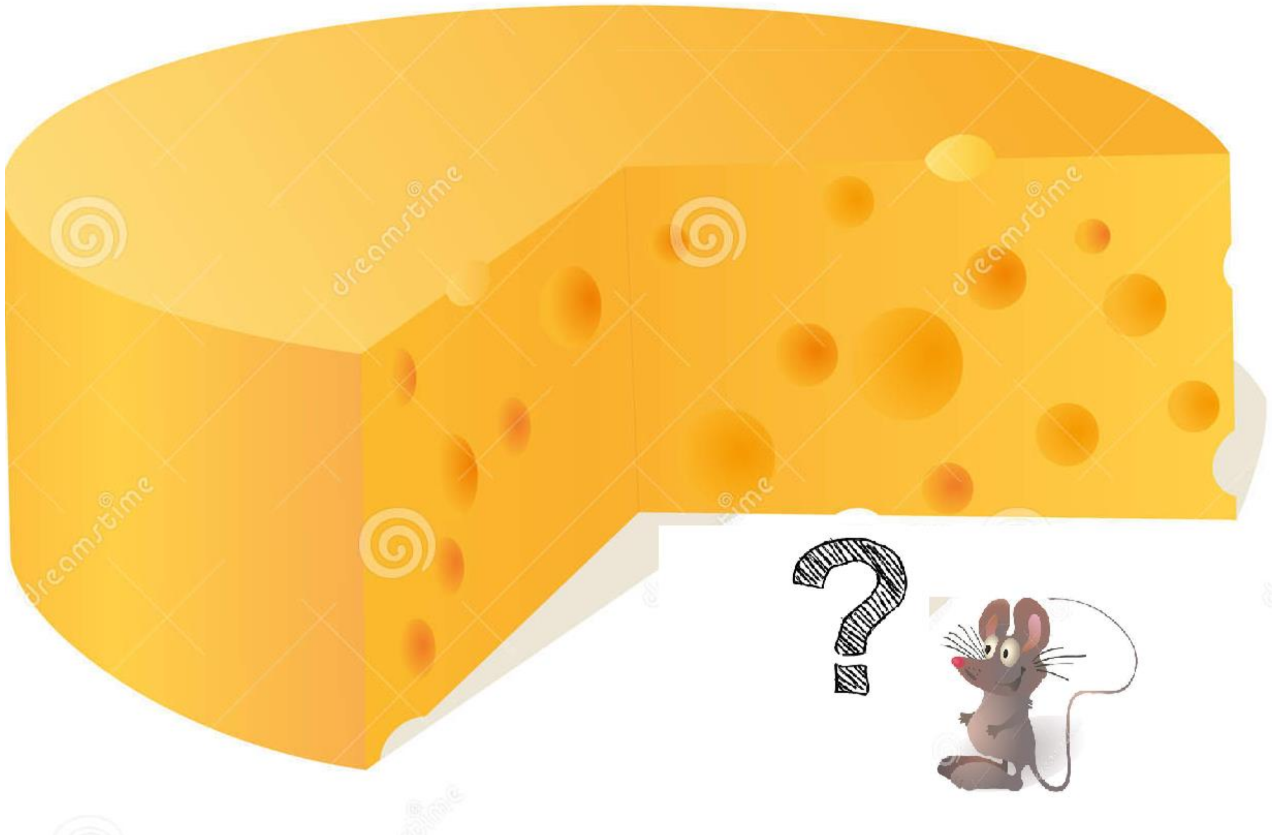
Big Data Class



LECTURER: DAN FELDMAN

TEACHING ASSISTANTS:

IBRAHIM JUBRAN

SOLIMAN NASSER

אוניברסיטת חיפה
University of Haifa
جامعة حيفا

Reminder - Definitions (k-centers)

Let P be an input set of n elements, X be a query space and $dist: P \times X \to [0, \infty)$.
For every $p \in P$ and $Y \subseteq X$ define $dist(p, Y) = \min_{y \in Y} dist(p, y)$.

- $OPT_k = \min_{Y \subseteq X, |Y| = k} \sum_{p \in P} dist(p, Y)$.
- Y' is an α_k-approximation if $|Y'| = k$ and $\sum_{p \in P} dist(p, Y') \leq \alpha \cdot OPT_k$.
- $Y \subseteq X$ is a β_k-approximation if $|Y| = \beta k$ and $\sum_{p \in P} dist(p, Y) \leq OPT_k$.
- $Y' \subseteq X$ is an $(\alpha, \beta)_k$-approximation if $|Y'| = \beta k$ and $\sum_{p \in P} dist(p, Y) \leq \alpha \cdot OPT_k$

Define $Closest(P, Y, \gamma)$ to be the $\lceil (1 - \gamma) n \rceil$ points $p \in P$ with smallest value $dist(p, Y)$.

- $\gamma\text{-}Robust\text{-}OPT_k = \min_{Y \subseteq X, |Y| = k} \sum_{p \in Closest(P, Y, \gamma)} dist(p, Y)$.

Reminder - Definitions (k-centers)

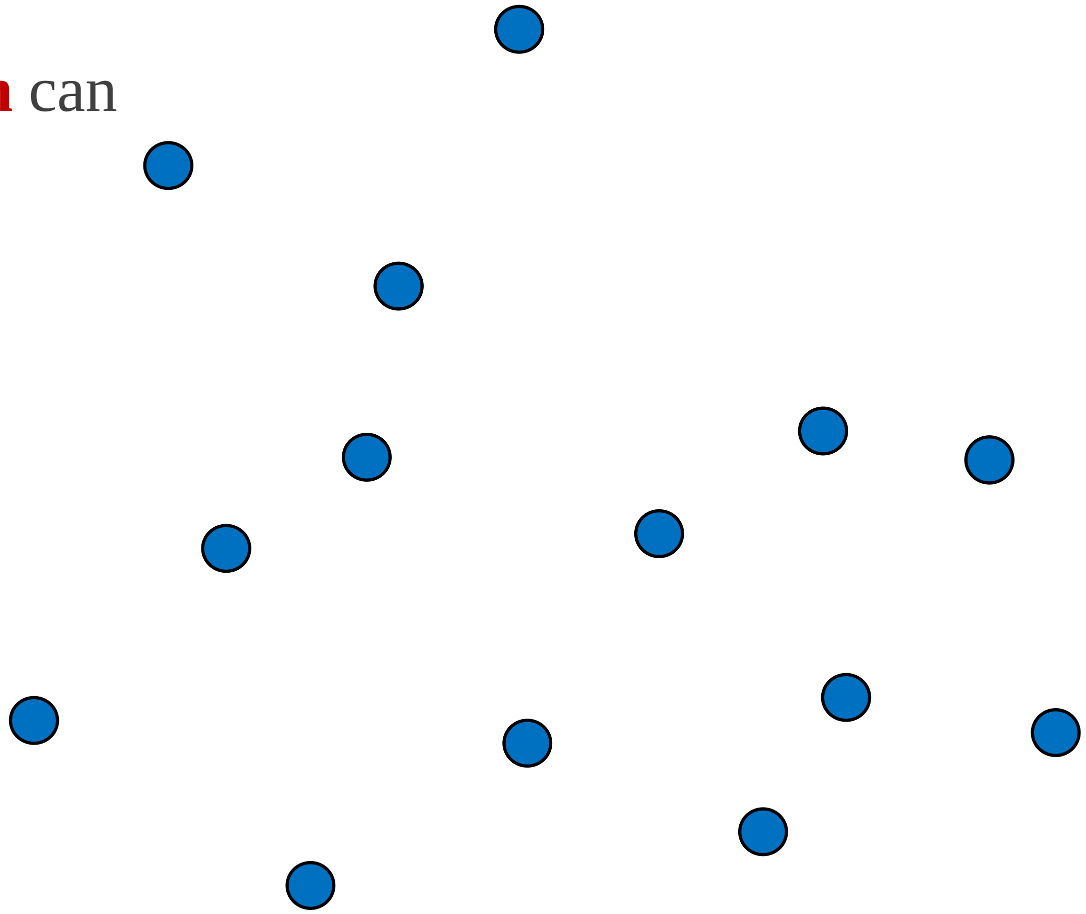
Define $Closest(P, Y, \gamma)$ to be the $\lceil(1-\gamma)n\rceil$ points $p \in P$ with smallest value $dist(p, Y)$.

- $\gamma\text{-}Robust\text{-}OPT_k = \min_{Y \subseteq X, |Y|=k} \sum_{p \in Closest(P,Y,\gamma)} dist(p, Y)$.
- $Y' \subseteq X$ is a $(\gamma, \alpha, \beta)_k$-approximation if $|Y'| = \beta k$ and
$$\sum_{p \in Closest(P,Y',\gamma)} dist(p, Y') \leq \alpha \cdot (\gamma - Robust - OPT_k)$$
- $Y' \subseteq X$ is a $(\gamma, \epsilon, \alpha, \beta)_k$-approximation if $|Y'| = \beta k$ and
$$\sum_{p \in Closest(P,Y',(1-\epsilon)\gamma)} dist(p, Y') \leq \alpha \cdot (\gamma - Robust - OPT_k)$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Claim:

A **$(\boldsymbol{\gamma}, \mathbf{0}, \mathbf{2}^{\boldsymbol{r}}, \mathbf{1})_{\boldsymbol{k}}$-approximation** can be computed in $O(n^k)$ time.

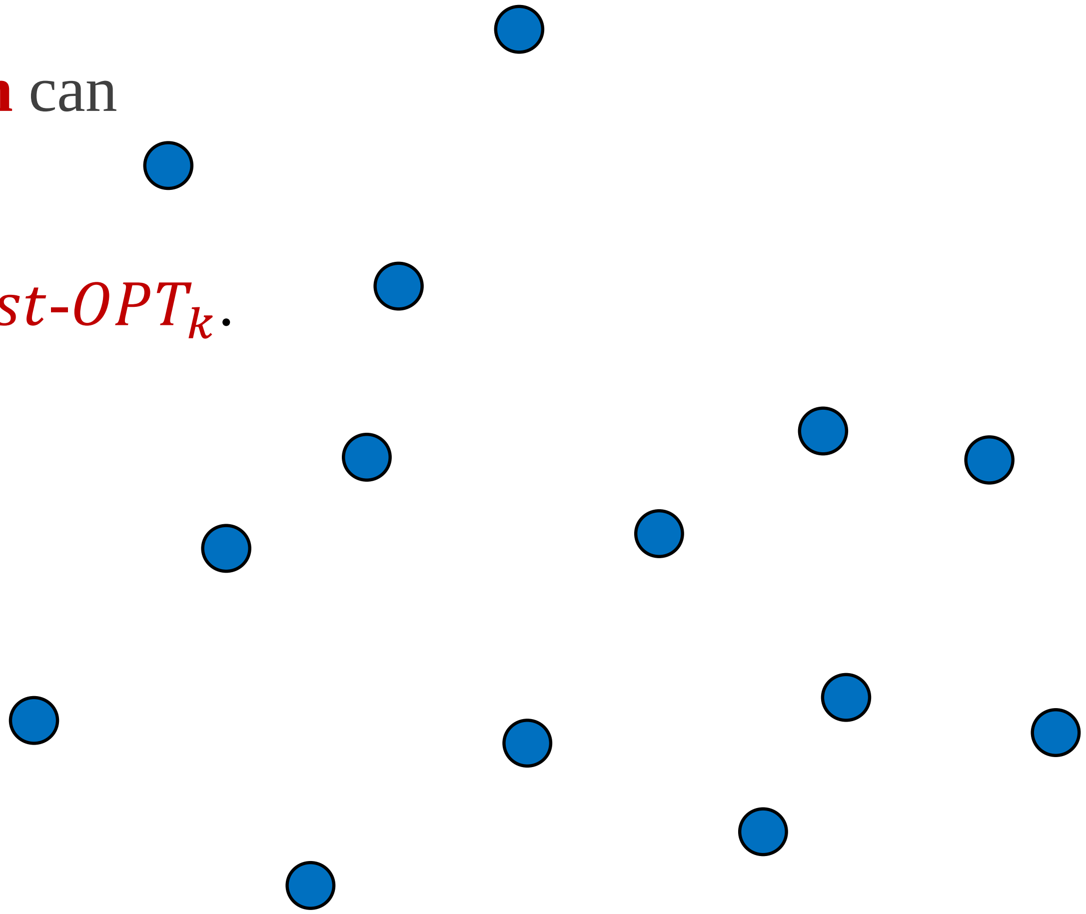


Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Claim:

A $\boldsymbol{(\gamma, 0, 2^r, 1)_k}$**-approximation** can be computed in $O\left(n^k\right)$ time.

Let $\boldsymbol{Y}^* = \{\boldsymbol{y_1}, \boldsymbol{y_2}\}$ be a γ-*Robust-OPT*$_k$.

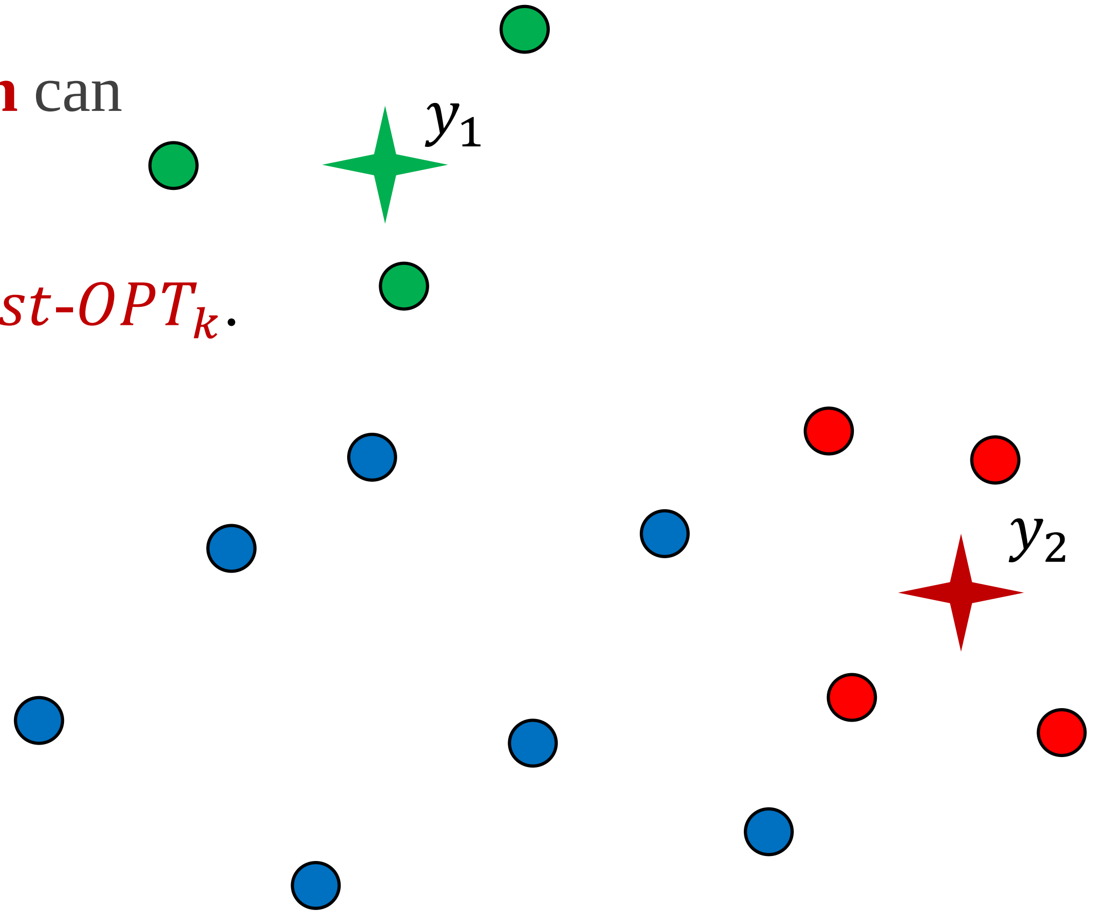


Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Claim:

A $\boldsymbol{(\gamma, 0, 2^r, 1)_k}$**-approximation** can be computed in $O(n^k)$ time.

Let $\boldsymbol{Y^*} = \{\boldsymbol{y_1}, \boldsymbol{y_2}\}$ be a γ*-Robust-OPT*$_k$.



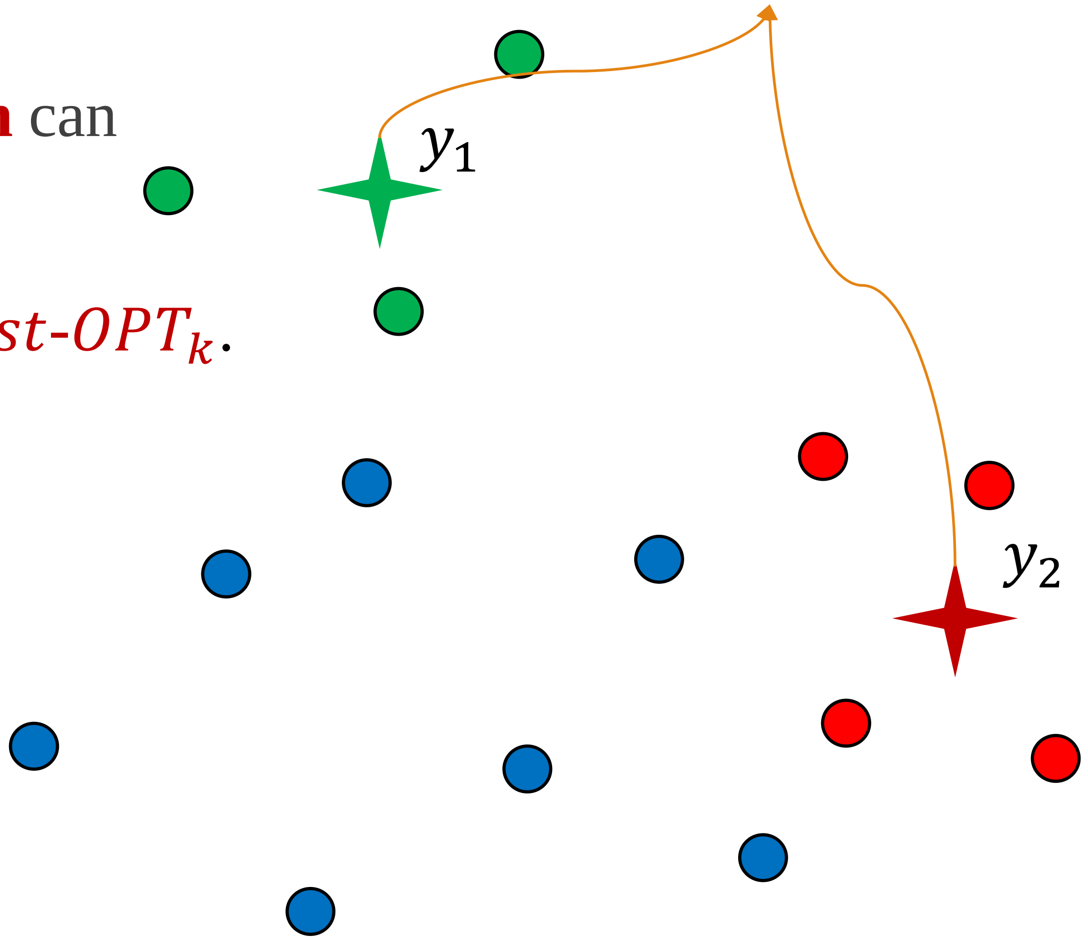
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

$$Y^* = arg \min_{Y \subseteq X, |Y|=k} \sum_{p \in closest(P,Y,\gamma)} dist(p, Y)$$

Claim:

A $(\gamma, 0, 2^r, 1)_k$**-approximation** can be computed in $O(n^k)$ time.

Let $Y^* = \{y_1, y_2\}$ be a γ-*Robust-OPT*$_k$.



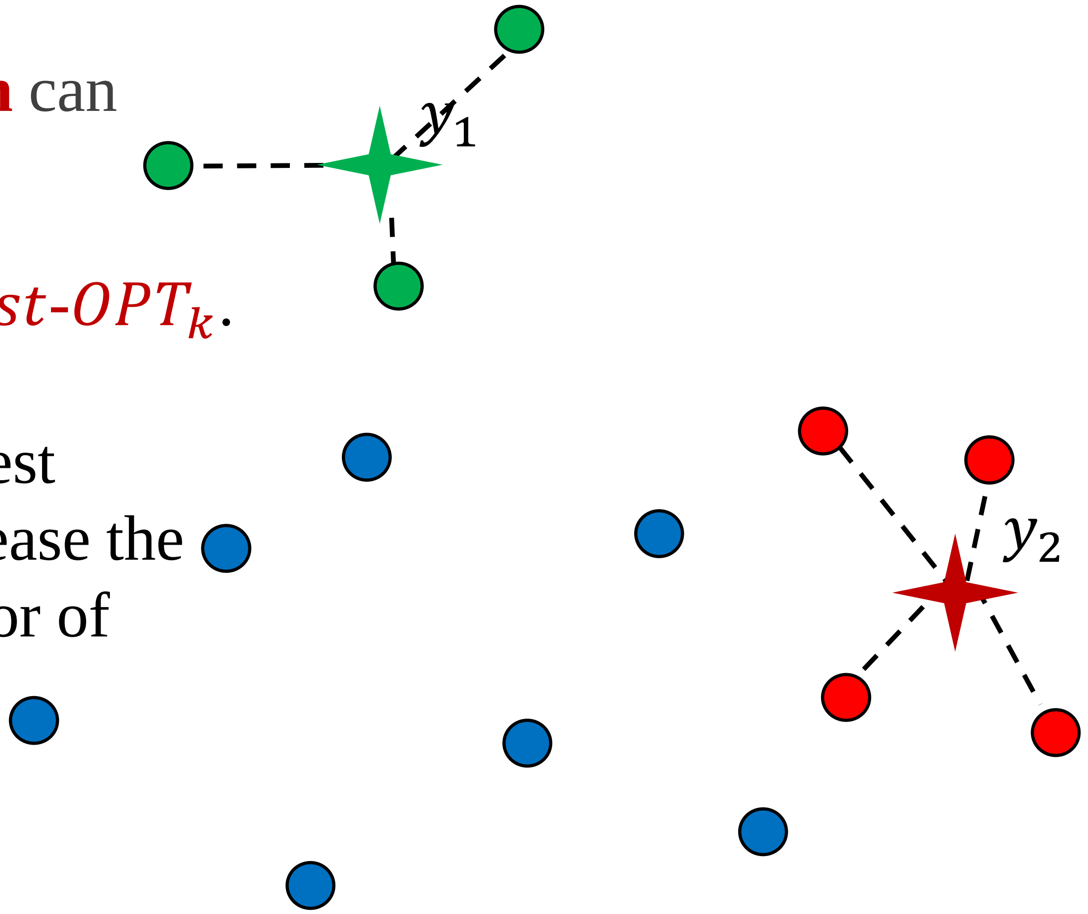
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Claim:

A $(\boldsymbol{\gamma}, \mathbf{0}, \mathbf{2}^{r}, \mathbf{1})_{k}$**-approximation** can be computed in $O\left(n^{k}\right)$ time.

Let $\boldsymbol{Y}^{*} = \{\boldsymbol{y_1}, \boldsymbol{y_2}\}$ be a γ-$Robust$-OPT_k.

Moving each $y \in Y^{*}$ to its closest $p \in closest(P, Y^{*}, \gamma)$ will increase the distance of each point by a factor of $\alpha = 2^{r}$ (worst case).



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

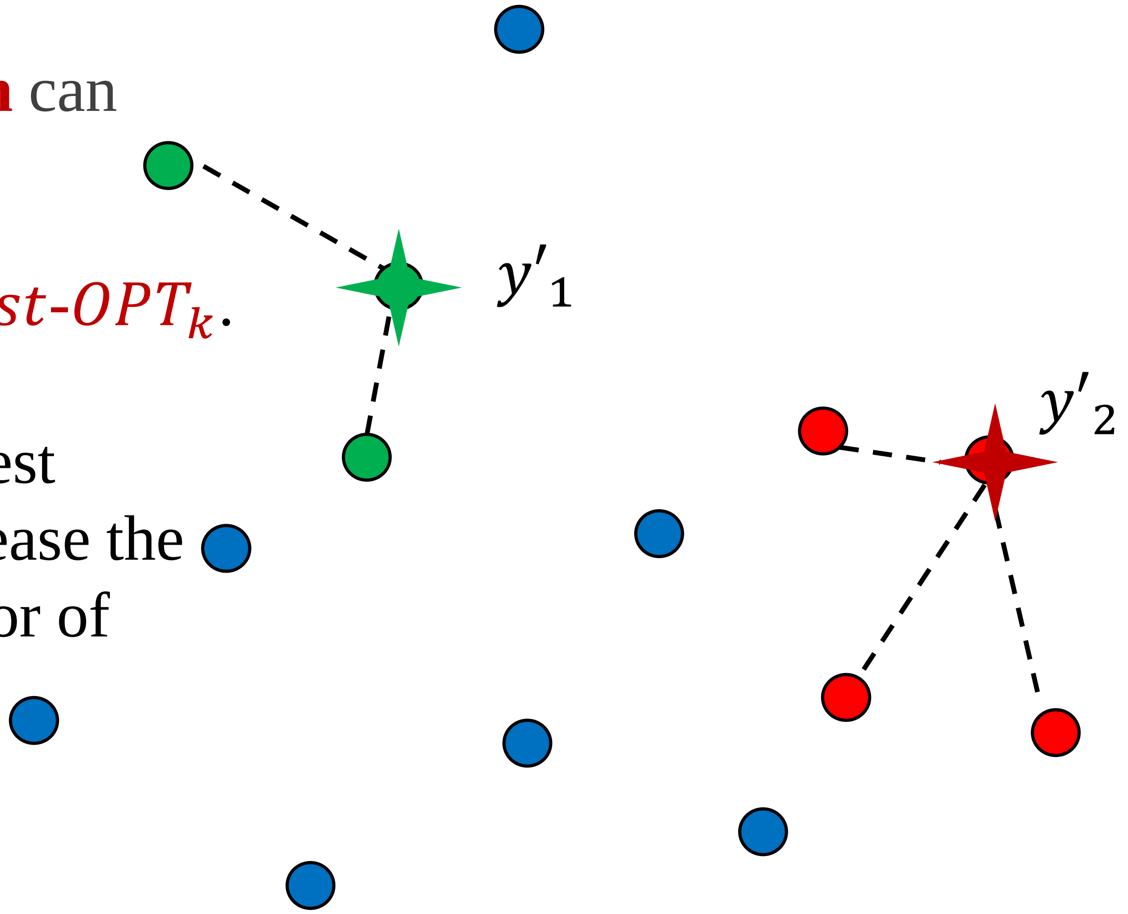
Claim:

A $(\boldsymbol{\gamma}, \mathbf{0}, \mathbf{2}^{r}, \mathbf{1})_{k}$**-approximation** can be computed in $O(n^k)$ time.

Let $\boldsymbol{Y}^* = \{\boldsymbol{y_1}, \boldsymbol{y_2}\}$ be a γ-$Robust$-OPT_k.

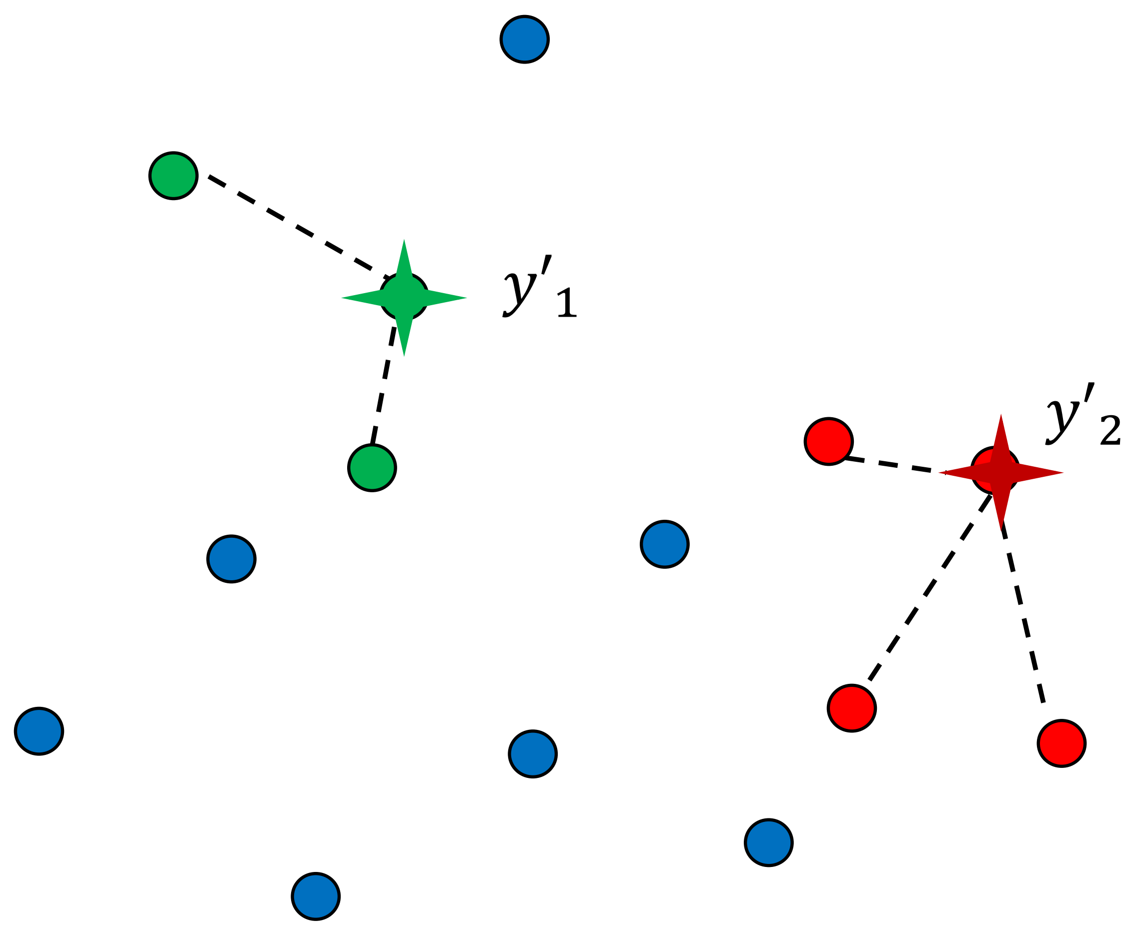
Moving each $y \in Y^*$ to its closest $p \in closest(P, Y^*, \gamma)$ will increase the distance of each point by a factor of $\alpha = 2^r$ (worst case).

$\boldsymbol{Y}' = \{\boldsymbol{y'}_1, \boldsymbol{y'}_2\}$



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

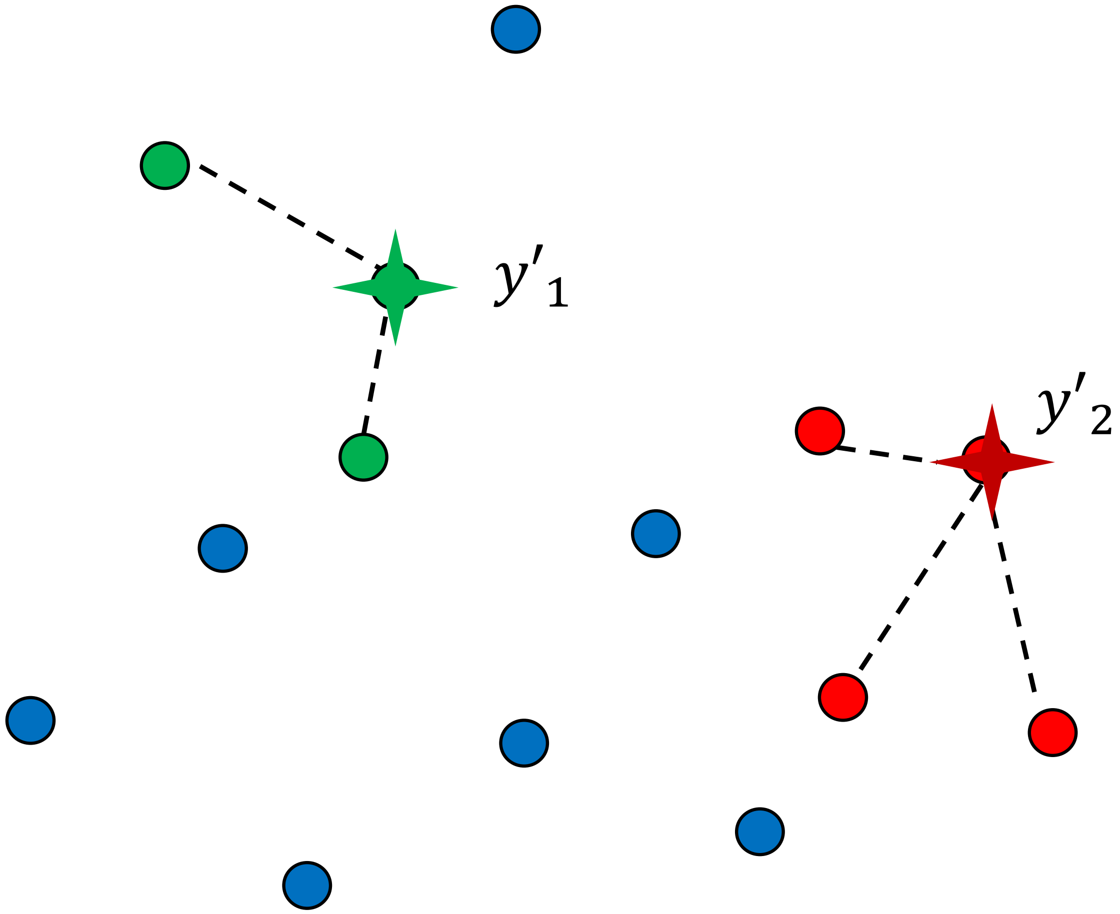
$$\sum_{p \in Closest(P, Y', \gamma)} dist(p, Y')$$



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

$$\sum_{p \in Closest(P, Y', \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, Y')$$

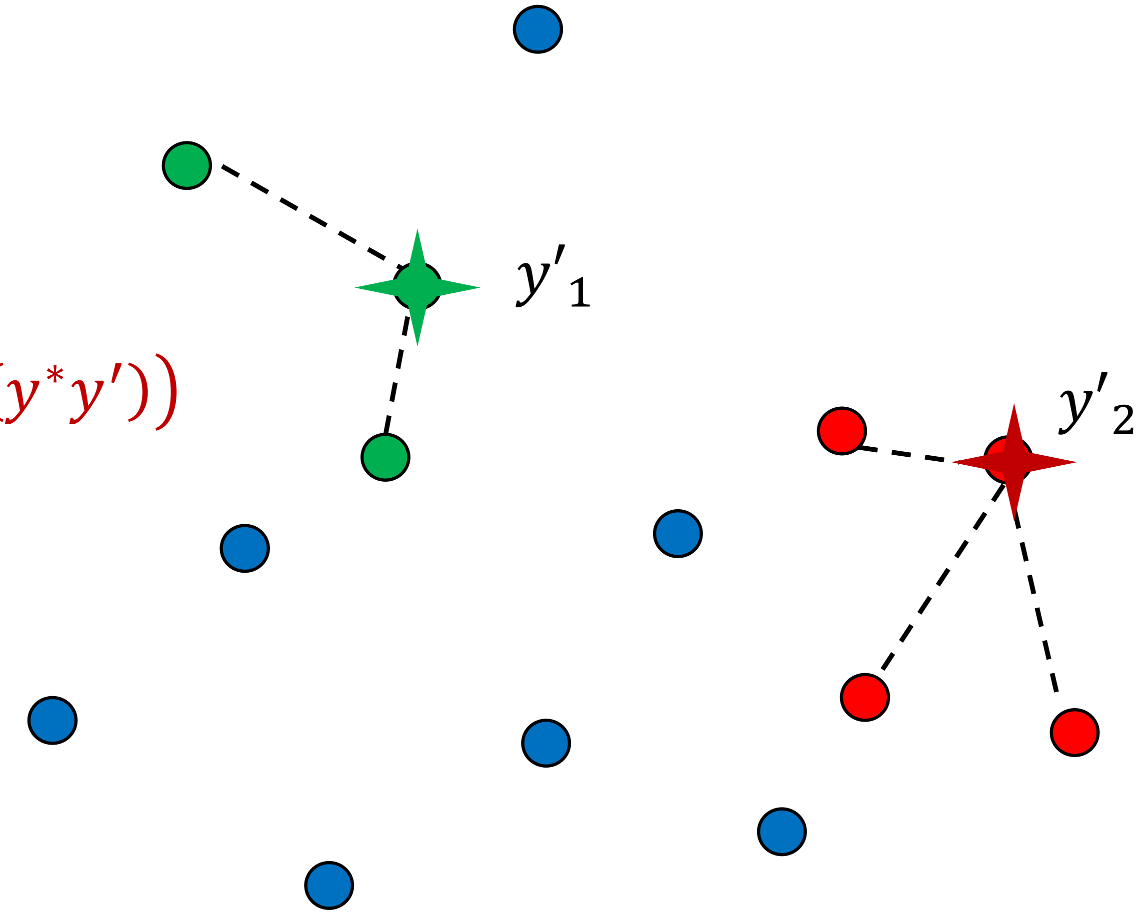


Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

$$\sum_{p \in Closest(P, Y', \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho\big(dist(p, y^*) + dist(y^* y')\big)$$



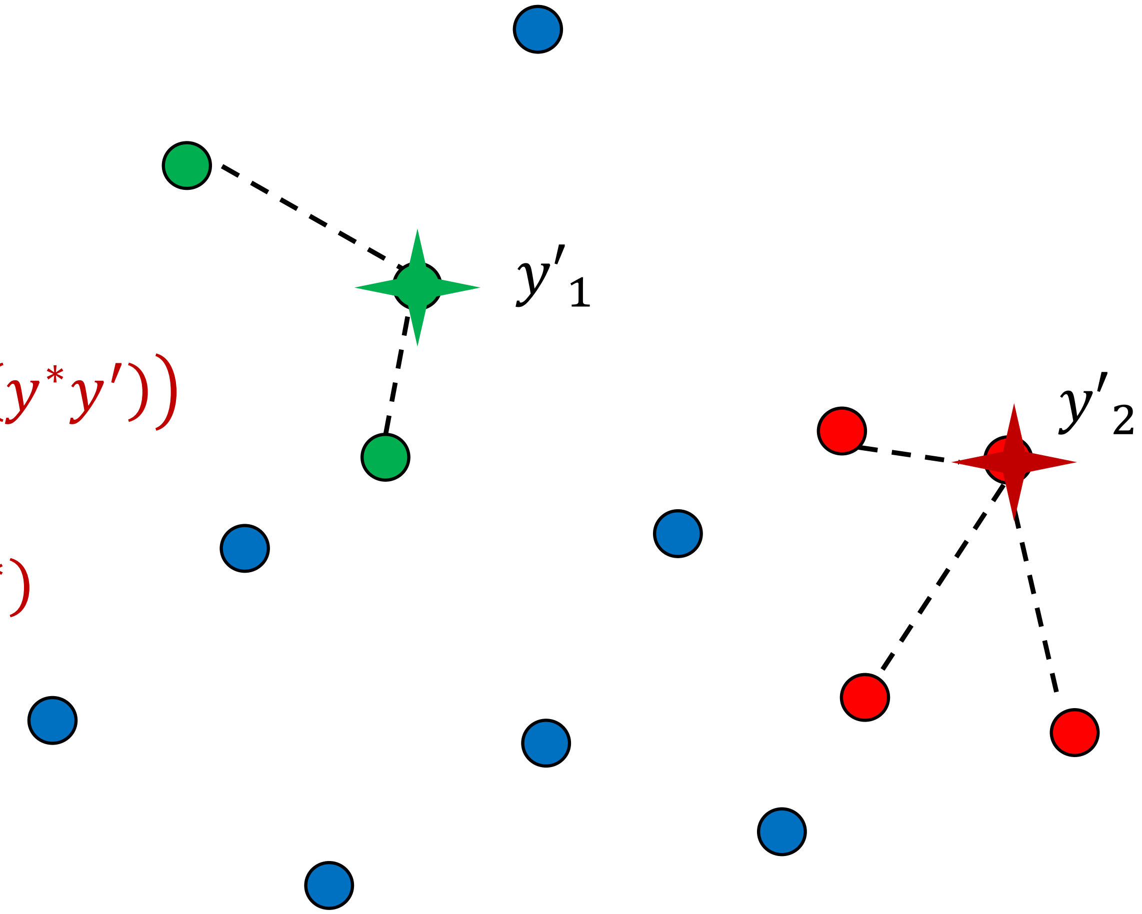
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

$$\sum_{p \in Closest(P, Y', \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho\left(dist(p, y^*) + dist(y^* y')\right)$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho \cdot 2 \cdot dist(p, y^*)$$



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

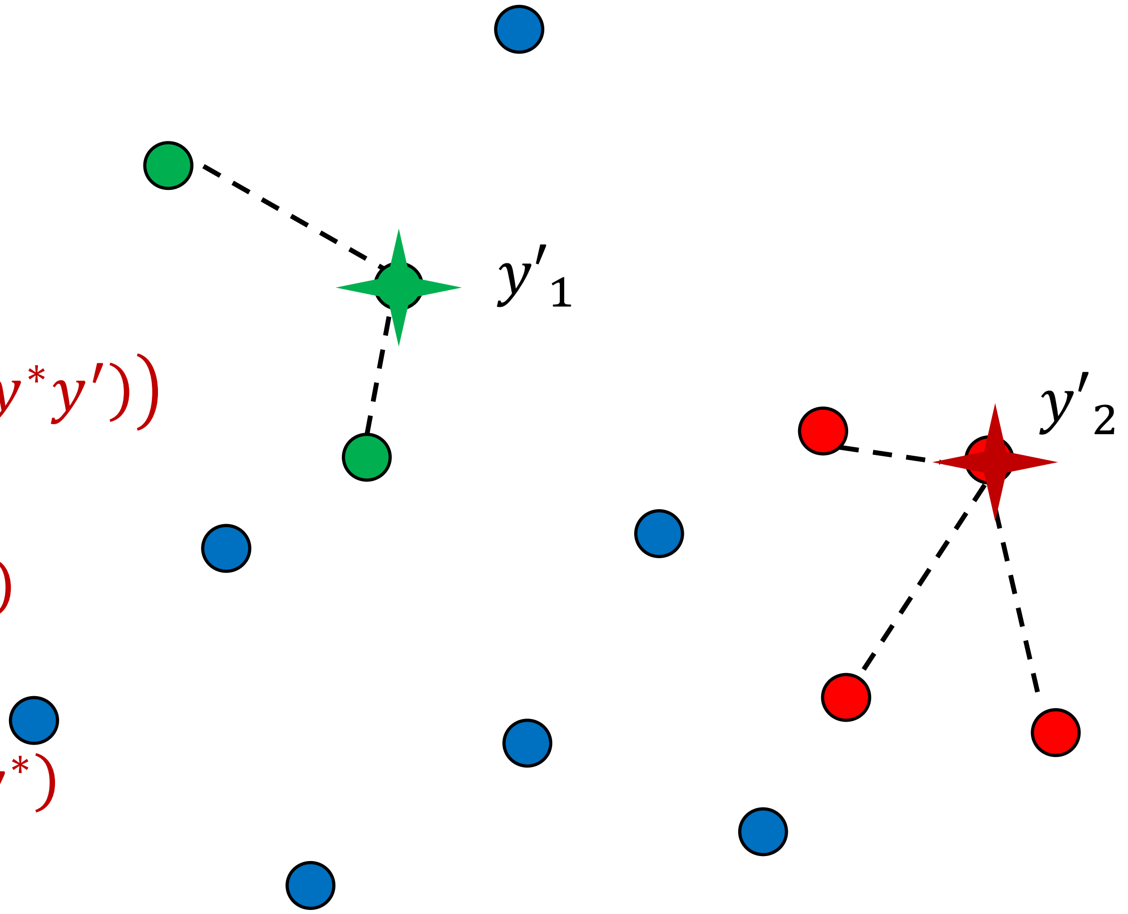
$$\sum_{p \in Closest(P, Y', \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho\big(dist(p, y^*) + dist(y^* y')\big)$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho \cdot 2 \cdot dist(p, y^*)$$

$$\leq 2^{r-1} \cdot 2 \cdot \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, y^*)$$



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

$$\sum_{p \in Closest(P, Y', \gamma)} dist(p, Y')$$

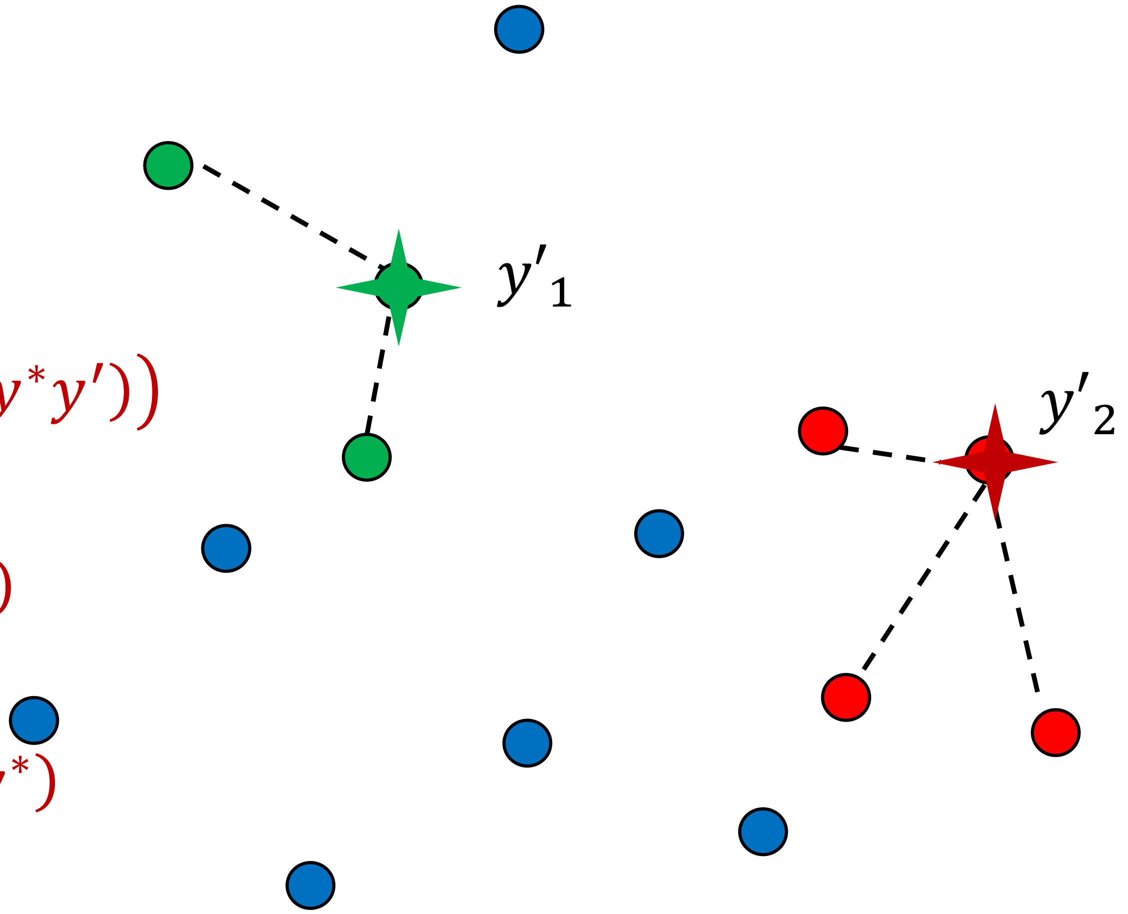
$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, Y')$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho\big(dist(p, y^*) + dist(y^* y')\big)$$

$$\leq \sum_{p \in Closest(P, Y^*, \gamma)} \rho \cdot 2 \cdot dist(p, y^*)$$

$$\leq 2^{r-1} \cdot 2 \cdot \sum_{p \in Closest(P, Y^*, \gamma)} dist(p, y^*)$$

$$= 2^r \cdot \gamma\text{–}Robust\text{–}OPT_k$$



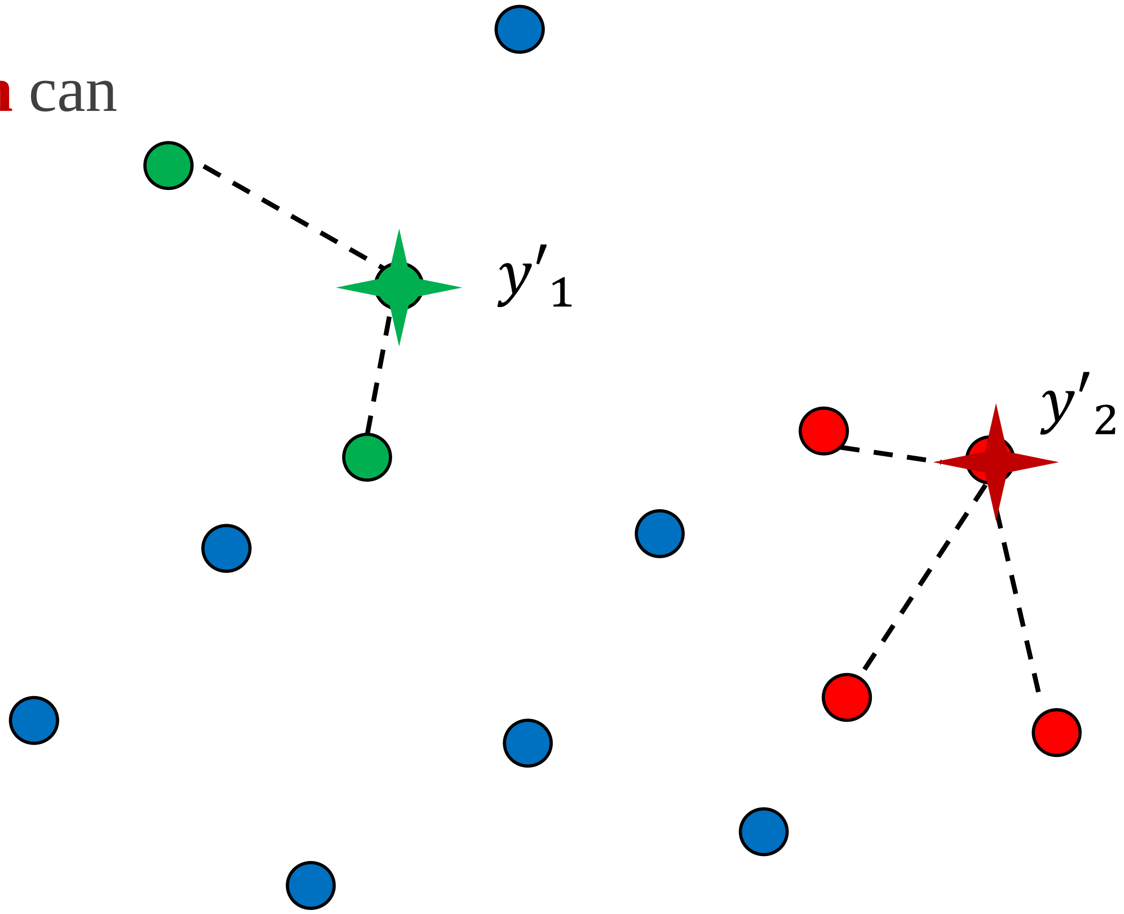
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Claim:

A $\boldsymbol{(\gamma, 0, 2^r, 1)_k}$**-approximation** can be computed in $O(n^k)$ time.

Algorithm:

Exhaustive search over all the k-tuples of points in P.



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

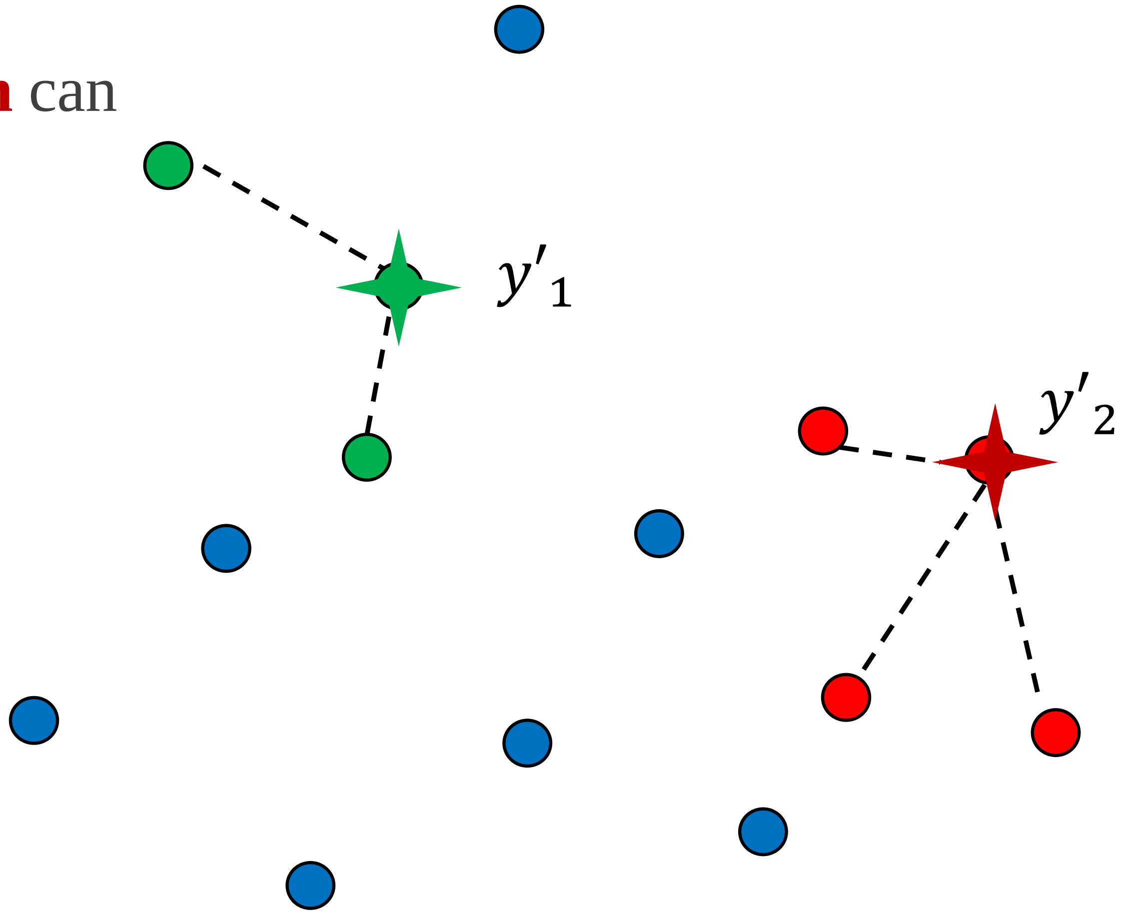
Claim:

A $(\boldsymbol{\gamma}, \mathbf{0}, \mathbf{2}^{\boldsymbol{r}}, \mathbf{1})_{\boldsymbol{k}}$**-approximation** can be computed in $O(n^k)$ time.

Algorithm:

Exhaustive search over all the k-tuples of points in P.

⇨ Inefficient!



(γ, ϵ)-coreset

Definition:

Let $\epsilon \in \left(0, \frac{1}{2}\right)$ and $\gamma \in (0,1]$. Let $P, S \subseteq D$ be two sets of elements.
Let $dist: D \times X \to [0, \infty)$. For every $x \in X$:

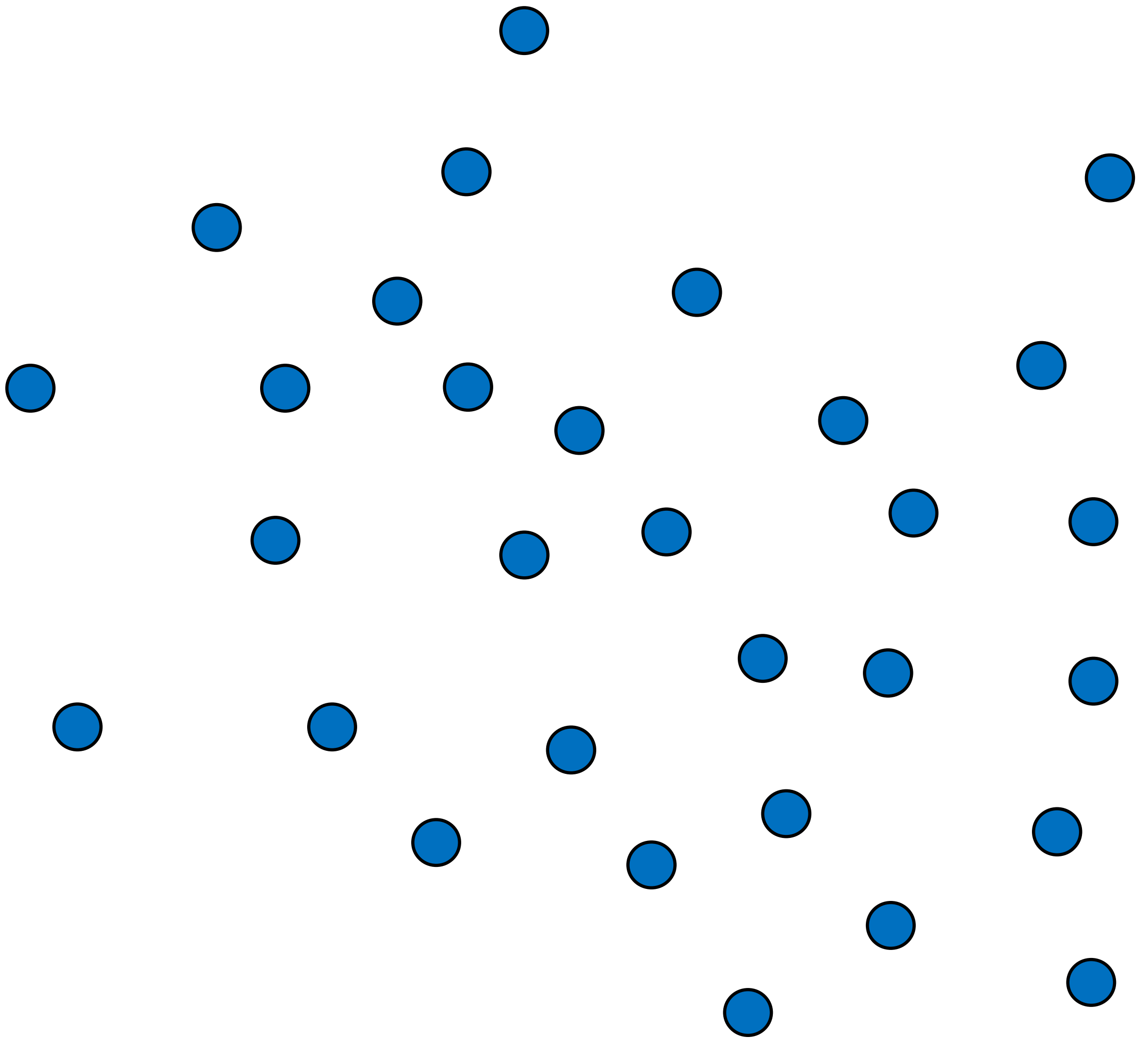
- P_x denotes the $\lceil \gamma |P| \rceil$ elements $p \in P$ with smallest value $dist(p, x)$.
- S_x denotes the $\lceil (1-\epsilon)\gamma |S| \rceil$ elements $p \in S$ with smallest value $dist(p, x)$.
- $G_x \subseteq P_x$ denotes the $\lceil (1-2\epsilon)\gamma |P| \rceil$ elements $p \in P$ with smallest value $dist(p, x)$.

➢ The set S is (γ, ϵ)-good for P if:

$$\forall x \in X: (1-\epsilon) \cdot \frac{cost(G_x, x)}{|G_x|} \leq \frac{\text{cost}(S_x, x)}{|S_x|} \leq \frac{cost(P_x, x)}{|P_x|} \cdot (1+\epsilon)$$

➢ The set S is a (γ, ϵ)-coreset of P if for every $\gamma' \in [\gamma, 1], \epsilon' \in \left[\epsilon, \frac{1}{2}\right)$ we have that S is (γ', ϵ')-good for P.

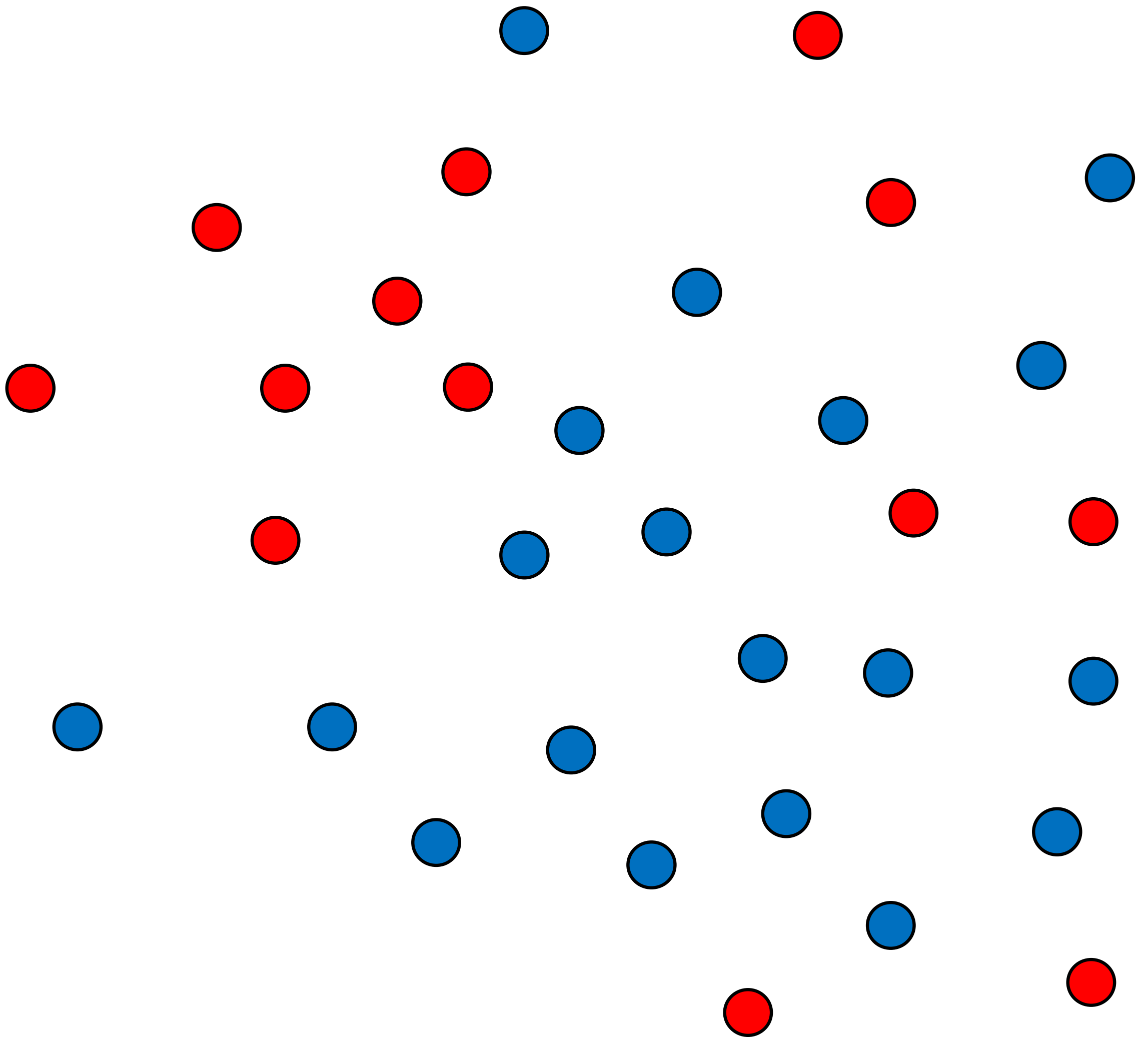
(γ, ϵ)-coreset



(γ, ϵ)-coreset

$\gamma = \frac{2}{3}, \quad \epsilon = \frac{1}{2}$

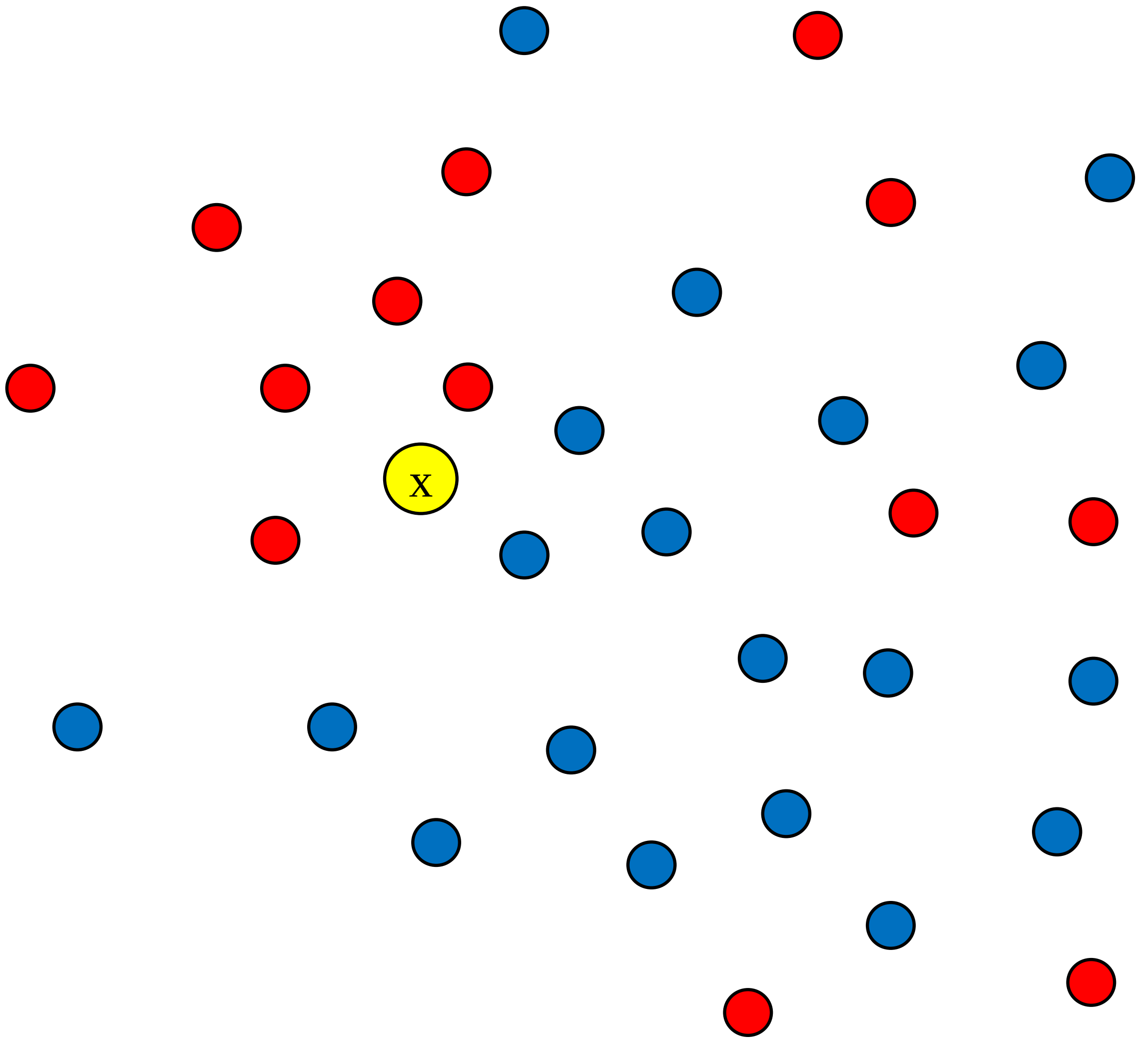
S is a (γ, ϵ)-coreset



(γ, ϵ)-coreset

S is a (γ, ϵ)-coreset

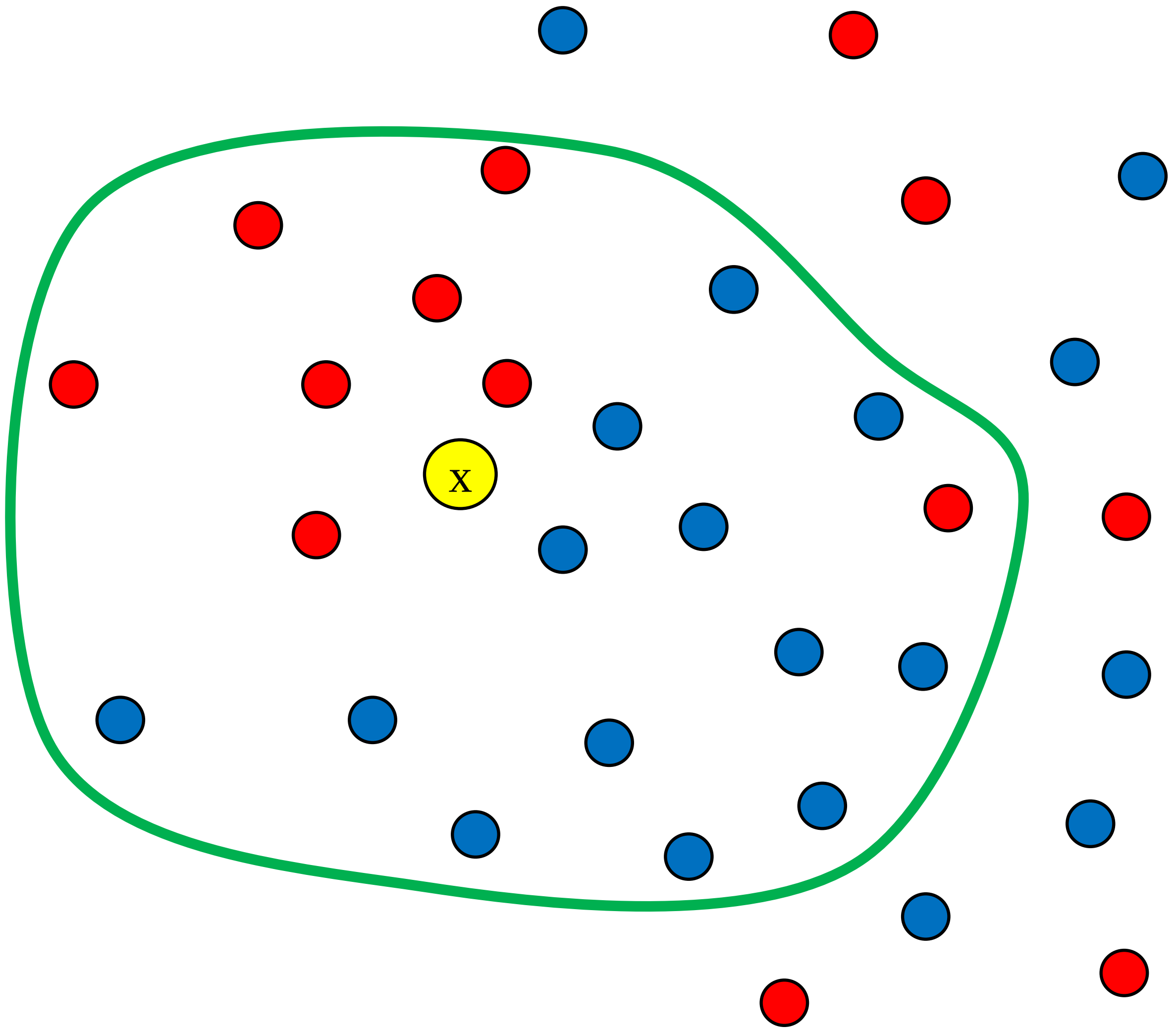
$$\boldsymbol{\gamma = \frac{2}{3}}, \qquad \boldsymbol{\epsilon = \frac{1}{2}}$$



(γ, ϵ)-coreset

S is a (γ, ϵ)-coreset

$$\gamma = \frac{2}{3}, \qquad \epsilon = \frac{1}{2}$$

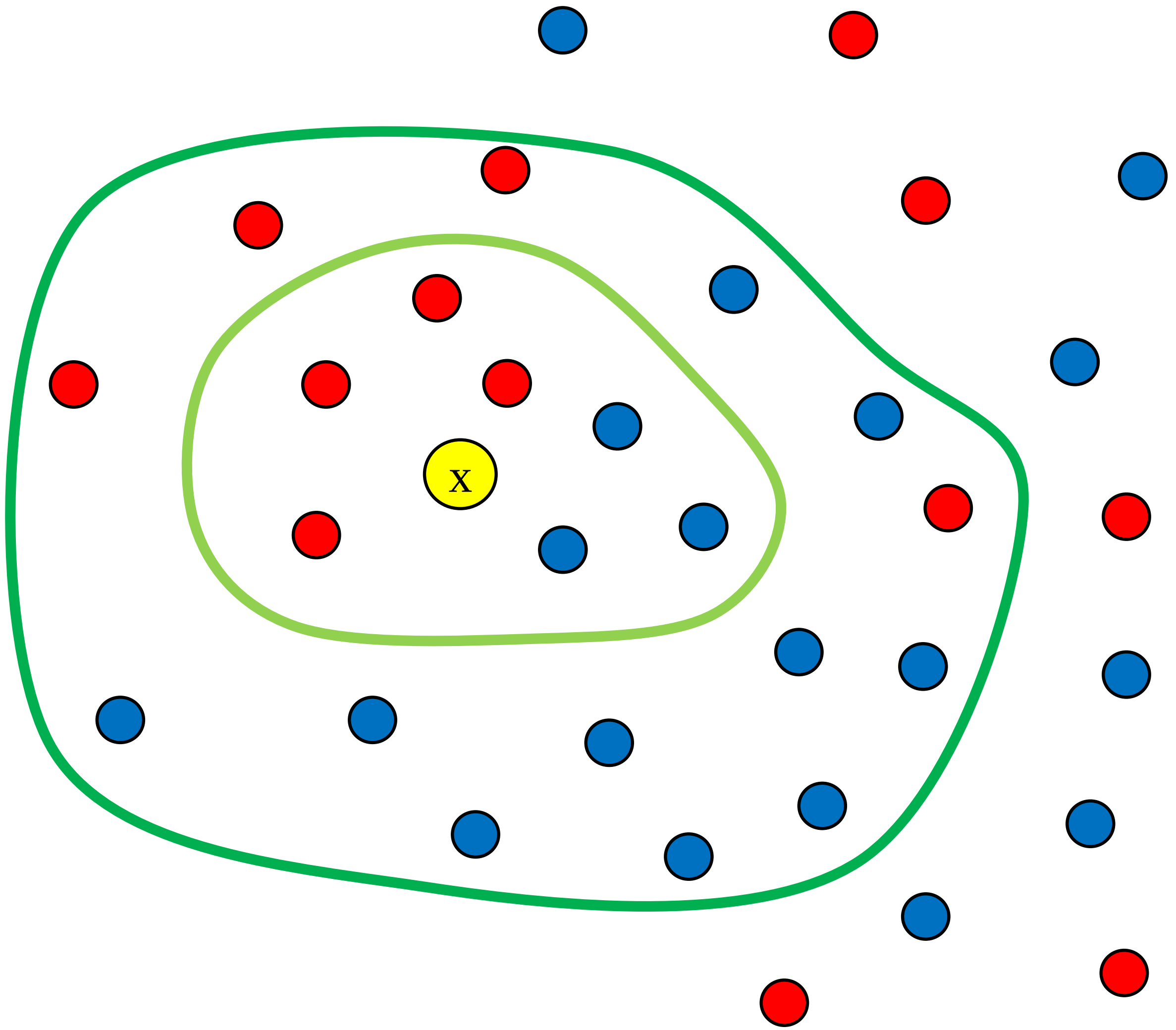


$P_x \rightarrow \lceil \gamma |P| \rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

(γ, ϵ)-coreset

S is a (γ, ϵ)-coreset

$$\gamma = \frac{2}{3}, \qquad \epsilon = \frac{1}{2}$$



$P_x \rightarrow \lceil \gamma |P| \rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

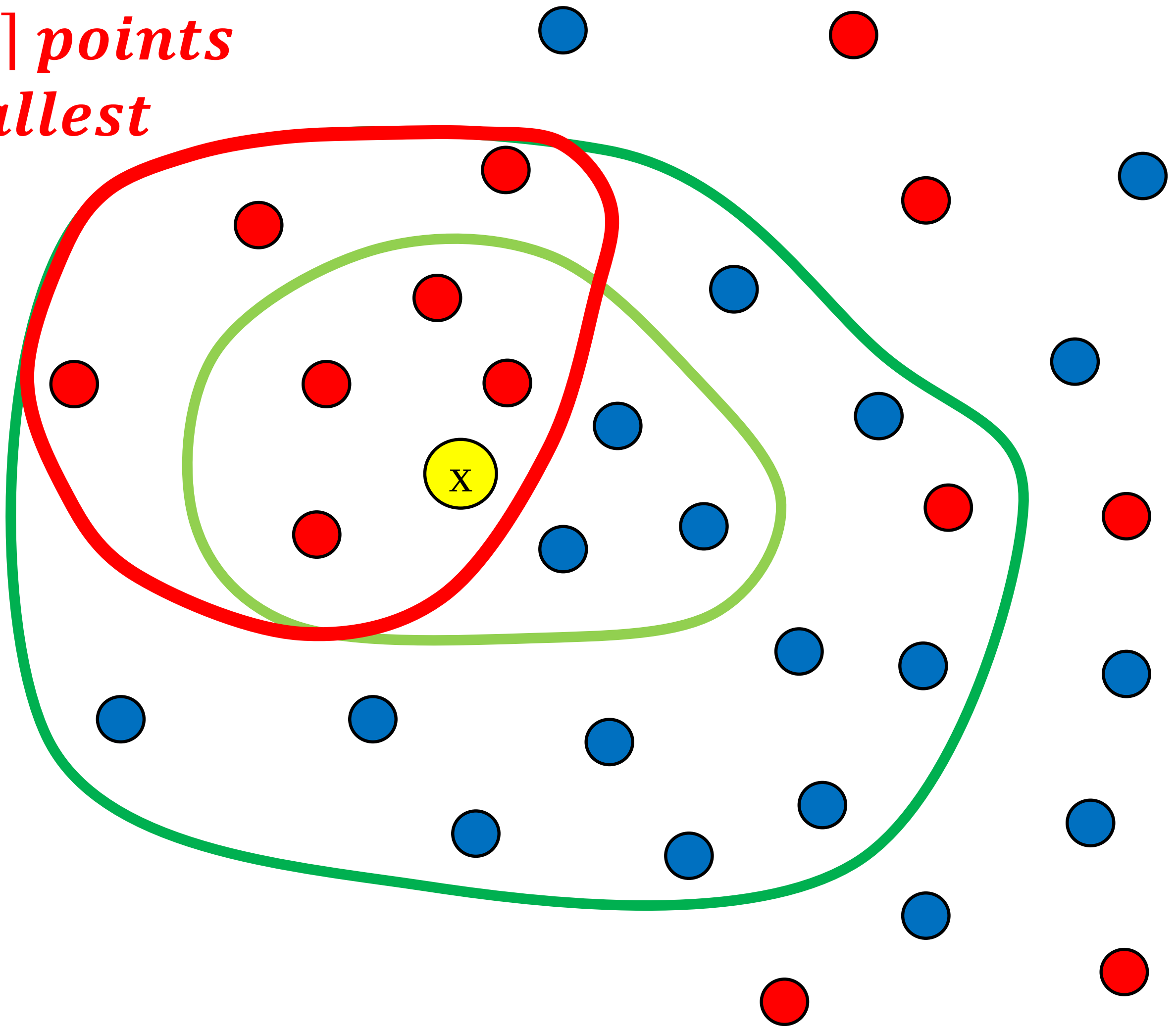
$G_x \rightarrow \lceil (1 - 2\epsilon)\gamma |P| \rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

(γ, ϵ)-coreset

$\gamma = \frac{2}{3}, \qquad \epsilon = \frac{1}{2}$

S is a (γ, ϵ)-coreset

$S_x \rightarrow \lceil (1-\epsilon)\gamma|P| \rceil$ *points* $s \in S$ *with smallest* $dist(s, x)$



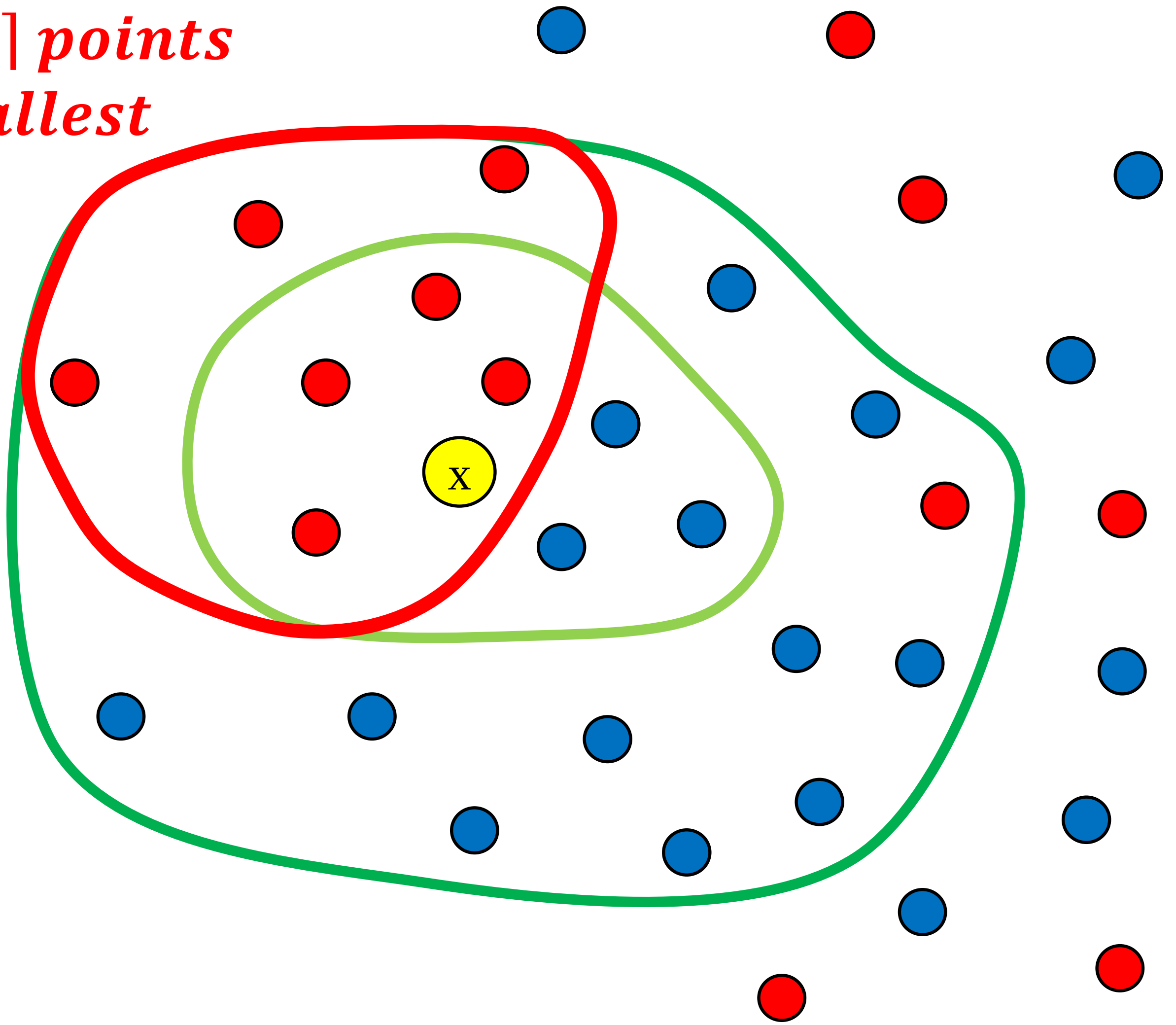
$P_x \rightarrow \lceil \gamma|P| \rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

$G_x \rightarrow \lceil (1-2\epsilon)\gamma|P| \rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

(γ, ϵ)-coreset

$$(1-\epsilon)\cdot\frac{cost(G_x, x)}{|G_x|} \le \frac{cost(S_x, x)}{|S_x|} \le \frac{cost(P_x, x)}{|P_x|}\cdot(1+\epsilon)$$

$S_x \rightarrow \lceil(1-\epsilon)\gamma|P|\rceil$ *points* $s \in S$ *with smallest* $dist(s, x)$



$P_x \rightarrow \lceil\gamma|P|\rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

$G_x \rightarrow \lceil(1-2\epsilon)\gamma|P|\rceil$ *points* $p \in P$ *with smallest* $dist(p, x)$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Construct a $\gamma\epsilon$-sample

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

Prove that a $\frac{\gamma\epsilon^2}{63}$-sample of P
is a (γ, ϵ)-coreset for P

Construct a
$\gamma\epsilon$-sample

Proof 2

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

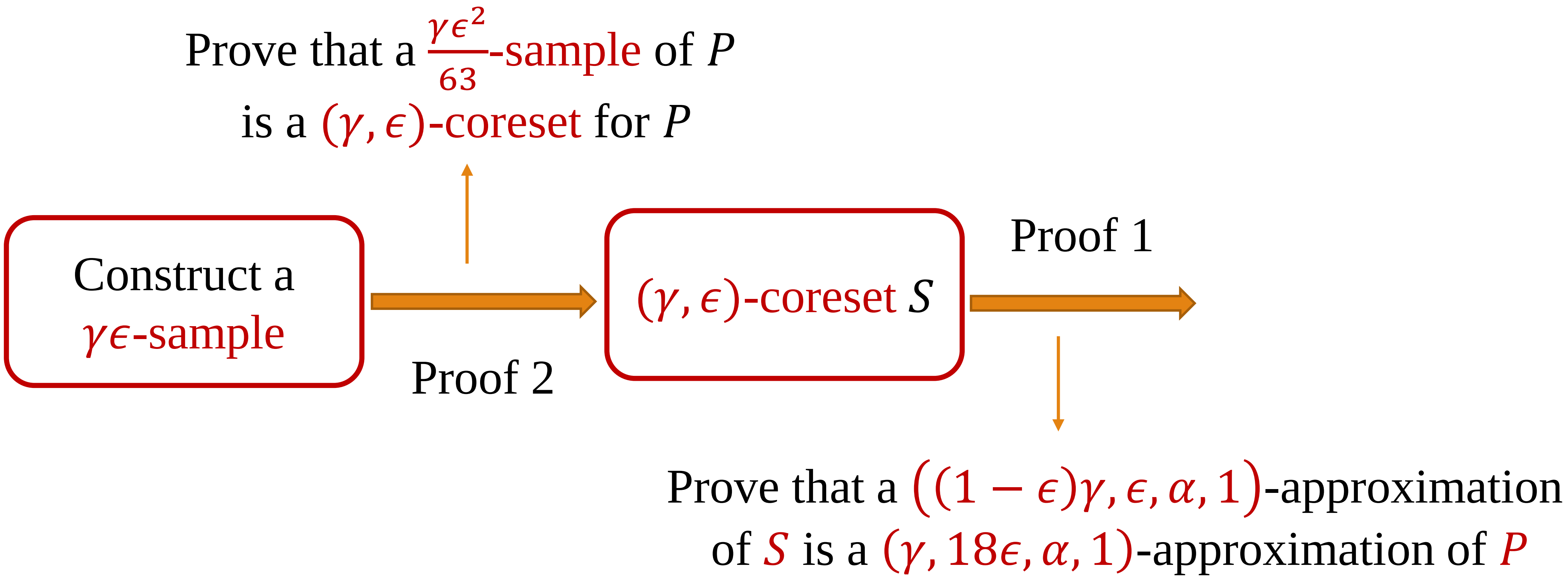
Prove that a $\frac{\gamma\epsilon^2}{63}$-sample of P
is a (γ, ϵ)-coreset for P

Construct a
$\gamma\epsilon$-sample

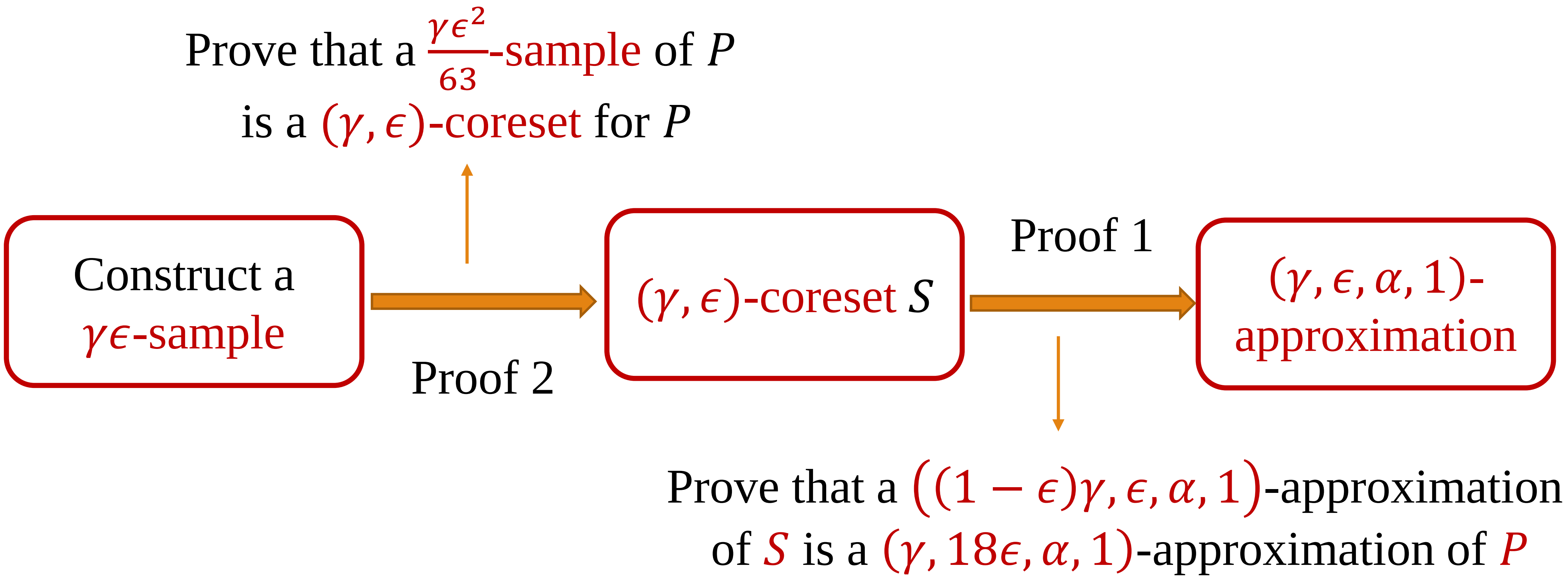
Proof 2

(γ, ϵ)-coreset S

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Claim 1:**

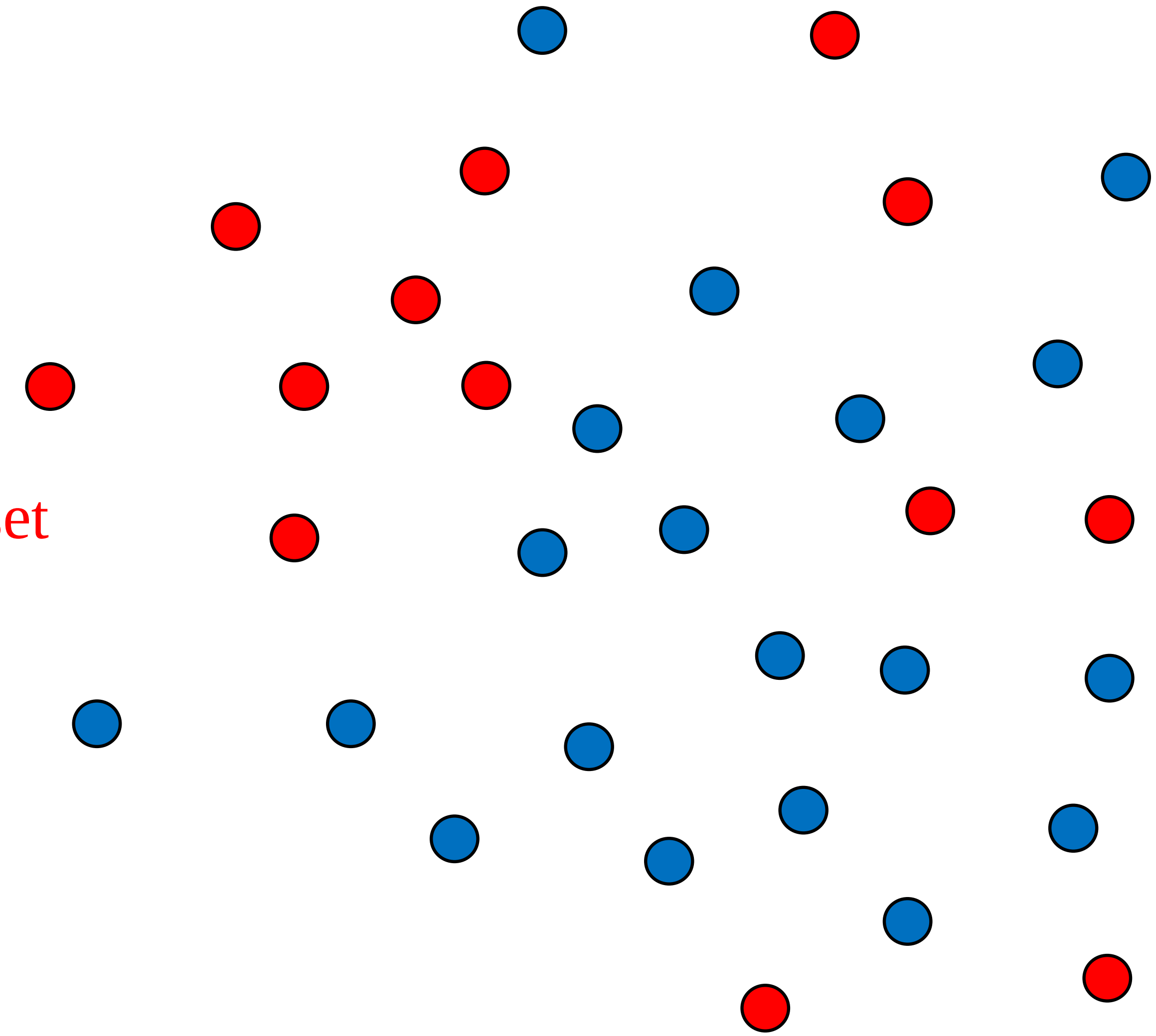
Let P be a set of points and $dist: P \times X \to [0, \infty)$. Let $\epsilon \in \left(0, \frac{1}{10}\right), \gamma \in (0,1]$. Suppose that S is a (γ, ϵ)-coreset of P and that $|P| \geq |S| \geq \frac{2}{\epsilon\gamma}$. Let $\alpha > 0$. Then a $\big((1-\epsilon)\gamma, \epsilon, \alpha\big)$-approximation of S is also a $(\gamma, 18\epsilon, \alpha)$-approximation of P.

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

Definitions:

S is a (γ, ϵ)-coreset



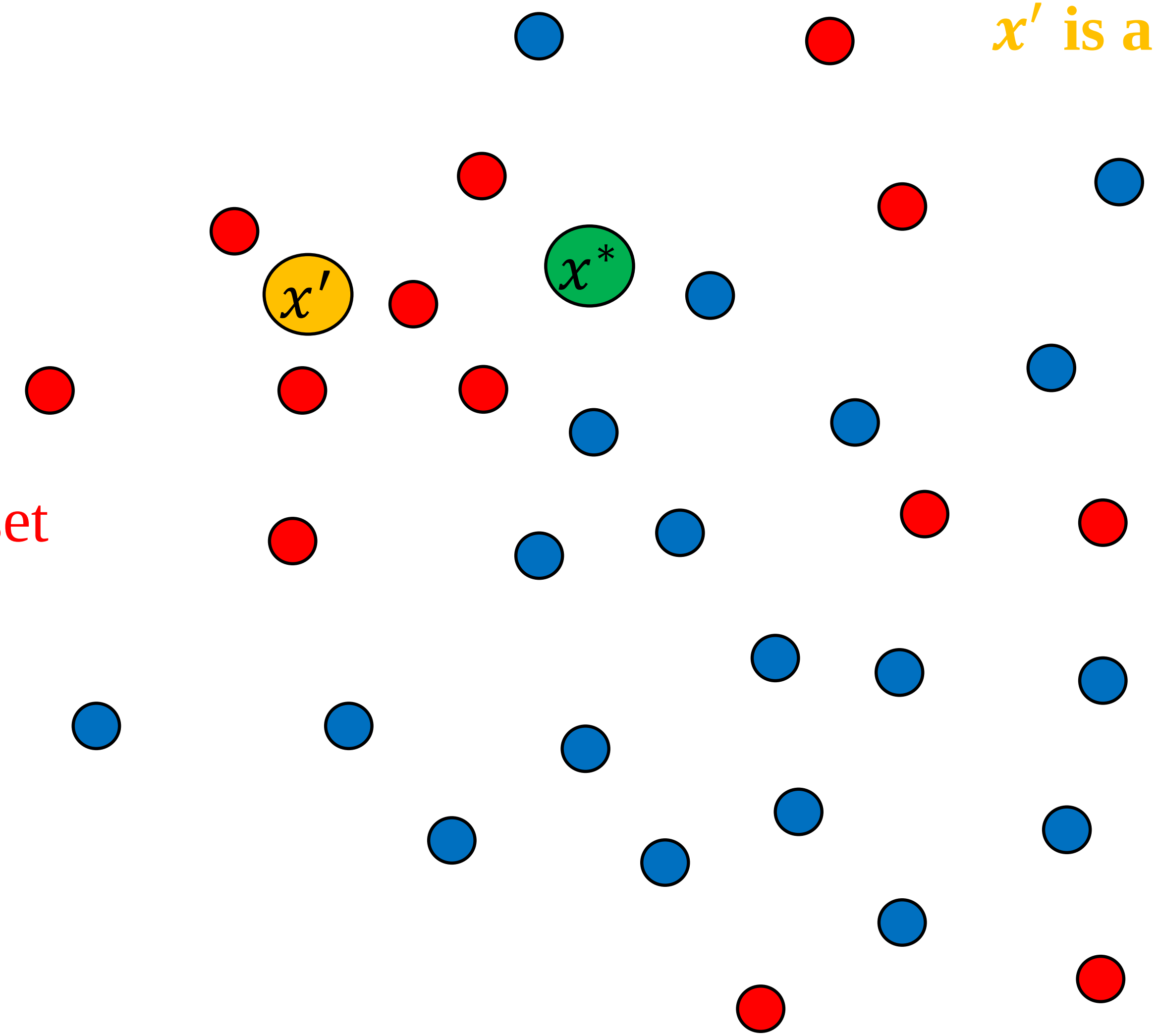
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

x^* is a $(\gamma, 0, 0)$-approx for P

x' is a $\left((1-\epsilon)\gamma, \epsilon, \alpha\right)$-approx for S

- **<u>Proof 1:</u>**

<u>Definitions:</u>



S is a (γ, ϵ)-coreset

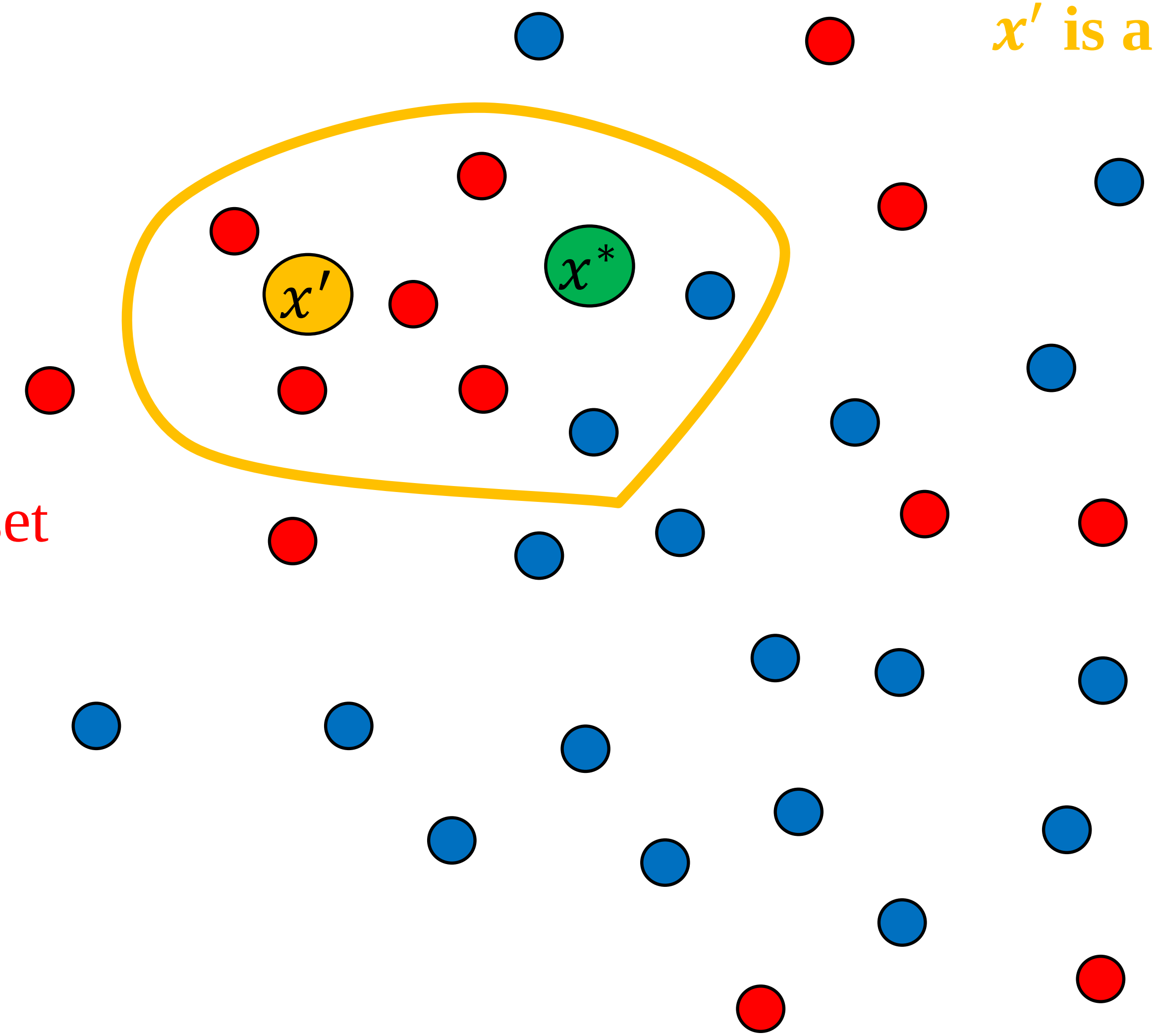
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

x^* **is a** $(\gamma, 0, 0)$**-approx for** P

x' **is a** $((1-\epsilon)\gamma, \epsilon, \alpha)$**-approx for** S

- **<u>Proof 1:</u>**

<u>Definitions:</u>



$G \rightarrow \lceil (1-8\epsilon)\gamma |P| \rceil$ *points* $p \in P$ *with smallest* $dist(p, x')$

S is a (γ, ϵ)-coreset

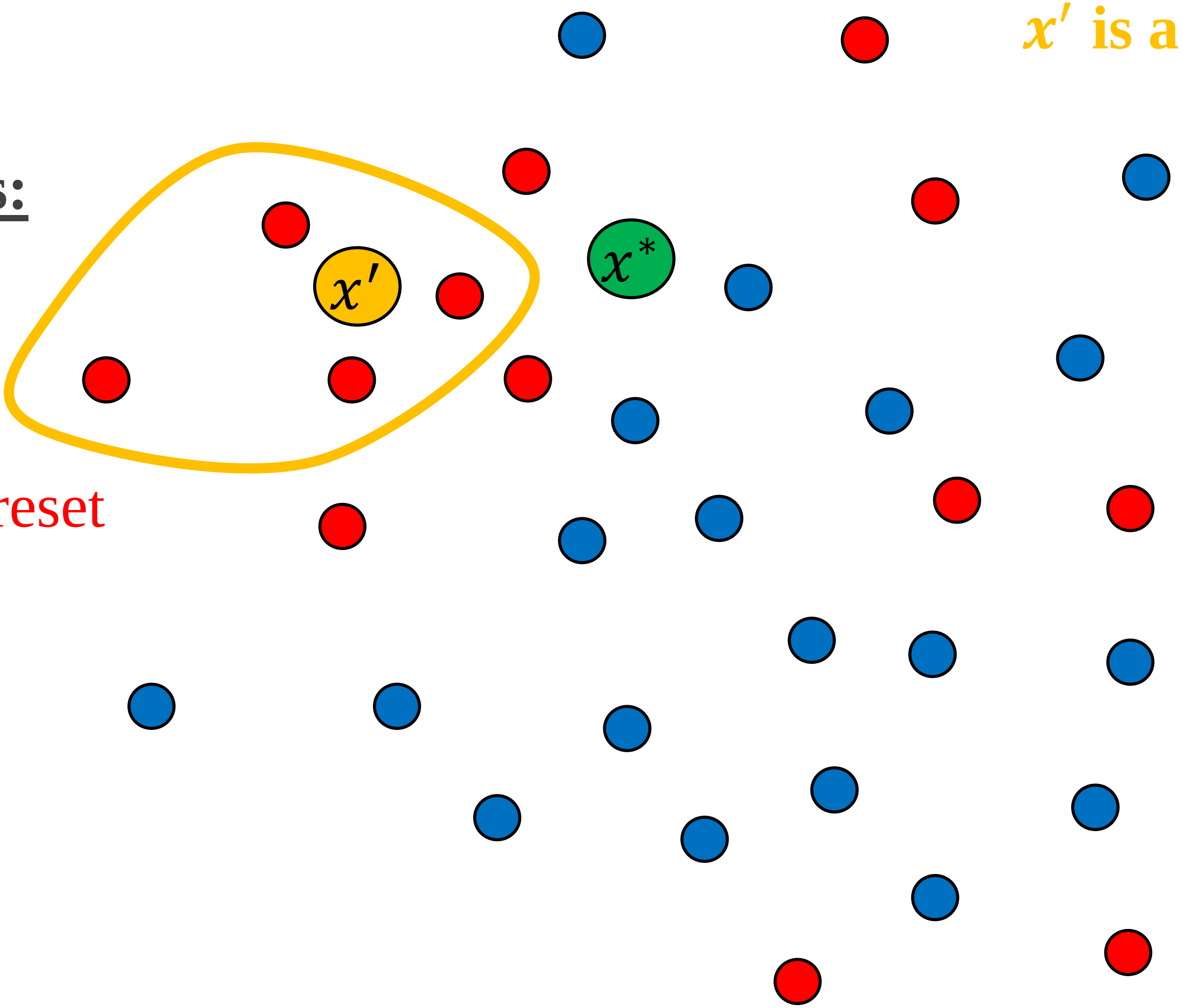
Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

x^* is a $(\gamma, 0, 0)$-approx for P

x' is a $\left((1-\epsilon)\gamma, \epsilon, \alpha\right)$-approx for S

- **<u>Proof 1:</u>**

<u>Definitions:</u>



S is a (γ, ϵ)-coreset

$S' \to \lceil (1-4\epsilon)\gamma|S| \rceil$ *points* $s \in S$ *with smallest* $dist(s, x')$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

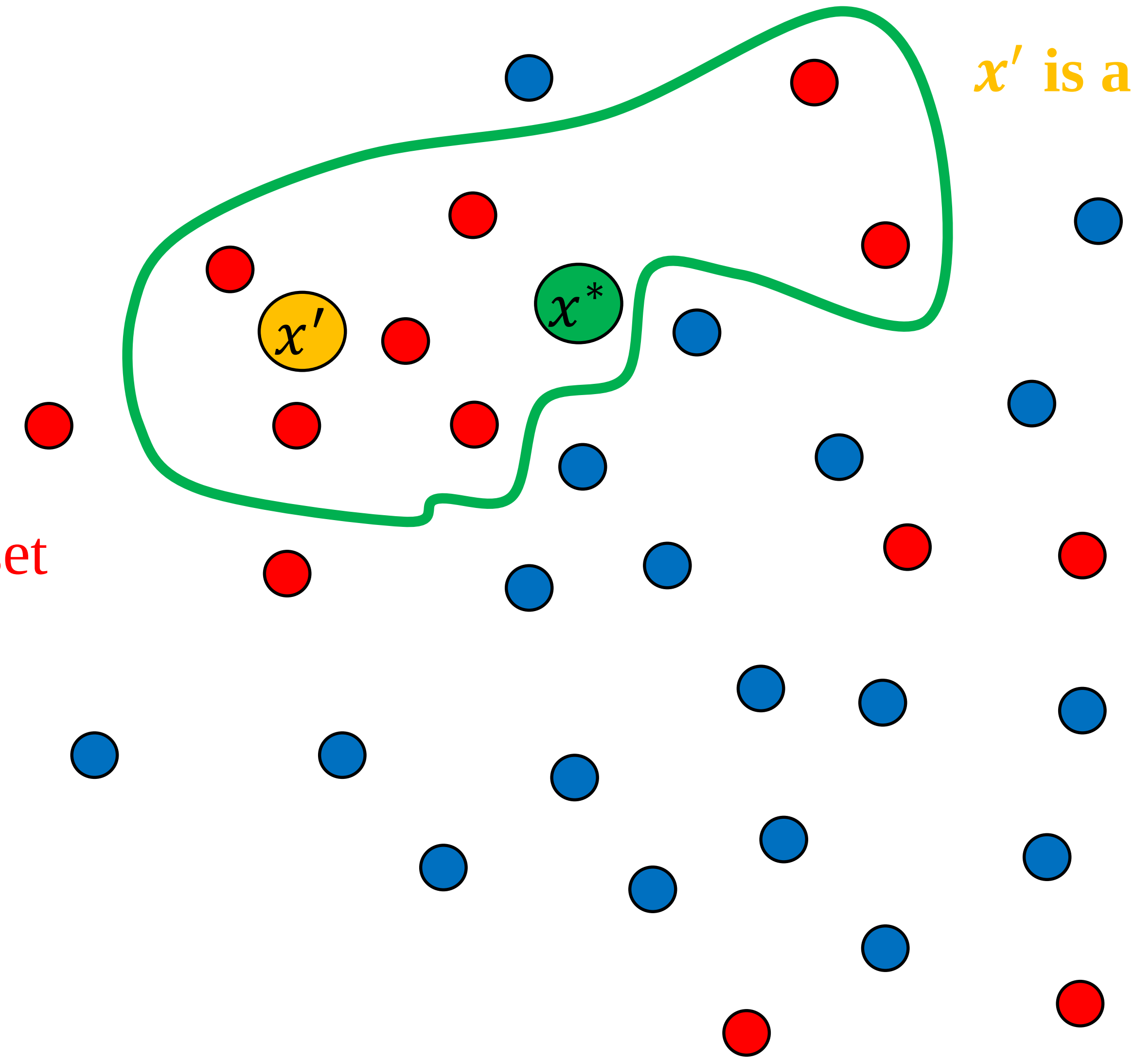
x^* is a $(\gamma, 0, 0)$-approx for P

x' is a $\left((1-\epsilon)\gamma, \epsilon, \alpha\right)$-approx for S

- **Proof 1:**

Definitions:

S is a (γ, ϵ)-coreset



$S^* \rightarrow \lceil (1-\epsilon)\gamma|S| \rceil$ *points* $s \in S$ *with smallest* $dist(s, x^*)$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

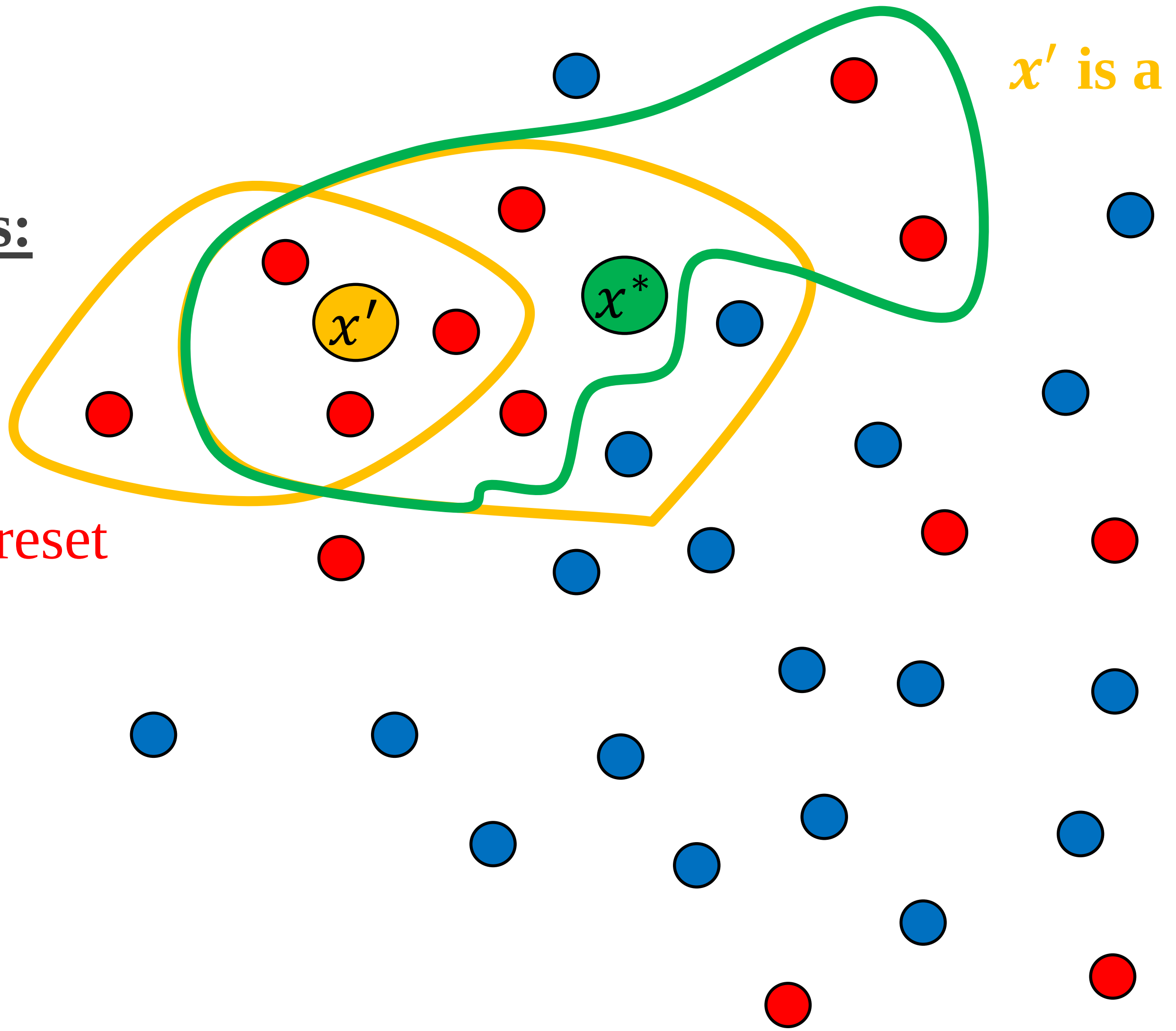
- **Proof 1:**

Definitions:

x^* is a $(\gamma, 0, 0)$-approx for P

x' is a $\left((1-\epsilon)\gamma, \epsilon, \alpha\right)$-approx for S

S is a (γ, ϵ)-coreset



$G \rightarrow \lceil (1-8\epsilon)\gamma|P| \rceil$ *points $p \in P$ with smallest dist(p, x')*

$S' \rightarrow \lceil (1-4\epsilon)\gamma|S| \rceil$ *points $s \in S$ with smallest dist(s, x')*

$S^* \rightarrow \lceil (1-\epsilon)\gamma|S| \rceil$ *points $s \in S$ with smallest dist(s, x^*)*

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon}$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $$\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\big((1-\epsilon)\gamma|S|\big)}{1-\epsilon}$$

$$\downarrow$$

$$|S^*| - 1 \leq (1-\epsilon)\gamma|S|$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \leq (1-4\epsilon)\gamma|S| \leq |S'|$

$|S^*| - 1 \leq (1-\epsilon)\gamma|S|$

$|S'| \geq \lceil (1-4\epsilon)\gamma|S| \rceil$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \leq (1-4\epsilon)\gamma|S| \leq |S'|$

$$\rightarrow |S^*| \leq \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \leq (1-4\epsilon)\gamma|S| \leq |S'|$

$$\rightarrow \boldsymbol{|S^*| \leq \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1}$$

2) $|S^*| \leq$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \leq (1-4\epsilon)\gamma|S| \leq |S'|$

$$\rightarrow |S^*| \leq \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1$$

2) $|S^*| \leq \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1$

By (1)

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \le \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \le (1-4\epsilon)\gamma|S| \le |S'|$

$$\to |\boldsymbol{S}^*| \le \frac{(\mathbf{1}-\boldsymbol{\epsilon})|\boldsymbol{S}'|}{\mathbf{1}-\mathbf{4}\boldsymbol{\epsilon}} + \mathbf{1}$$

2) $|S^*| \le \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1 \le (1+8\epsilon)(1-\epsilon)|S'| + (1+8\epsilon)\epsilon|S'|$

By (1)

1. $\frac{1}{1-4\epsilon} \sim 1 + 4\epsilon \le 1 + 8\epsilon$
2. $1 \le (1+8\epsilon)\epsilon(1-4\epsilon)\gamma|S| \le (1+8\epsilon)\epsilon|S'|$

$|S| \ge \frac{2}{\epsilon\gamma}$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \leq (1-4\epsilon)\gamma|S| \leq |S'|$

$$\rightarrow |\boldsymbol{S}^*| \leq \frac{(\mathbf{1}-\boldsymbol{\epsilon})|\boldsymbol{S}'|}{\mathbf{1}-\mathbf{4}\boldsymbol{\epsilon}} + \mathbf{1}$$

2) $|S^*| \leq \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1 \leq (1+8\epsilon)(1-\epsilon)|S'| + (1+8\epsilon)\epsilon|S'| = (1+8\epsilon)|S'|$

By (1)

1. $\frac{1}{1-4\epsilon} \sim 1 + 4\epsilon \leq 1 + 8\epsilon$
2. $1 \leq (1+8\epsilon)\epsilon(1-4\epsilon)\gamma|S| \leq (1+8\epsilon)\epsilon|S'|$

$|S| \geq \frac{2}{\epsilon\gamma}$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

1) $\frac{(1-4\epsilon)(|S^*|-1)}{1-\epsilon} \leq \frac{(1-4\epsilon)\left((1-\epsilon)\gamma|S|\right)}{1-\epsilon} \leq (1-4\epsilon)\gamma|S| \leq |S'|$

$$\rightarrow |\boldsymbol{S}^*| \leq \frac{(\mathbf{1}-\boldsymbol{\epsilon})|\boldsymbol{S}'|}{\mathbf{1}-\mathbf{4}\boldsymbol{\epsilon}} + \mathbf{1}$$

2) $|S^*| \leq \frac{(1-\epsilon)|S'|}{1-4\epsilon} + 1 \leq (1+8\epsilon)(1-\epsilon)|S'| + (1+8\epsilon)\epsilon|S'| = (1+8\epsilon)|S'|$

$$\rightarrow \frac{|\boldsymbol{S}^*|}{\mathbf{1}+\mathbf{8}\boldsymbol{\epsilon}} \leq |\boldsymbol{S}'|$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1 - 4\epsilon)\gamma|S| + 1$

$$|S'| = \lceil (\mathbf{1} - \mathbf{4}\boldsymbol{\epsilon})\boldsymbol{\gamma}|\boldsymbol{S}| \rceil$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1 - 4\epsilon)\gamma|S| + 1 \leq (1 - 4\epsilon)\gamma|S| + 2\epsilon\gamma|S|$

$|S'| = \lceil(\mathbf{1 - 4\epsilon)\gamma|S|}\rceil$ $\quad$ $1 \leq 2\epsilon\gamma|S|$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S|$

$|S'| = \lceil (\mathbf{1} - \mathbf{4}\boldsymbol{\epsilon})\boldsymbol{\gamma}|\boldsymbol{S}| \rceil$ $\quad$ $1 \leq 2\epsilon\gamma|S|$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$|S'| = \lceil (\mathbf{1} - \mathbf{4}\boldsymbol{\epsilon})\boldsymbol{\gamma}|\boldsymbol{S}| \rceil$ $\quad$ $1 \leq 2\epsilon\gamma|S|$ $\quad$ $(1-2\epsilon) \leq (1-\epsilon)^2$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow \boldsymbol{|S'| \leq (1-\epsilon)^2\gamma|S|}$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow \boldsymbol{|S'| \leq (1-\epsilon)^2\gamma|S|}$$

4)

$cost(S', x')$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow \boldsymbol{|S'| \leq (1-\epsilon)^2\gamma|S|}$$

4)

$$cost(S', x') = \sum_{s \in Closest(S, x', (1-4\epsilon)\gamma)} dist(s, x')$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow \boldsymbol{|S'| \leq (1-\epsilon)^2\gamma|S|}$$

4)

$$cost(S', x') = \sum_{s \in Closest(S, x', (1-4\epsilon)\gamma)} dist(s, x') \leq \sum_{s \in Closest(S, x', (1-\epsilon)^2\gamma)} dist(s, x')$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow \boldsymbol{|S'| \leq (1-\epsilon)^2\gamma|S|}$$

4)

$$cost(S', x') = \sum_{s \in Closest(S, x', (1-4\epsilon)\gamma)} dist(s, x') \leq \sum_{s \in Closest(S, x', (1-\epsilon)^2\gamma)} dist(s, x')$$

Def. of x'

$$\leq \alpha \cdot \sum_{s \in Closest(S, x_S^*, (1-\epsilon)\gamma)} dist(s, x_S^*)$$

x_S^* is a $(1-\epsilon)\gamma$-robust-opt

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \le (1-4\epsilon)\gamma|S| + 1 \le (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \le (1-\epsilon)^2\gamma|S|$

$$\rightarrow |S'| \le (1-\epsilon)^2\gamma|S|$$

4)

$$cost(S', x') = \sum_{s \in Closest(S, x', (1-4\epsilon)\gamma)} dist(s, x') \le \sum_{s \in Closest(S, x', (1-\epsilon)^2\gamma)} dist(s, x')$$

Def. of x'

$$\le \alpha \cdot \sum_{s \in Closest(S, x_S^*, (1-\epsilon)\gamma)} dist(s, x_S^*) \le \sum_{s \in Closest(S, x^*, (1-\epsilon)\gamma)} dist(s, x^*)$$

x_S^* is a $(1-\epsilon)\gamma$-robust-opt

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow \boldsymbol{|S'| \leq (1-\epsilon)^2\gamma|S|}$$

4)

$$cost(S', x') = \sum_{s \in Closest(S, x', (1-4\epsilon)\gamma)} dist(s, x') \leq \sum_{s \in Closest(S, x', (1-\epsilon)^2\gamma)} dist(s, x')$$

Def. of x'

$$\leq \alpha \cdot \sum_{s \in Closest(S, x_S^*, (1-\epsilon)\gamma)} dist(s, x_S^*) \leq \alpha \sum_{s \in Closest(S, x^*, (1-\epsilon)\gamma)} dist(s, x^*)$$

$$= \alpha \cdot cost(S^*, x^*)$$

x_S^* is a $(1-\epsilon)\gamma$-robust-opt

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

3) $|S'| \leq (1-4\epsilon)\gamma|S| + 1 \leq (1-4\epsilon)\gamma|S| + 2\epsilon\gamma|S| = (1-2\epsilon)\gamma|S| \leq (1-\epsilon)^2\gamma|S|$

$$\rightarrow |S'| \leq (1-\epsilon)^2\gamma|S|$$

4)

$$cost(S', x') = \sum_{s \in Closest(S, x', (1-4\epsilon)\gamma)} dist(s, x') \leq \sum_{s \in Closest(S, x', (1-\epsilon)^2\gamma)} dist(s, x')$$

$$\leq \alpha \cdot \sum_{s \in Closest(S, x_S^*, (1-\epsilon)\gamma)} dist(s, x_S^*) \leq \alpha \sum_{s \in Closest(S, x^*, (1-\epsilon)\gamma)} dist(s, x^*)$$

$$= \alpha \cdot cost(S^*, x^*)$$

$$\rightarrow cost(S', x') \leq \alpha \cdot cost(S^*, x^*)$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $(1-4\epsilon) \cdot \frac{cost(G,x')}{|G|}$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $(1-4\epsilon) \cdot \frac{cost(G, x')}{|G|} \leq \frac{cost(S', x')}{|S'|}$

↓

S is a (γ, ϵ)-coreset

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $(1-4\epsilon) \cdot \frac{cost(G,x')}{|G|} \leq \frac{cost(S',x')}{|S'|} \leq \frac{\alpha \cdot cost(S^*,x^*)}{|S'|}$

S is a (γ, ϵ)-coreset By (4)

$$\boldsymbol{cost(S', x') \leq \alpha \cdot cost(S^*, x^*)}$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $$(1-4\epsilon)\cdot\frac{cost(G,x')}{|G|} \le \frac{cost(S',x')}{|S'|} \le \frac{\alpha\cdot cost(S^*,x^*)}{|S'|} \le \frac{\alpha(1+8\epsilon)cost(S^*,x^*)}{|S^*|}$$

S is a (γ, ϵ)-coreset By (4) By (2)

$$\frac{|S^*|}{1+8\epsilon} \le |S'|$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $$(1-4\epsilon)\cdot\frac{cost(G,x')}{|G|} \leq \frac{cost(S',x')}{|S'|} \leq \frac{\alpha \cdot cost(S^*,x^*)}{|S'|} \leq \frac{\alpha(1+8\epsilon)cost(S^*,x^*)}{|S^*|}$$

$$\leq \frac{\alpha(1+8\epsilon)(1+\epsilon)cost(P^*,x^*)}{|P^*|}$$

↓

S is a (γ, ϵ)-coreset

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $$(1-4\epsilon)\cdot\frac{cost(G,x')}{|G|} \le \frac{cost(S',x')}{|S'|} \le \frac{\alpha\cdot cost(S^*,x^*)}{|S'|} \le \frac{\alpha(1+8\epsilon)cost(S^*,x^*)}{|S^*|}$$

$$\le \frac{\alpha(1+8\epsilon)(1+\epsilon)cost(P^*,x^*)}{|P^*|} \le \frac{\alpha(1+10\epsilon)cost(P^*,x^*)}{|P^*|}$$

S is a (γ, ϵ)-coreset

$(1+8\epsilon)(1+\epsilon) \le (1+10\epsilon)$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $(1-4\epsilon) \cdot \frac{cost(G,x')}{|G|} \leq \frac{cost(S',x')}{|S'|} \leq \frac{\alpha \cdot cost(S^*,x^*)}{|S'|} \leq \frac{\alpha(1+8\epsilon)cost(S^*,x^*)}{|S^*|}$

$$\leq \frac{\alpha(1+8\epsilon)(1+\epsilon)cost(P^*,x^*)}{|P^*|} \leq \frac{\alpha(1+10\epsilon)cost(P^*,x^*)}{|P^*|}$$

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

5) $(1-4\epsilon) \cdot \frac{cost(G,x')}{|G|} \le \frac{cost(S',x')}{|S'|} \le \frac{\alpha \cdot cost(S^*,x^*)}{|S'|} \le \frac{\alpha(1+8\epsilon)cost(S^*,x^*)}{|S^*|}$

$$\le \frac{\alpha(1+8\epsilon)(1+\epsilon)cost(P^*,x^*)}{|P^*|} \le \frac{\alpha(1+10\epsilon)cost(P^*,x^*)}{|P^*|}$$

$\rightarrow$ **x' is a $\left(\gamma, 4\epsilon, \alpha(1+10\epsilon)\right)$-approximation of P.**

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Proof 1:**

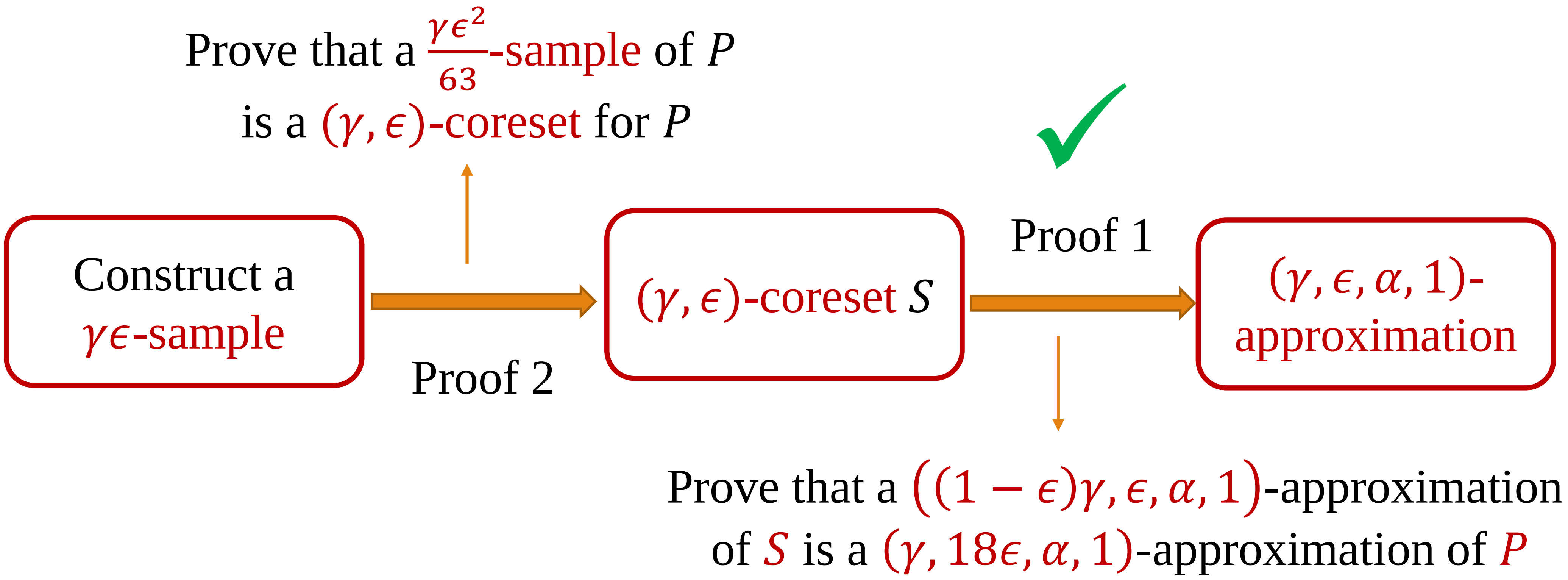
5) $(1-4\epsilon)\cdot\frac{cost(G,x')}{|G|} \le \frac{cost(S',x')}{|S'|} \le \frac{\alpha\cdot cost(S^*,x^*)}{|S'|} \le \frac{\alpha(1+8\epsilon)cost(S^*,x^*)}{|S^*|}$

$$\le \frac{\alpha(1+8\epsilon)(1+\epsilon)cost(P^*,x^*)}{|P^*|} \le \frac{\alpha(1+10\epsilon)cost(P^*,x^*)}{|P^*|}$$

$\rightarrow$ **x' is a $\left(\gamma, 4\epsilon, \alpha(1+10\epsilon)\right)$-approximation of P.**

➢ If we leave in G only the closest $\frac{|G|}{1+10\epsilon} > (1-18\epsilon)\gamma|P|$ points, then $cost(G,x')$ is reduced by a factor of at least $\frac{1}{1+10\epsilon}$ $\rightarrow$ x' is a $(\gamma, 18\epsilon, \alpha)$-approximation for P. ∎

Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation



Computing a $(\gamma, \epsilon, \alpha, \beta)$-approximation

- **Claim 2:**

Let P be a set of points and $dist: P \times X \to [0, \infty)$. Let $\epsilon \in \left(0, \frac{1}{4}\right), \gamma \in (0,1]$. Let S be an $\left(\frac{\epsilon^2 \gamma}{63}\right)$-approximation of P such that $|P|, |S| \geq \frac{5}{\epsilon^2 \gamma}$. Then S is a (γ, ϵ)-coreset of P.

- **Proof 2:**

We will not prove this Claim.

Summary

We want to solve:

(α, β)-approx for P

Summary

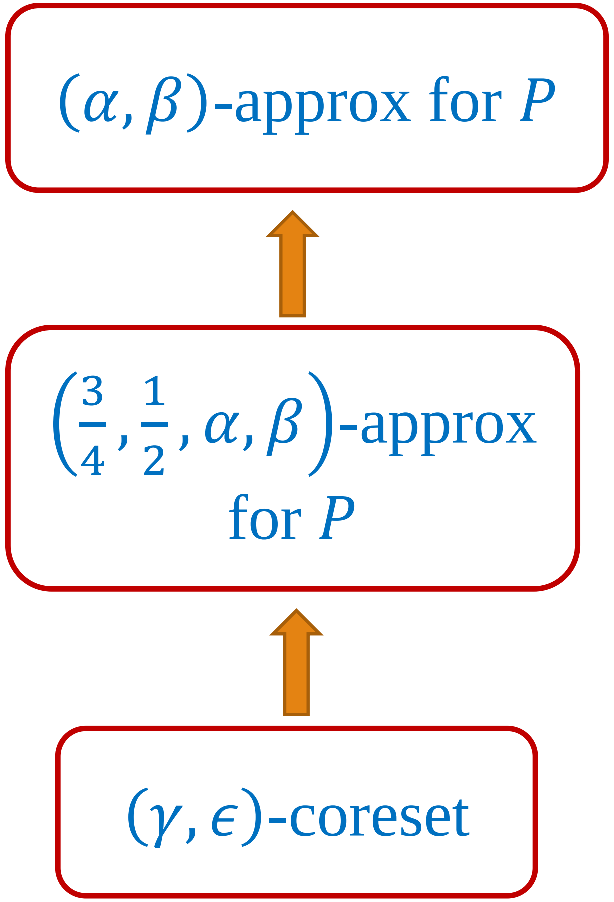
We want to solve:

(α, β)-approx for P

$\left(\frac{3}{4}, \frac{1}{2}, \alpha, \beta\right)$-approx for P

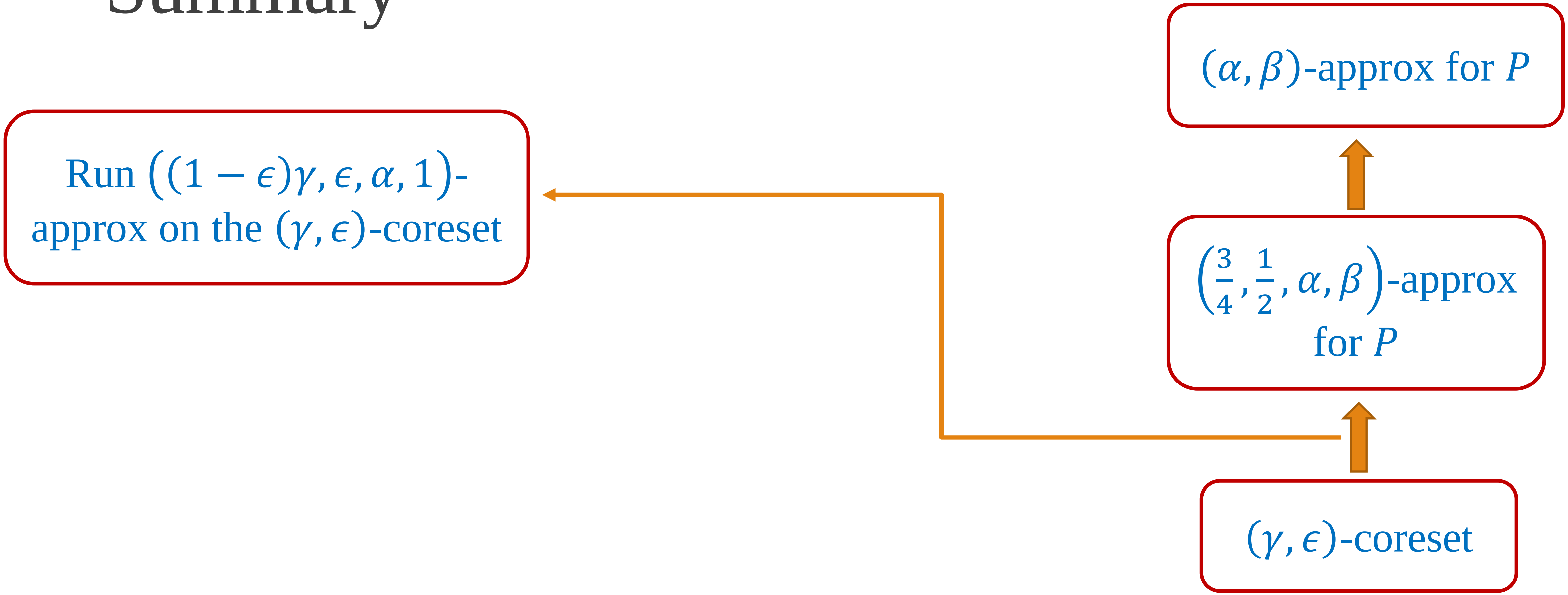
Summary

We want to solve:



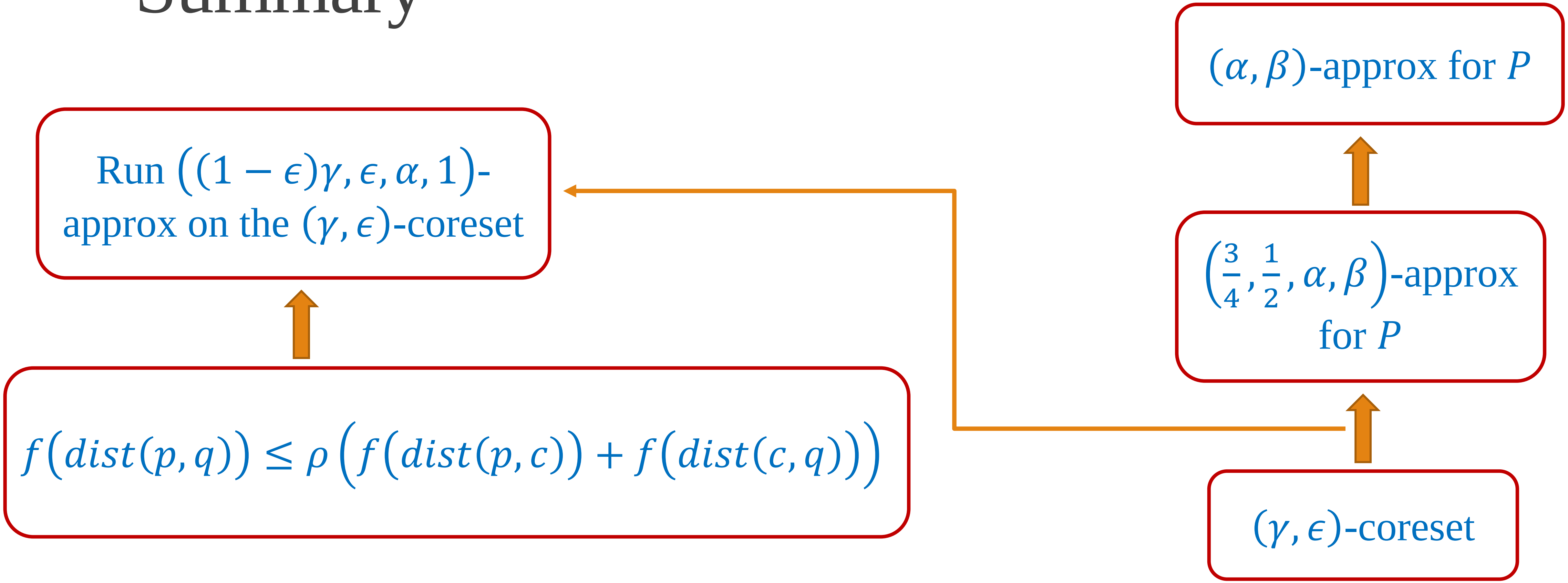
Summary

We want to solve:

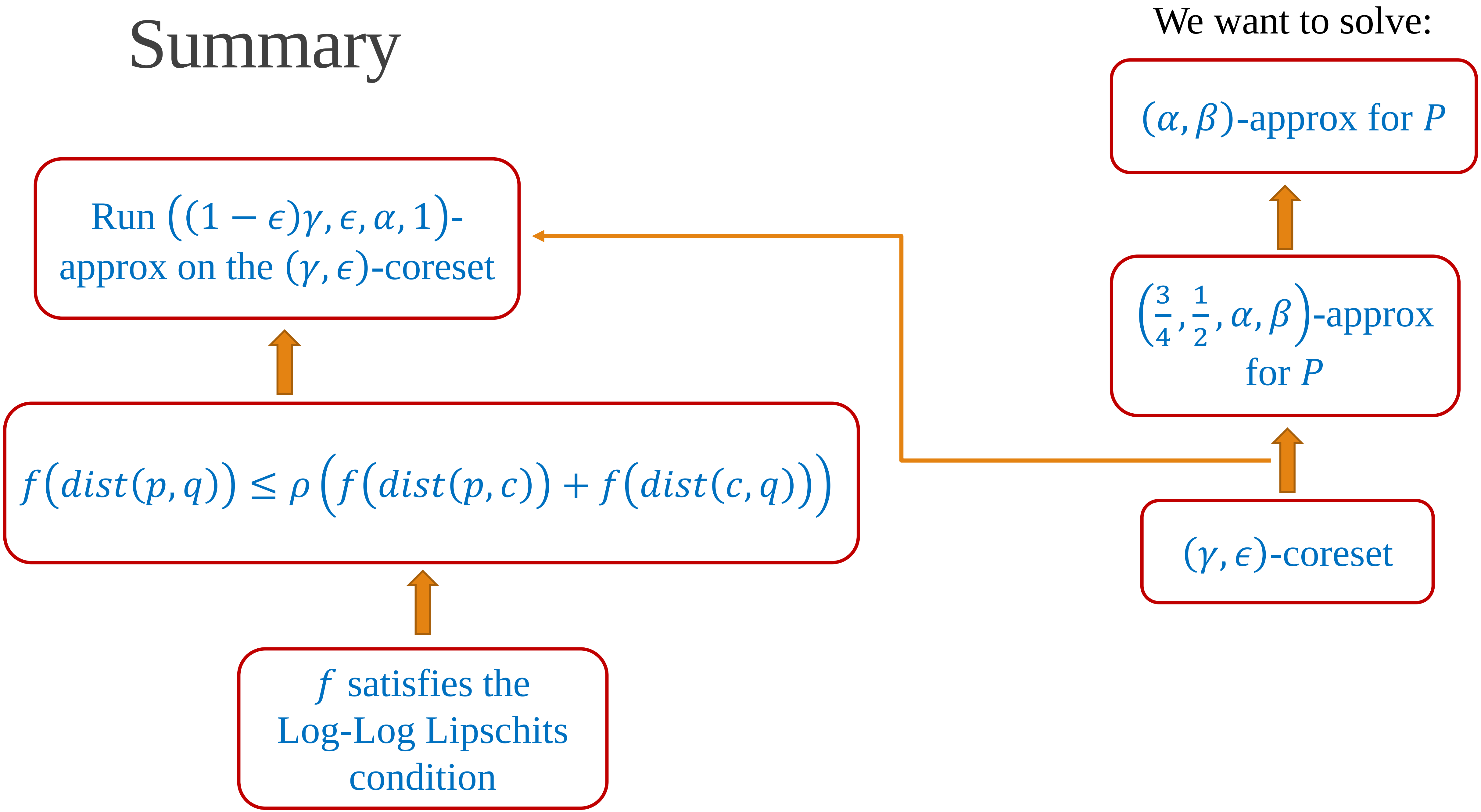


Summary

We want to solve:

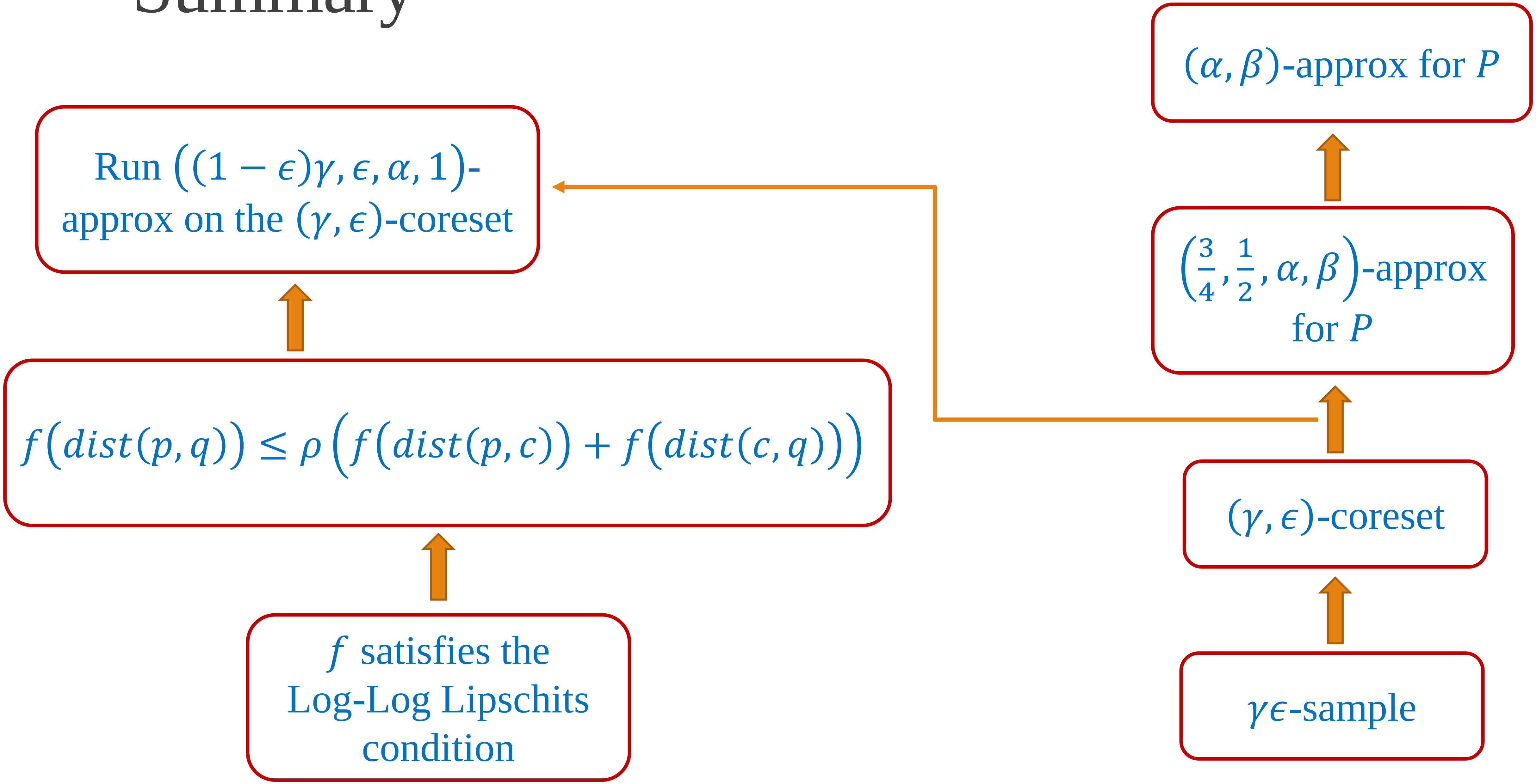


Summary


We want to solve:
(α, β)-approx for P
$\left(\frac{3}{4}, \frac{1}{2}, \alpha, \beta\right)$-approx for P
(γ, ϵ)-coreset
Run $\big((1-\epsilon)\gamma, \epsilon, \alpha, 1\big)$-approx on the (γ, ϵ)-coreset
$f\big(dist(p,q)\big) \le \rho\Big(f\big(dist(p,c)\big) + f\big(dist(c,q)\big)\Big)$
f satisfies the Log-Log Lipschits condition

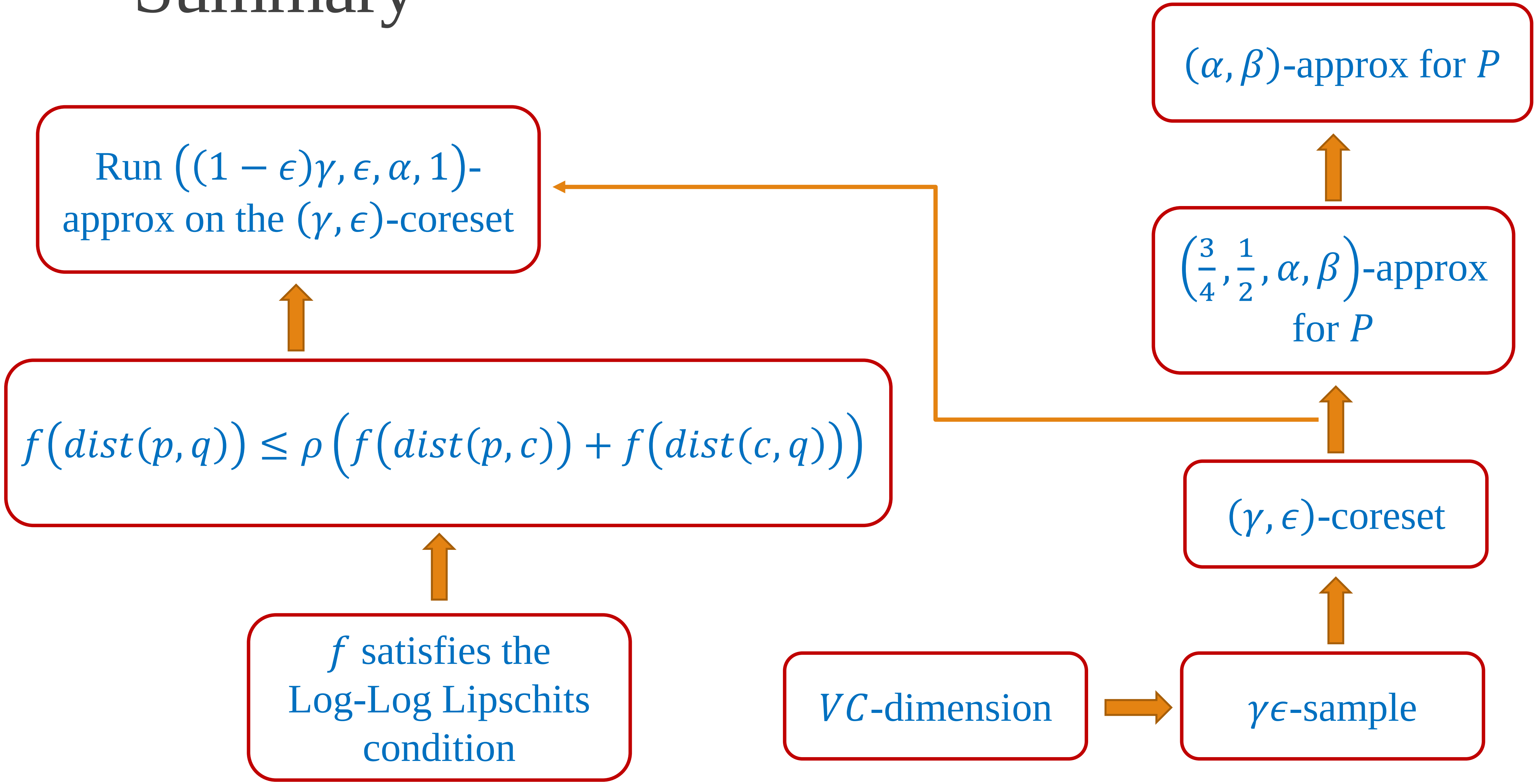
Summary

We want to solve:



Summary

We want to solve:



Summary

We want to solve:

(α, β)-approx for P

$\left(\frac{3}{4}, \frac{1}{2}, \alpha, \beta\right)$-approx for P

(γ, ϵ)-coreset

$\gamma\epsilon$-sample

Will learn soon

***VC*-dimension**

Run $\big((1-\epsilon)\gamma, \epsilon, \alpha, 1\big)$-approx on the (γ, ϵ)-coreset

$$f\big(dist(p,q)\big) \leq \rho\Big(f\big(dist(p,c)\big) + f\big(dist(c,q)\big)\Big)$$

***f* satisfies the Log-Log Lipschits condition**