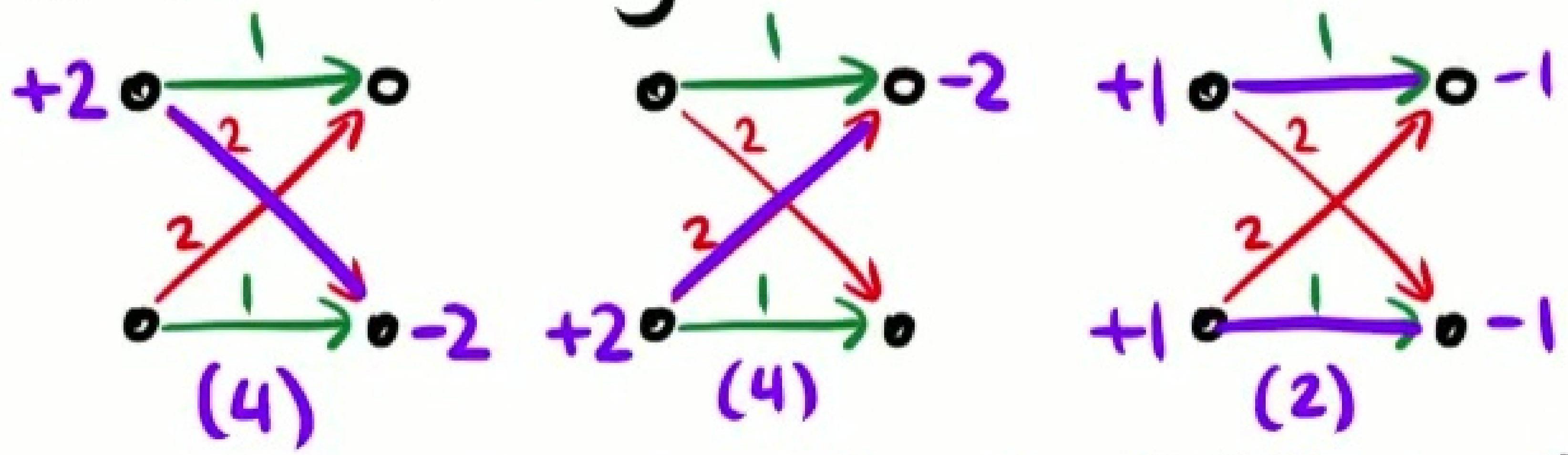
total demand $|C|$

$$g(d) := \text{MinCostFlow}(\text{demands}, F)$$

$-\text{cap}(f)$ demand at each $f \in F$

$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

- Convexity of MinCostFlow:



$$\Rightarrow g(\mathbb{E}[d]) \le \mathbb{E}[g(d)]$$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\le} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

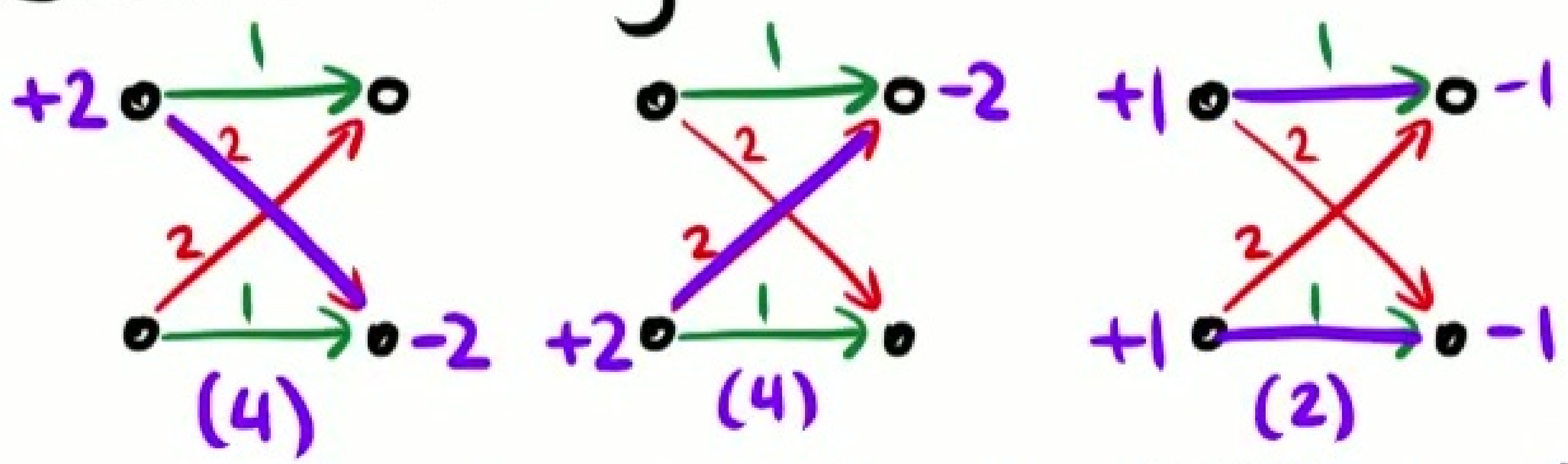
$g(d)$: $\left.\begin{array}{l} \bullet \text{ demand } +d_v \text{ at each } v \in C \\ \bullet \text{ demand } |C| - \sum_v d_v \text{ at center } c \end{array}\right\}$ total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\text{cap}(f)$ demand at each $f \in F$

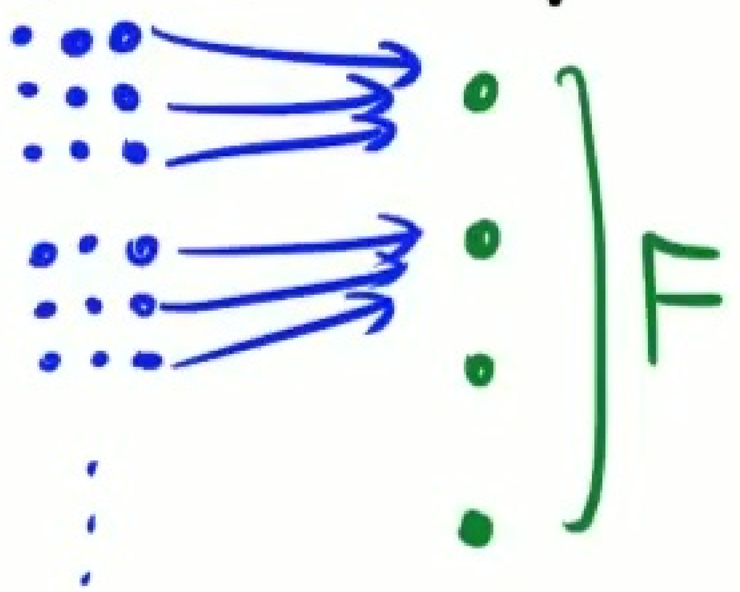
$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

- Convexity of MinCostFlow:



$$\Rightarrow g(\mathbb{E}[d]) \leq \mathbb{E}[g(d)]$$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\leq} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d
- Look at opt flow for $\mathbb{E}[d] = \mathbb{1}$:



Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

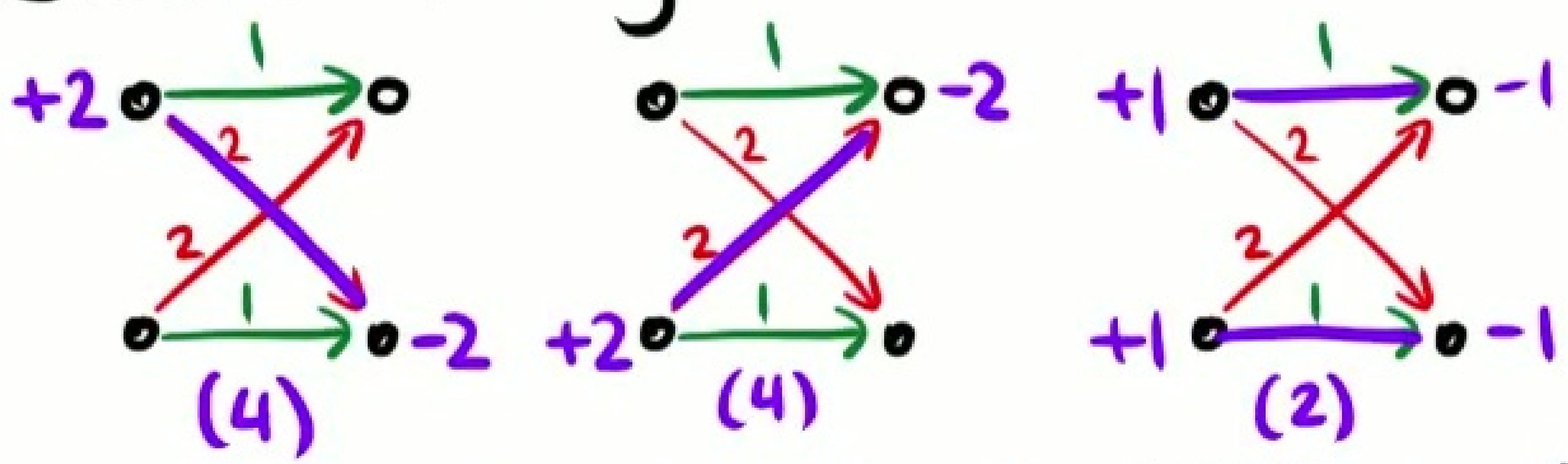
total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\text{cap}(f)$ demand at each $f \in F$

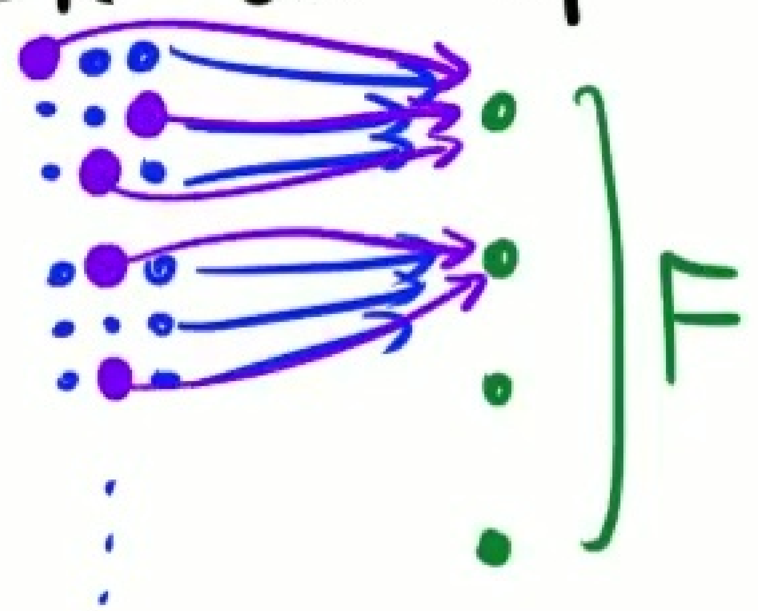
$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$

- Convexity of MinCostFlow:



$$\Rightarrow g(\mathbb{E}[d]) \leq \mathbb{E}[g(d)]$$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\leq} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d
- Look at opt flow for $\mathbb{E}[d] = \mathbb{1}$:



Route sampled clients to same facility in F

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

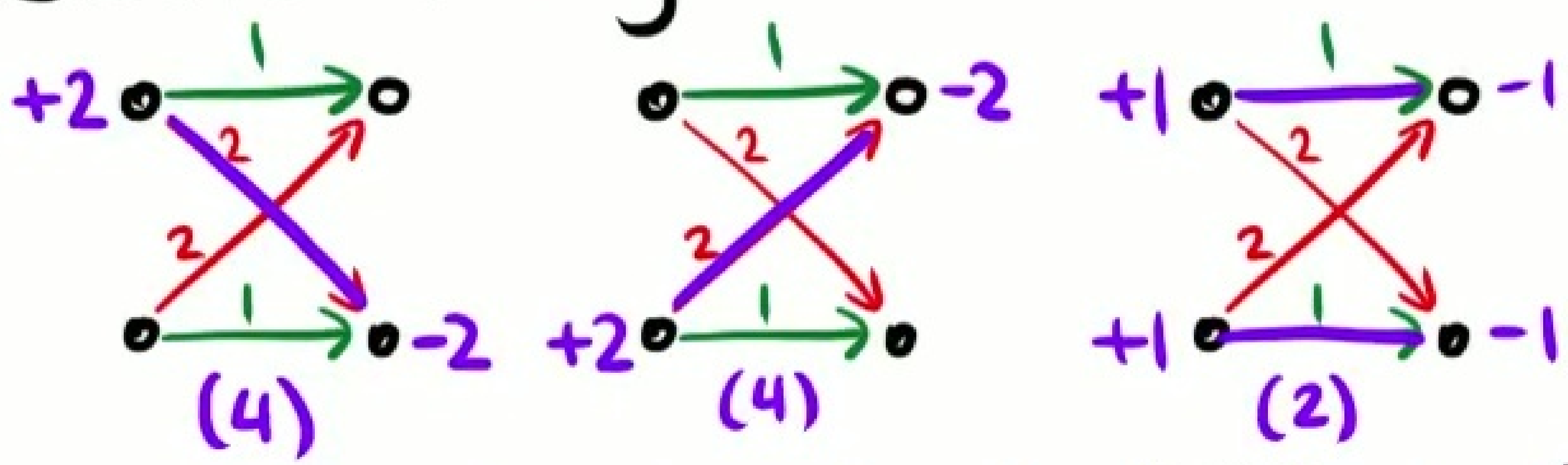
total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\mathrm{cap}(f)$ demand at each $f \in F$

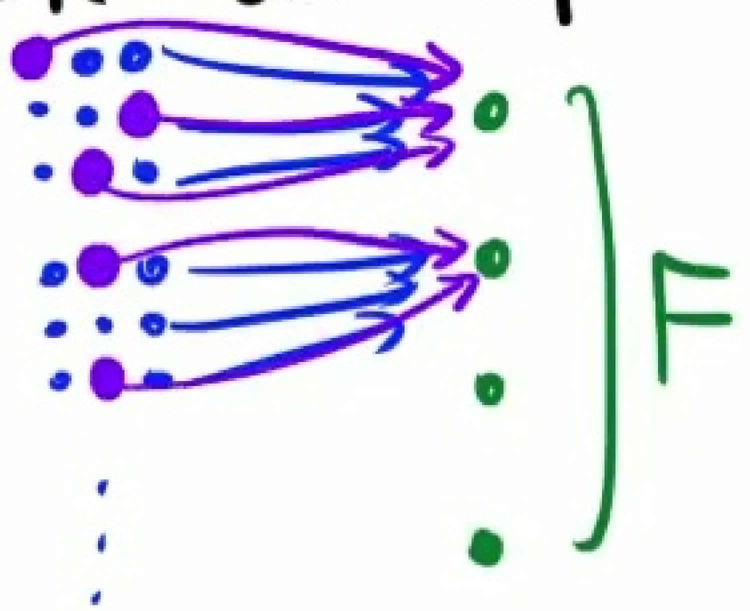
$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

- Convexity of MinCostFlow:



$$\Rightarrow g(\mathbb{E}[d]) \le \mathbb{E}[g(d)]$$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\lesssim} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d
- Look at opt flow for $\mathbb{E}[d] = \mathbb{1}$:



Route sampled clients to same facility in F

$(\pm\varepsilon|C|R)$

for most d (concentration)

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

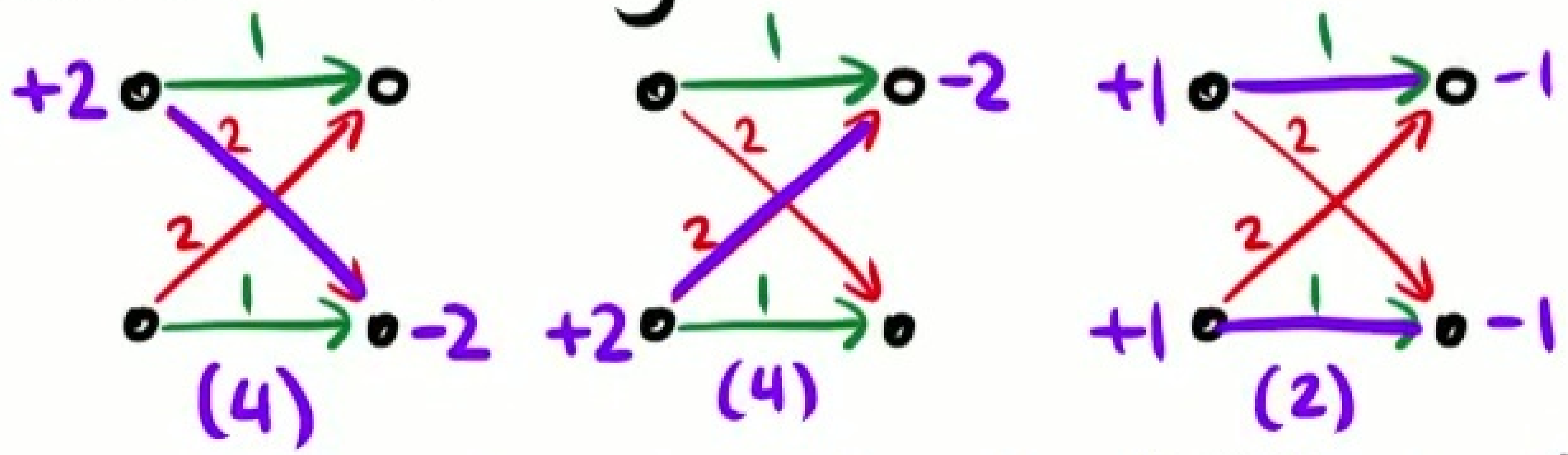
total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\mathrm{cap}(f)$ demand at each $f \in F$

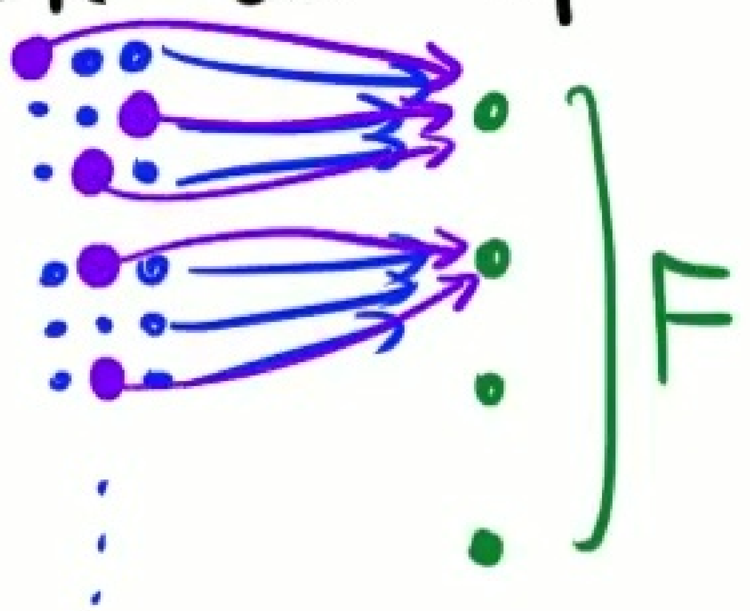
$$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$$

- Convexity of MinCostFlow:

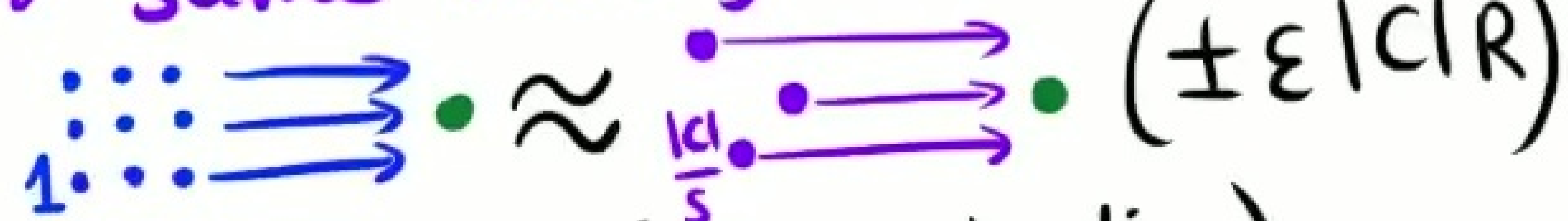


$\Rightarrow g(\mathbb{E}[d]) \le \mathbb{E}[g(d)]$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\le} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d
- Look at opt flow for $\mathbb{E}[d] = \mathbb{1}$:



Route sampled clients to same facility in F



for most d (concentration)

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

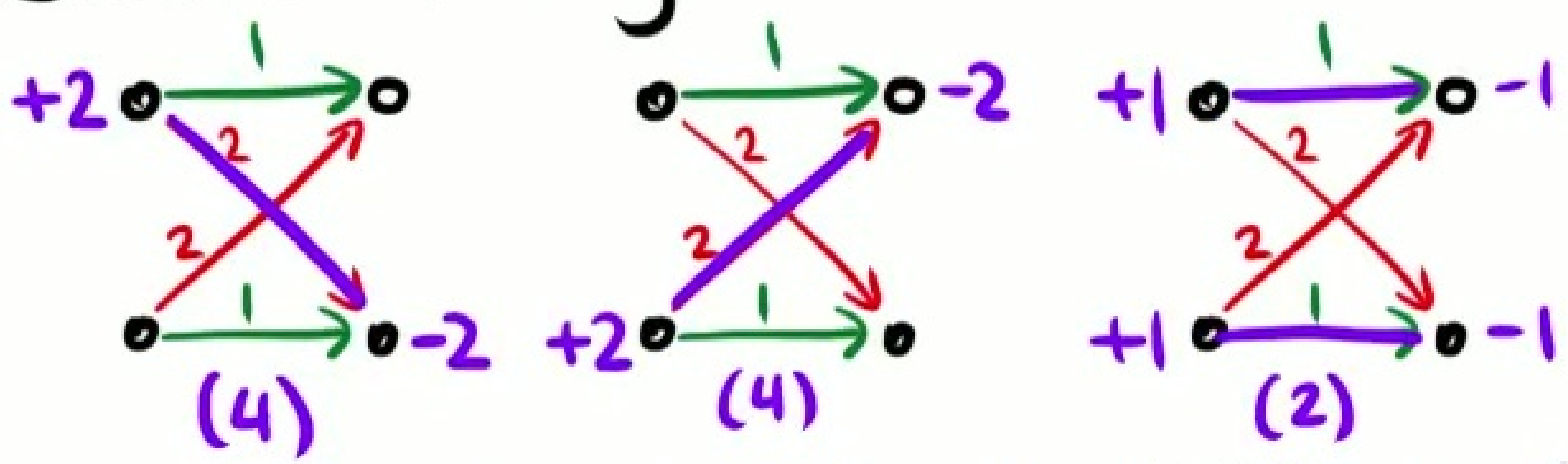
$g(d) :=$ MinCostFlow(demands, F)

$-\text{cap}(f)$ demand at each $f \in F$

- For most d, sampled $\approx$ size(⁘)$\cdot\frac{s}{|C|}$ many $\Rightarrow$ demands almost preserved

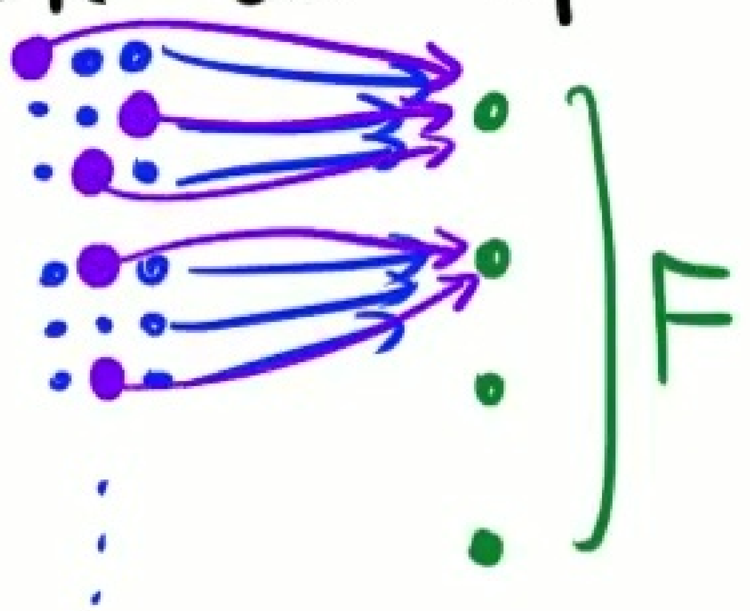
$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$

- Convexity of MinCostFlow:



$\Rightarrow g(\mathbb{E}[d]) \leq \mathbb{E}[g(d)]$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\lesssim} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d
- Look at opt flow for $\mathbb{E}[d] = \mathbb{1}$:



Route sampled clients to same facility in F

$(\pm \varepsilon|C|R)$

for most d (concentration)

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$: • demand $+d_v$ at each $v \in C$
• demand $|C| - \sum_v d_v$ at center c $\Big\}$ total demand $|C|$

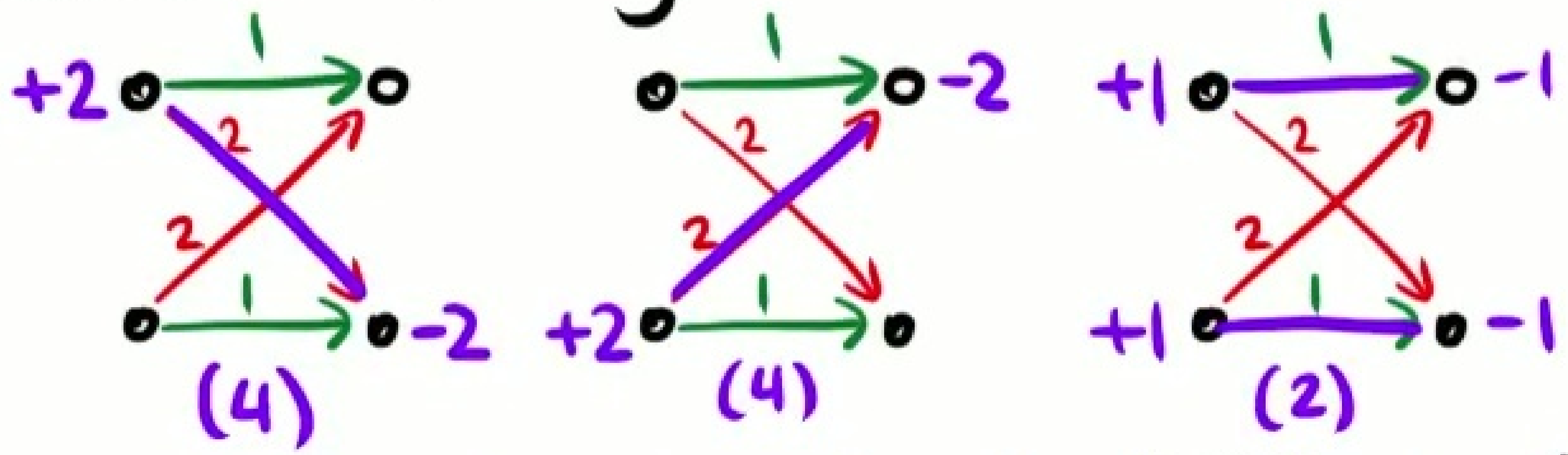
$g(d) := \text{MinCostFlow}(\text{demands}, F)$

$-\text{cap}(f)$ demand at each $f \in F$

- For most d, sampled $\approx \text{size}(\because\because) \cdot \frac{s}{|C|}$ many $\Rightarrow$ demands almost preserved
- Re-route surplus in demand ($\leq \varepsilon|C|$) to different • (cost $\leq \varepsilon|C|R$)

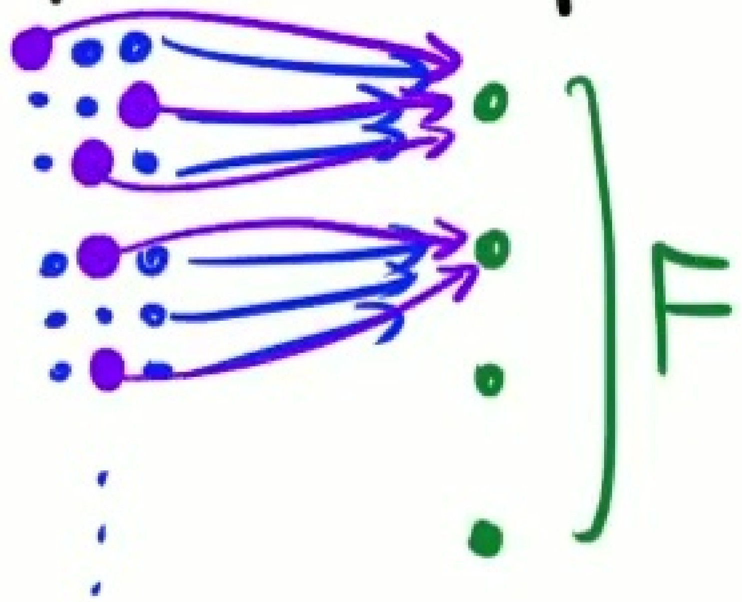
$\mathbb{E}[g(d)] \approx g(\mathbb{E}[d]) = g(\mathbb{1})$

- Convexity of MinCostFlow:



$\Rightarrow g(\mathbb{E}[d]) \leq \mathbb{E}[g(d)]$

- Other direction: $\mathbb{E}[g(d)] \overset{?}{\leq} g(\mathbb{E}[d]) + \varepsilon|C|R$
- Idea: look at $g(\mathbb{E}[d])$; construct flow not much worse for "most" d
- Look at opt flow for $\mathbb{E}[d] = \mathbb{1}$:



Route sampled clients to same facility in F

$(\pm\varepsilon|C|R)$

for most d (concentration)

Suppose sample each $v \in C$ w.p. $\frac{s}{|C|}$ independently

Demand vector $d \in \mathbb{R}_+^C$:

$$d_v = \frac{|C|}{s} \text{ if } v \text{ sampled, else } d_v = 0$$

$g(d)$:
- demand $+d_v$ at each $v \in C$
- demand $|C| - \sum_v d_v$ at center c

total demand $|C|$

$g(d) :=$ MinCostFlow(demands, F)

$-\mathrm{cap}(f)$ demand at each $f \in F$

- For most d, sampled $\approx \mathrm{size}(\because\!\because) \cdot \frac{s}{|C|}$ many $\Rightarrow$ demands almost preserved
- Re-route surplus in demand $(\leq \varepsilon|C|)$ to different • (cost $\leq \varepsilon|C|R$)
- Total error $\leq O(\varepsilon|C|R) \cdot k$ over all •

Reset $\varepsilon \leftarrow \Theta(\varepsilon/k)$

Open problems

- Improve (3+eps)-approx in FPT?
  (Lower bound 1+2/e even for
  k-median in FPT)
- Hardness > (1+2/e), even for polytime?
- More problems where FPT improves
  approximation?