Tight FPT Approximations for k-Median and k-Means

Jason Li

Joint work with
Vincent Cohen-Addad, Anupam Gupta, Amit Kumar, Euiwoong Lee

ICALP 2019

k-median problem

- Metric space (V, d)
- Clients $C \subseteq V$, Facilities $\mathbb{F} \subseteq V$
- Find set F of k facilities minimizing

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

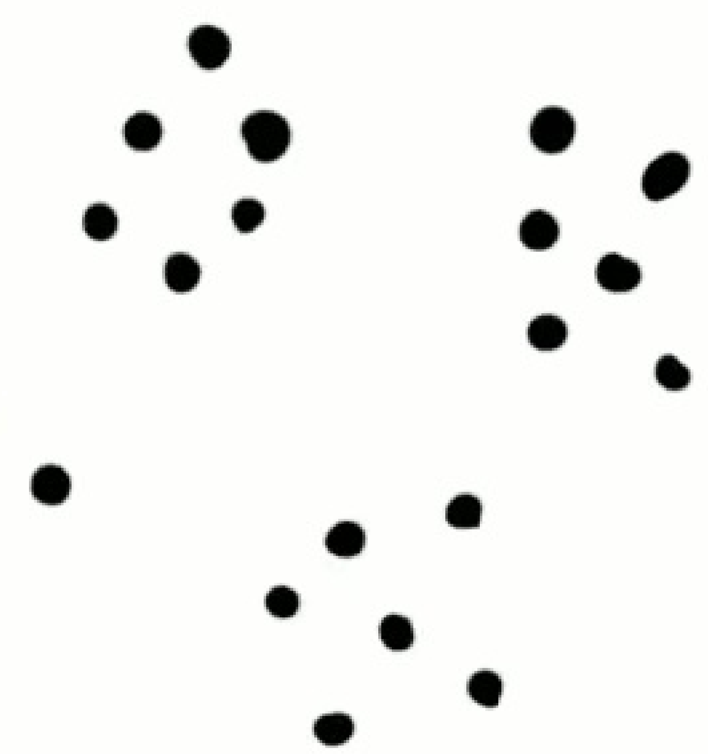
$$= \min_{f \in F} d(v, f)$$

k-median problem

- Metric space (V, d)
- Clients $C \subseteq V$, Facilities $\mathbb{F} \subseteq V$
- Find set F of k facilities minimizing

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

- Clustering:



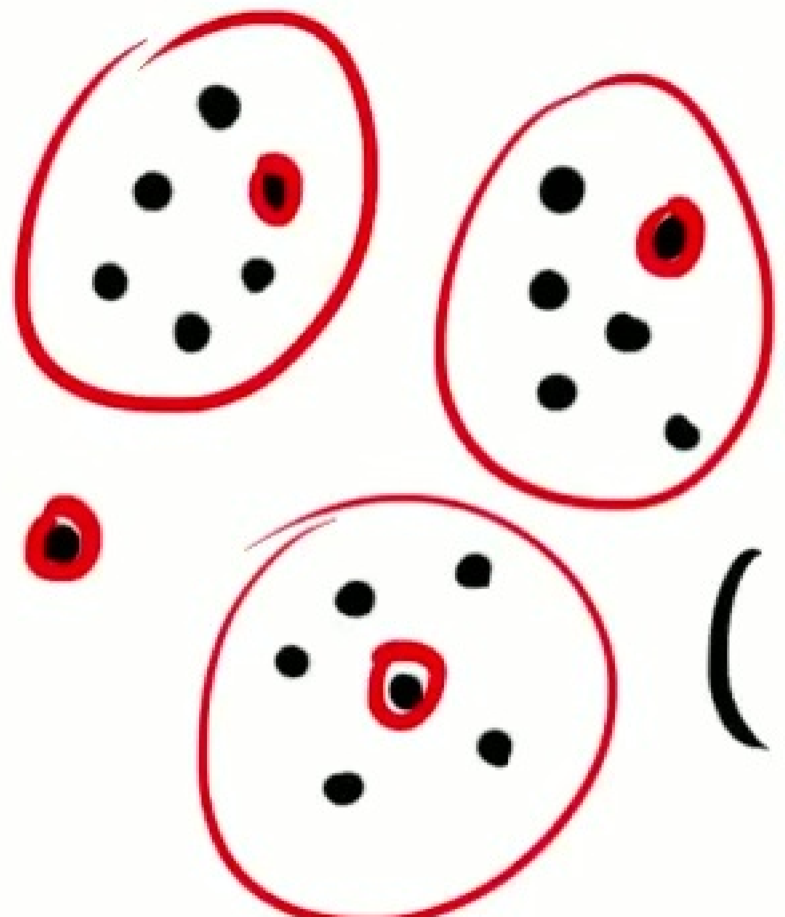
k-median problem

- Metric space (V, d)
- Clients $C \subseteq V$, Facilities $\mathbb{F} \subseteq V$
- Find set F of k facilities minimizing

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

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

- Clustering:



$k=4$

$(C = \mathbb{F})$

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v,F)$$

- NP-hard (dominating set $\leq$ k-median)

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v,F)$$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v,F)$$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]
- $(1+2/e)$-apx hard from set cover [Guha+Khuller]

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v,F)$$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]
- $(1+2/e)$-apx hard from set cover [Guha + Khuller]
- FPT? Runtime $f(k)n^{O(1)} \ll$ trivial n^k

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v,F)$$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]
- $(1+2/e)$-apx hard from set cover [Guha+Khuller]
- FPT? Runtime $f(k)n^{O(1)} \ll$ trivial n^k
  - W[2]-hard (k-dominating set)

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v, F)$$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]
- $(1+2/e)$-apx hard from set cover [Guha + Khuller]
- FPT? Runtime $f(k)n^{O(1)} \ll$ trivial n^k
  - W[2]-hard (k-dominating set)
  - $(1+2/e+\varepsilon)$-apx in FPT time

k-median problem

$$\min_{|F|=k} \sum_{v \in C} d(v,F)$$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]
- $(1+2/e)$-apx hard from set cover [Guha+Khuller]
- FPT? Runtime $f(k)n^{O(1)} \ll$ trivial n^k
  - W[2]-hard (k-dominating set)
  - $(1+2/e+\varepsilon)$-apx in FPT time
  - $(1+2/e-\varepsilon)$-apx hard, even in FPT time

k-median problem

$\min_{|F|=k} \sum_{v \in C} d(v,F)$

- NP-hard (dominating set $\leq$ k-median)
- $(3+\varepsilon)$-apx local search [Arya et al]
- 2.611-apx [Byrka et al]
- $(1+2/e)$-apx hard from set cover [Guha + Khuller]
- FPT? Runtime $f(k)n^{O(1)} \ll$ trivial n^k
  - W[2]-hard (k-dominating set)
  - $\boxed{(1+2/e+\varepsilon)\text{-apx in FPT time}}$ ← this talk
  - $(1+2/e-\varepsilon)$-apx hard, even in FPT time

Core-sets

Def[Coreset]: a subset $S \subseteq C$ with weights $w(v): v \in S$

s.t.

$\forall F$ set of k facilities:

$$\sum_{v \in S} w(v) \cdot d(v,F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v,F)$$

Core-sets

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

$\forall F$ set of k facilities:

$$\sum_{v \in S} w(v) \cdot d(v, F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v, F)$$

$$\Longrightarrow \min_{|F|=k} \sum_{v \in S} w(v)\, d(v, F) \in (1 \pm \varepsilon) \min_{|F|=k} \sum_{v \in C} d(v, F)$$

Core-sets

Def[Coreset]: a subset $S \subseteq C$ with weights $w(v): v \in S$

s.t.

$\forall F$ set of k facilities:

$$\sum_{v \in S} w(v) \cdot d(v, F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v, F)$$

$$\Rightarrow \min_{|F|=k} \sum_{v \in S} w(v)\, d(v, F) \in (1 \pm \varepsilon) \min_{|F|=k} \sum_{v \in C} d(v, F)$$

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

Core-sets

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

$\forall F$ set of k facilities:

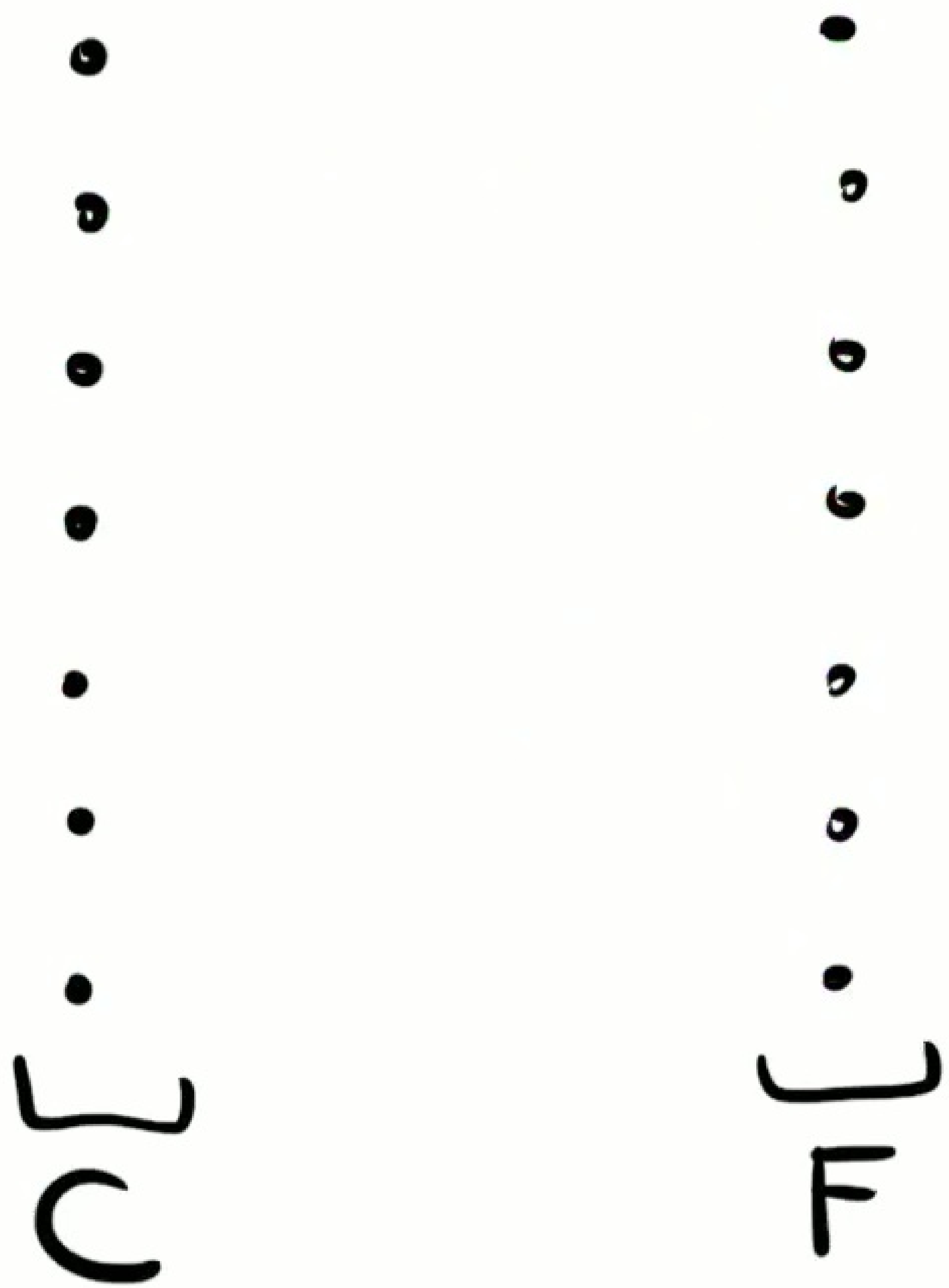
$$\sum_{v \in S} w(v) \cdot d(v,F) \in (1 \pm \varepsilon) \sum_{v \in C} d(v,F)$$

$$\Rightarrow \min_{|F|=k} \sum_{v \in S} w(v)\, d(v,F) \in (1 \pm \varepsilon) \min_{|F|=k} \sum_{v \in C} d(v,F)$$

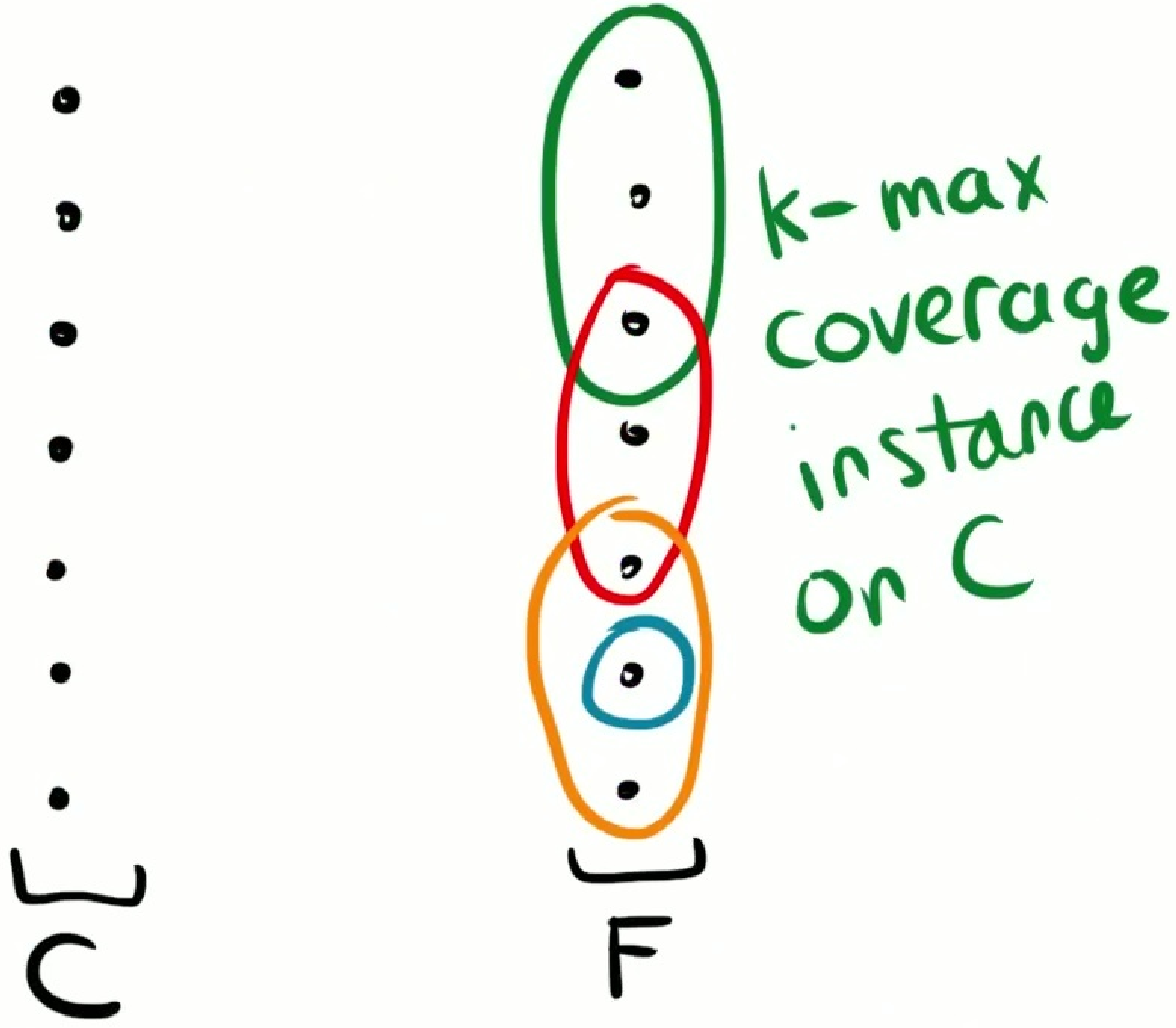
Thm[Chen; Feldman-Langberg] $\exists$ coreset size $\mathrm{poly}(\varepsilon^{-1} k \log n)$.

At loss of $(1+\varepsilon)$, can assume $\mathrm{poly}(\varepsilon^{-1} k \log n)$ (weighted) clients

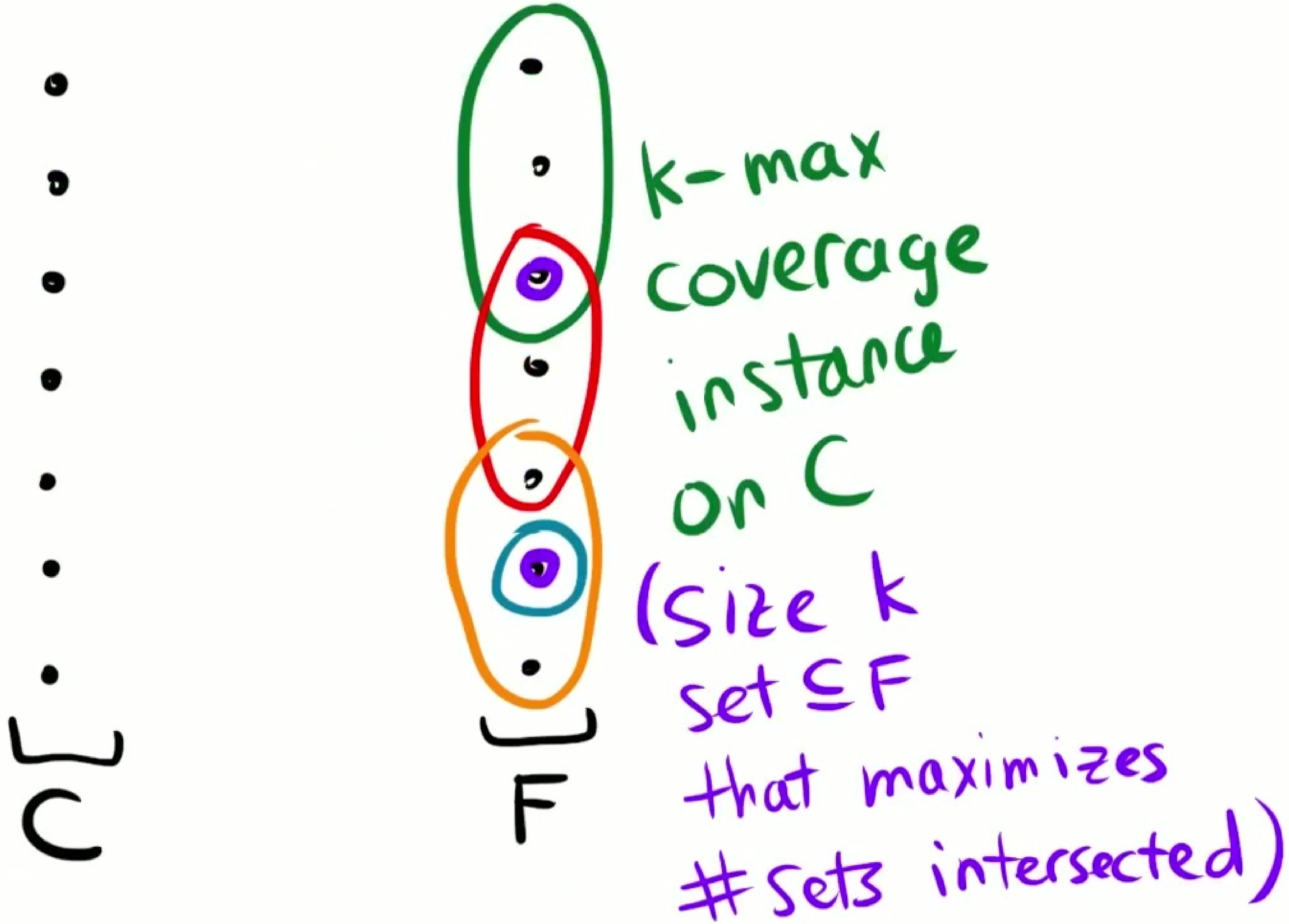
Insight from hardness: why 1+2/e?



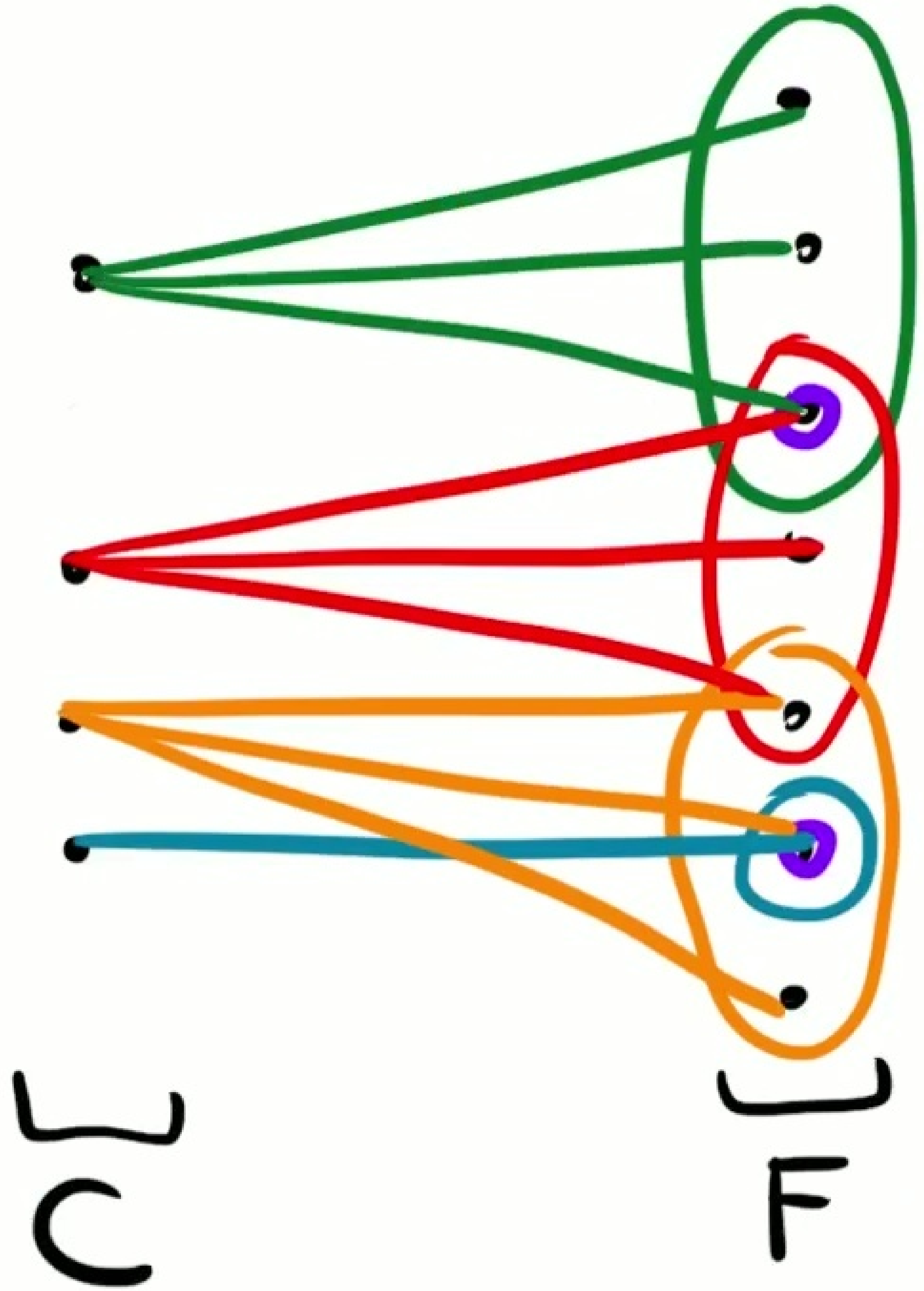
Insight from hardness: why 1+2/e?



Insight from hardness: why 1+2/e?



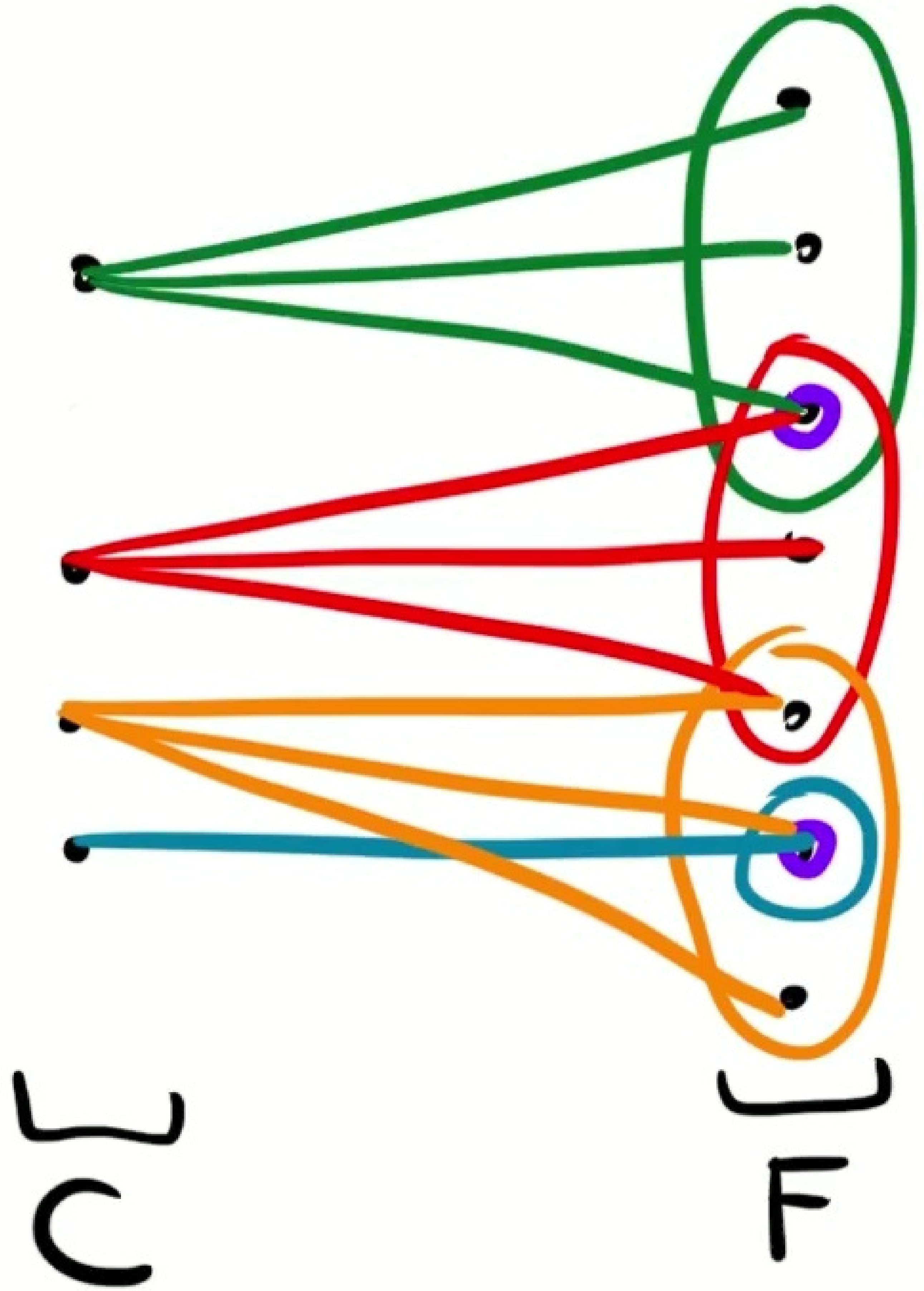
Insight from hardness: why 1+2/e?



k-max coverage instance on C

(size k set ⊆ F that maximizes #sets intersected)

Insight from hardness: why 1+2/e?

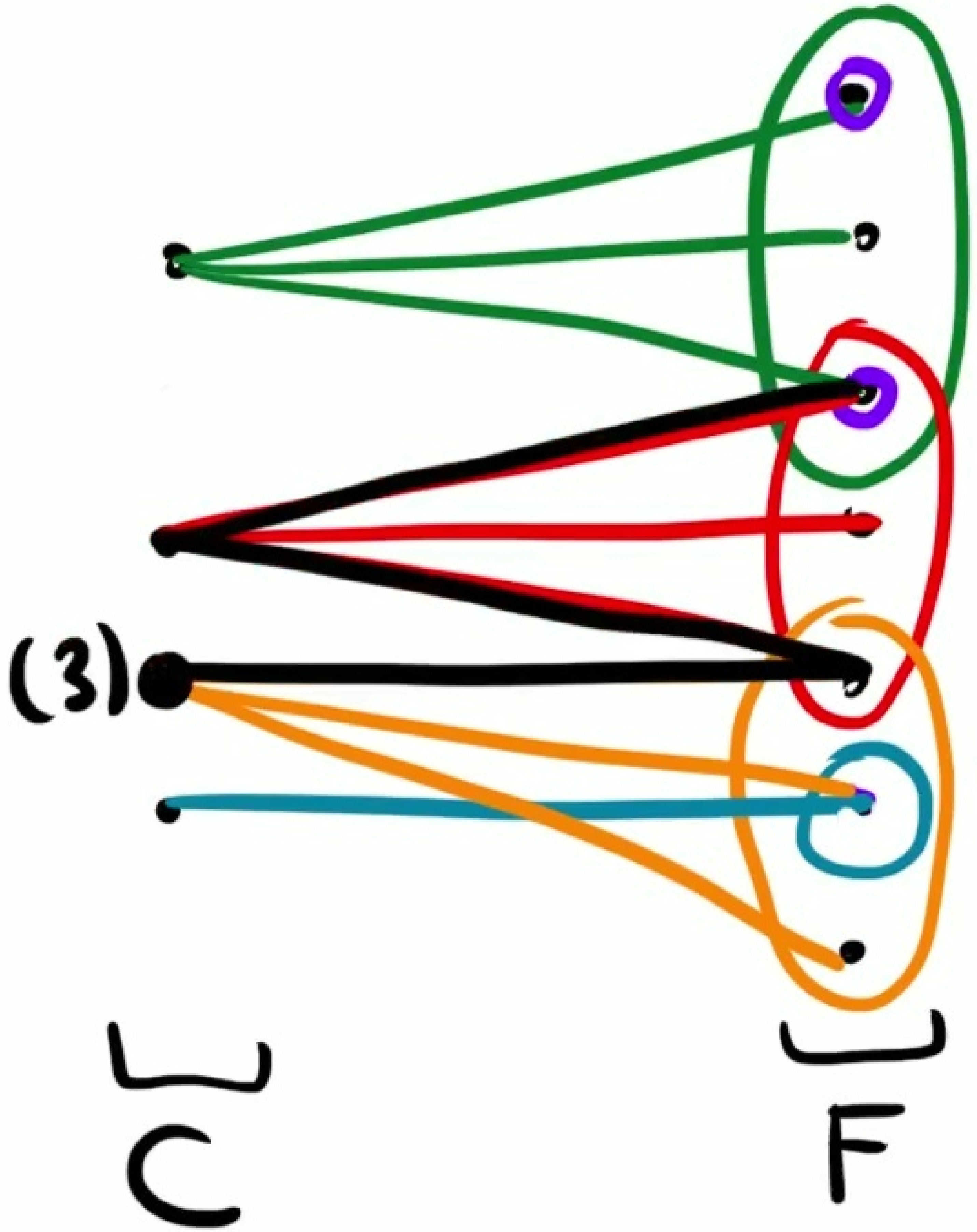


k-max coverage instance on C

(size k set $\subseteq F$ that maximizes #sets intersected)

- set covered $\Longleftrightarrow$ client dist $= 1$

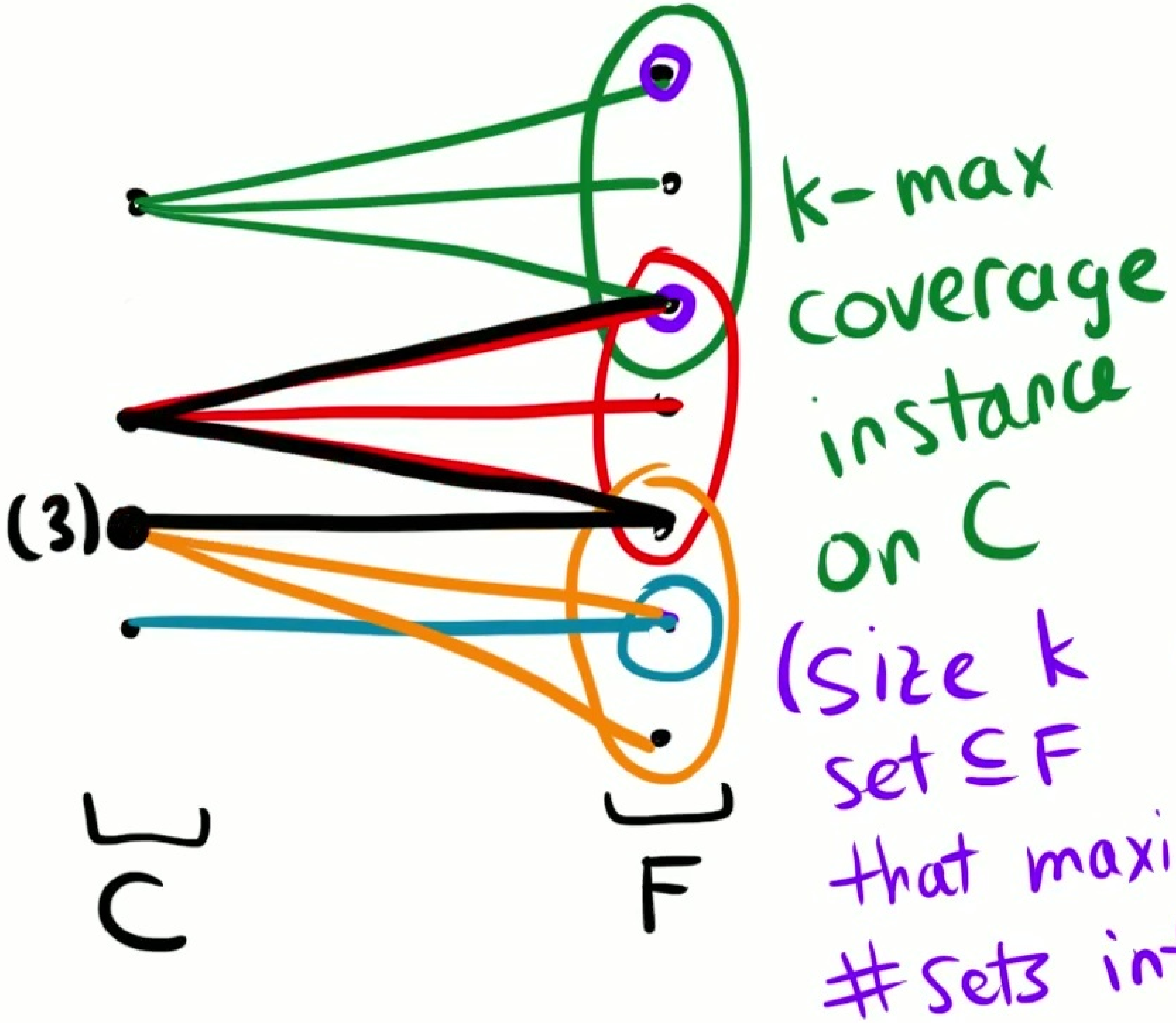
Insight from hardness: why 1+2/e?



k-max coverage instance on C
(size k set $\subseteq F$ that maximizes #sets intersected)

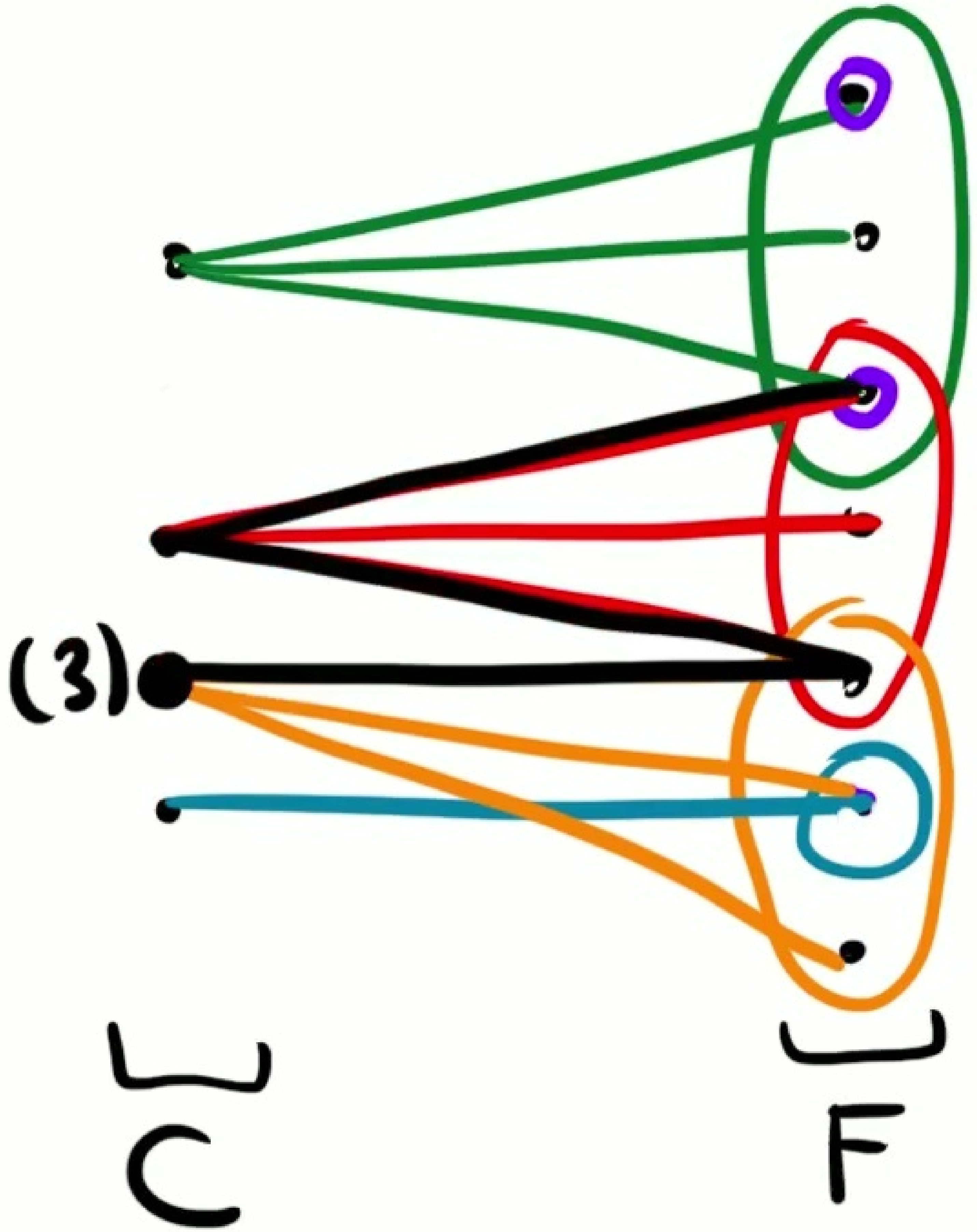
- set covered $\Longleftrightarrow$ client dist $= 1$
- set uncovered $\Longleftrightarrow$ client dist ≥ 3

Insight from hardness: why 1+2/e?



- set covered $\Longleftrightarrow$ client dist = 1
- set uncovered $\Longleftrightarrow$ client dist ≥ 3
- k-set-coverage $(1-1/e)$-apx hard

Insight from hardness: why 1+2/e?



k-max coverage instance on C

(size k set ⊆ F that maximizes # sets intersected)

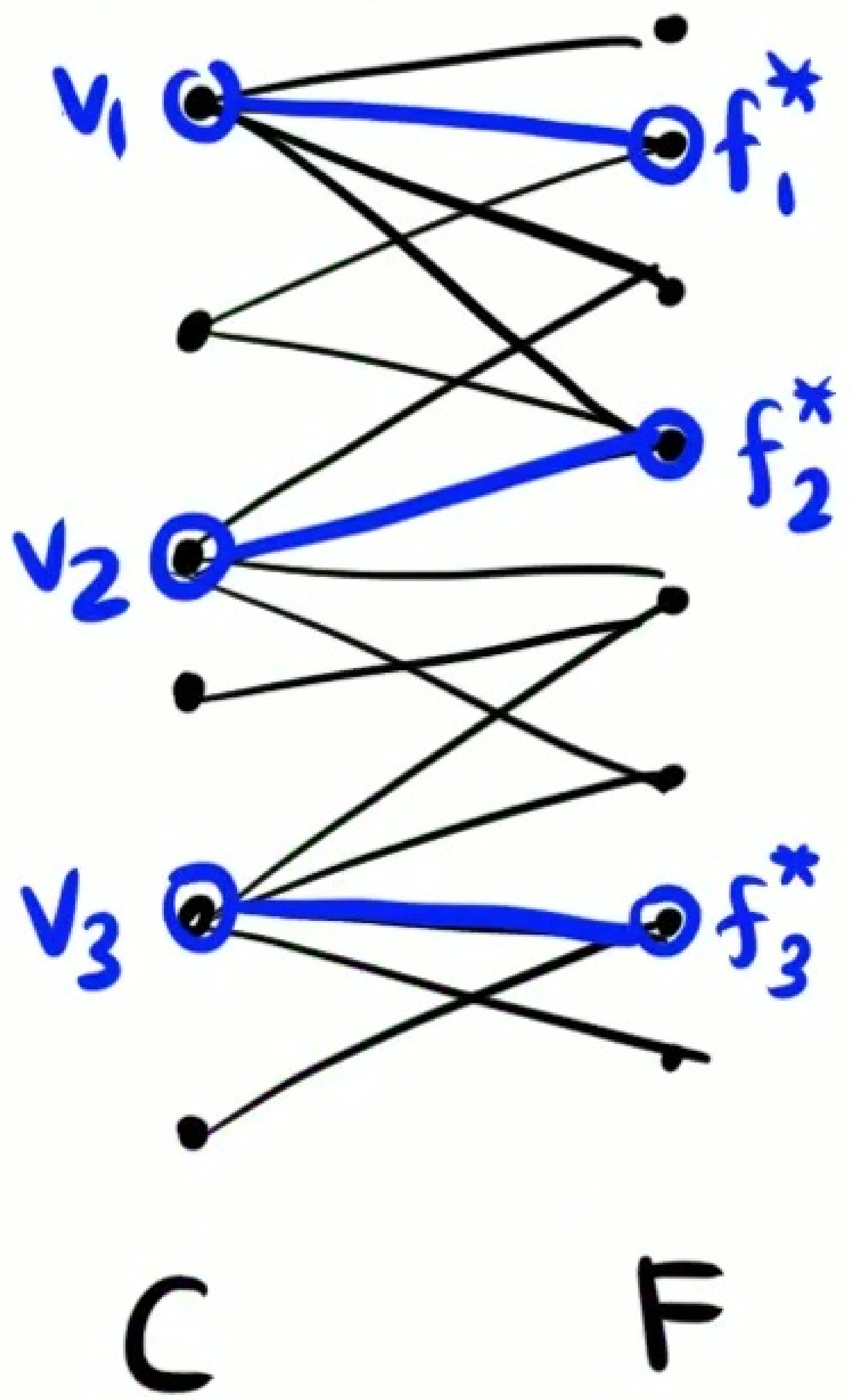
- set covered $\Longleftrightarrow$ client dist = 1
- set uncovered $\Longleftrightarrow$ client dist ≥ 3
- k-set-coverage $(1-1/e)$-apx hard

$\Longrightarrow$ k-median $(1-1/e)\cdot 1+\frac{1}{e}\cdot 3$

$=(1+2/e)$-apx hard

Transformation to Algorithm

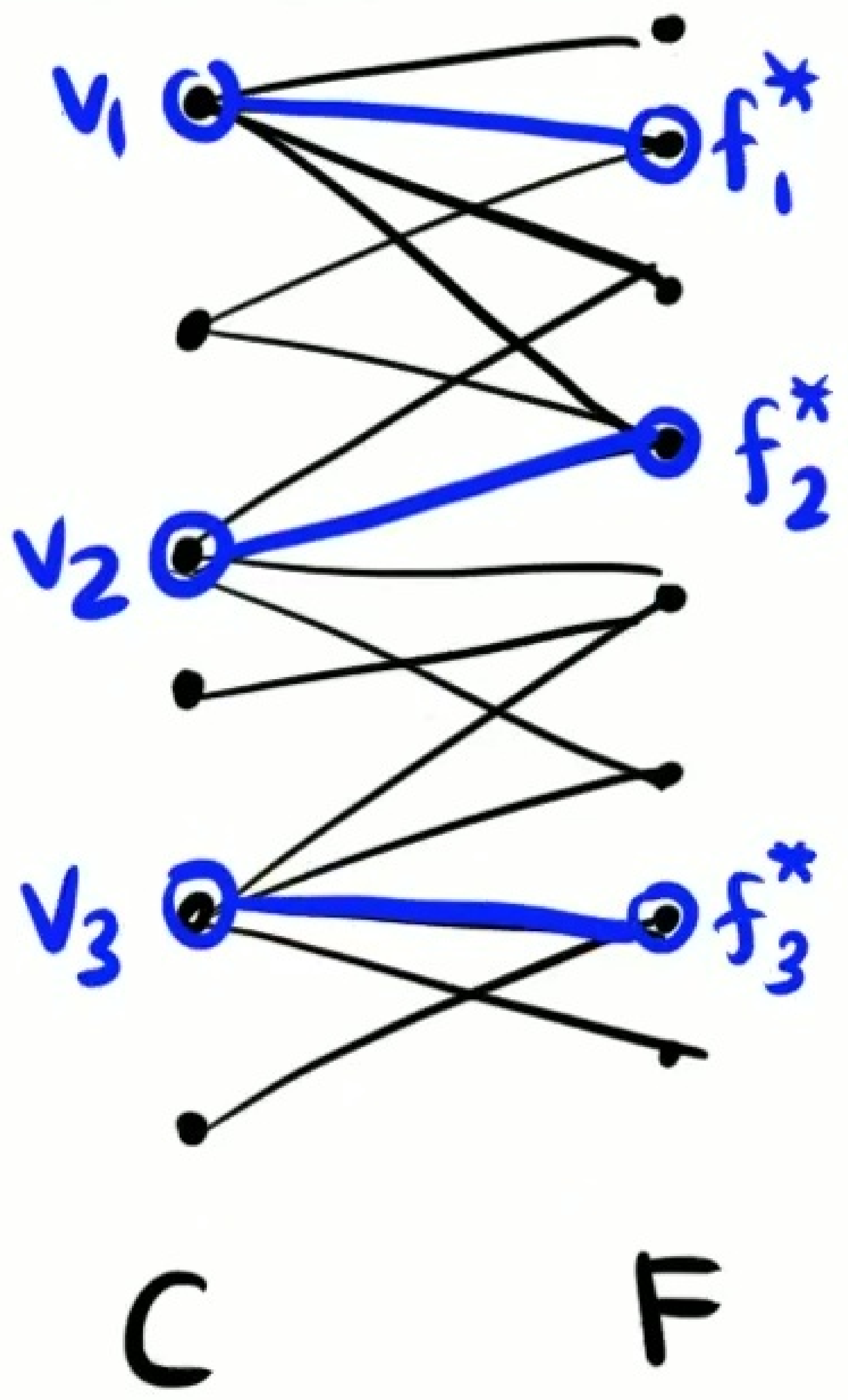
- Guess k centers $v_1 \dots v_k$
  s.t. $f_i^* \sim v_i \ \forall i$
  $[\mathrm{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)



Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$
  s.t. $f_i^* \sim v_i \ \forall i$
  $[\mathrm{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

- Optimize ________
  subject to $f_i \in N(v_i) \ \forall i$

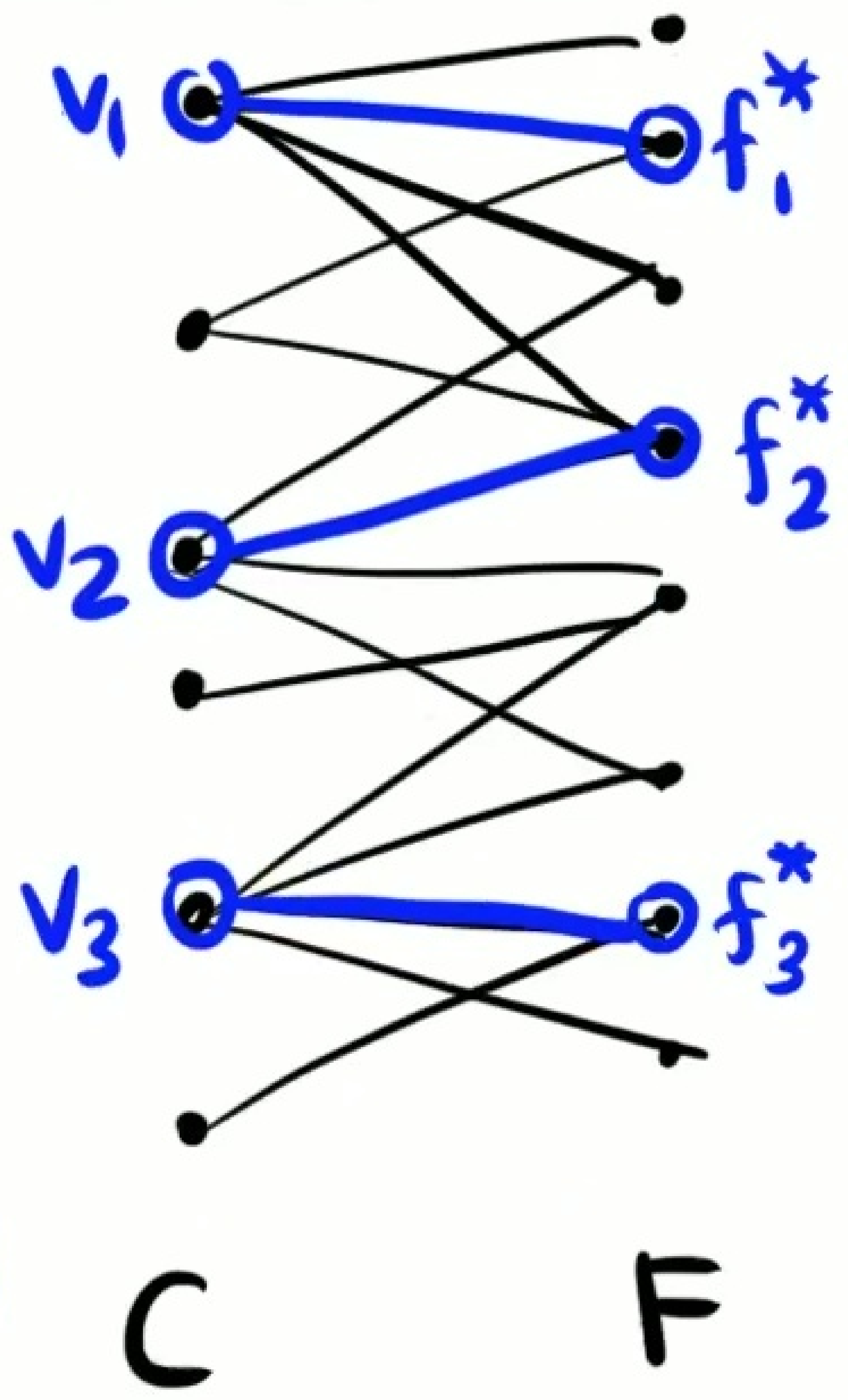


Transformation to Algorithm

- Guess k centers $v_1 \ldots v_k$
  s.t. $f_i^* \sim v_i \ \forall i$
  $[\mathrm{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

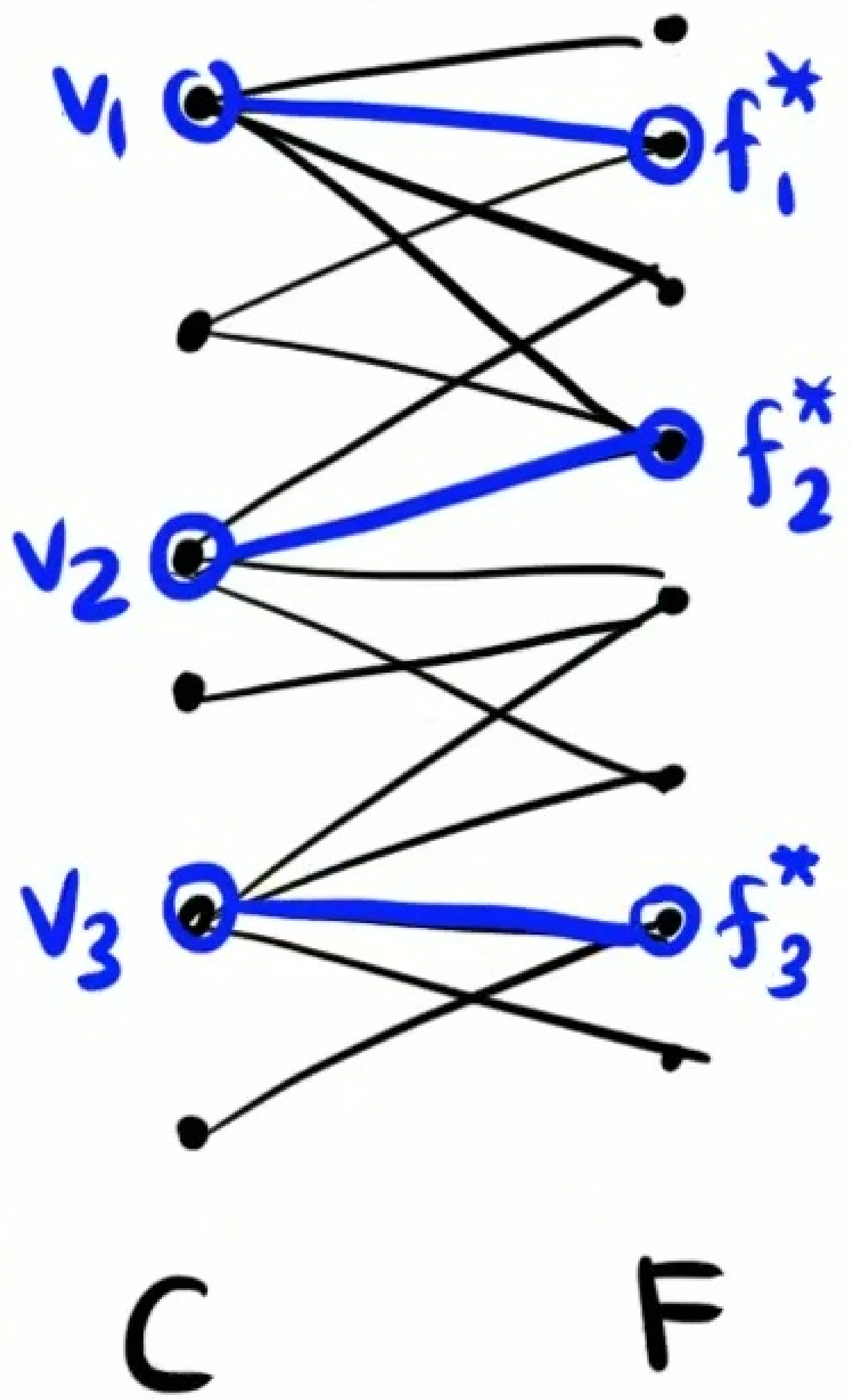
- Optimize ________
  subject to $f_i \in N(v_i) \ \forall i$

- Claim: $d(v,F) \leq 3\, d(v,F^*)$
  $\forall v \in C.$



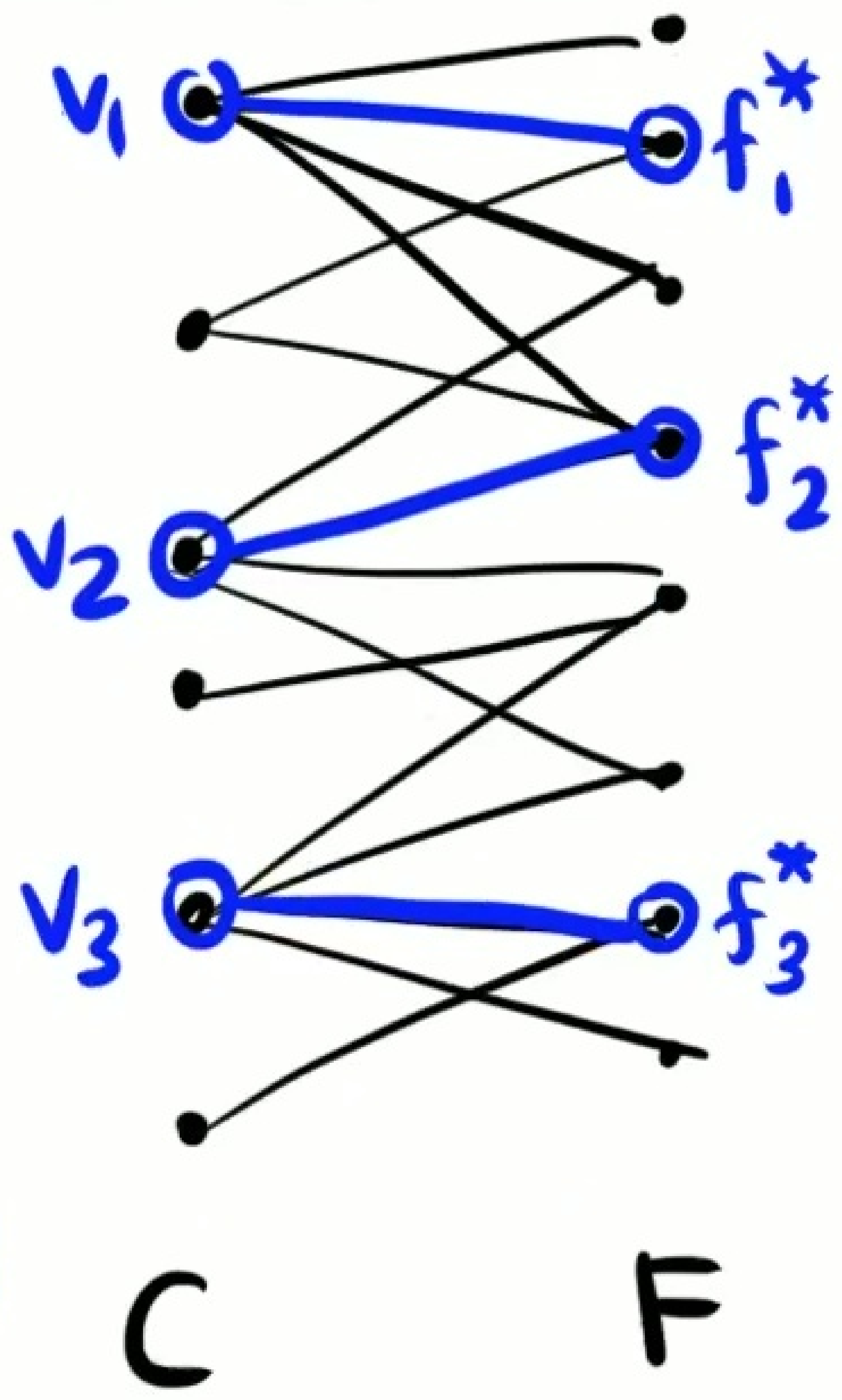
Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$ s.t. $f_i^* \sim v_i \ \forall i$
  $[\mathrm{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)
- Optimize ________ subject to $f_i \in N(v_i) \ \forall i$
- Claim: $d(v,F) \leq 3\, d(v,F^*)$ $\forall v \in C$.
- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$



Transformation to Algorithm

- Guess k centers $v_1 \ldots v_k$
  s.t. $f_i^* \sim v_i \ \forall i$
  $[\mathrm{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)



- Optimize __________
  subject to $f_i \in N(v_i) \ \forall i$

- Claim: $d(v,F) \le 3\, d(v,F^*) \ \forall v \in C$.

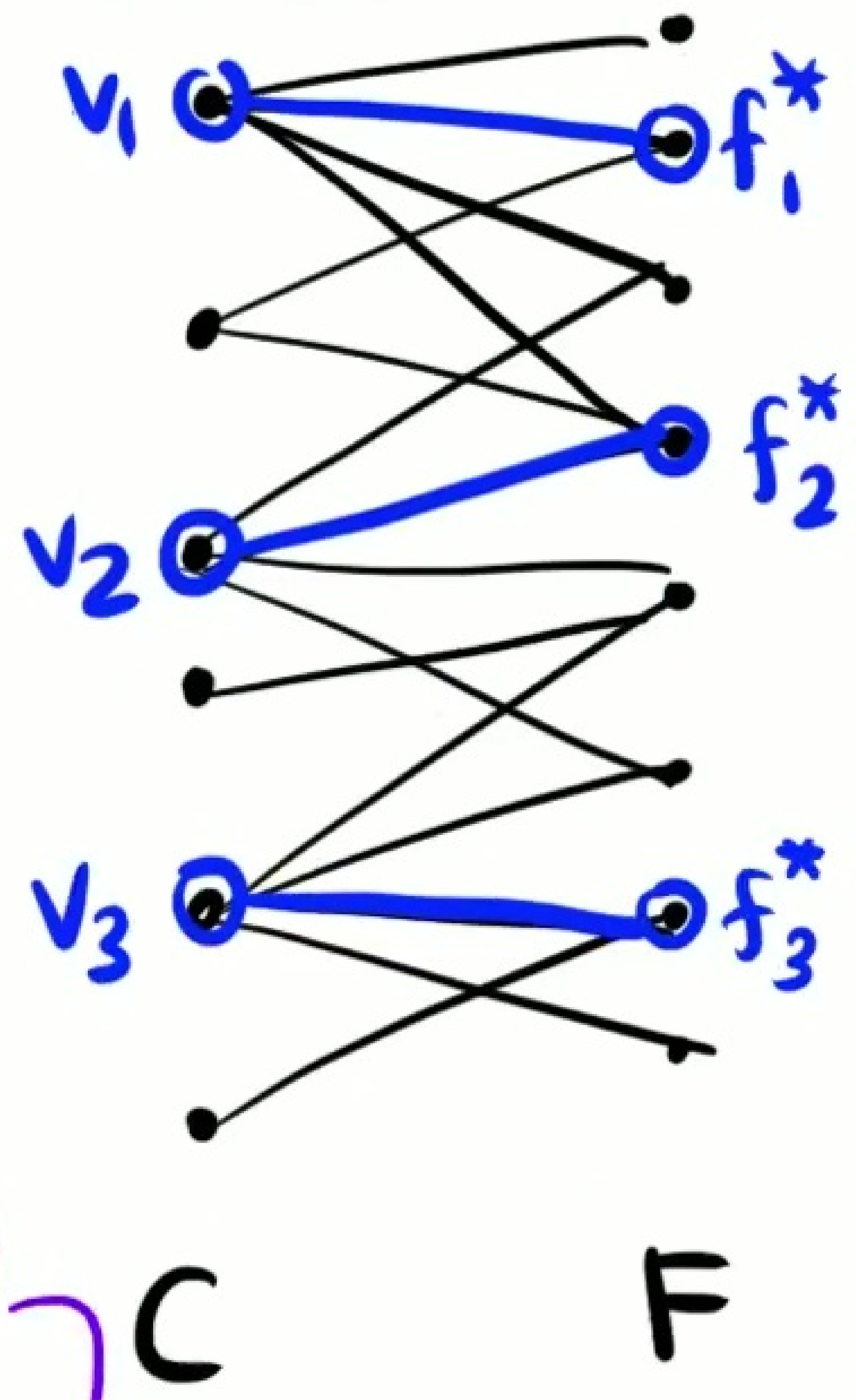
- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$
  $f_i, f_i^* \in N(v_i) \Rightarrow d(f_i^*, f_i) \le 2$
  $d(v,f_i) \le d(v,f_i^*) + d(f_i^*, f_i) \le d(v,F^*) + 2 \le 3\, d(v,F^*)$.

Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$
  s.t. $f_i^* \sim v_i \ \forall i$
  $[\text{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

- Optimize ________
  subject to $f_i \in N(v_i) \ \forall i$

- Claim: $d(v,F) \leq 3\, d(v,F^*)$
  $\forall v \in C$.



3-approx.
How to get $(1+2/e)$?

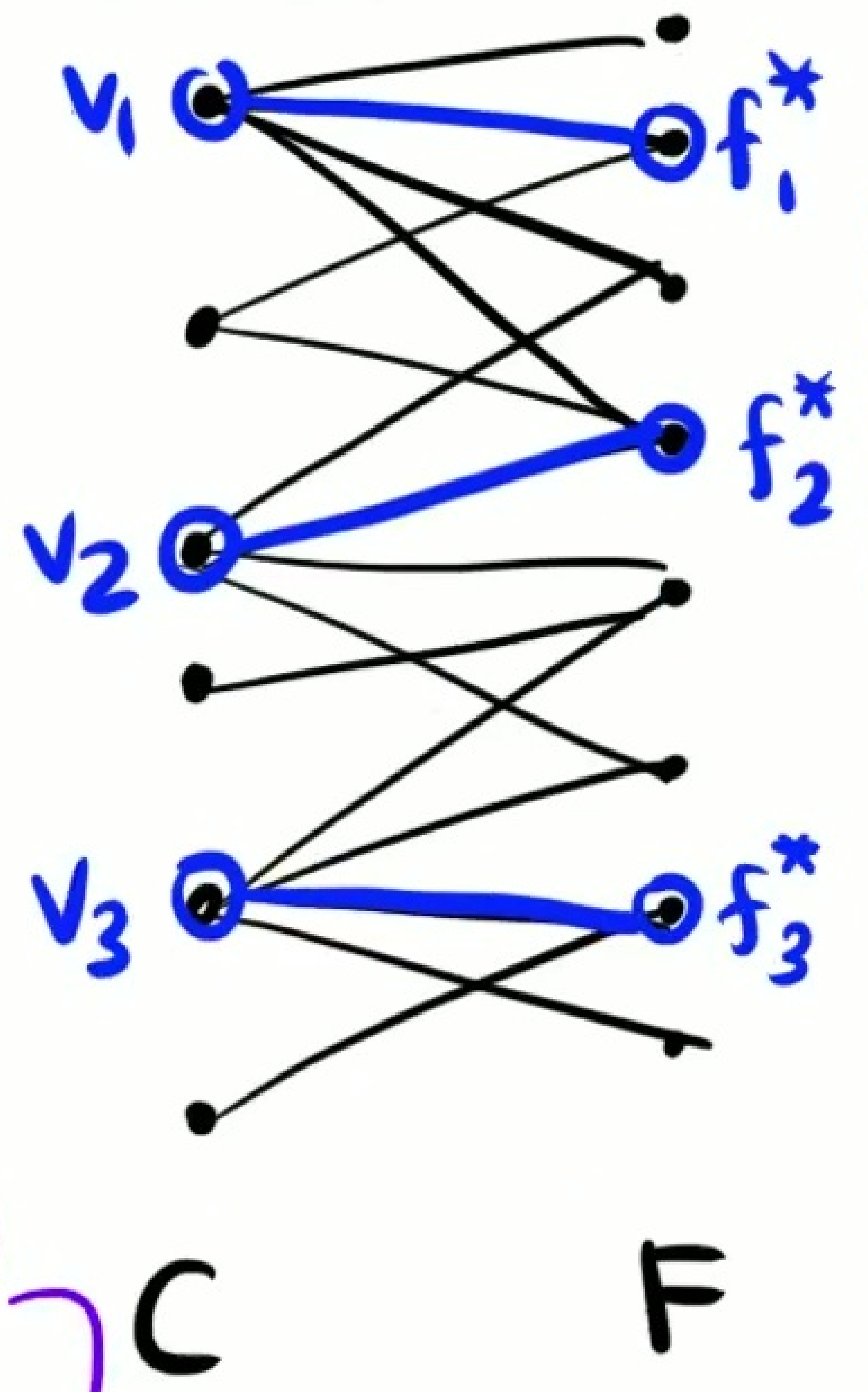
- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$
  $f_i, f_i^* \in N(v_i) \Rightarrow d(f_i^*, f_i) \leq 2$
  $d(v,f_i) \leq d(v,f_i^*) + d(f_i^*, f_i) \leq d(v,F^*) + 2 \leq 3\, d(v,F^*)$.

Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$
  s.t. $f_i^* \sim v_i \ \forall i$
  $[\text{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

- Optimize ______
  subject to $f_i \in N(v_i) \ \forall i$

- Claim: $d(v,F) \le 3\, d(v,F^*)$
  $\forall v \in C$.



3-approx.
How to get $(1+2/e)$?

Want $d(v,F)=d(v,F^*)$
for $(1-1/e)$ fraction

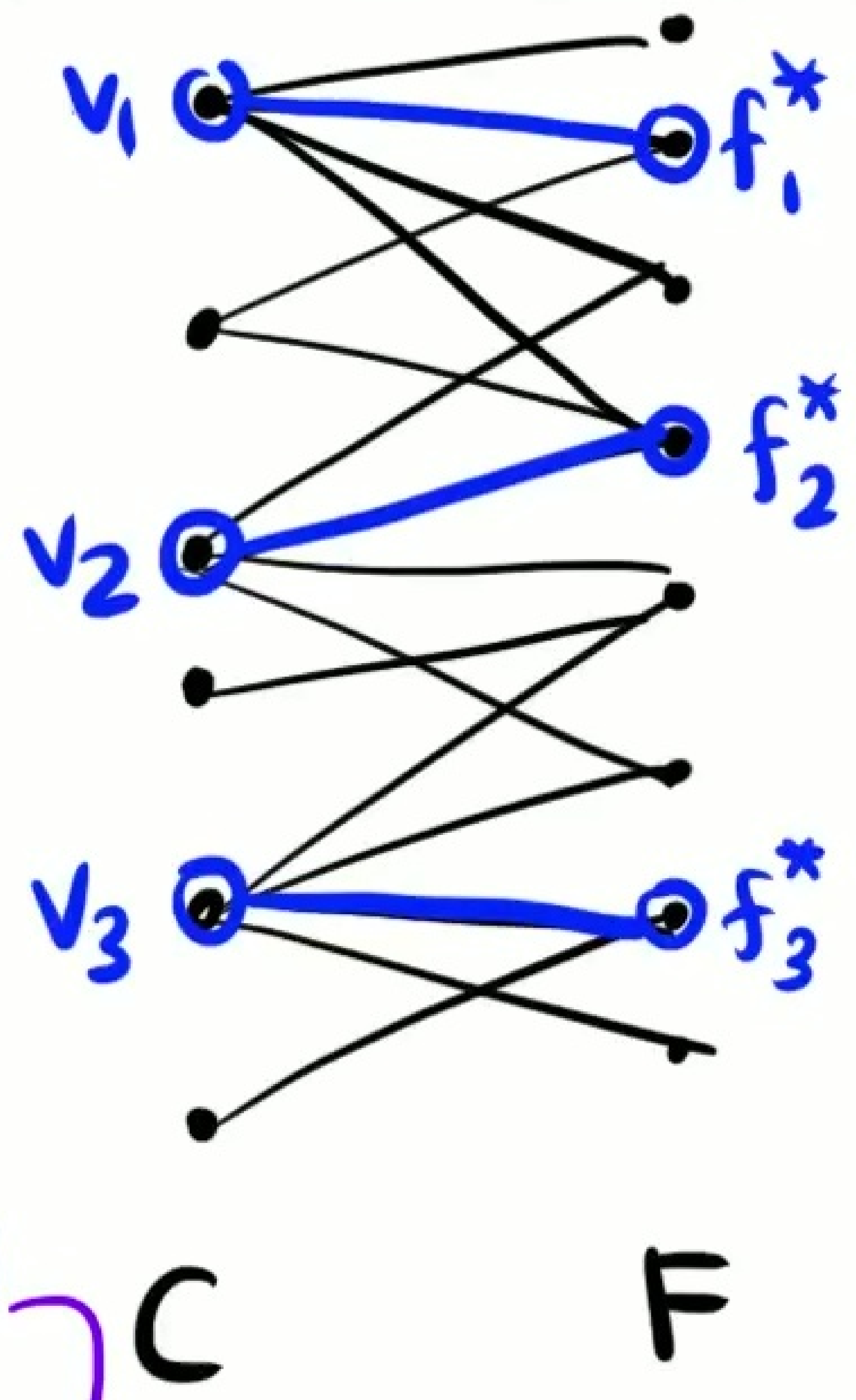
- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$
  $f_i, f_i^* \in N(v_i) \Rightarrow d(f_i^*, f_i) \le 2$
  $d(v,f_i) \le d(v,f_i^*) + d(f_i^*, f_i) \le d(v,F^*) + 2 \le 3\, d(v,F^*)$.

Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$
  s.t. $f_i^* \sim v_i \; \forall i$
  $[\text{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

- Optimize ______
  subject to $f_i \in N(v_i) \; \forall i$

- Claim: $d(v,F) \le 3\, d(v,F^*)$
  $\forall v \in C$.



→ 3-approx.
How to get $(1+2/e)$?

Want $d(v,F)=d(v,F^*)$ for $(1-1/e)$ fraction

For rest, can fallback on $d(v,F) \le 3 d(v,F^*)$

- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$
  $f_i, f_i^* \in N(v_i) \Rightarrow d(f_i^*, f_i) \le 2$
  $d(v,f_i) \le d(v,f_i^*) + d(f_i^*, f_i) \le d(v,F^*) + 2 \le 3\, d(v,F^*)$.

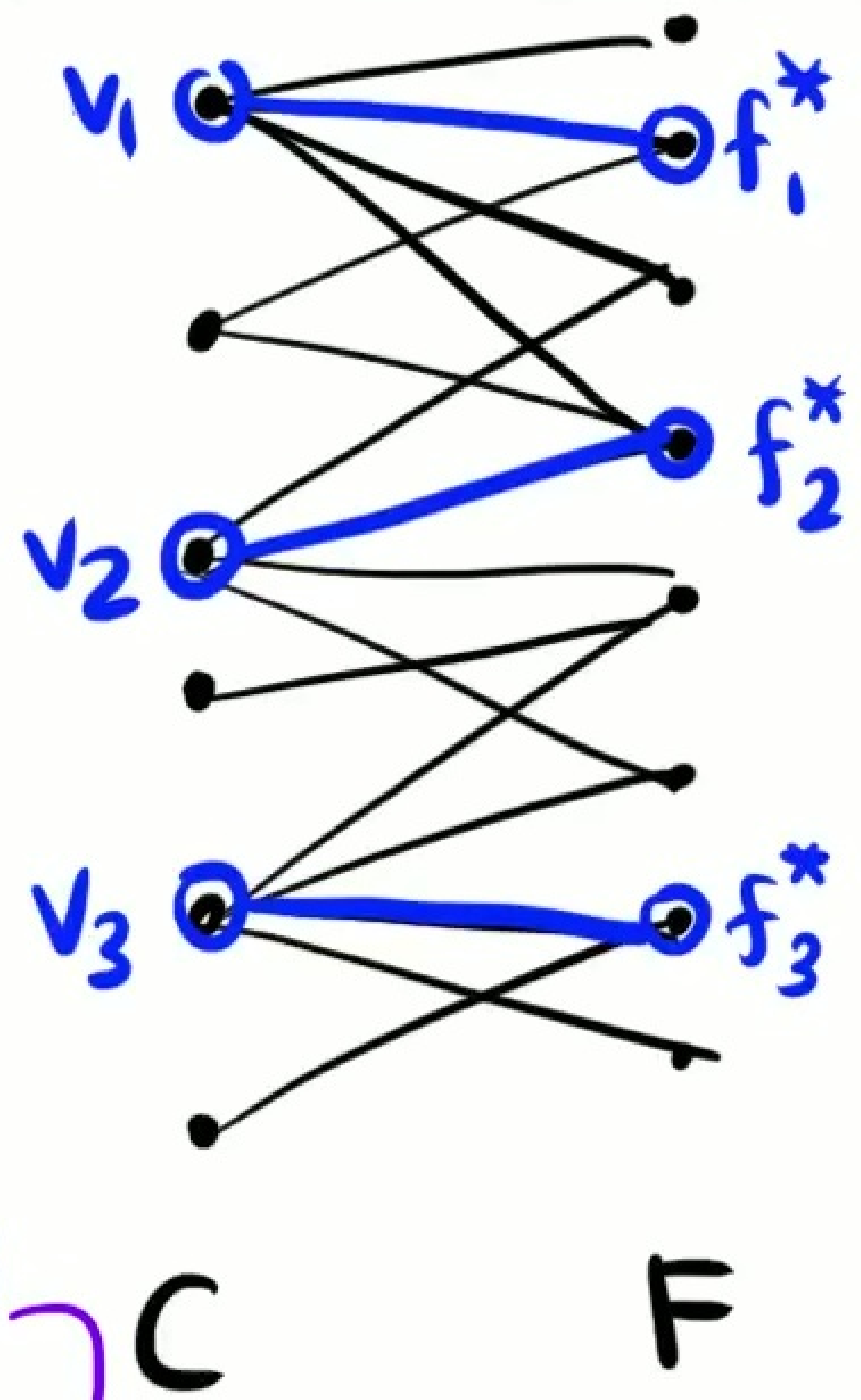
Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$ s.t. $f_i^* \sim v_i \ \forall i$

  $[\text{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

- Optimize ______ subject to $f_i \in N(v_i) \ \forall i$

- Claim: $d(v,F) \le 3\, d(v,F^*) \ \forall v \in C$.



→ 3-approx.

How to get $(1+2/e)$?

Want $d(v,F)=d(v,F^*)$ for $(1-1/e)$ fraction

For rest, can fallback on $d(v,F) \le 3\, d(v,F^*)$

No such guarantee in poly-time k-med algos!

- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$

  $f_i, f_i^* \in N(v_i) \Rightarrow d(f_i^*, f_i) \le 2$

  $d(v,f_i) \le d(v,f_i^*) + d(f_i^*, f_i) \le d(v,F^*) + 2 \le 3\, d(v,F^*)$.

Transformation to Algorithm

- Guess k centers $v_1 \dots v_k$ s.t. $f_i^* \sim v_i \ \forall i$

  $[\text{poly}(\varepsilon^{-1} k \log n)]^k$ time (FPT!)

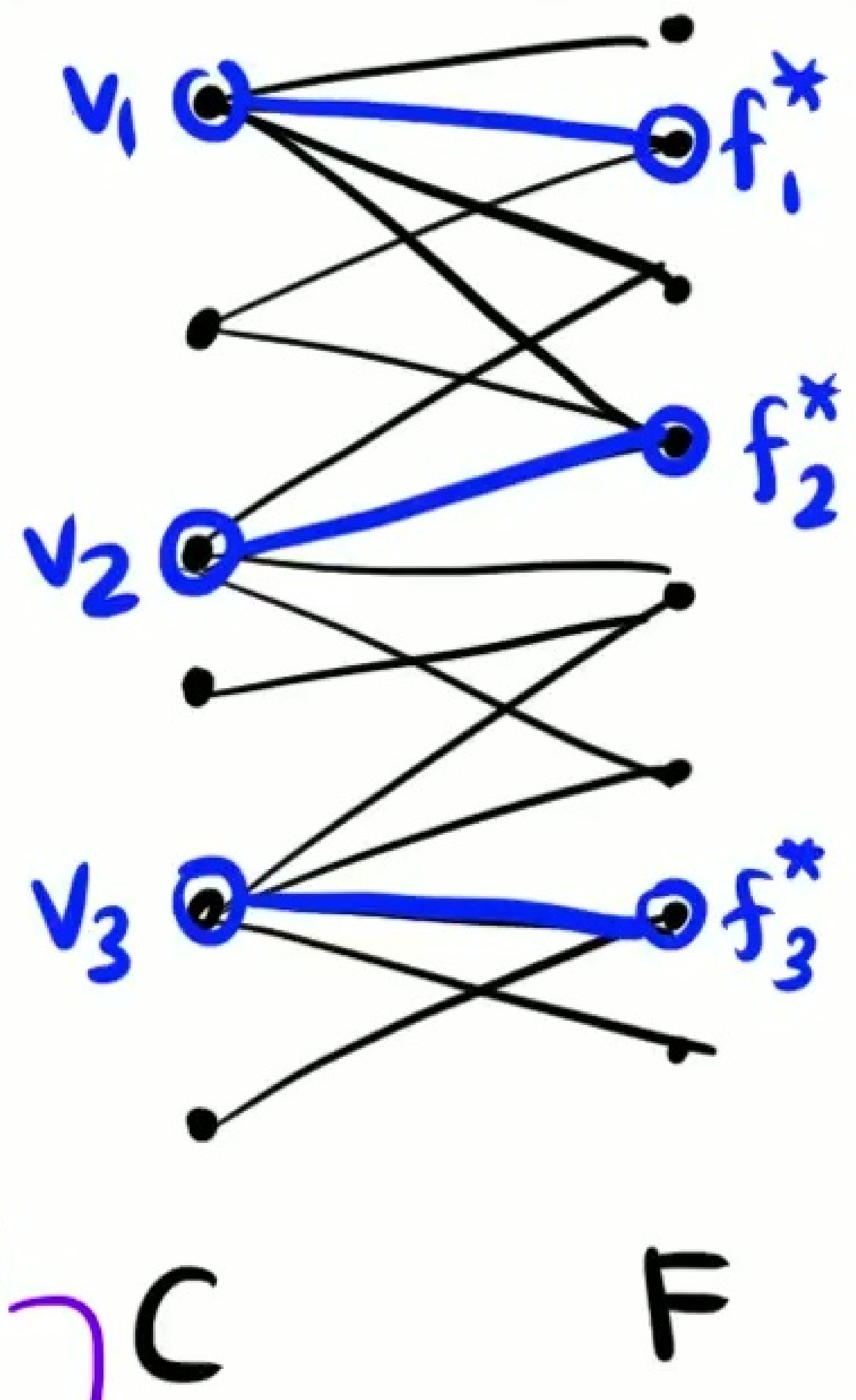
- Optimize ______ subject to $f_i \in N(v_i) \ \forall i$

- Claim: $d(v,F) \le 3\, d(v,F^*) \ \forall v \in C$.

- Pf: Let $f_i^* \in F^*$ s.t. $d(v,F^*) = d(v,f_i^*)$

  $f_i, f_i^* \in N(v_i) \Rightarrow d(f_i^*, f_i) \le 2$

  $d(v,f_i) \le d(v,f_i^*) + d(f_i^*, f_i) \le d(v,F^*) + 2 \le 3\, d(v,F^*)$.



→ 3-approx.

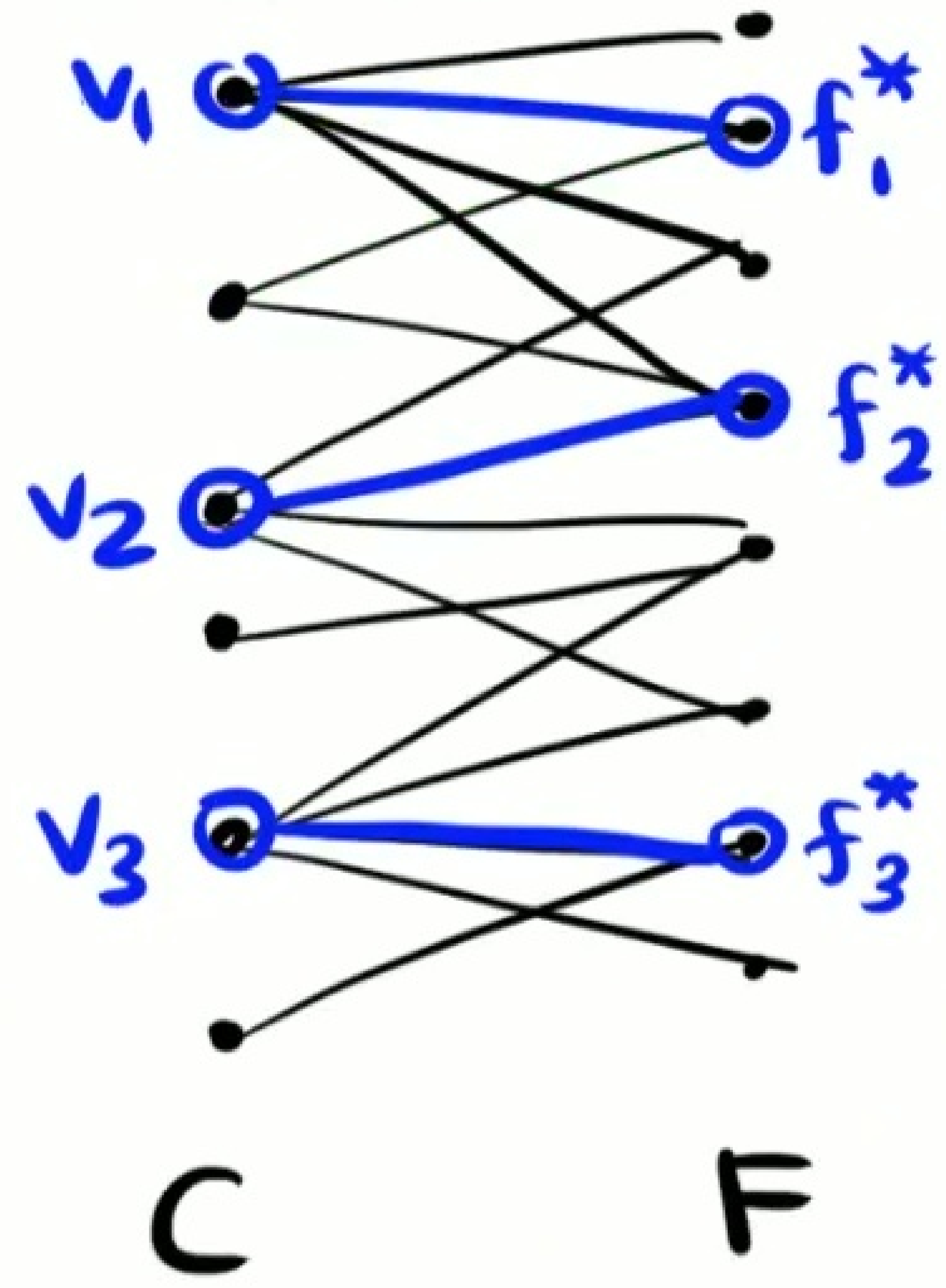
How to get $(1+2/e)$? don't know $d(v,F^*)$...

Want $d(v,F) = d(v,F^*)$ for $(1-1/e)$ fraction

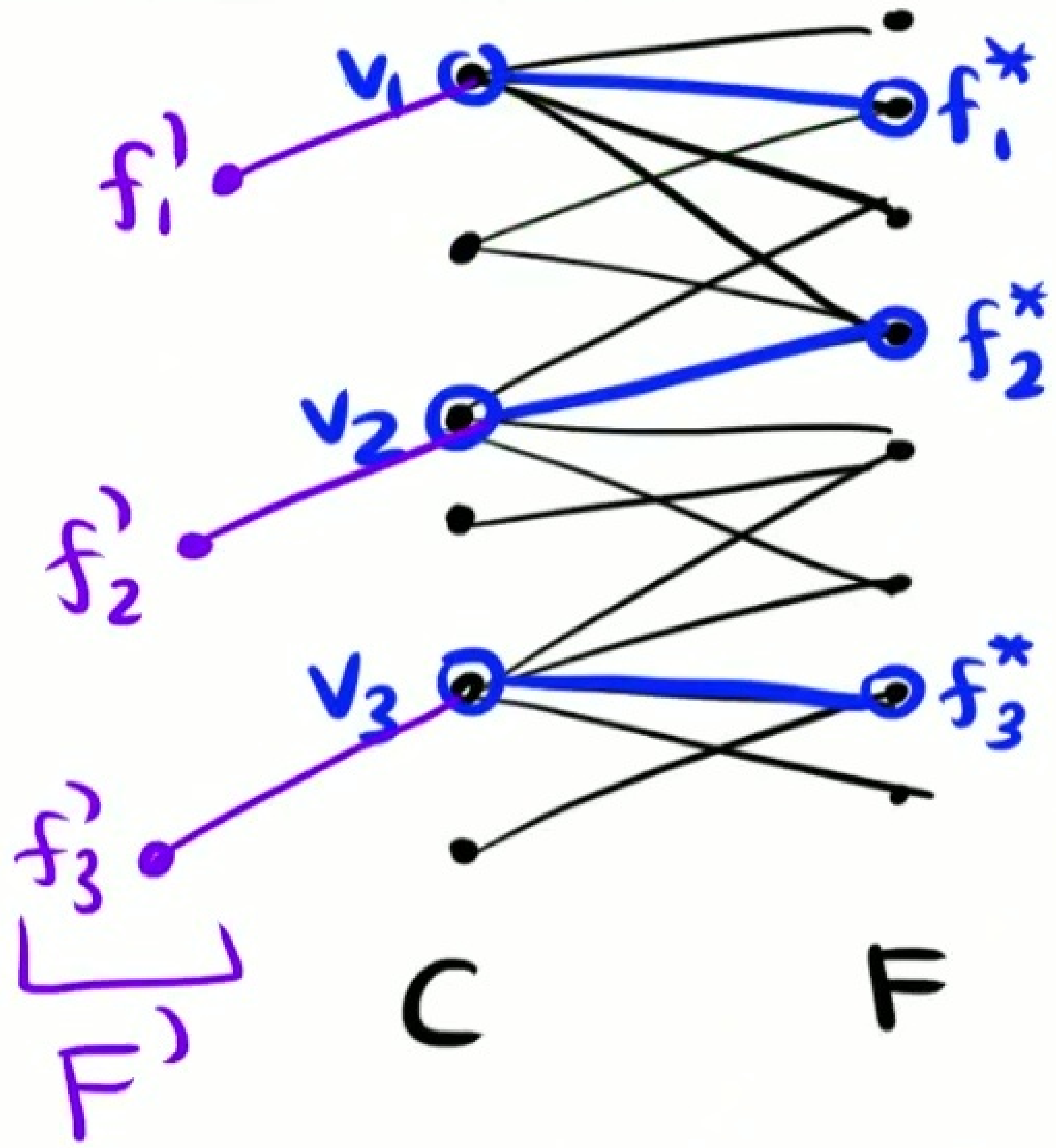
For rest, can fallback on $d(v,F) \le 3\, d(v,F^*)$

No such guarantee in poly-time k-med algos!

Optimization function



Optimization function

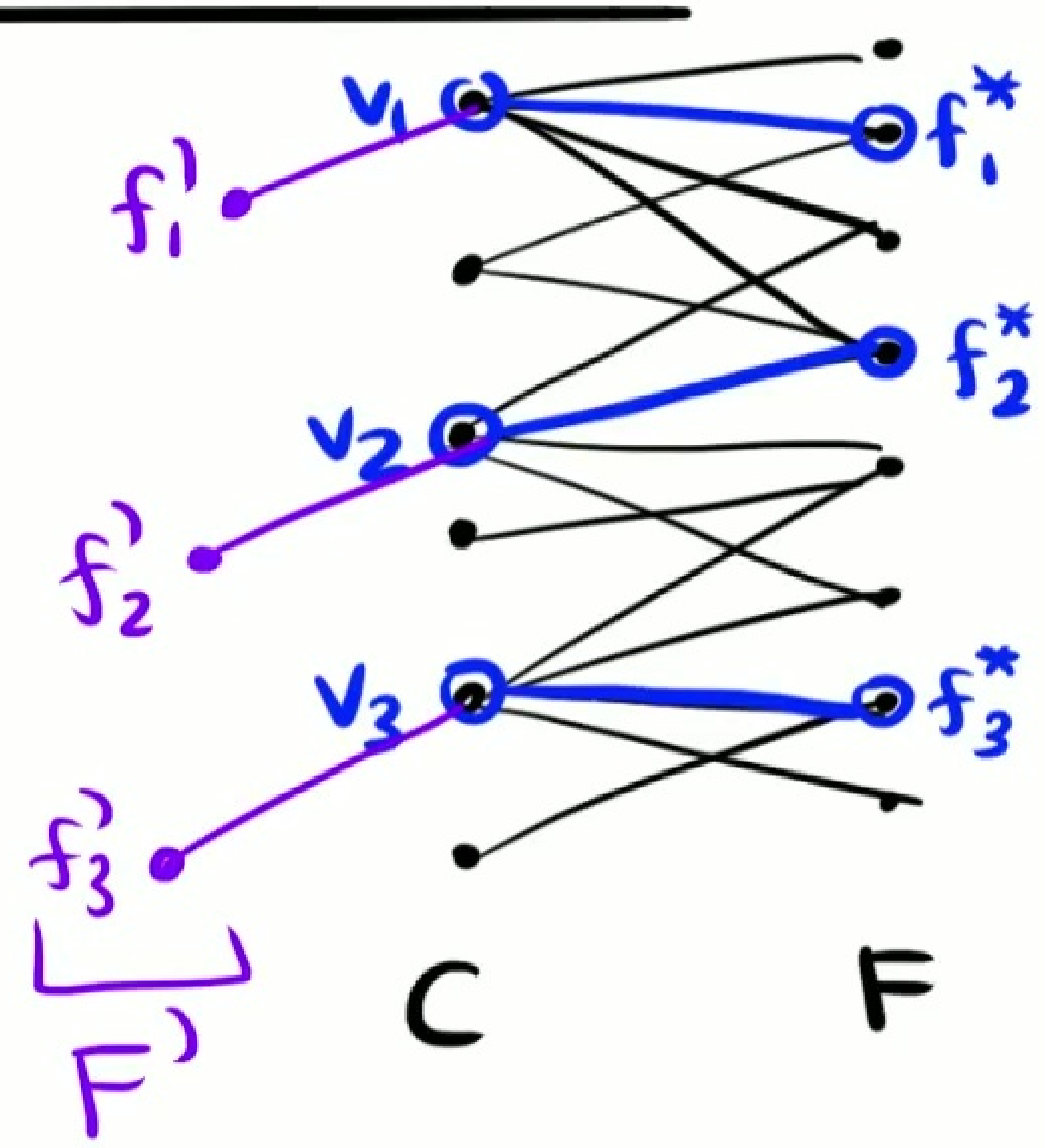


Optimization function

maximize

$$\mathrm{impr}(S) := \sum_{v} w(v)\Big(d(v, F') - d(v, F' \cup S)\Big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$



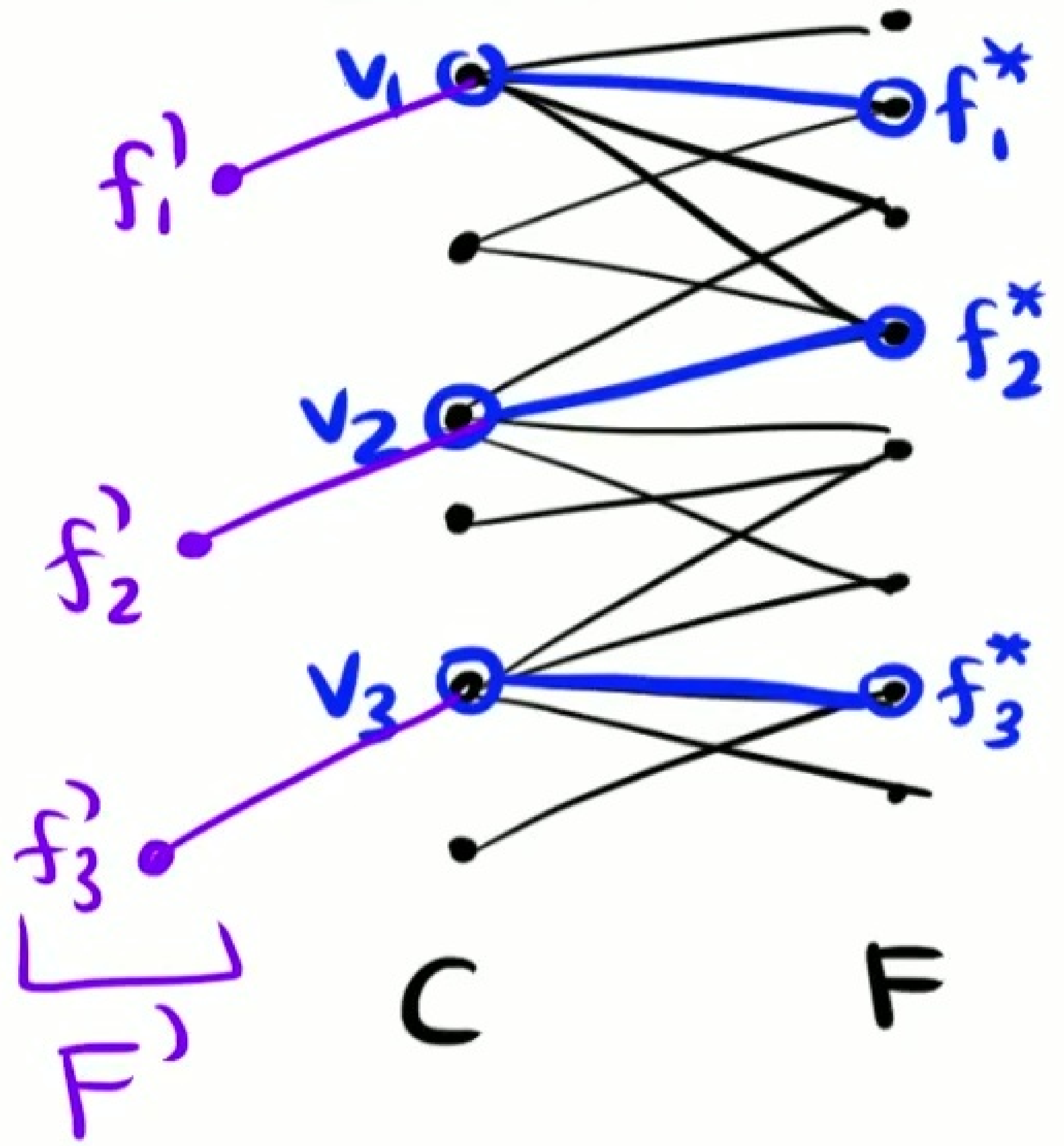
Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\Big(d(v,F') - d(v,F' \cup S)\Big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is monotone submodular with $\mathrm{impr}(\emptyset) = 0$.



Optimization function

maximize

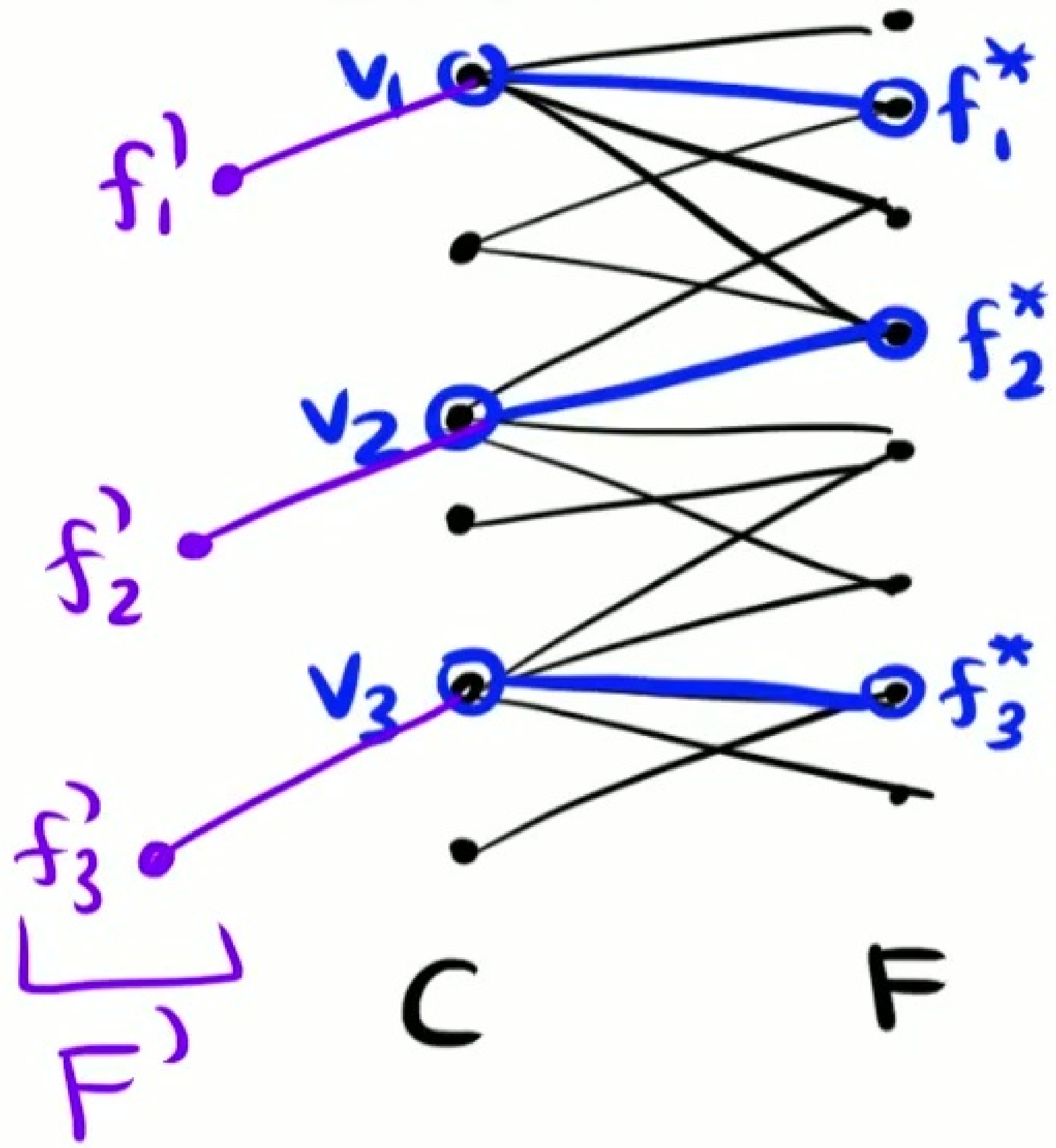
$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - d(v, F' \cup S)\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone submodular with $\mathrm{impr}(\emptyset) = 0$.

① Adding into S only decreases

$\Rightarrow$ increases impr



Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\Big(d(v,F') - d(v,F' \cup S)\Big)$$

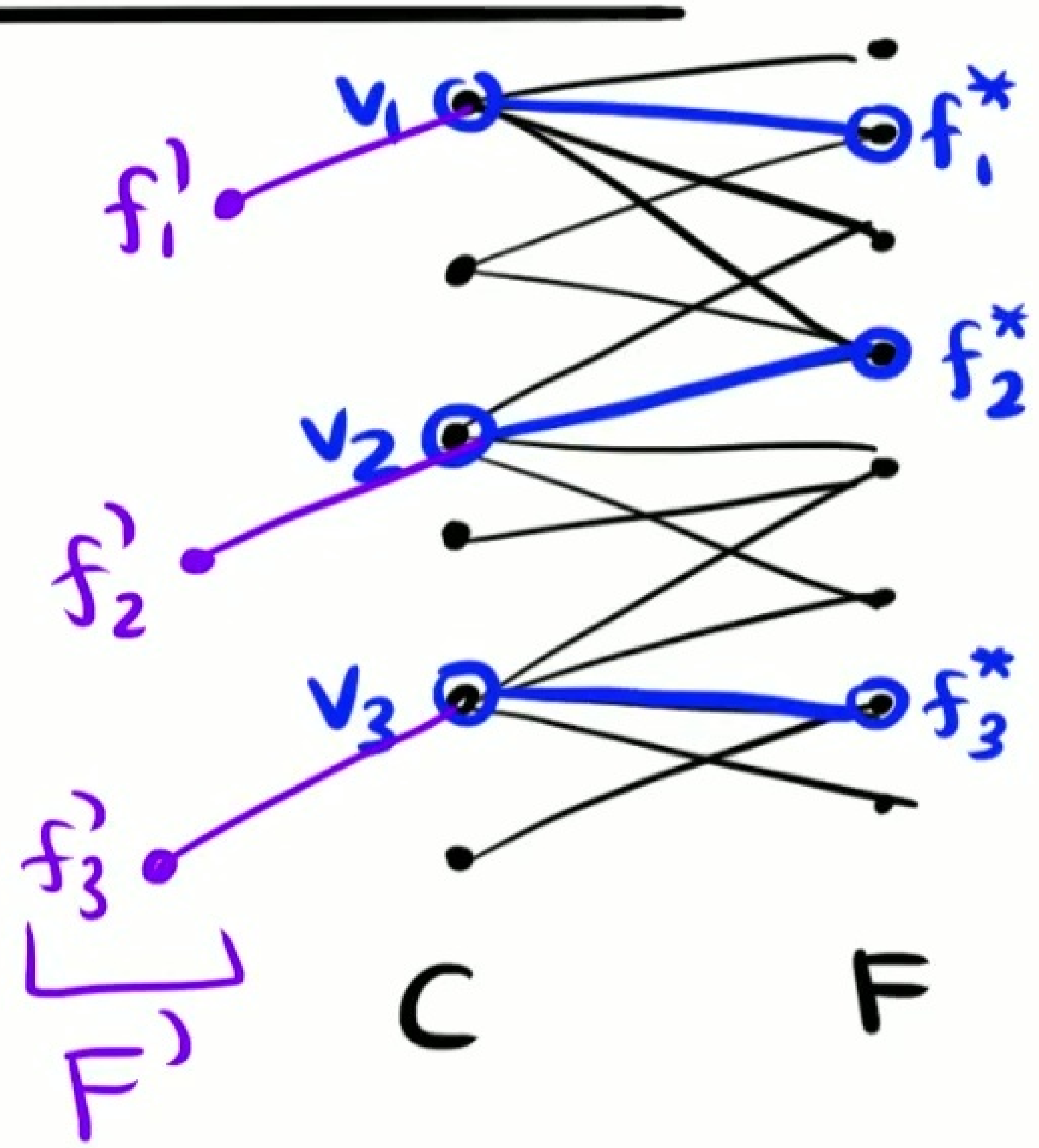
over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone submodular with ② $\mathrm{impr}(\emptyset) = 0$.

① Adding into S only decreases

$\Longrightarrow$ increases impr

② easy.



Optimization function

maximize

$$\operatorname{impr}(S) := \sum_v w(v)\left(d(v, F') - d(v, F' \cup S)\right)$$

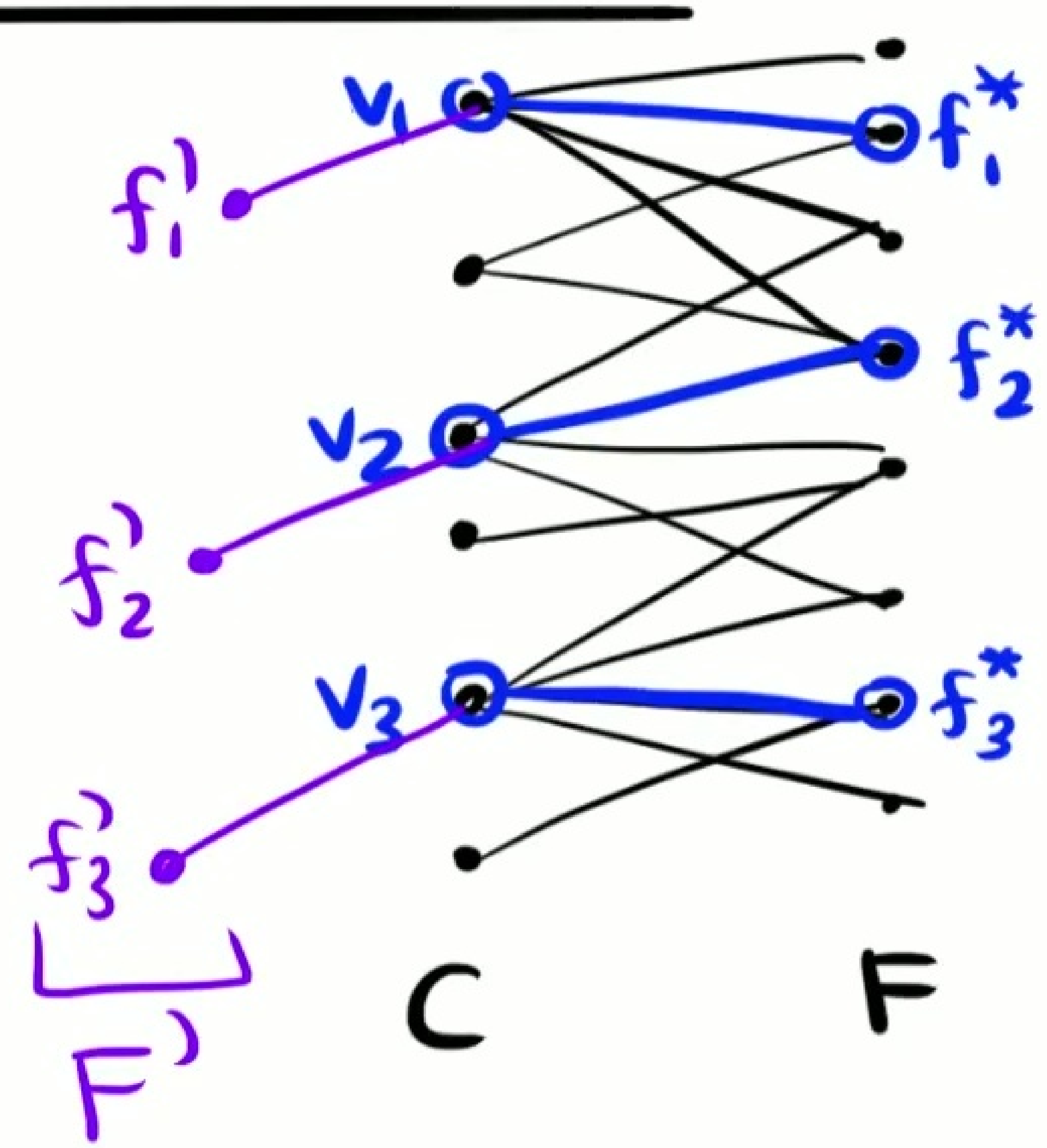
over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: impr(S) is ① monotone ③ submodular with ② $\operatorname{impr}(\emptyset) = 0$.

① Adding into S only decreases $\Rightarrow$ increases impr

② easy.

③ Submodular: $\forall S \subseteq T, f$: $\operatorname{impr}(S \cup f) - \operatorname{impr}(S) \geq \operatorname{impr}(T \cup f) - \operatorname{impr}(T)$

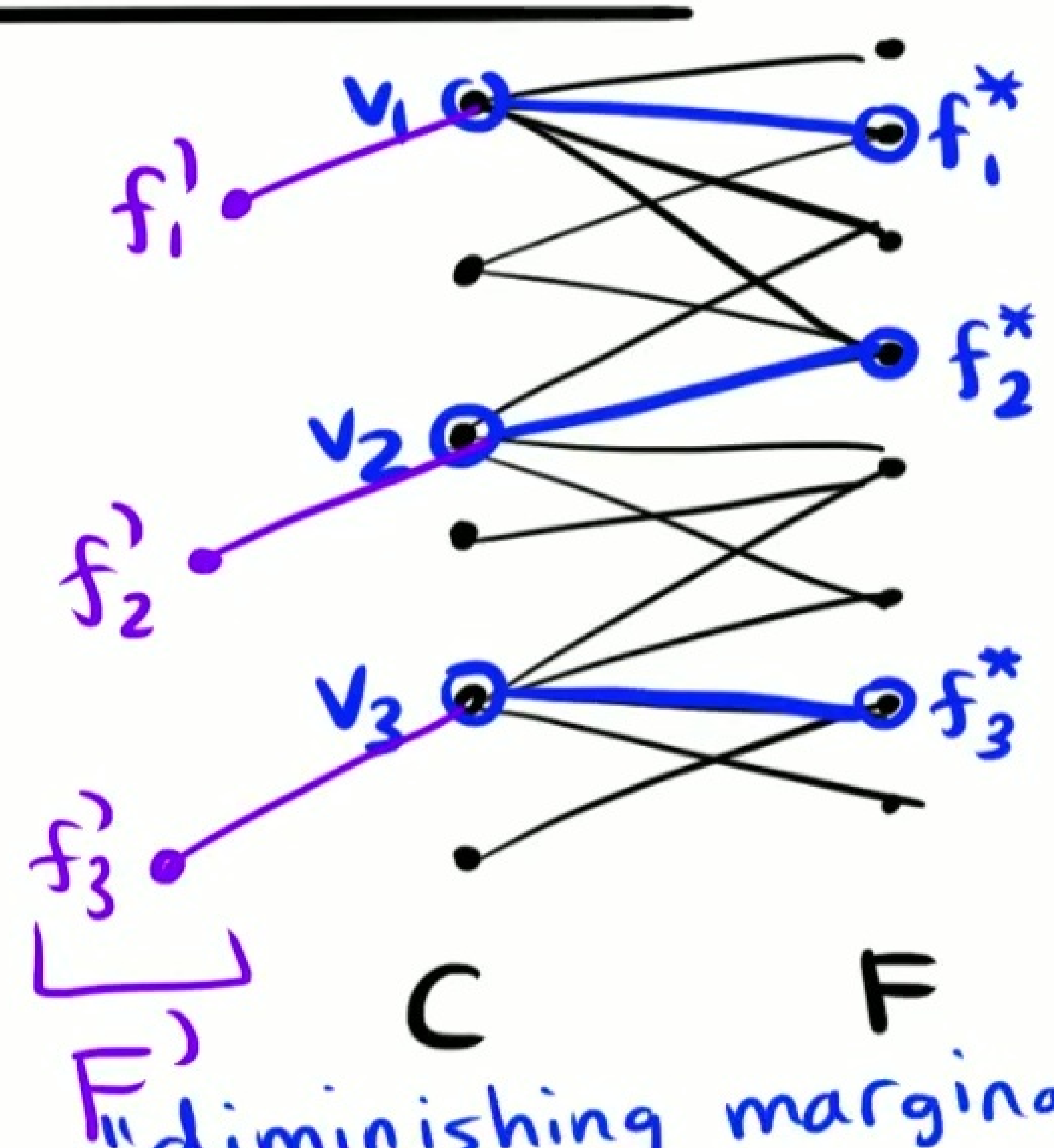


Optimization function

maximize

$$\text{impr}(S) := \sum_v w(v)\Big(d(v, F') - d(v, F' \cup S)\Big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$



Lemma: impr(S) is ① monotone ③ submodular with ② $\text{impr}(\emptyset) = 0$.

① Adding into S only decreases

$\Rightarrow$ increases impr

② easy.

"diminishing marginal returns"

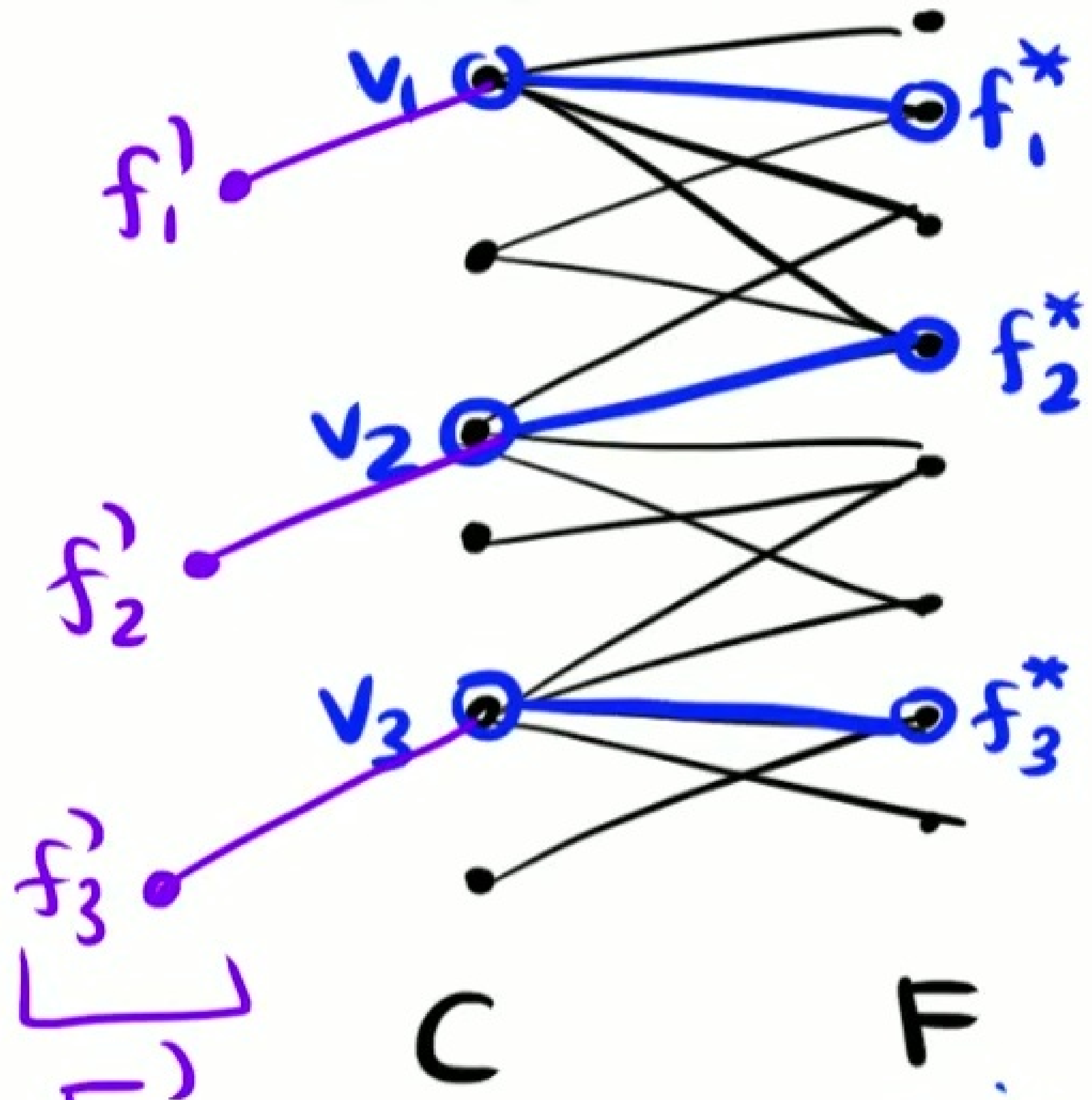
③ Submodular: $\forall S \subseteq T, f$: $\text{impr}(S \cup f) - \text{impr}(S) \geq \text{impr}(T \cup f) - \text{impr}(T)$

Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - d(v, F' \cup S)\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$



Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.

① Adding into S only decreases

$\Rightarrow$ increases impr

② easy.

"diminishing marginal returns"

③ Submodular: $\forall S \subseteq T, f$: $\mathrm{impr}(S \cup f) - \mathrm{impr}(S) \geq \mathrm{impr}(T \cup f) - \mathrm{impr}(T)$

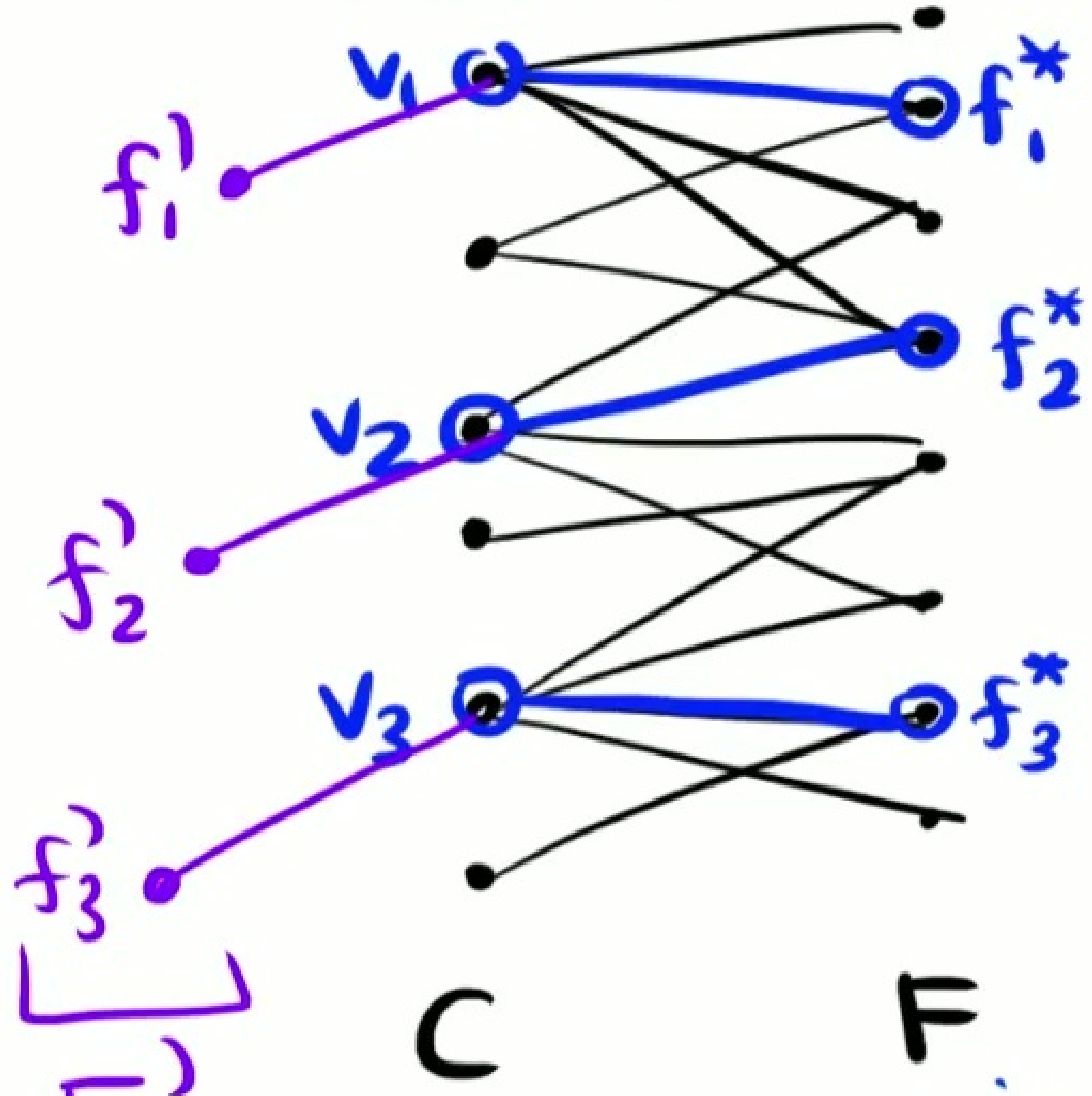
$\Leftarrow \forall v: -d(v, F' \cup S \cup f) + d(v, F' \cup S) \geq -d(v, F' \cup T \cup f) + d(v, F' \cup T)$

Optimization function

maximize

$$\text{impr}(S) := \sum_v w(v)\left(d(v,F') - d(v,F' \cup S)\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$



Lemma: $\text{impr}(S)$ is ① monotone ③ submodular with ② $\text{impr}(\emptyset) = 0$.

① Adding into S only decreases $\Rightarrow$ increases impr

② easy.

③ Submodular: $\forall S \subseteq T,\ f:\ \text{impr}(S \cup f) - \text{impr}(S) \geq \text{impr}(T \cup f) - \text{impr}(T)$ "diminishing marginal returns"

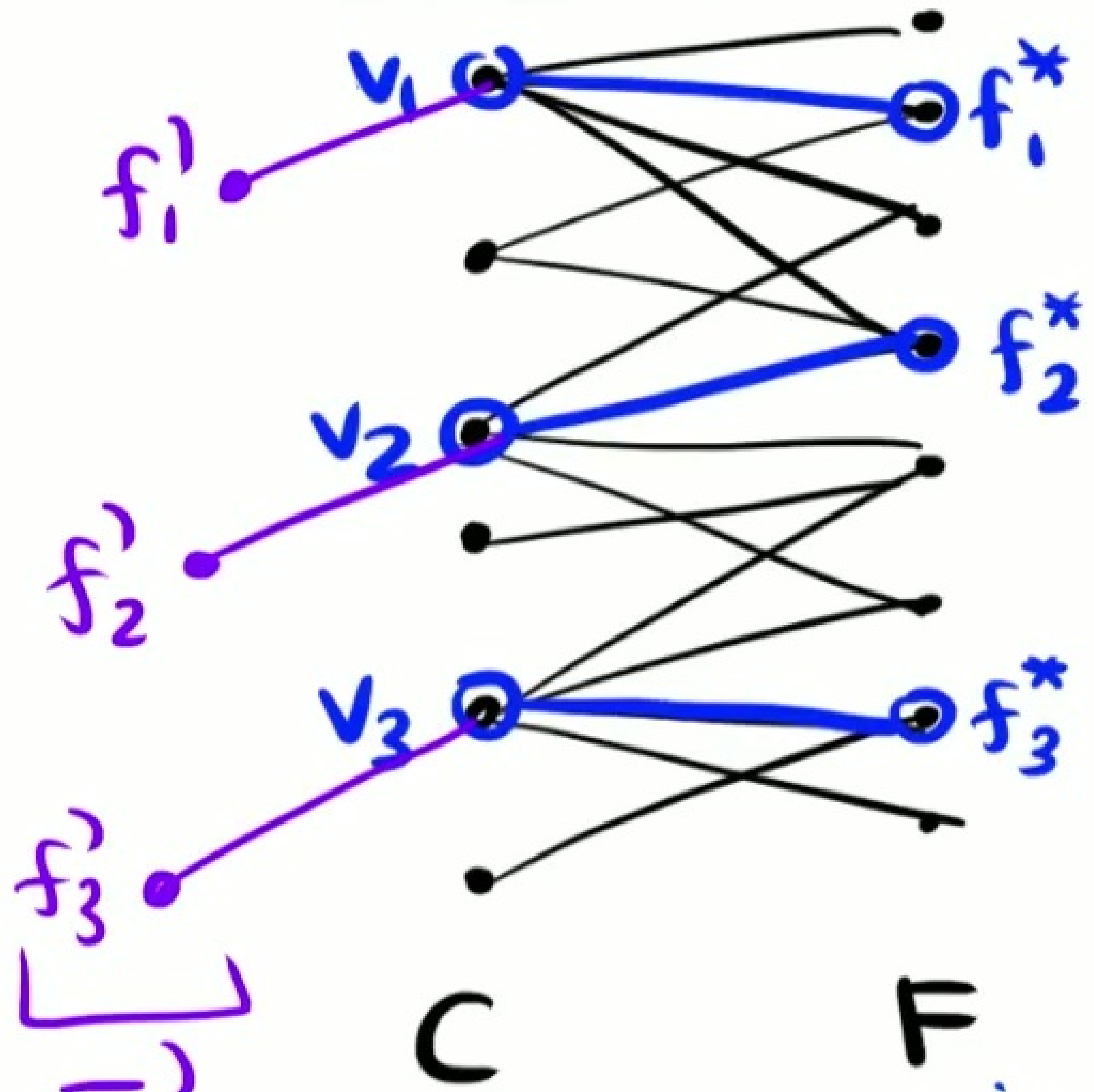
$$\Leftarrow \forall v: -\underbrace{d(v, F' \cup S \cup f)}_{=\min(d(v,F' \cup S),\ d(v,f))} + d(v, F' \cup S) \geq -\underbrace{d(v, F' \cup T \cup f)}_{=\min(d(v,F' \cup T),\ d(v,f))} + d(v, F' \cup T)$$

Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - d(v,F'\cup S)\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$



Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.

① Adding into S only decreases $\Rightarrow$ increases impr

② easy.

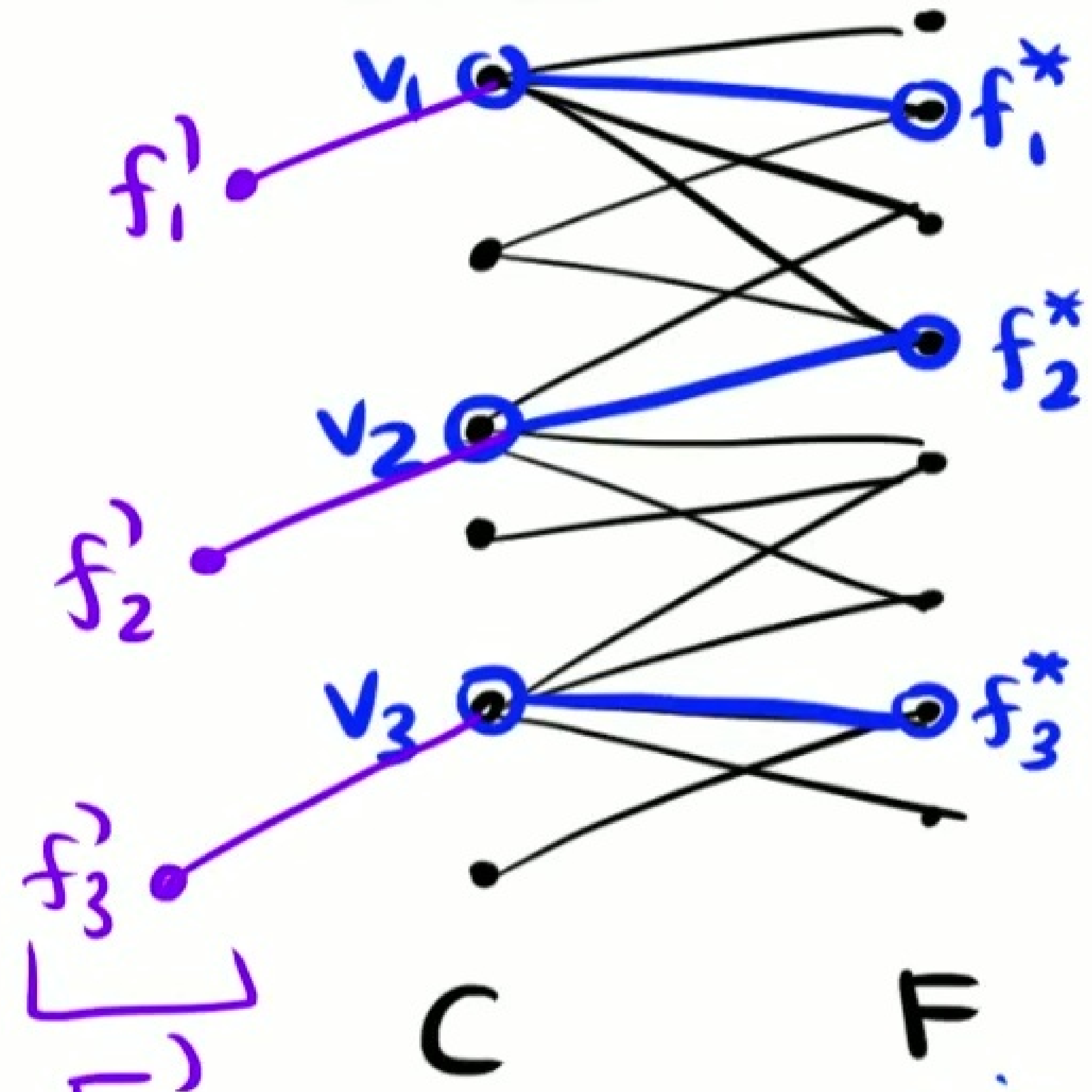
"diminishing marginal returns"

③ Submodular: $\forall S \subseteq T, f:\ \mathrm{impr}(S \cup f) - \mathrm{impr}(S) \geq \mathrm{impr}(T \cup f) - \mathrm{impr}(T)$

$$\Leftarrow \forall v: -\underbrace{d(v, F'\cup S \cup f)}_{=\min(d(v,F'\cup S),\ d(v,f))} + d(v, F'\cup S) \geq -\underbrace{d(v, F'\cup T \cup f)}_{=\min(d(v,F'\cup T),\ d(v,f))} + d(v, F'\cup T)$$

$$\Leftrightarrow \max(0,\ d(v,F'\cup S) - d(v,f)) \geq \max(0, d(v,F'\cup T) - d(v,f))$$

Optimization function



maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - d(v, F' \cup S)\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.

① Adding into S only decreases
$\Rightarrow$ increases impr

② easy.

"diminishing marginal returns"

③ Submodular: $\forall S \subseteq T, f$: $\mathrm{impr}(S \cup f) - \mathrm{impr}(S) \geq \mathrm{impr}(T \cup f) - \mathrm{impr}(T)$

$$\Leftarrow \forall v: -\underbrace{d(v, F' \cup S \cup f)}_{= \min(d(v,F' \cup S),\ d(v,f))} + d(v, F' \cup S) \geq -\underbrace{d(v, F' \cup T \cup f)}_{= \min(d(v,F' \cup T),\ d(v,f))} + d(v, F' \cup T)$$

$$\Leftrightarrow \max(0, d(v, F' \cup S) - d(v,f)) \geq \max(0, d(v, F' \cup T) - d(v,f))$$

$$d(v, F' \cup S) \geq d(v, F' \cup T)$$

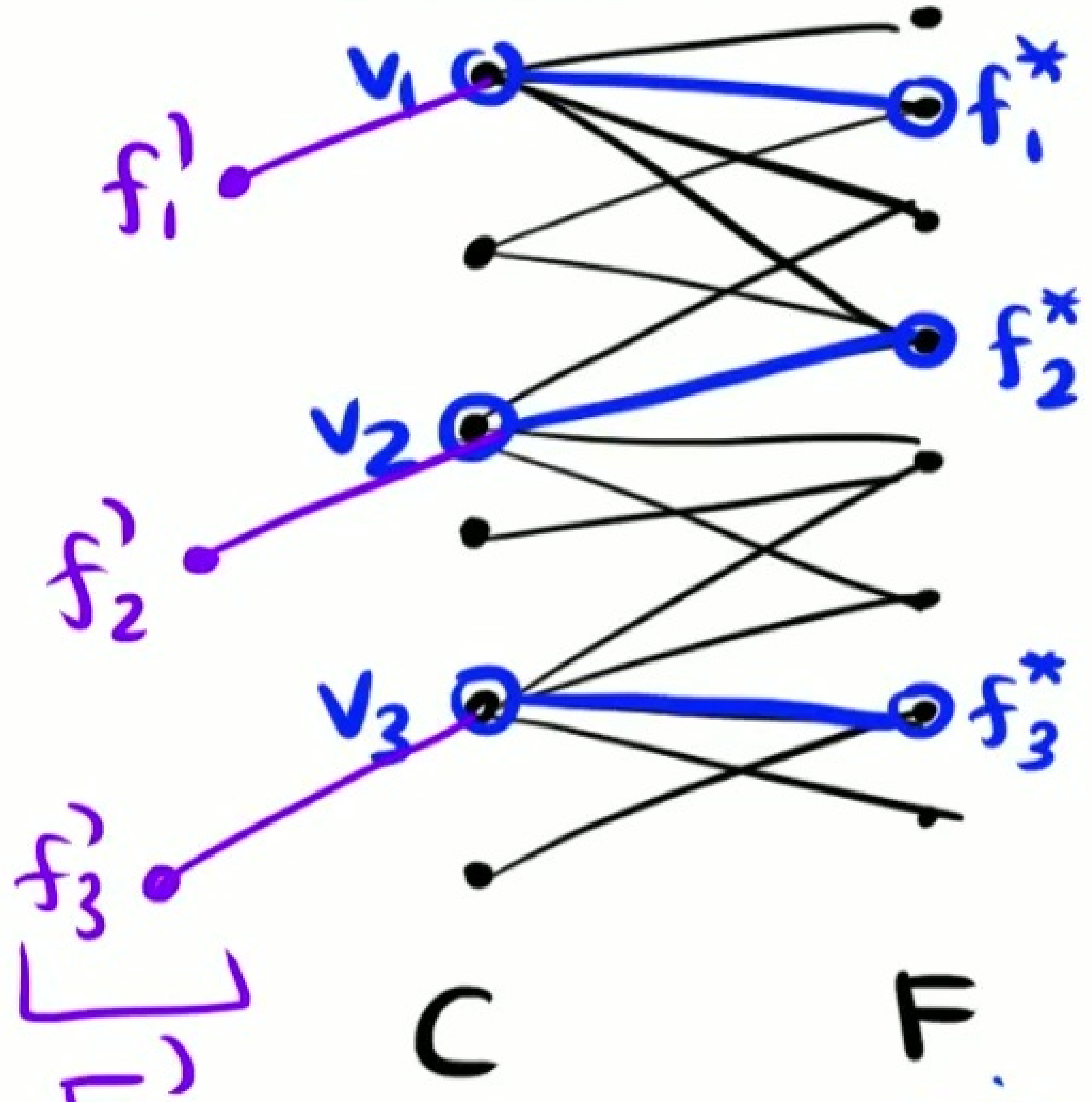
Optimization function

maximize

$$\text{impr}(S) := \sum_v w(v)\big(d(v,F') - d(v, F' \cup S)\big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

Lemma: impr(S) is ① monotone ③ submodular with ② $\text{impr}(\emptyset) = 0$.



Thm [CCPV] can $(1-1/e)$-approximate monotone submodular.

"diminishing marginal returns"

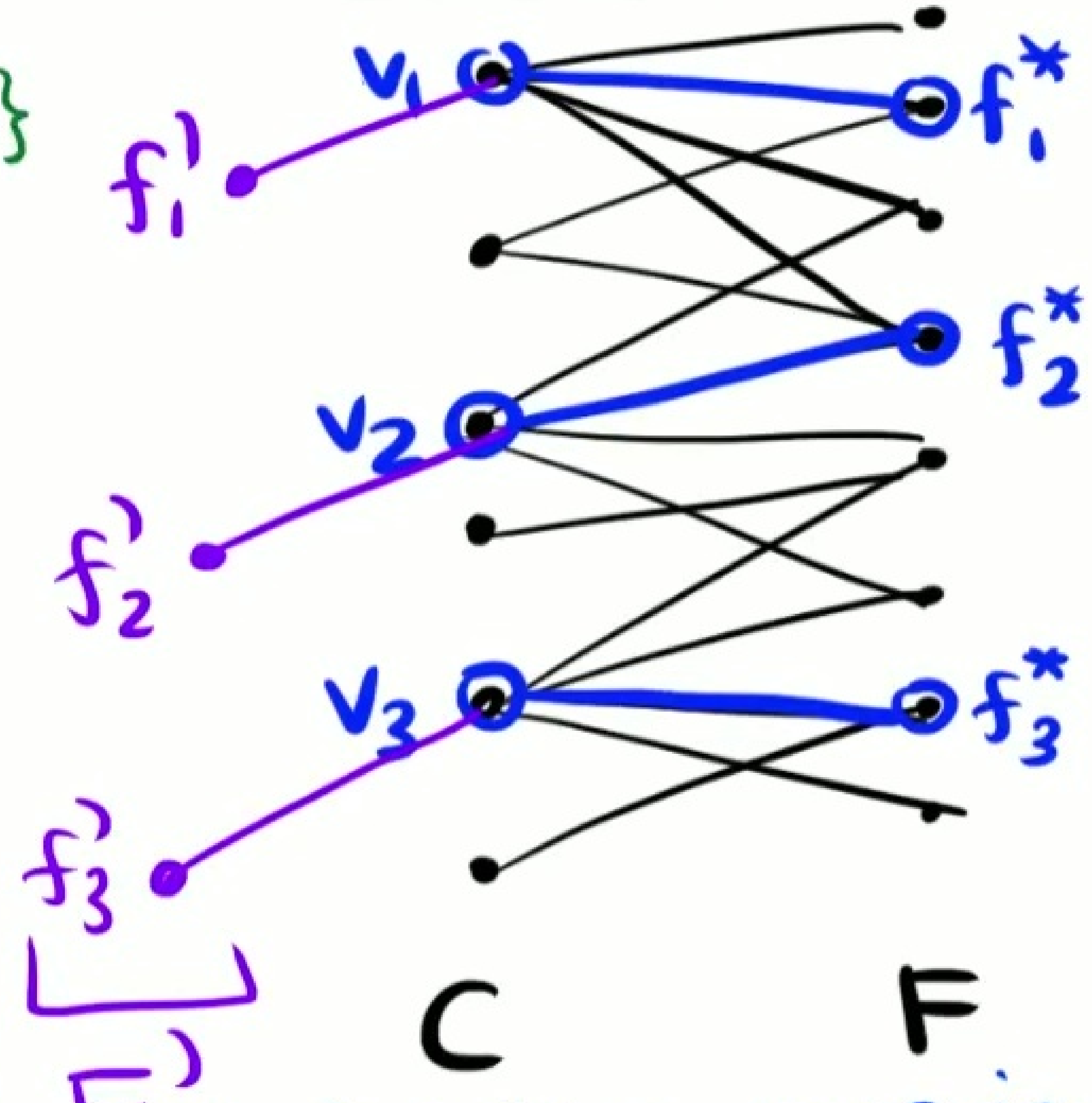
Optimization function

maximize

$$\text{impr}(S) := \sum_v w(v)\Big(d(v,F') - \overbrace{d(v,F'\cup S)}^{= d(v,S) \text{ when } S=\{f_1..f_k,\ f_i\in N(v_i)\}}\Big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\text{impr}(S)$ is ① monotone ③ submodular with ② $\text{impr}(\emptyset) = 0$.



Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

"diminishing marginal

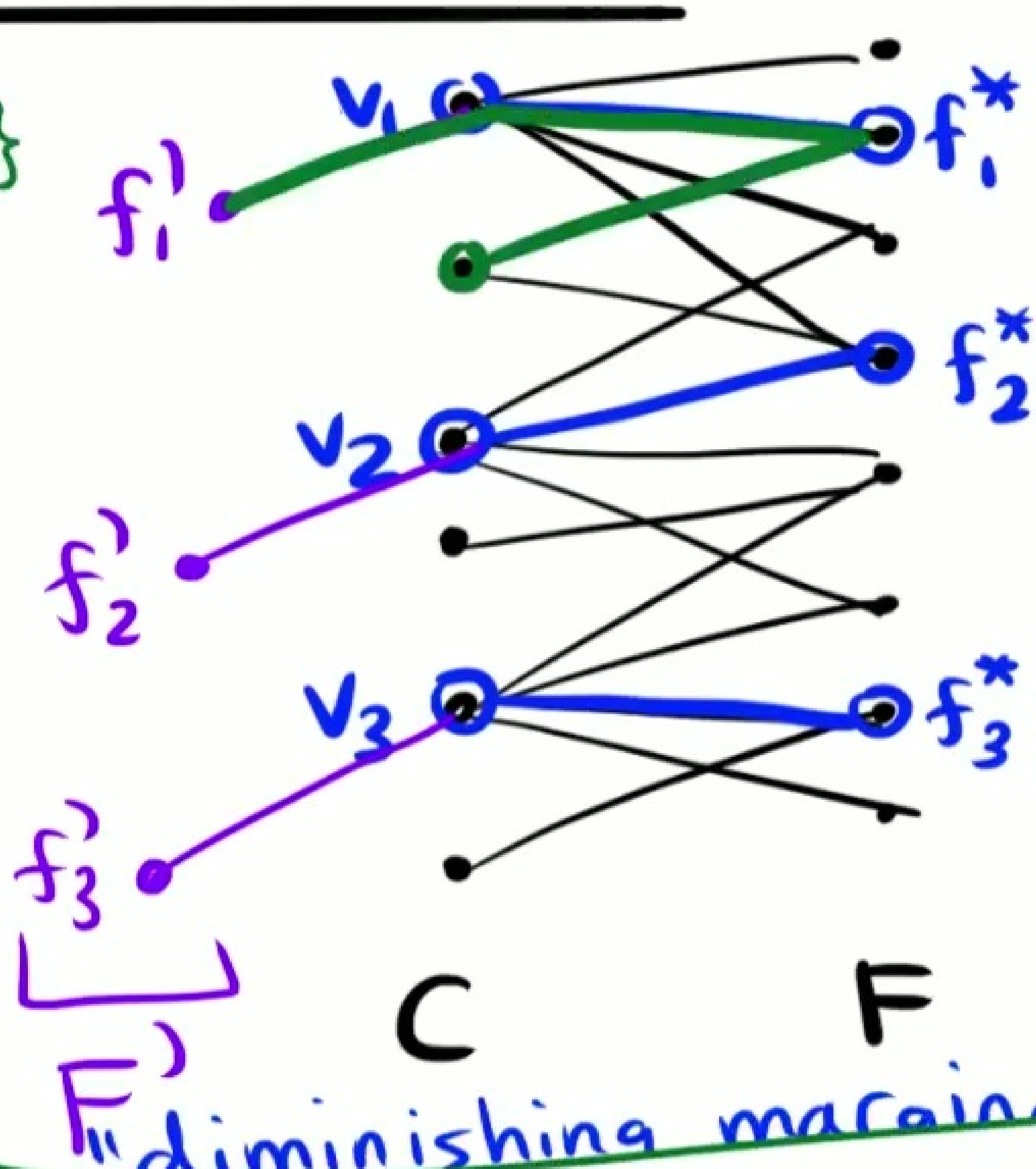
Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\Big(d(v,F') - \overbrace{d(v, F' \cup S)}^{\substack{= d(v,S) \text{ when} \\ S=\{f_1 .. f_k,\ f_i \in N(v_i)\}}}\Big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.



Obs: $\forall v$, $d(v, F') \le 3\, d(v, F^*)$ (same as before)

Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

"diminishing marginal returns"

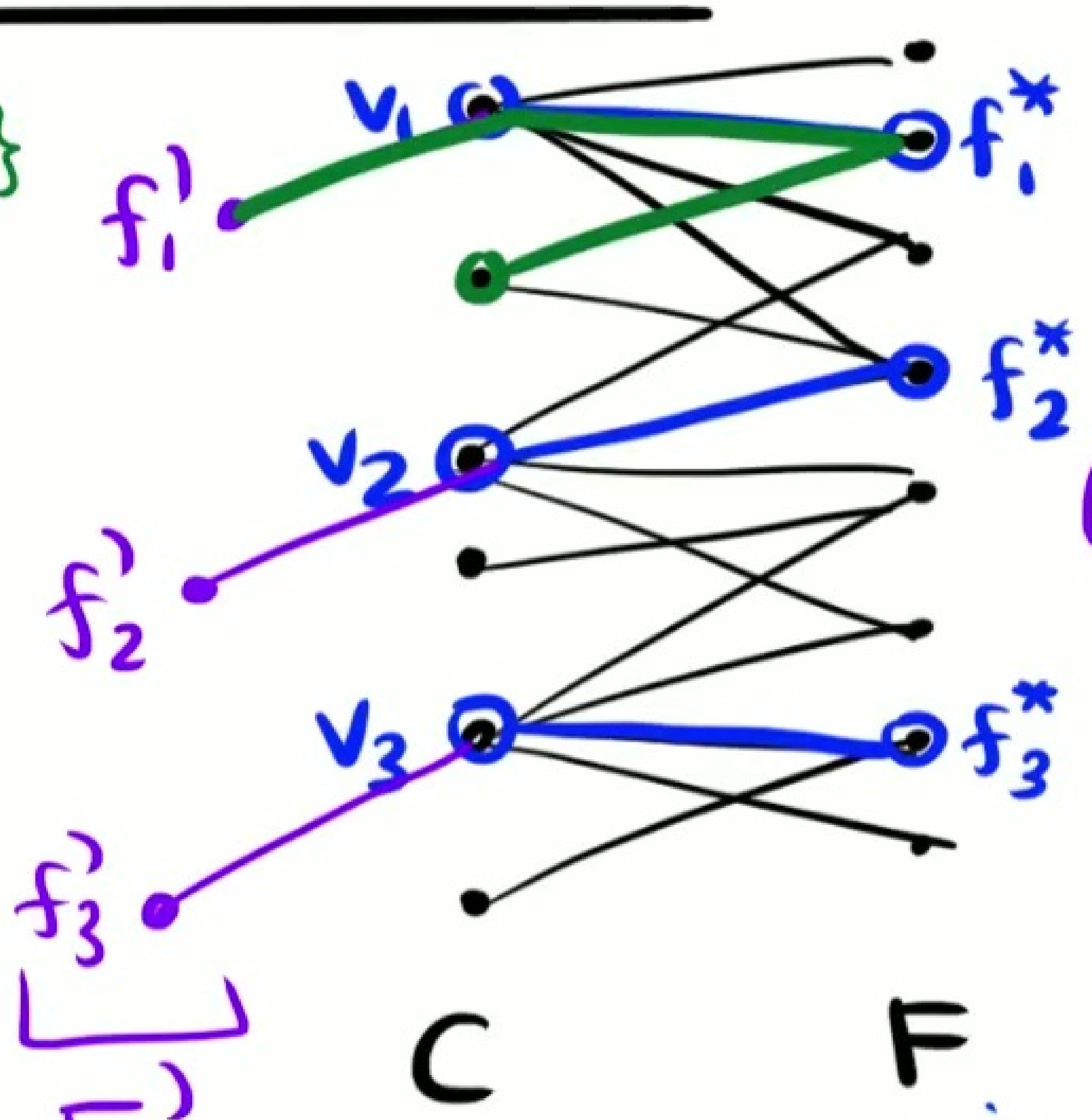
Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\Big(d(v,F') - \overbrace{d(v, F'\cup S)}^{=d(v,S) \text{ when } S=\{f_1..f_k,\ f_i\in N(v_i)\}}\Big)$$

over $S \subseteq \mathbb{F}$, $S=\{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset)=0$.



Obs: $\forall v$, $d(v,F') \le 3\, d(v,F^*)$ (same as before)

"diminishing marginal returns"

Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

Compute S s.t. $\mathrm{impr}(S) \ge (1-1/e)\,\mathrm{impr}(F^*)$.

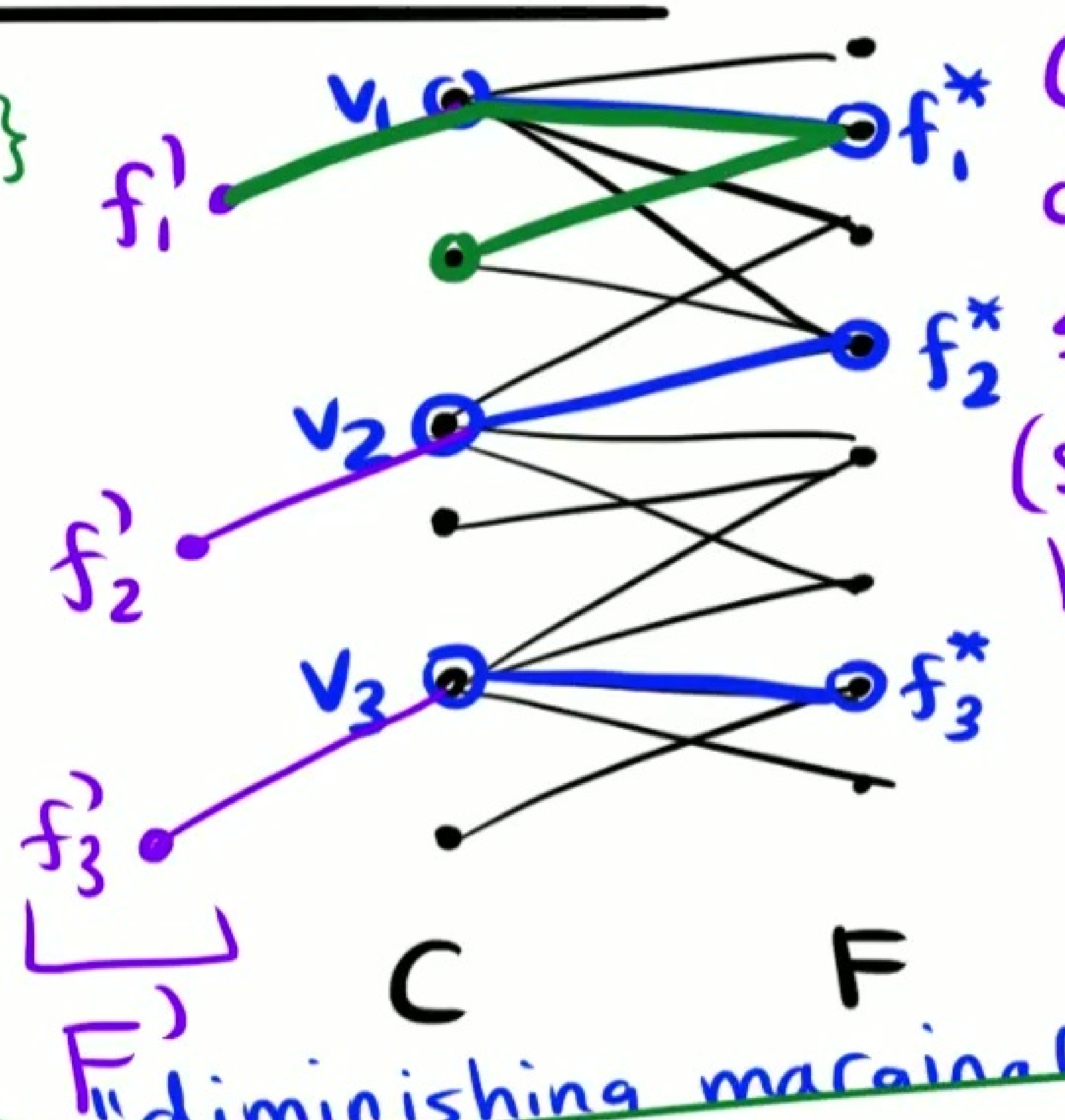
Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - \overbrace{d(v, F' \cup S)}^{= d(v,S) \text{ when } S=\{f_1 \ldots f_k,\ f_i \in N(v_i)\}}\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.



Obs: $\forall v$, $d(v, F') \le 3\, d(v, F^*)$ (same as before)

Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

"diminishing marginal returns"

Compute S s.t. $\mathrm{impr}(S) \ge (1-1/e)\,\mathrm{impr}(F^*)$.

$$\mathrm{cost}(S) = \sum_v w(v)\, d(v,S) = \sum_v w(v)\, d(v,F') - \mathrm{impr}(S)$$

Optimization function

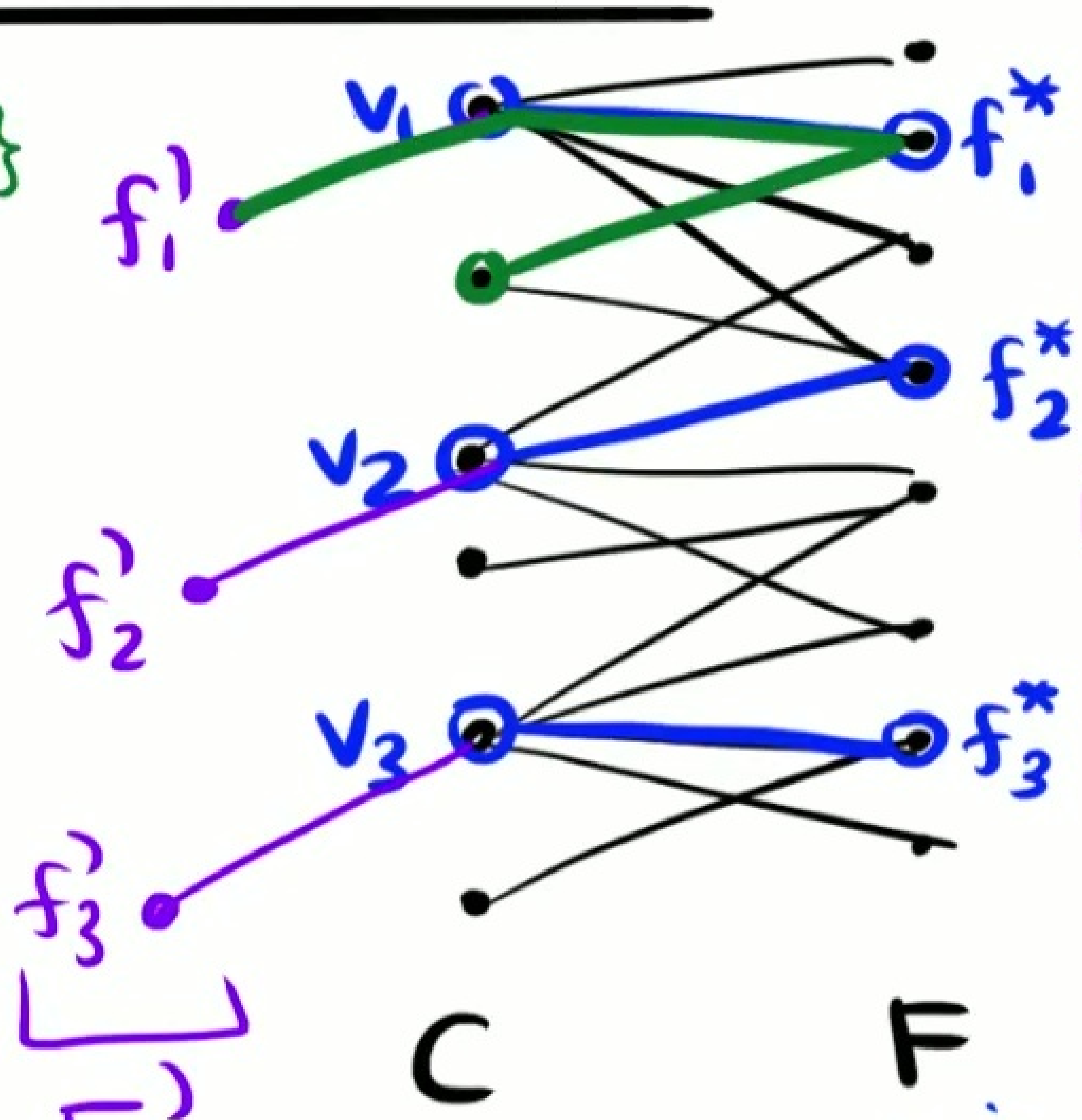
$= d(v,S)$ when $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - d(v, F' \cup S)\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.



Obs: $\forall v$, $d(v,F') \leq 3\, d(v,F^*)$ (same as before)

Thm[CCPV] Can $(1-1/e)$-approximate monotone submodular.

"diminishing marginal returns"

Compute S s.t. $\mathrm{impr}(S) \geq (1-1/e)\,\mathrm{impr}(F^*)$.

$$\mathrm{cost}(S) = \sum_v w(v)\, d(v,S) = \sum_v w(v)\, d(v,F') - \mathrm{impr}(S) \leq \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\mathrm{impr}(F^*)$$

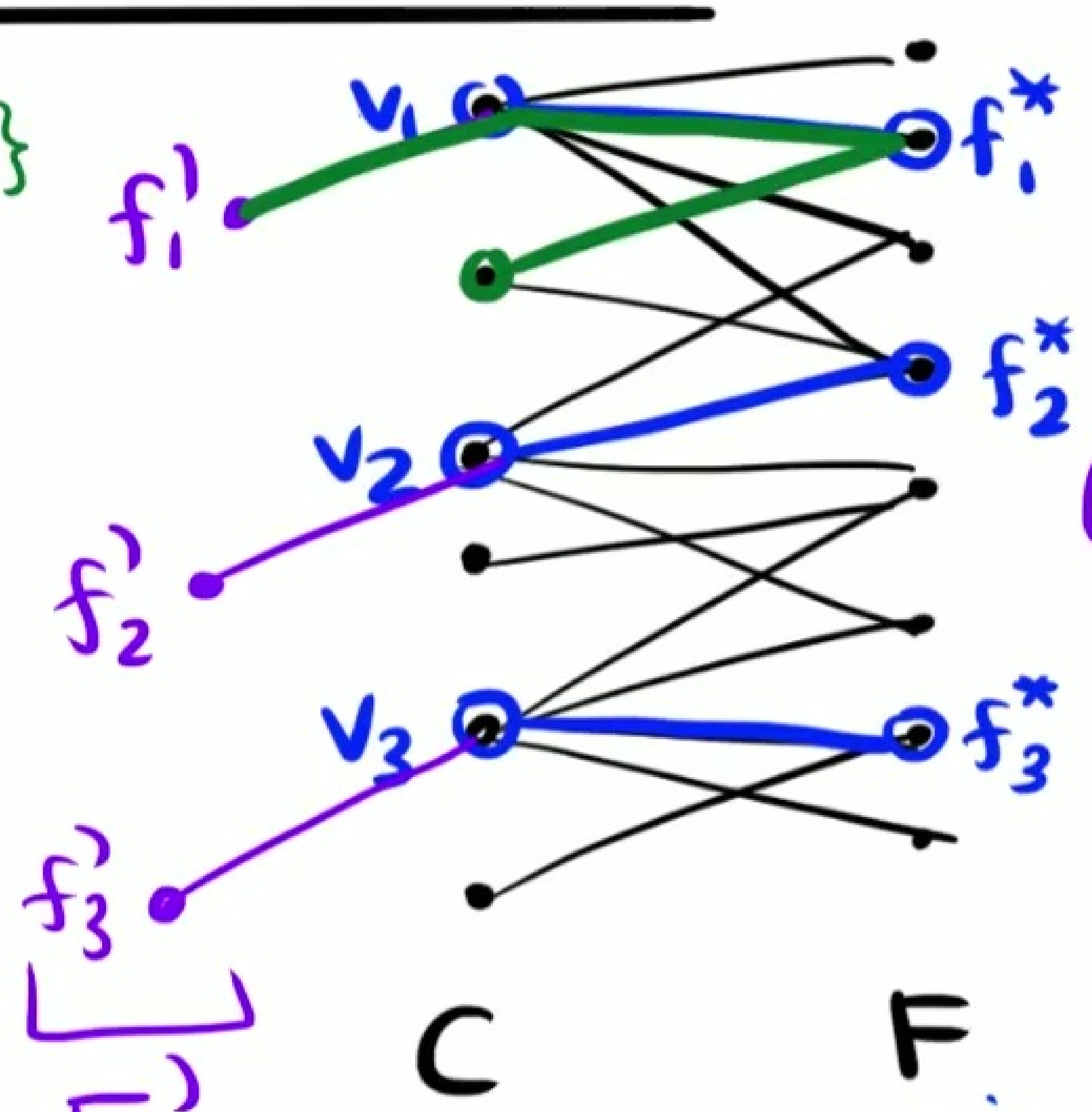
Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - \overbrace{d(v,F'\cup S)}^{= d(v,S) \text{ when } S=\{f_1\ldots f_k,\ f_i\in N(v_i)\}}\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.



Obs: $\forall v$, $d(v,F') \le 3\, d(v,F^*)$ (same as before)

"diminishing marginal returns"

Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

Compute S s.t. $\mathrm{impr}(S) \ge (1-1/e)\,\mathrm{impr}(F^*)$.

$$\mathrm{cost}(S) = \sum_v w(v)\, d(v,S) = \sum_v w(v)\, d(v,F') - \mathrm{impr}(S) \le \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\mathrm{impr}(F^*)$$

$$= \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\left(\sum_v w(v)\left(d(v,F') - d(v,F^*)\right)\right)$$

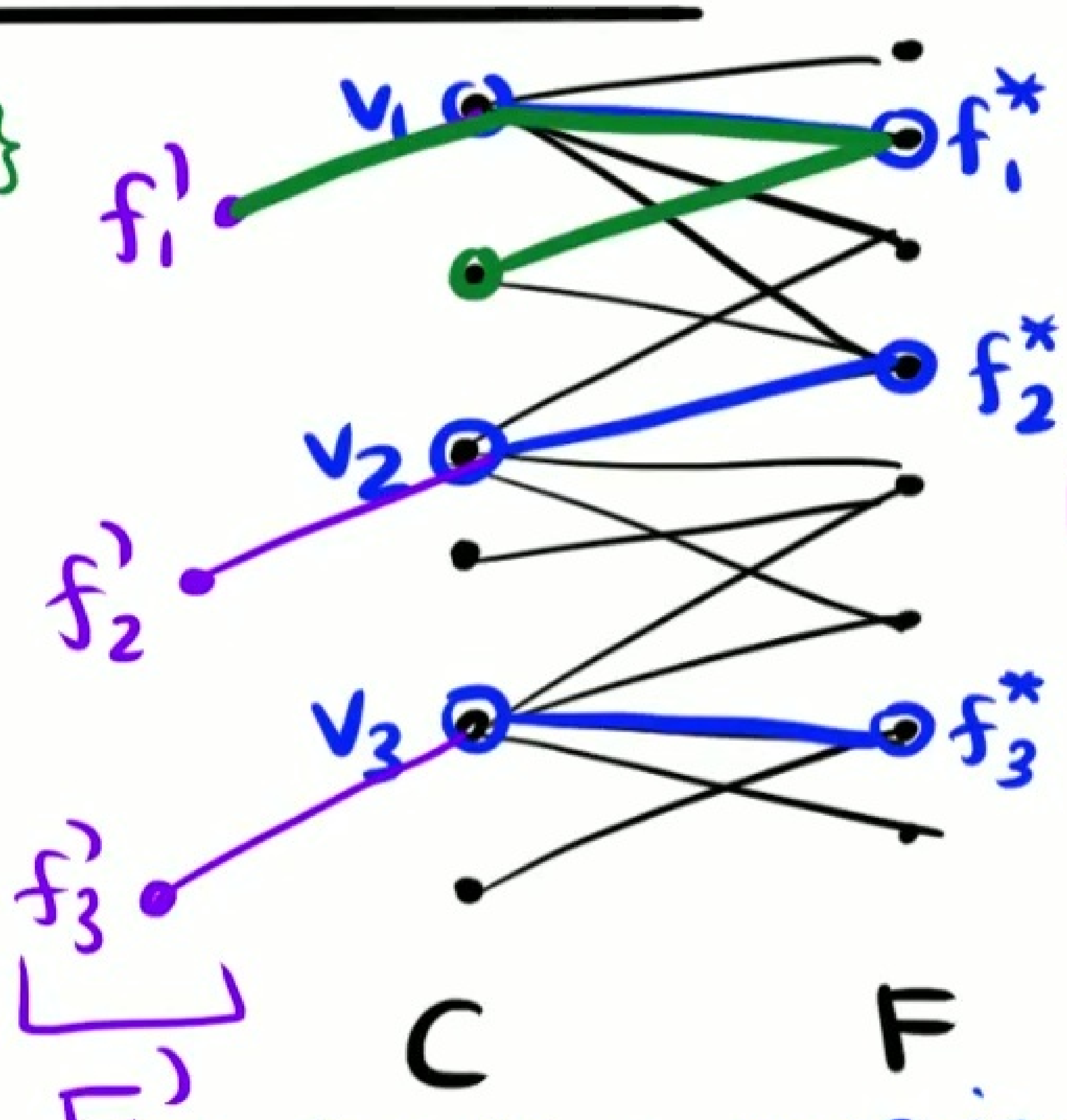
Optimization function

maximize

$$\mathrm{impr}(S) := \sum_v w(v)\left(d(v,F') - \overbrace{d(v,F'\cup S)}^{= d(v,S) \text{ when } S=\{f_1..f_k,\ f_i\in N(v_i)\}}\right)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \dots f_k,\ f_i \in N(v_i)\}$

Lemma: $\mathrm{impr}(S)$ is ① monotone ③ submodular with ② $\mathrm{impr}(\emptyset) = 0$.



Obs: $\forall v$, $d(v,F') \le 3\, d(v,F^*)$ (same as before)

"diminishing marginal returns"

Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

Compute S s.t. $\mathrm{impr}(S) \ge (1-1/e)\,\mathrm{impr}(F^*)$.

$$\mathrm{cost}(S) = \sum_v w(v)\, d(v,S) = \sum_v w(v)\, d(v,F') - \mathrm{impr}(S) \le \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\mathrm{impr}(F^*)$$

$$= \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\left(\sum_v w(v)\left(d(v,F') - d(v,F^*)\right)\right)$$

$$= \tfrac{1}{e}\sum w(v) d(v,F') + \left(1-\tfrac{1}{e}\right)\sum w(v)\, d(v,F^*)$$

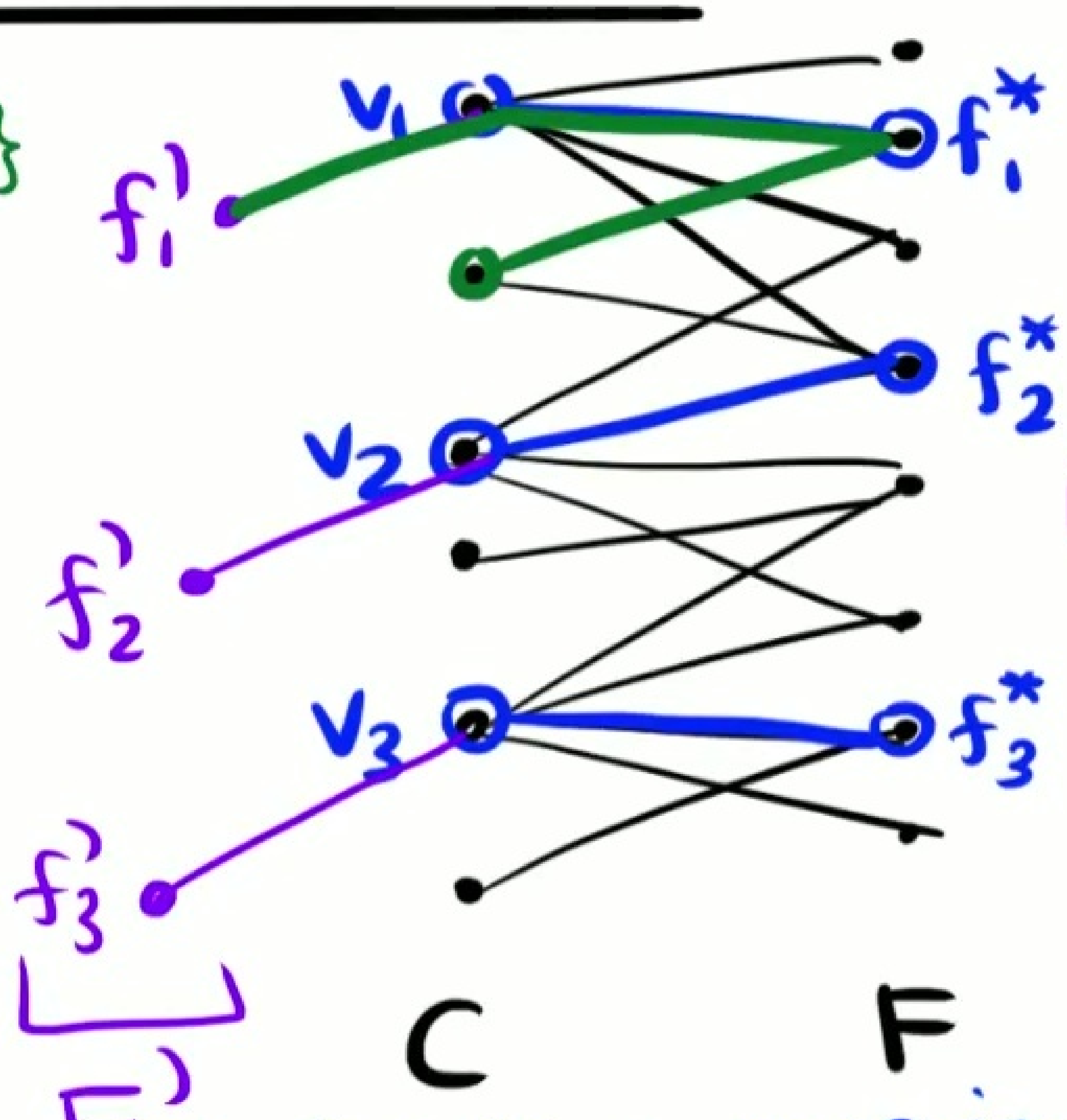
Optimization function

maximize

$$\operatorname{impr}(S) := \sum_v w(v)\Big(d(v,F') - \overbrace{d(v,F'\cup S)}^{= d(v,S) \text{ when } S=\{f_1 \ldots f_k,\ f_i \in N(v_i)\}}\Big)$$

over $S \subseteq \mathbb{F}$, $S = \{f_1 \ldots f_k,\ f_i \in N(v_i)\}$

Lemma: $\operatorname{impr}(S)$ is ① monotone ③ submodular with ② $\operatorname{impr}(\emptyset) = 0$.



Obs: $\forall v$, $d(v,F') \le 3\, d(v,F^*)$ (same as before)

"diminishing marginal returns"

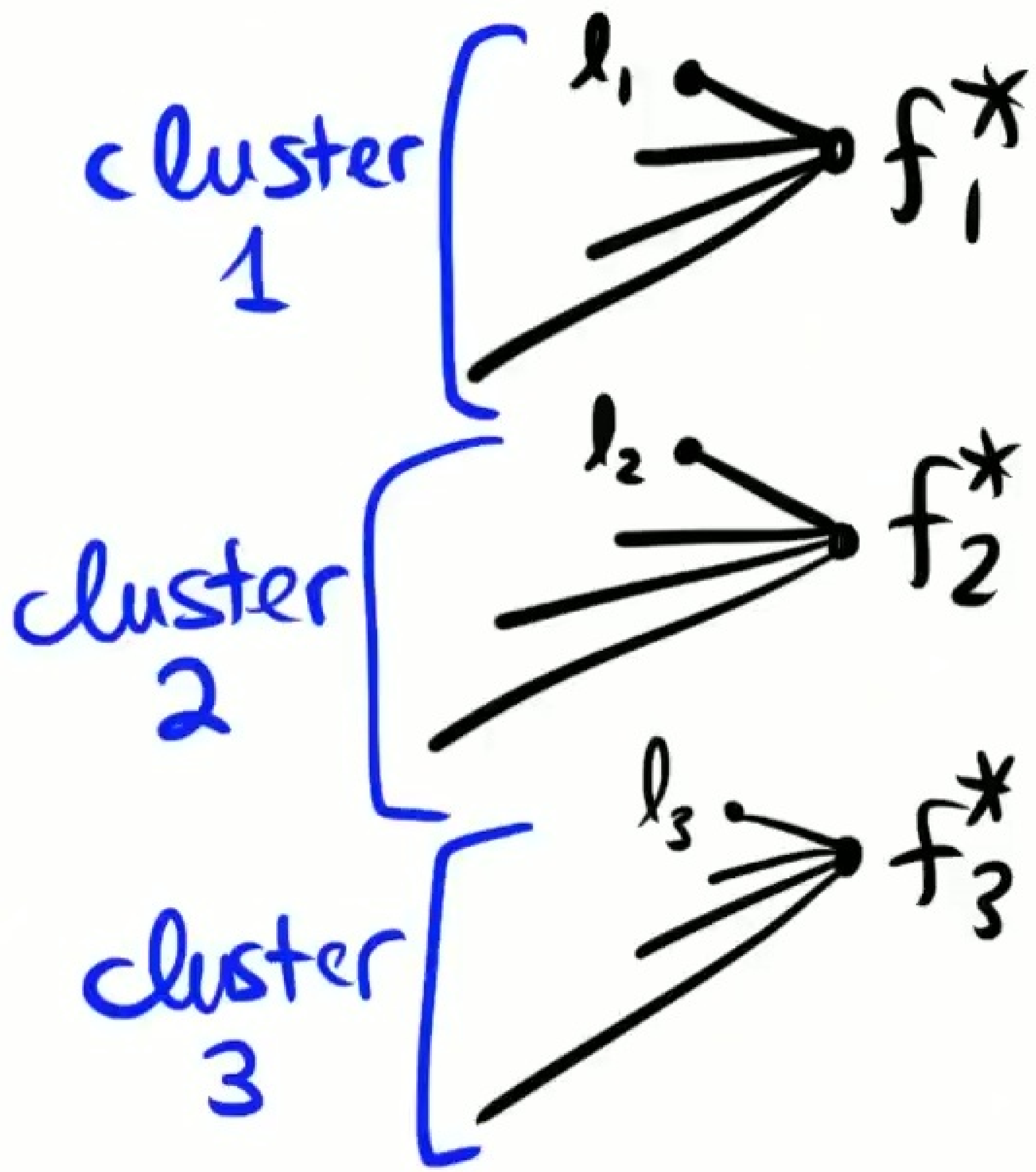
Thm [CCPV] Can $(1-1/e)$-approximate monotone submodular.

Compute S s.t. $\operatorname{impr}(S) \ge (1-1/e)\operatorname{impr}(F^*)$.

$$\begin{aligned}
\operatorname{cost}(S) = \sum_v w(v)\, d(v,S) &= \sum_v w(v)\, d(v,F') - \operatorname{impr}(S) \le \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\operatorname{impr}(F^*) \\
&= \sum_v w(v)\, d(v,F') - \left(1-\tfrac{1}{e}\right)\Big(\sum_v w(v)\big(d(v,F') - d(v,F^*)\big)\Big) \\
&= 1/e \sum w(v) d(v,F') + \left(1-\tfrac{1}{e}\right)\sum w(v)\, d(v,F^*) \\
&\le 3/e \sum w(v) d(v,F^*) + \left(1-\tfrac{1}{e}\right)\sum w(v)\, d(v,F^*)
\end{aligned}$$

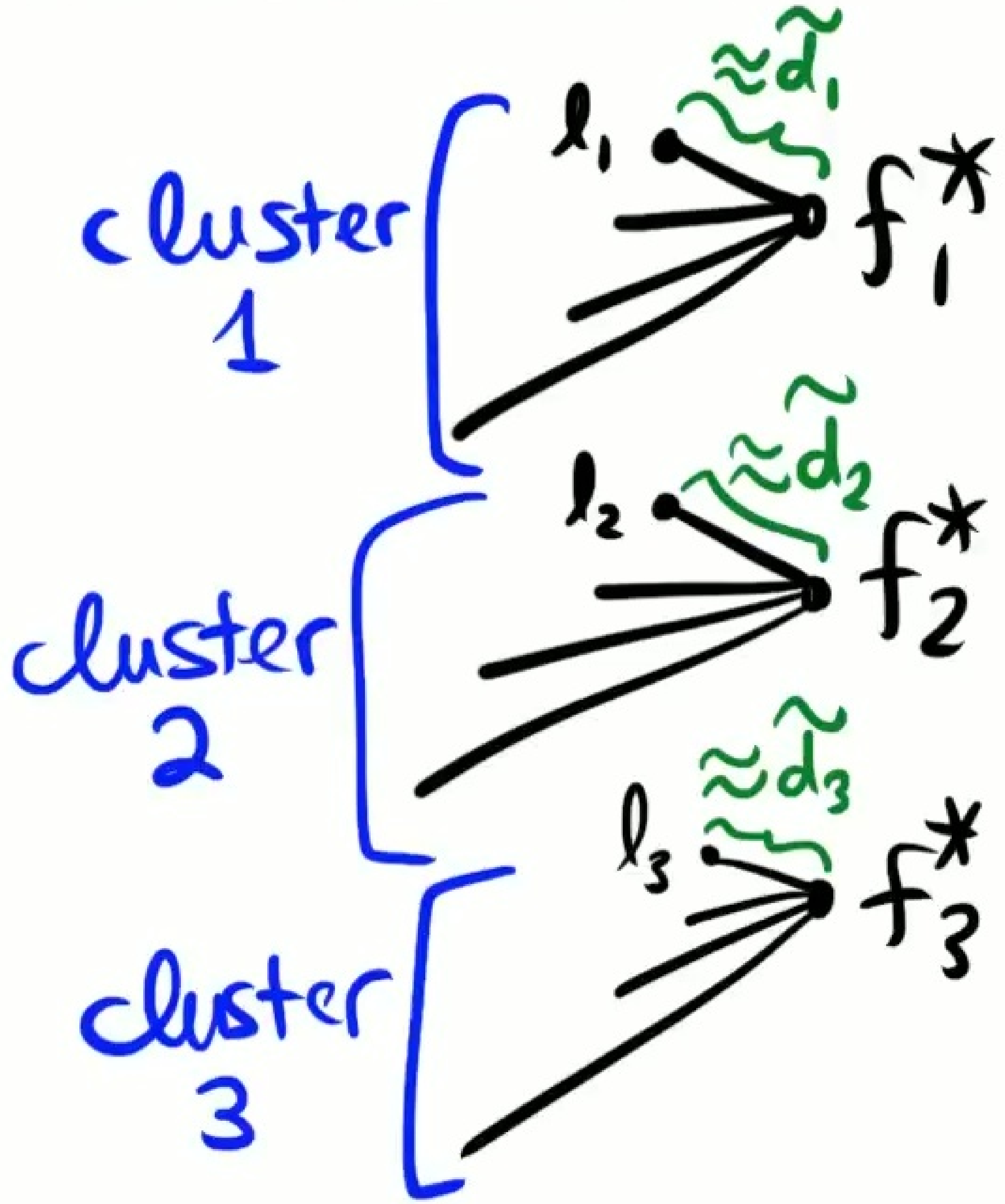
General case

- cluster i : all clients assigned to f_i^* in OPT
- l_i := closest client to f_i^* in cluster i



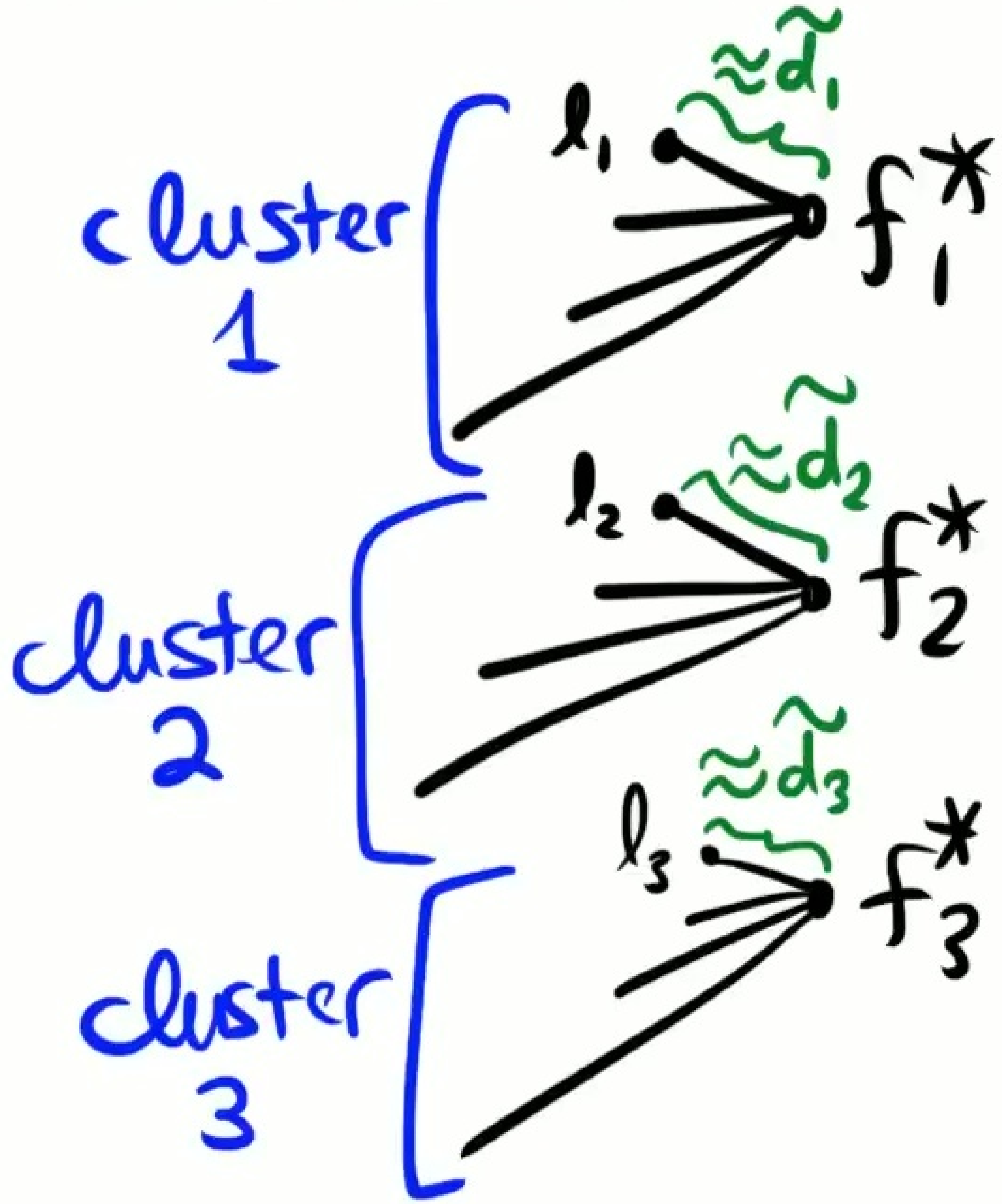
General case

- cluster i: all clients assigned to f_i^* in OPT
- $l_i :=$ closest client to f_i^* in cluster i
- Guess $l_1 \dots l_K$ ((coreset size)^K choices)
- Guess $\tilde{d}_i \approx d(l_i, f_i^*)$ up to factor $(1+\varepsilon)$



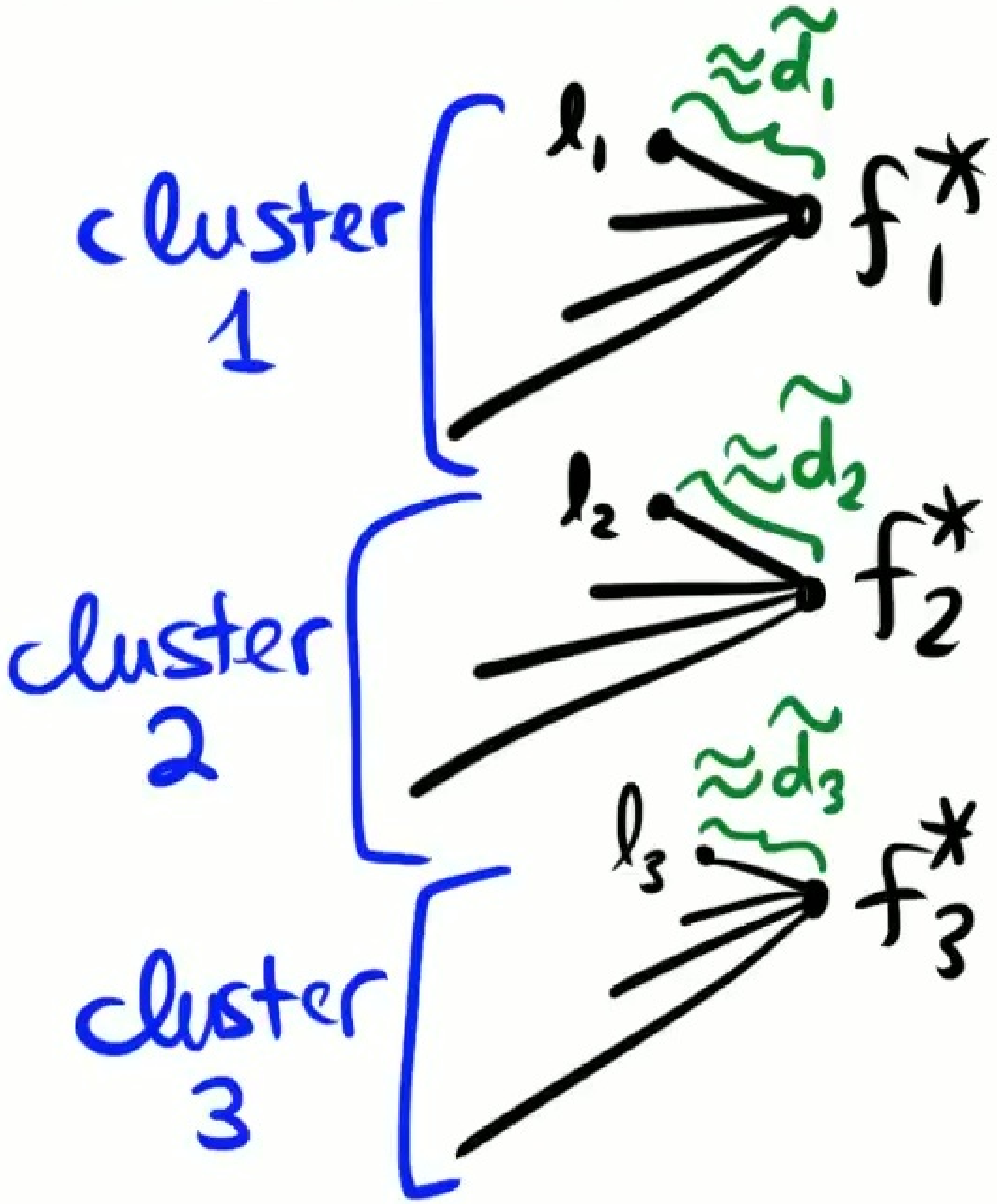
General case

- cluster i: all clients assigned to f_i^* in OPT
- $l_i :=$ closest client to f_i^* in cluster i
- Guess $l_1 \dots l_K$ ($(\text{coreset size})^K$ choices)
- Guess $\tilde{d}_i \approx d(l_i, f_i^*)$ up to factor $(1+\varepsilon)$
- Define $N(l_i) := \{ f \in \mathbb{F} : d(l_i, f) \underset{(1+\varepsilon)}{\approx} \tilde{d}_i \}$
  [so $f_i^* \in N(l_i)$]



General case

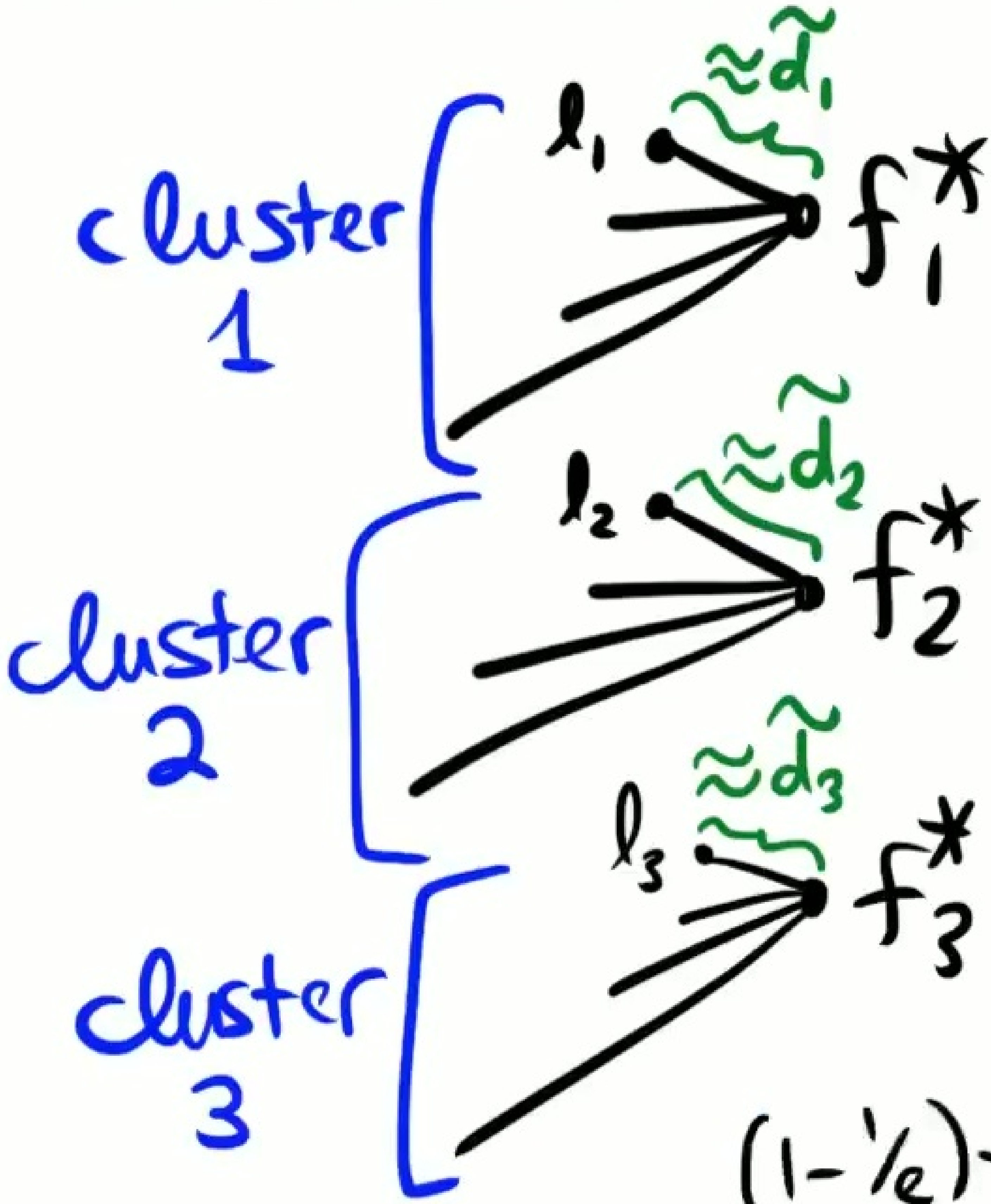
- cluster i: all clients assigned to f_i^* in OPT
- $\ell_i :=$ closest client to f_i^* in cluster i
- Guess $\ell_1 \dots \ell_K$ ($(\text{coreset size})^K$ choices)
- Guess $\tilde{d}_i \approx d(\ell_i, f_i^*)$ up to factor $(1+\varepsilon)$
- Define $N(\ell_i) := \{f \in \mathbb{F} : d(\ell_i, f) \underset{(1+\varepsilon)}{\approx} \tilde{d}_i\}$
  [so $f_i^* \in N(\ell_i)$]



- Same $\text{impr}(S) := \sum w(v)\big(d(v,F') - d(v, F' \cup S)\big)$
  - $\text{impr}(S)$ is monotone submodular
  - $d(v, F') \le 3\, d(v, F^*)$
  - if $S = \{f_1 \dots f_K\}$ where $f_i \in N(\ell_i)$, then $d(v, F' \cup S) = d(v, S)$

General case

- cluster i: all clients assigned to f_i^* in OPT
- $l_i :=$ closest client to f_i^* in cluster i
- Guess $l_1 \dots l_K$ ((coreset size)^K choices)
- Guess $\tilde{d}_i \approx d(l_i, f_i^*)$ up to factor $(1+\varepsilon)$
- Define $N(l_i) := \{ f \in \mathbb{F} : d(l_i, f) \underset{(1+\varepsilon)}{\approx} \tilde{d}_i \}$
  [so $f_i^* \in N(l_i)$]



- Same $\mathrm{impr}(S) := \sum w(v)\big(d(v, F') - d(v, F' \cup S)\big)$
  - $\mathrm{impr}(S)$ is monotone submodular
  - $d(v, F') \le 3\, d(v, F^*)$
  - if $S = \{f_1 \dots f_K\}$ where $f_i \in N(l_i)$, then $d(v, F' \cup S) = d(v, S)$

$(1 - 1/e)$-apx to $\mathrm{impr}(S)$

$\Rightarrow (1 + 2/e + \varepsilon)$-apx to k-median

Open problems

- Can FPT techniques be used for <u>polytime</u> k-median?
  - Fallback of 3-approx, submod opt
- Other variants?
  - Capacitated k-median: upper bound 3+eps
    Lower bound 1+2/e
  - Matroid median: upper bound 2+eps
    Lower bound 1+2/e