EC'24 Tutorial

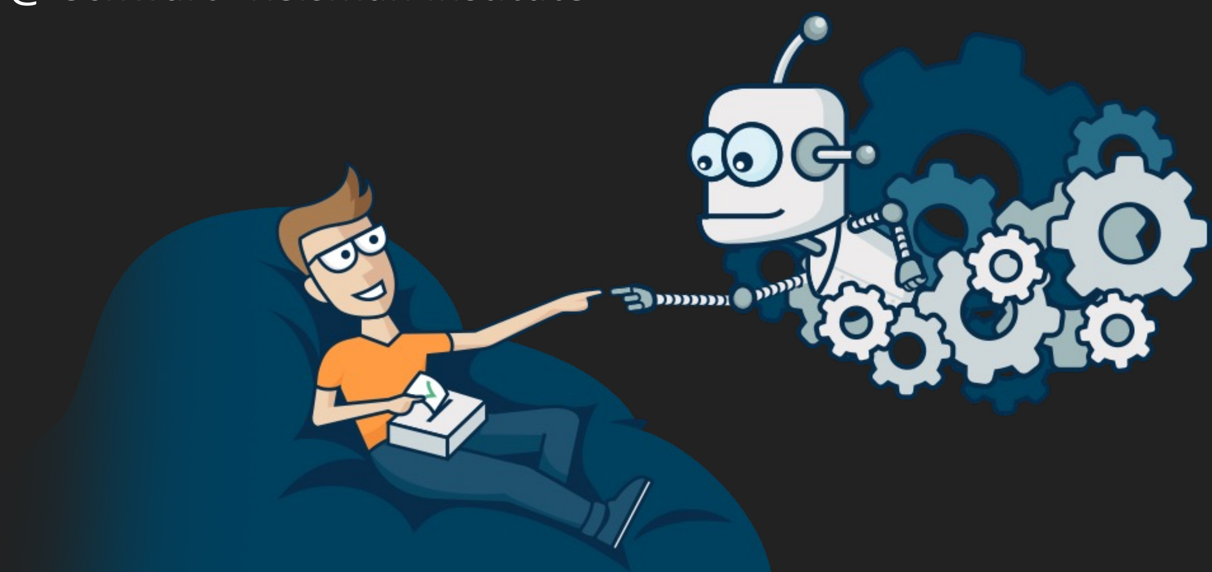
Fairness in AI/ML via Social Choice

Evi Micha

Postdoctoral Fellow @ Harvard University
Assistant Professor (soon) @ University of Southern California

Nisarg Shah

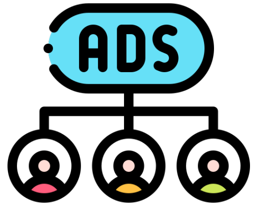
Associate Professor @ University of Toronto
Research Lead @ Schwartz Reisman Institute

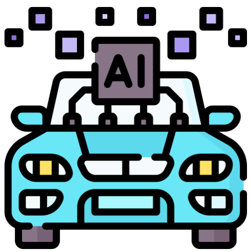


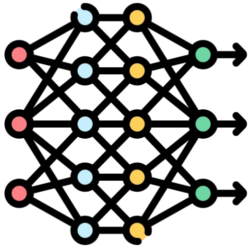
Disclaimers

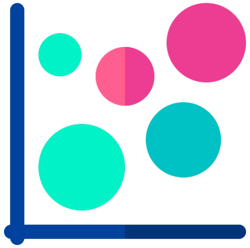
- Focused on conceptual applications
  - No proofs
- Covers many different settings
  - Can’t define them all super formally
- Covers multidisciplinary work
  - Simplifications galore
  - (And possible errors)

Why fairness?

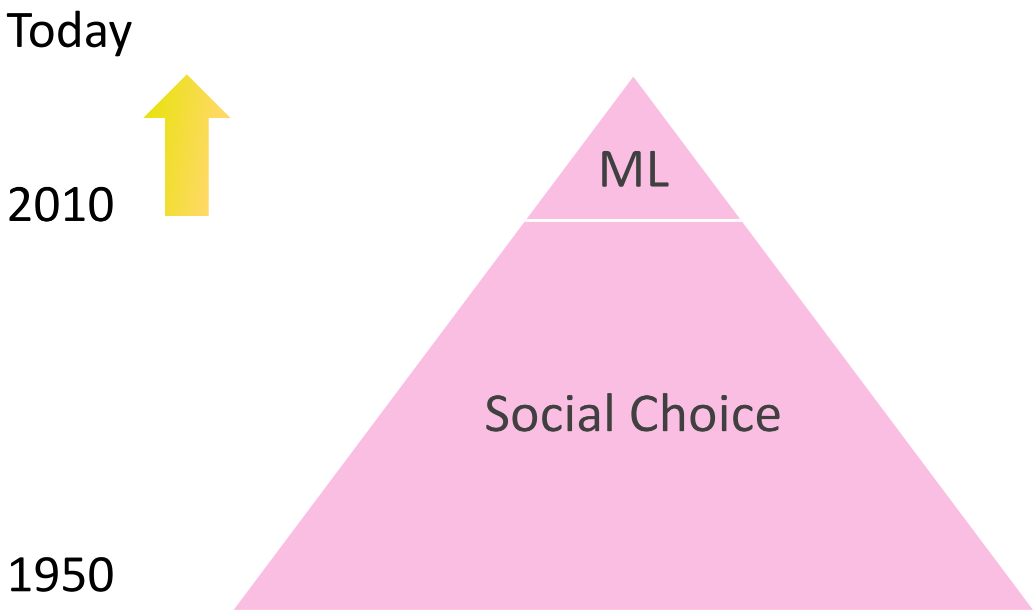


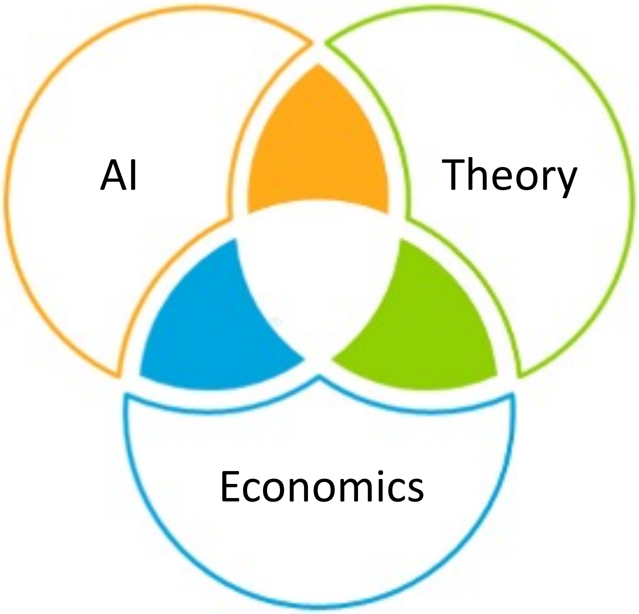






Fairness research





Social Choice Theory

Aggregating individual preferences into “good” collective decisions

Outline

- Introduction
  - Fairness in social choice
- Envy-freeness
  - Classification, recommender systems, clustering
- Nash social welfare
  - Multi-armed bandits, rankings, classification
- Core
  - Federated learning, clustering

Fairness in Social Choice

Example: Allocating Divisible Resources

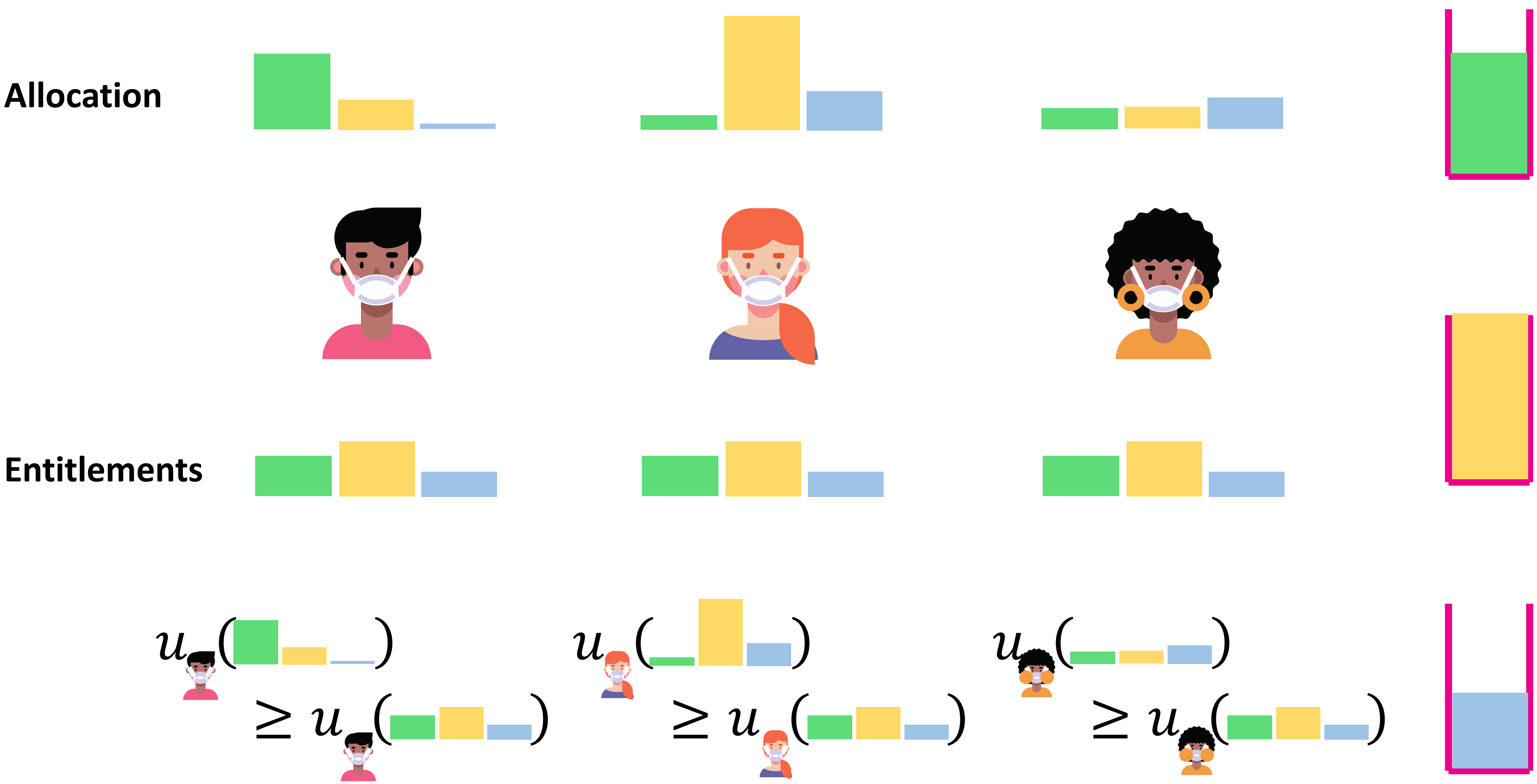
- Set of agents N
- Set of divisible resources M
- Each agent $i \in N$ has an additive linear utility function $u_i: 2^M \to \mathbb{R}$
  - For $X \in [0,1]^M$, $u_i(X) = \sum_{g \in M} u_{i,g} \cdot X_g$
- Non-negative utilities $u_{i,g} \geq 0$ vs non-positive utilities $u_{i,g} \leq 0$ (i.e., costs)
  - Often stronger results for the former in social choice, but the latter more prevalent in ML

Example: Allocating Divisible Resources

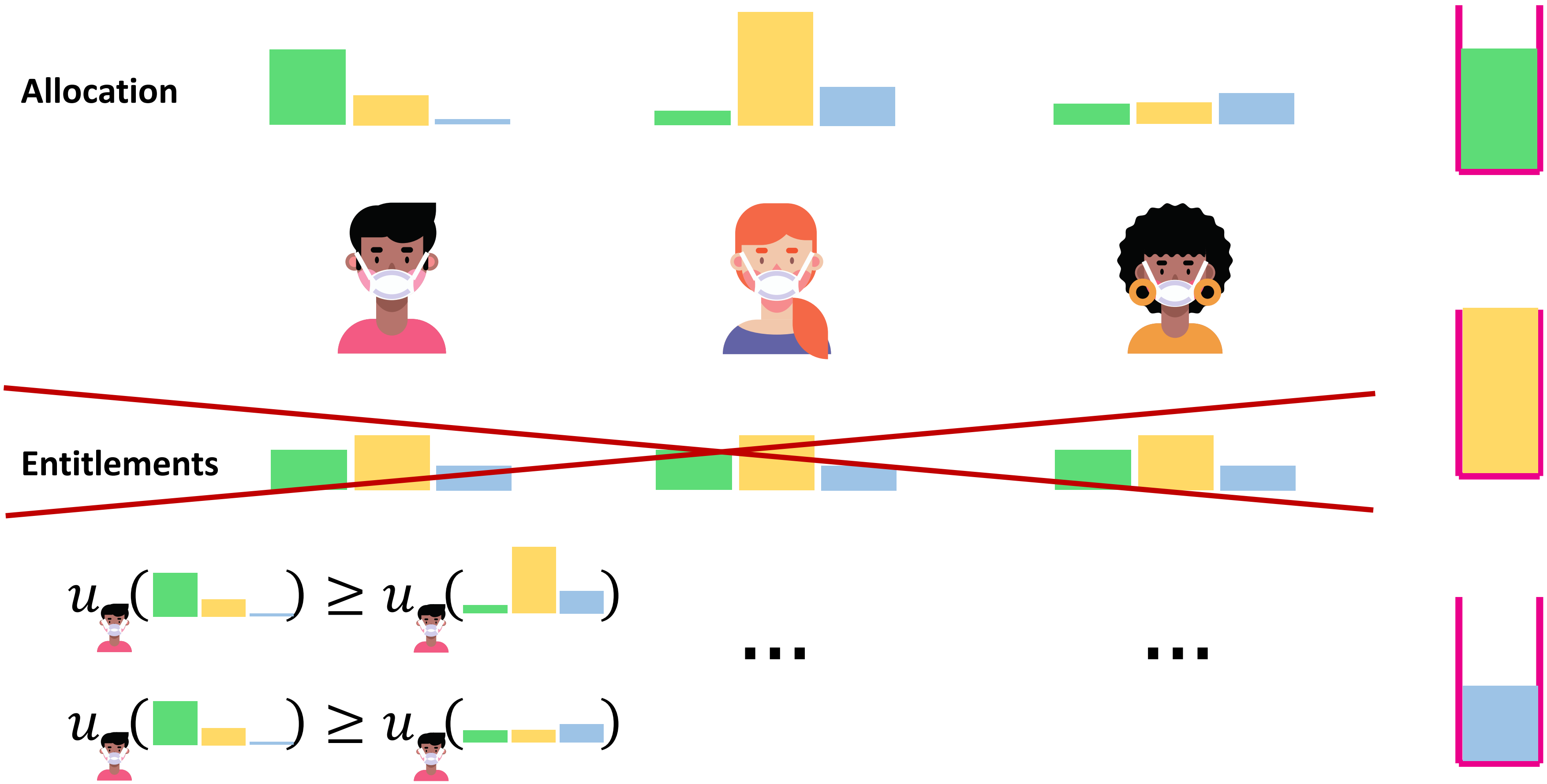
- Allocation A
  - $A_{i,g}$ = fraction of good g allocated to agent i
  - $\sum_i A_{i,g} \le 1, \forall g \in M$
  - Complete allocation: $\sum_i A_{i,g} = 1, \forall g \in M$

- Utility to agent i under allocation A is $u_i(A_i)$
  - But for various fairness definitions, other terms such as $u_i(A_j)$, $u_i(\cup_{j \in S} A_j)$, etc. may also matter

Proportionality



Envy-Freeness



Proportionality and Envy-Freeness

- Proportionality
  - "My utility for my allocation should be at least as much as my utility for my entitled (equal) share of the resources"
  - $u_i(A_i) \geq u_i\left(\frac{1}{n} \cdot M\right) = \frac{1}{n} \cdot u_i(M), \forall i \in N$

- Envy-freeness
  - "My utility for my allocation should be at least as much as my utility for anyone else's allocation"
  - $u_i(A_i) \geq u_i(A_j), \forall i, j \in N$

Proportionality and Envy-Freeness

- Proportionality
  - $u_i(A_i) \geq u_i\left(\frac{1}{n} \cdot M\right) = \frac{1}{n} \cdot u_i(M), \forall i \in N$

- Envy-freeness
  - $u_i(A_i) \geq u_i(A_j), \forall i, j \in N$

- Question: For an allocation A, which of the following is always true?
  - *a)* A is proportional $\Rightarrow$ A is envy-free
  - *b)* A is envy-free $\Rightarrow$ A is proportional
  - c) Both (equivalent)
  - d) Neither (incomparable)

Proportionality and Envy-Freeness

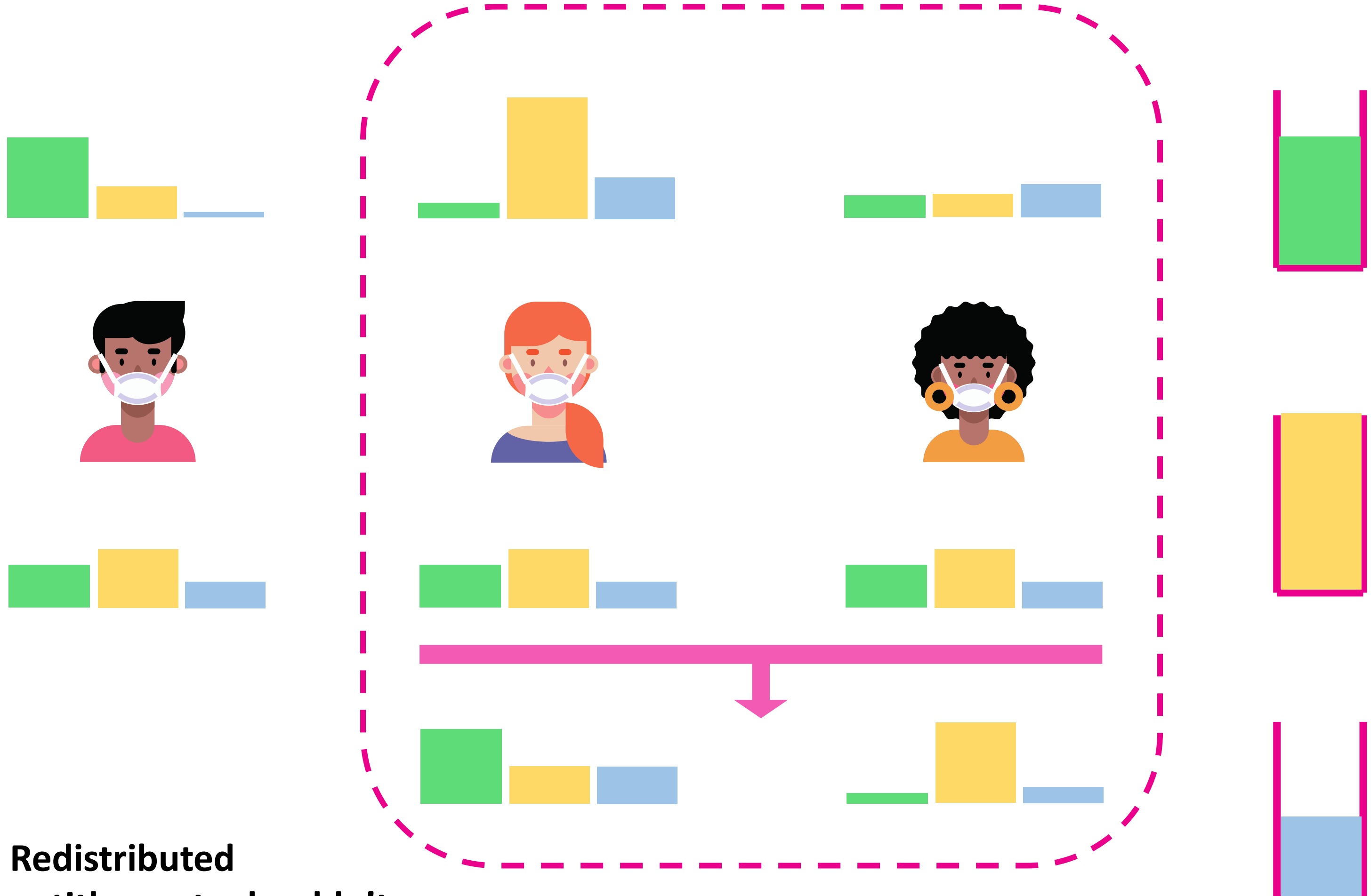
- Proportionality
  - $u_i(A_i) \geq u_i\left(\frac{1}{n} \cdot M\right) = \frac{1}{n} \cdot u_i(M), \forall i \in N$
- Envy-freeness
  - $u_i(A_i) \geq u_i(A_j), \forall i, j \in N$
- Question: For a complete allocation A, which of the following is always true?
  - $a)$ A is proportional $\Rightarrow$ A is envy-free
  - $b)$ A is envy-free $\Rightarrow$ A is proportional
  - c) Both (equivalent)
  - d) Neither (incomparable)

The Core

Allocation

Entitlements

Any groups can redistribute its entitlements



Redistributed entitlements shouldn't be a Pareto improvement

The Core

- Resource-scaling version
  - "No group of agents S should be able to find any allocation B of their proportionally entitled share of the resources that is a Pareto improvement"
  - $\nexists S \subseteq N, \frac{|S|}{|N|} \cdot M \to B : u_i(B_i) > u_i(A_i), \forall i \in S$
- Utility-scaling version
  - "No group of agents S should be able to find any allocation B of the resources that is a Pareto improvement even after proportional utility-scaling"
  - $\nexists S \subseteq N, M \to B : \frac{|S|}{|N|} \cdot u_i(B_i) > u_i(A_i), \forall i \in S$
- Comparison
  - The two versions are equivalent for our example setting
  - But they're different when utilities aren't linear additive
  - Resource-scaling version may not be defined if there is no "scalable" resource, but when defined, it's often more appealing

Nash Social Welfare

- Nash Social Welfare
  - $\mathrm{NSW}(A) = \left(\prod_{i \in N} u_i(A_i)\right)^{1/|N|}$
  - Geometric mean of agent utilities
  - Often more appealing fairness guarantees than other popular welfare functions
    - Utilitarian social welfare: $\mathrm{USW}(A) = \frac{1}{|N|} \cdot \sum_{i \in N} u_i(A_i)$
    - Egalitarian social welfare: $\mathrm{ESW}(A) = \min_{i \in N} u_i(A_i)$
- Theorem [Varian '74]:
  - Any allocation maximizing the Nash social welfare is envy-free and in the core.
- Theorem [Orlin '10]:
  - An allocation maximizing the Nash social welfare can be computed in strongly polynomial time.

Core ⇒ Committee Selection

- Setup
  - Set of voters N, set of candidates M
  - Each agent i approves a subset of candidates $A_i \subseteq M$
    - For any $W \subseteq M$, $u_i(W) = |W \cap A_i|$ ("number of candidates I approve")
  - Goal: Find $W \subseteq M$ with $|W| \leq k$ (where k is given)

- Resource-scaling version
  - W is in the core if…
  - there is no $S \subseteq N$ and $T \subseteq M$ with $|T| \leq \frac{|S|}{|N|} \cdot k$ such that…
  - $u_i(T) > u_i(W), \forall i \in S$

- Open question: Does a committee in the core always exist?
  - A variety of constant approximations provably exist even in more general settings

Advantages

- Key advantages of social choice fairness criteria

- Broadly defined
  - Often depend only on the definition of *who* the agents are and *what* their preferences are
  - Applicable to any setting as long as you define these two pieces of information

- They respect the preferences of the agents to whom we wish to be fair
  - As a consequence, they are often defined beyond just binary decisions

- Notions such as the core achieve group fairness to all possible groups
  - No need to pre-specify the groups
  - The strength of the guarantee scales automatically with the group size and cohesiveness, without having to subjectively choose free parameter values

Envy-Freeness in ML

Classification

- Model
  - Population of individuals given by a distribution D over X
    - Individual i represented using data point $x_i \in X$
  - Classifier $f: X \rightarrow Y$ maps every individual to a classification outcome

- Types of classification outcomes
  - Hard binary classification: $Y = \{0,1\}$
  - Hard multiclass classification: $|Y| = p > 2$
  - Soft binary classification: $Y = [0,1]$
  - Soft multiclass classification: $Y \in \mathbb{R}^p, p > 2$

Classification

- Objective of the principal: minimize the loss $\mathbb{E}_{x\sim D}\left[\ell\left(x, f(x)\right)\right]$
  - If $f(x)$ is a distribution, $\ell\left(x, f(x)\right) = \mathbb{E}_{y\sim f(x)}[\ell(x, y)]$

- Utility function $u: X \times Y \to \mathbb{R}_{\geq 0}$
  - Utility to individual i is $u\left(x_i, f(x_i)\right)$

- Fairness is often modeled as a constraint that uses the utility function u

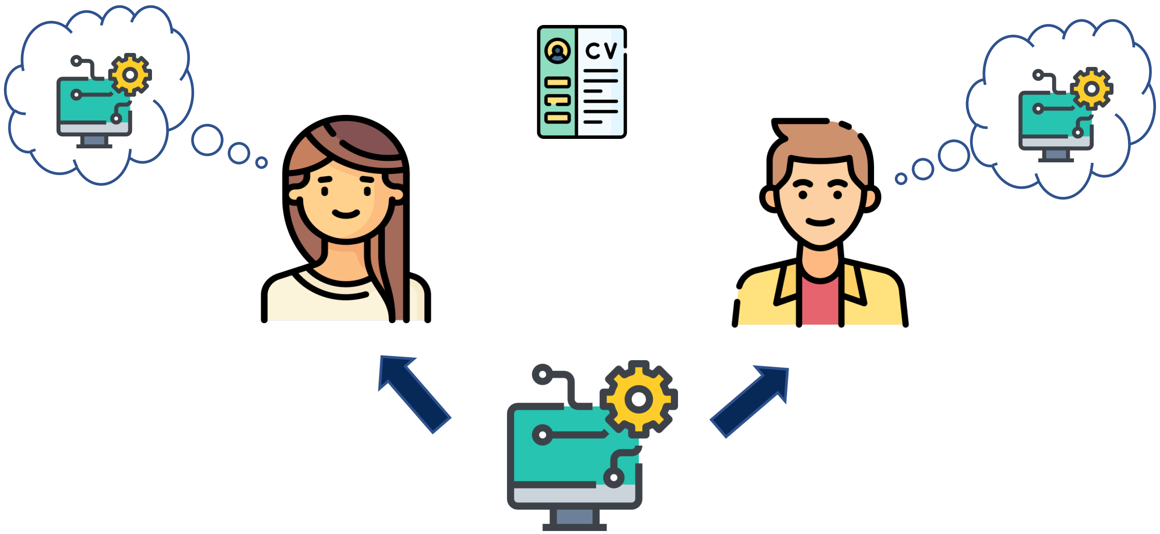
Individual Fairness

[Dwork, Hardt, Pitassi, Reingold, Zemel, 2012]

"Similar individuals should be treated similarly"

Classifier f is individual fair if:

$$\forall x, y \in N,\ D\big(f(x), f(y)\big) \le d(x, y)$$



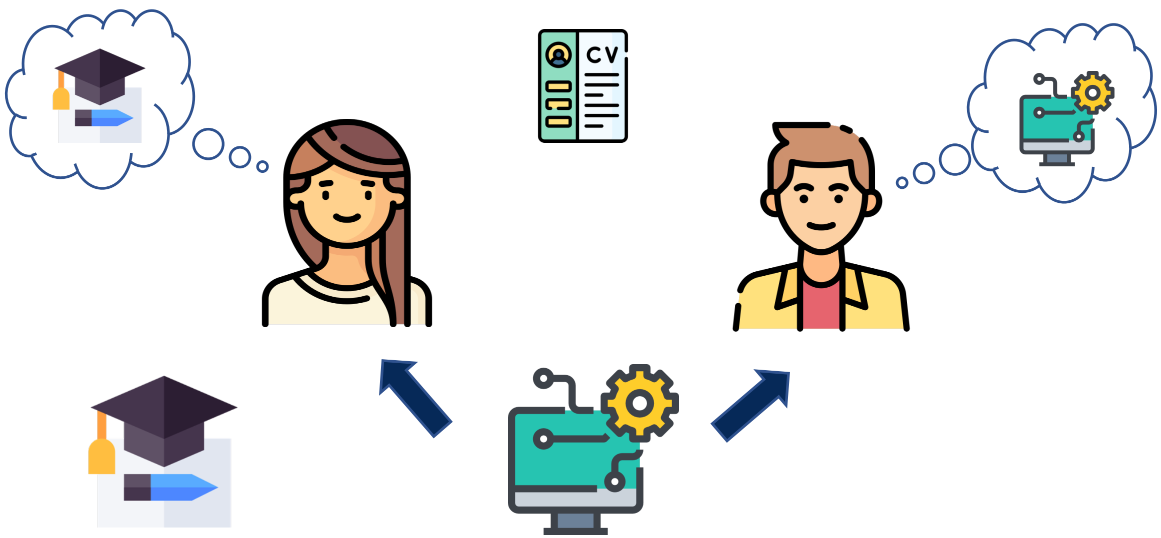
Individual Fairness

[Dwork, Hardt, Pitassi, Reingold, Zemel, 2012]

"Similar individuals should be treated similarly"

Classifier f is individual fair if:

$$\forall x, y \in N,\ \ D\big(f(x), f(y)\big) \leq d(x, y)$$



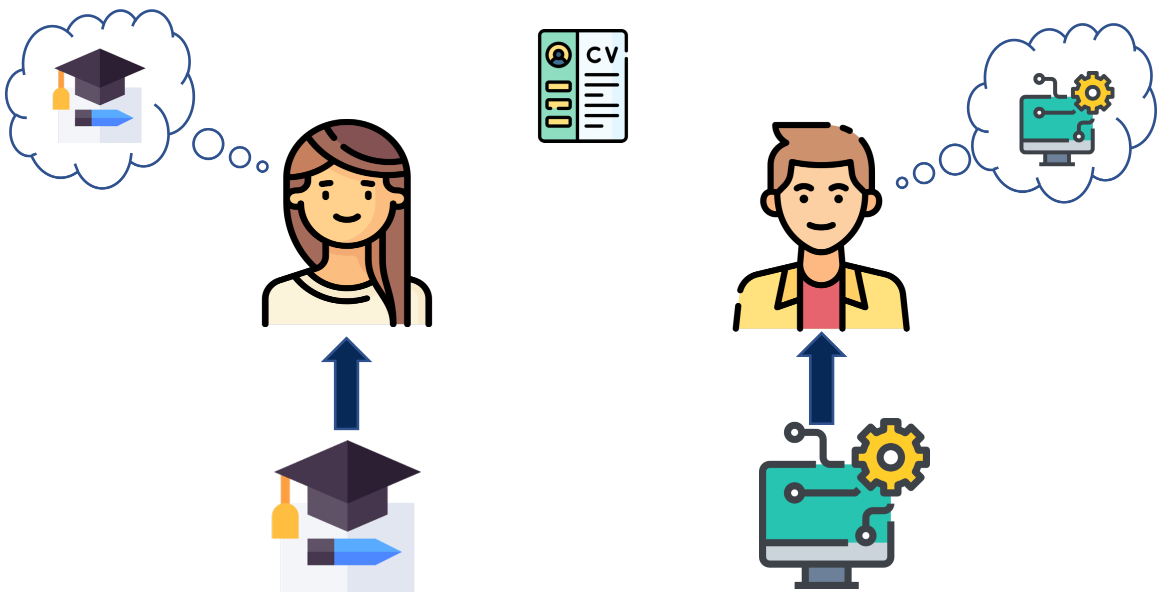
Envy-Freeness

[Balcan, Dick, Noothigattu, Procaccia, 2019]

"Equal individuals shouldn't envy each other"

Classifier f is envy-free if:

$$\forall x, y \in N,\ u_x(f(x)) \geq u_x(f(y))$$



Envy-Freeness

[Balcan, Dick, Noothigattu, Procaccia, 2019]

- Observation: Envy-freeness is too strong for deterministic classifiers
  - Loss of optimal deterministic EF classifier ≥ 1

Utilities

	1	0	$\boxed{0}$
	$\boxed{1/\gamma}$	1	0

Losses

	1	0	$\boxed{0}$
	$\boxed{0}$	1	1

Envy-Freeness

[Balcan, Dick, Noothigattu, Procaccia, 2019]

- Observation: Envy-freeness is too strong for deterministic classifiers
  - Loss of optimal randomized EF classifier $\leq {}^{1}/_{\gamma}$

Utilities

	1 (1/γ)	0	0 (1-1/γ)
	1/γ	1	0

Losses

	1 (1/γ)	0	0 (1-1/γ)
	0	1	1

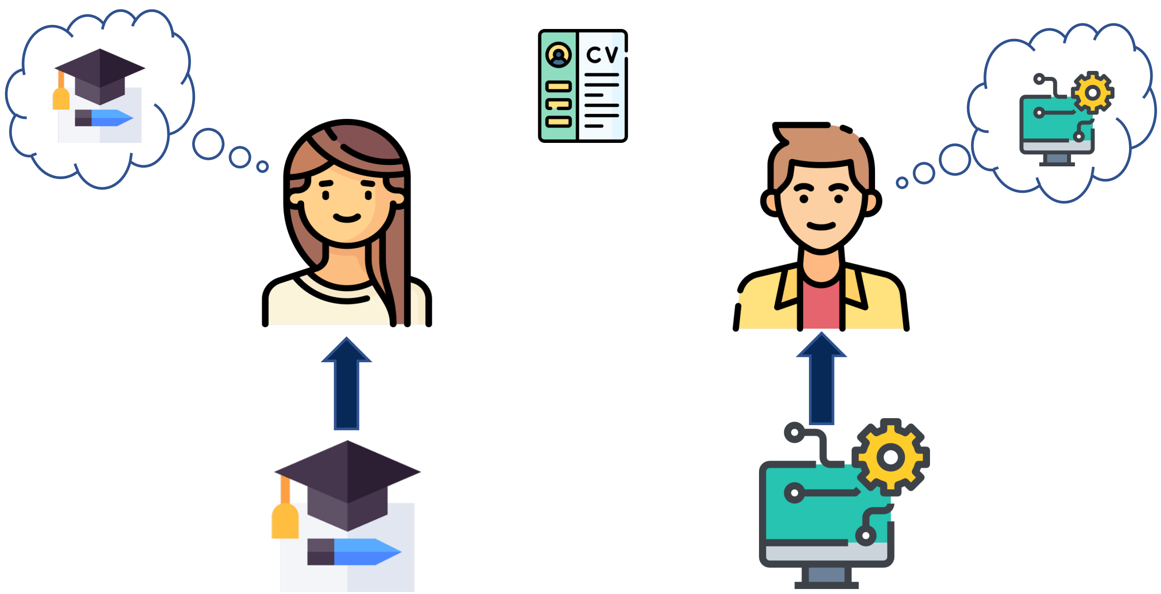
Preference-Informed IF

[Kim, Korolova, Rothblum, Yona, 2019]

“Similar individuals shouldn’t envy each other too much”

Classifier f is PIIF if:

$$\forall x, y \in N, \exists z \in Y, D\big(z, f(y)\big) \leq d(x, y) \wedge u_x\big(f(x)\big) \geq u_x(z)$$

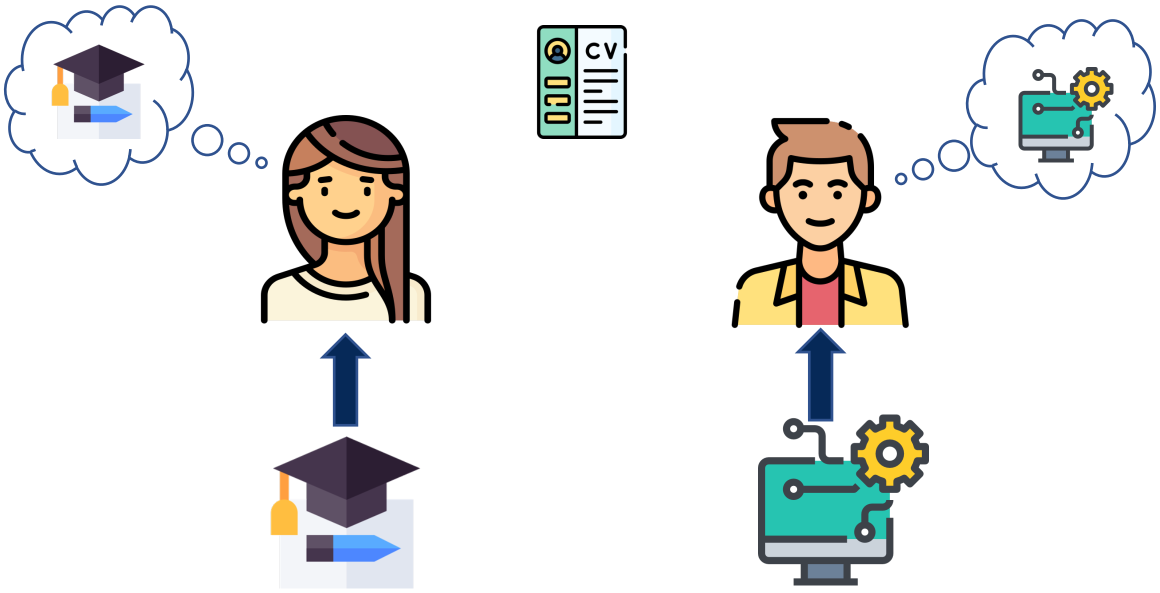


Approximate EF

"<u>Equal</u> individuals shouldn't envy each other <u>too much</u>"

Classifier f is approximately EF if:

$$\forall x, y \in N,\ u_x\big(f(x)\big) \geq u_x\big(f(y)\big) - \varepsilon$$

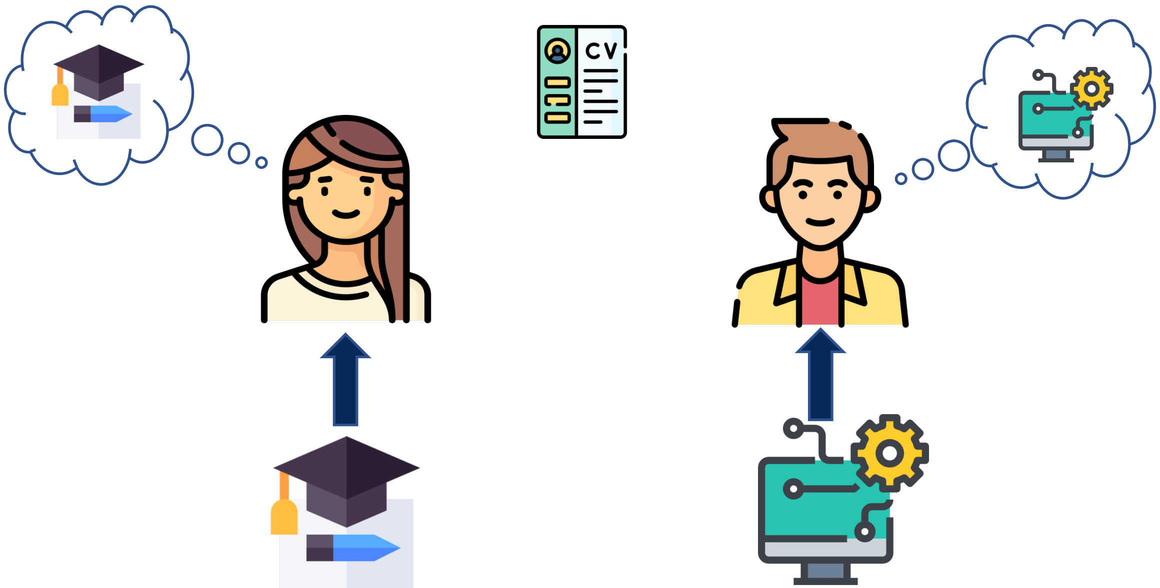


Approximate EF

"Similar individuals shouldn't envy each other too much"

Classifier f is approximately EF if:

$$\forall x, y \in N,\ u_x\big(f(x)\big) \geq u_x\big(f(y)\big) - d(x, y)$$



Average Group Envy-Freeness

[Hossain, Mladenovic, S, 2020]

"Equal groups shouldn't envy each other too much on average"

Classifier f is apx. average group EF w.r.t. groups G_1 and G_2 if:

$$\mathbb{E}_{x\sim G_1, y\sim G_2}\left[u_x\big(f(y)\big) - u_x\big(f(x)\big)\right] \leq 0$$

- Applicable for decision-making with limited resources
  - E.g., deciding on loan or bail applications
  - Envy cannot be prevented
- Can be imposed for several pairs of groups simultaneously
  - # training samples needed depends on the complexity of the family of classifiers and log(#pairs of groups)

Average Group Envy-Freeness

[Hossain, Mladenovic, S, 2020]

"Equal groups shouldn't envy each other too much on average"

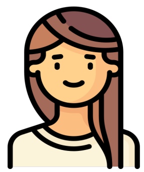
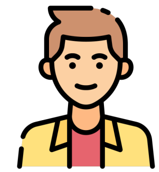
Classifier f is apx. average group EF w.r.t. groups G_1 and G_2 if:

$$\mathbb{E}_{x \sim G_1, y \sim G_2}\left[u_x\big(f(y)\big) - u_x\big(f(x)\big)\right] \leq 0$$

- Extends various traditional ML fairness definitions
  - G_b, G_w groups based on a sensitive attribute
  - '+' deserves positive treatment (utility 1), '-' does not
  - Demographic parity: $(G_w, G_b), (G_b, G_w)$
  - Equal opportunity: $(G_w^+, G_b^+), (G_b^+, G_w^+)$
  - Equalized odds: $(G_w^+, G_b^+), (G_b^+, G_w^+), (G_w^-, G_b^-), (G_b^-, G_w^-)$
  - Average group EF extends these notions from the limited case of binary classification + binary utilities to the general case

Envy-Freeness ⇒ Groups Revisited

[Ustun, Liu, Parkes, 2019]

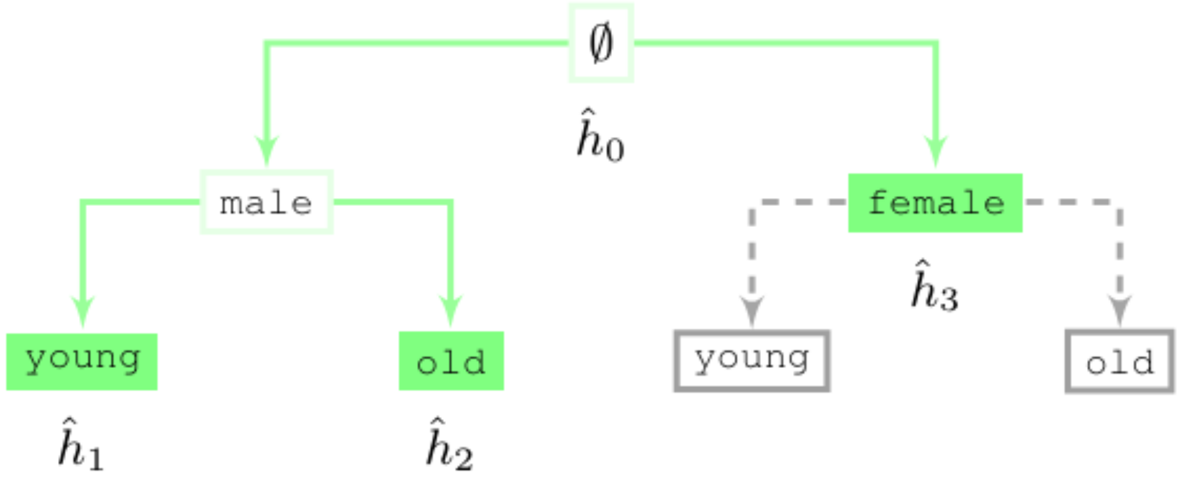
  

When different groups require different treatments

- Idea: Train a different classifier for each group
- Problem: It may harm groups from which we do not have sufficient data
- Goal: Collectively train decoupled classifiers (one for each group) such that each group prefers (in the average envy-freeness sense) its own classifier to
  - A pooled classifier that ignores group membership (individual rationality)
  - The classifier assigned to any other group (envy-freeness)

Envy-Freeness ⇒ Groups Revisited

[Ustun, Liu, Parkes, 2019]



- Example with three groups: (male, young), (male, old), (female)
  - No group should prefer $\hat{h}_0$ or the classifier of another group to their own

- Generalization: #training samples needed depends on the complexity of the family of classifiers and log(#pairs of groups)

Envy-Freeness ⇒ Recommendations

Envy-Freeness ⇒ Recommendations

[Do, Corbett-Davies, Atif, Usunier, 2023]

- Model
  - Individuals represented by data points in set X
  - A set items Y
  - A set of contexts C

- Recommendation policy π
  - $\pi_x(y|c)$ = probability of recommending item y to user x given a context c

- Utility function: $u_x(\pi_x) = \mathbb{E}_{c\sim C_m, y\sim \pi_x(\cdot|c)}[v_x(y|c)]$

- Envy-freeness: $\forall x, x' \in X,\ u_x(\pi_x) \geq u_x(\pi_{x'}) - \varepsilon$

Envy-Freeness ⇒ Recommendations

Two-Sided Fairness in Recommendations

[Biswas, Patro, Ganguly, Gummadi, Chakraborty, 2023]

- Many-to-many matching
  - Each user is recommended k products
  - Each product may be recommended to a different number of users
- Relevance of products to users given by $V: X \times Y \rightarrow \mathbb{R}$
- Recommendation policy π
  - Each user x is recommended $\pi_x \subseteq Y$ with $|\pi_x| = k$
  - Let Y_x^* be the top-k products for user x by relevance
- Utilities
  - Utility to user x given by $u_x(\pi_x) = \frac{\sum_{y \in \pi_x} V(x,y)}{\sum_{y \in \pi_x^*} V(x,y)}$
  - Utility to product y given by $E_y(\pi)$, the number of users y is exposed to

Two-Sided Fairness in Recommendations

[Biswas, Patro, Ganguly, Gummadi, Chakraborty, 2023]

- Two-sided fairness
  - Fairness for users: envy-freeness up to one (EF1)

$$\forall x, x' \in X, \exists y \in \pi_{x'}: u_x(\pi_x) \geq u_x(\pi_{x'} \setminus \{y\})$$

  - Fairness for products: minimum exposure $\bar{E}$

$$\forall y \in Y, E_y(\pi) \geq \bar{E}$$

- Theorem: There exists an efficient algorithm that achieves EF1 among all users and the minimum exposure guarantee among at least $m - k$ products.

- Future directions: Fairness to products in terms of the relevance, asymmetric entitlements of users

Two-Sided Fairness in Recommendations

[Freeman, M, S, 2021]

- Many-to-many matching
  - Each user is recommended k products
  - Each product is recommended to k users
- Relevance of products to users given by $V: X \times Y \to \mathbb{R}$
- Recommendation policy π
  - Each user x is recommended $\pi_x \subseteq Y$ with $|\pi_x| = k$
  - Each product y is recommended to $\pi_y \subseteq X$ with $|\pi_y| = k$
- Utilities
  - Utility to user x given by $u_x(\pi_x) = \sum_{y \in \pi_x} V(x, y)$
  - Utility to product y given by $u_y(\pi_y) = \sum_{\mathrm{x} \in \pi_{\mathrm{y}}} V(x, y)$

Two-Sided Fairness in Recommendations

[Freeman, M, S, 2021]

- Two-sided fairness
  - Fairness for users: envy-freeness up to one (EF1)

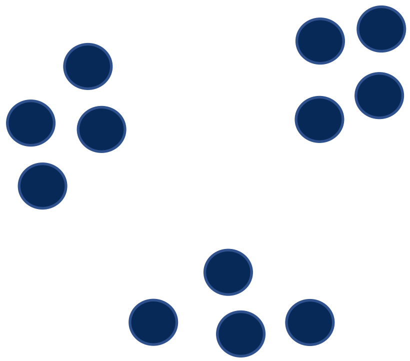
$$\forall x, x' \in X, \exists y \in \pi_{x'}: u_x(\pi_x) \geq u_x(\pi_{x'} \setminus \{y\})$$

  - Fairness for products: envy-freeness up to one (EF1)

$$\forall \mathrm{y}, \mathrm{y}' \in \mathrm{Y}, \exists x \in \pi_y: u_{\mathrm{y}}\left(\pi_{\mathrm{y}}\right) \geq u_{\mathrm{y}}\left(\pi_{y'} \setminus \{x\}\right)$$

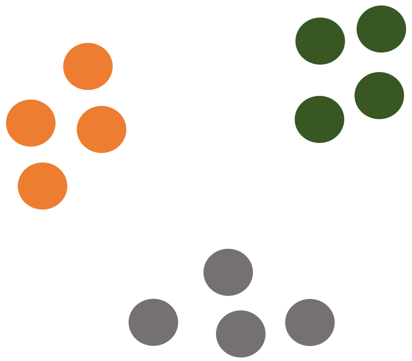
- Theorem: When each side agrees on the ranking of the other side by relevance, a policy that is EF1 w.r.t. both users and products exists and can be computed efficiently.
- Open question: Does a policy that is EF1 w.r.t. both sides always exist?
- Future directions: Non-stationary recommendations, different entitlements

Non-Centroid Clustering



- Goal: Partition the agents into k clusters, i.e., $C = (C_1, \dots, C_k)$

Non-Centroid Clustering



- Goal: Partition the agents into k clusters, i.e., $C = (C_1, \ldots, C_k)$

Non-Centroid Clustering

[Ahmadi, Awasthi, Khuller, Kleindessner, Morgenstern, Sukprasert, Vakilian, 2022]

[Aamand, Chen, Liu, Silwal, Sukprasert, Vakilian, Zhang, 2023]

- Model
  - Set of agents N partitioned into $C = (C_1, \ldots, C_k)$
  - Cluster containing agent i denoted by $C(i)$
  - Distance metric $d: N \times N \to \mathbb{R}_{\geq 0}$

- α-Envy-freeness: For each $i \in N$ and $j \in [k]$ with $i \notin C_j$, either $C(i) = \{i\}$ or

$$\frac{1}{|C(i)| - 1} \sum_{i' \in C(i)} d(i, i') \leq \frac{\alpha}{|C_j|} \sum_{i' \in C_j} d(i, i')$$

- Theorem: A 1-envy-free clustering does not always exist, but an $O(1)$-envy-free clustering always does (and can be computed efficiently).

Non-Centroid Clustering

[Li, M, Nikolov, S, 2023]

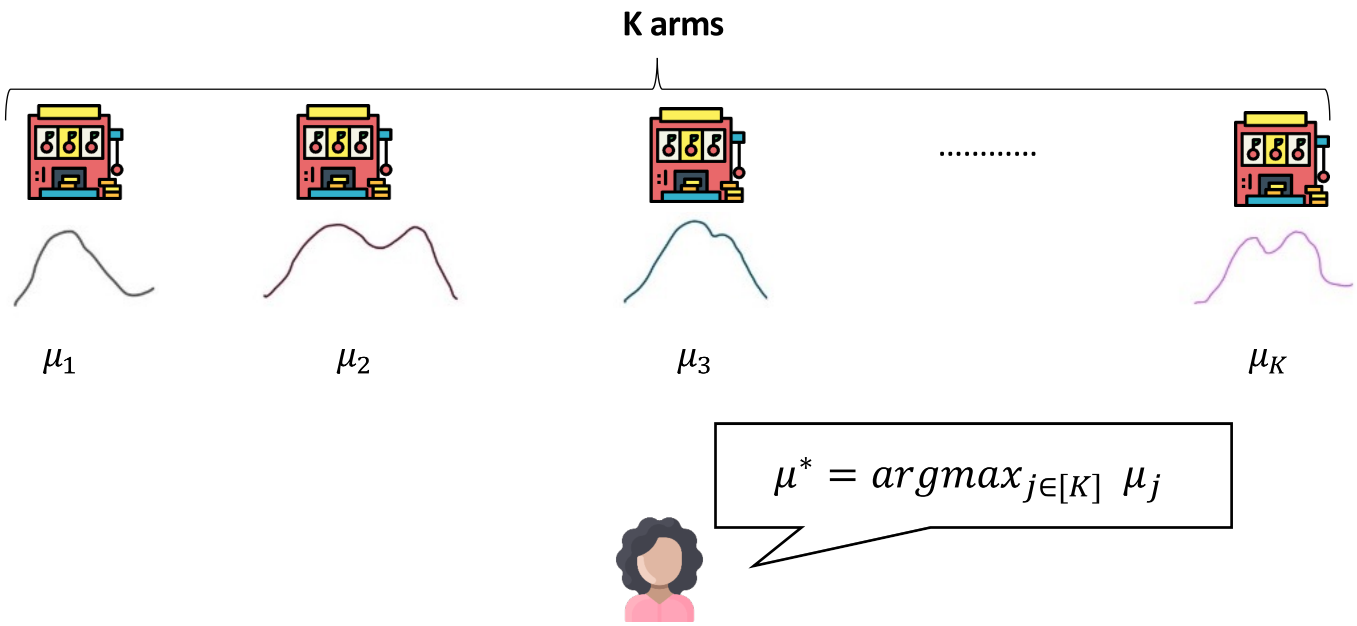
- Model
  - Set of agents N partitioned into $C = (C_1, \dots, C_k)$
  - Cluster containing agent i denoted by $C(i)$
  - Binary costs $d: N \times N \to \{0,1\}$

- Theorem: There exists a balanced clustering ($|C(i)| = |C(j)| \pm 1, \forall i, j$) such that for all $i \in N$ and $j \in [k]$, $\sum_{i' \in C(i)} d(i, i') \leq \sum_{i' \in C_j} d(i, i') + \tilde{O}\left(\sqrt{n/k}\right)$.

- If we divide by the sizes of the clusters ($\approx n/k$), then the additive term becomes $o(1)$

Nash Social Welfare in ML

Multi-Armed Bandits

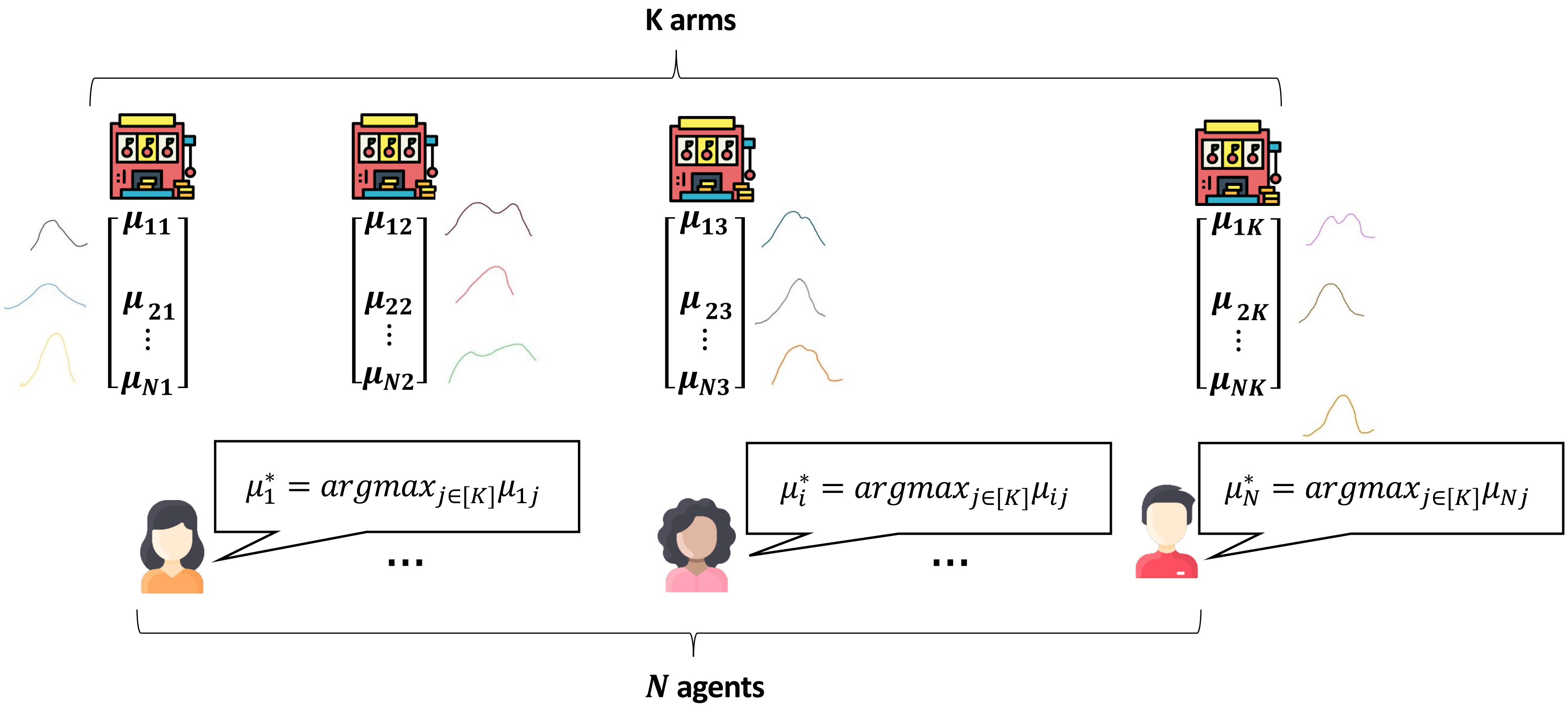


Exploration vs Exploitation

Regret: $R_T = T\mu^* - \sum_{t=1}^{T} \mu(t)$

Multi-Agent Multi-Armed Bandits

[Hossain, M, S, 2021]



What is a fair policy?

Multi-Agent Multi-Armed Bandits

[Hossain, M, S, 2021]

- Distribution $p = [p_1, \dots, p_K]$ gives expected reward $\sum_{j=1}^{K} p_j \cdot \mu_{ij}$ to agent i
- Maximizing welfare functions
  a) Utilitarian welfare $\sum_{i=1}^{N} \sum_{j=1}^{K} p_j \cdot \mu_{ij}$
  b) Egalitarian welfare $\min_{i \in N} \sum_{j=1}^{K} p_j \cdot \mu_{ij}$
  c) Nash welfare $\prod_{i=1}^{N} \sum_{j=1}^{K} \boldsymbol{p_j \cdot \mu_{ij}}$


$p_1^a = 1$ $\quad$ $p_2^a = 0$

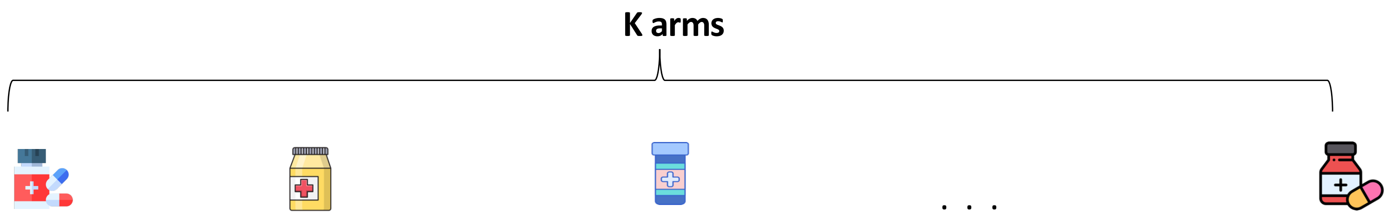
$p_1^b = 1/2$ $\quad \begin{bmatrix}1\\1\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\end{bmatrix} \quad$ $p_2^b = 1/2$

$p_1^c = 2/3$ $\quad$ $p_2^c = 1/3$

- Regret: $R_T = NSW(p^*, \mu) - \sum_{t=1}^{T} NSW(p(t), \mu)$
- Theorem: A variation of UCB achieves the optimal $\Theta(\sqrt{T})$ regret
  - Regret bound and computation improved by subsequent work [Jones, Nguyen, Nguyen, 2023]

Fair Exploration

[Barman, Khan, Maiti, Sawarni, 2023]

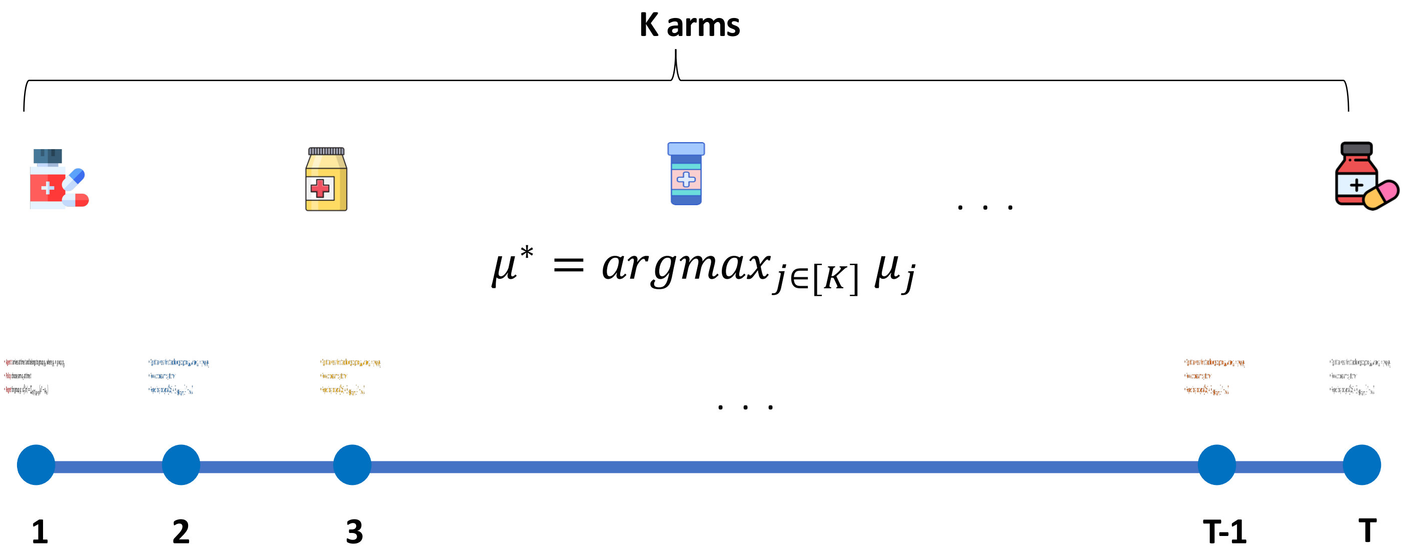


$$\mu^* = argmax_{j \in [K]}\, \mu_j$$

- Agent t arrives at time t
- We chose distribution P_t, which gives the agent utility $E_{j \backsim P_t}\left[\mu_j\right]$
- Regret: $R_T = \mu^* - \left(\Pi_{\mathrm{t}=1}^{\mathrm{T}} E_{j \backsim P_t}\left[\mu_j\right]\right)^{1/T}$
- Theorem: A variation of UCB achieves near-optimal regret in terms of T

Fair Exploration

[Baek, Farias, 2021]

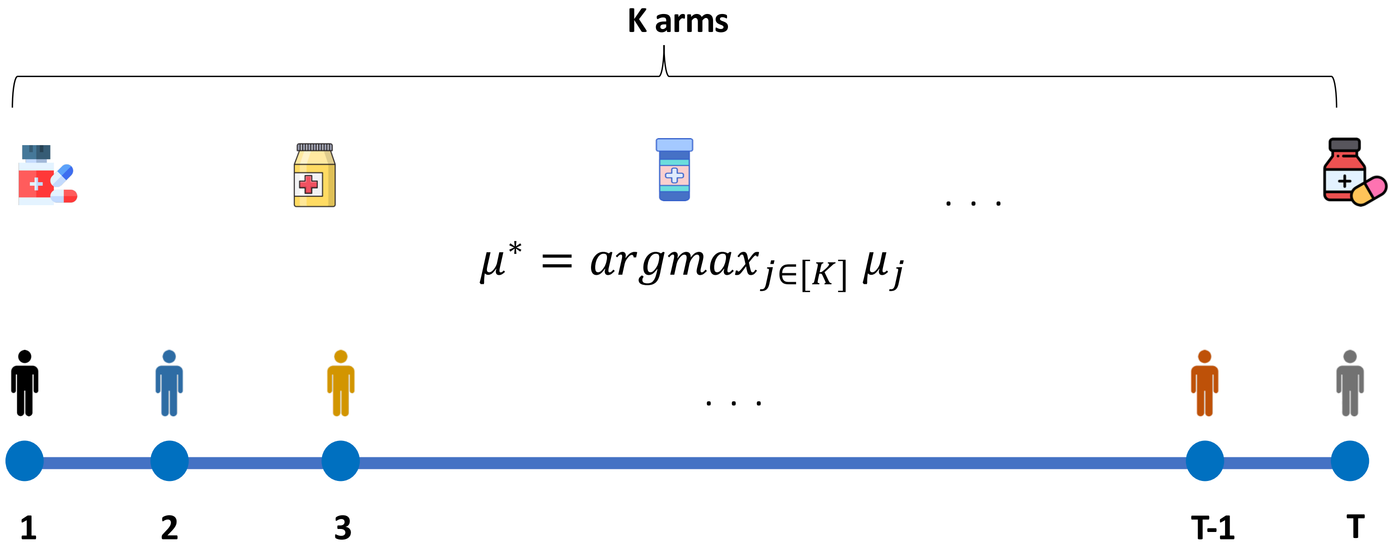


$$\mu^* = argmax_{j \in [K]}\, \mu_j$$

- Agent t arrives at time t and belongs to group g_t, where $g_t = g$ w.p. p_g
- Policy: choose arm a_t at time t
- Regret for group g: $u_T^g(\pi) = \sum_{t \in [T] : g_t = g} \left(\mu^* - \mu_{a_t}\right)$

Fair Exploration

[Baek, Farias, 2021]

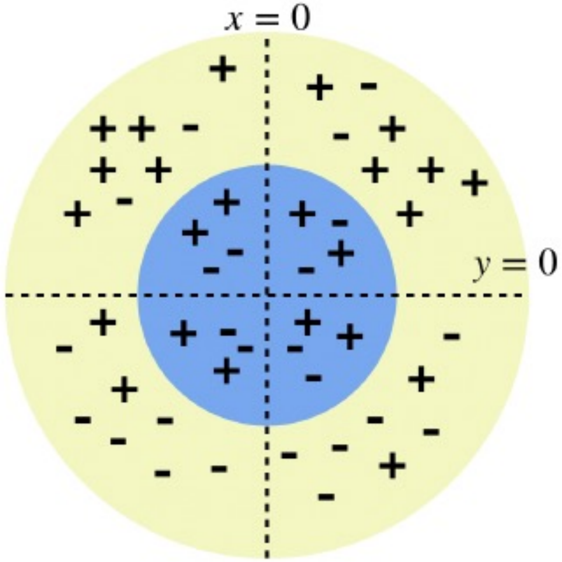


- Utility to group g: $u^g(\pi) = R_T^g(\pi^0) - R_T^g(\pi)$, where π^0 is the policy minimizing the overall regret ("default" or "outside" option)
- Nash social welfare objective: $NSW(\pi) = \prod_g u^g(\pi)$
- Theorem: A version of UCB exactly optimizes this NSW objective.

Classification

[Krishnaswamy, Jiang, Wang, Cheng, Munagala, 2021]

- **Standard Notion of Fairness:** Statistical Parity or Equalized odds



Can *every* group of individuals be treated at least as well as it can be classified in itself?

Classification

[Krishnaswamy, Jiang, Wang, Cheng, Munagala, 2021]

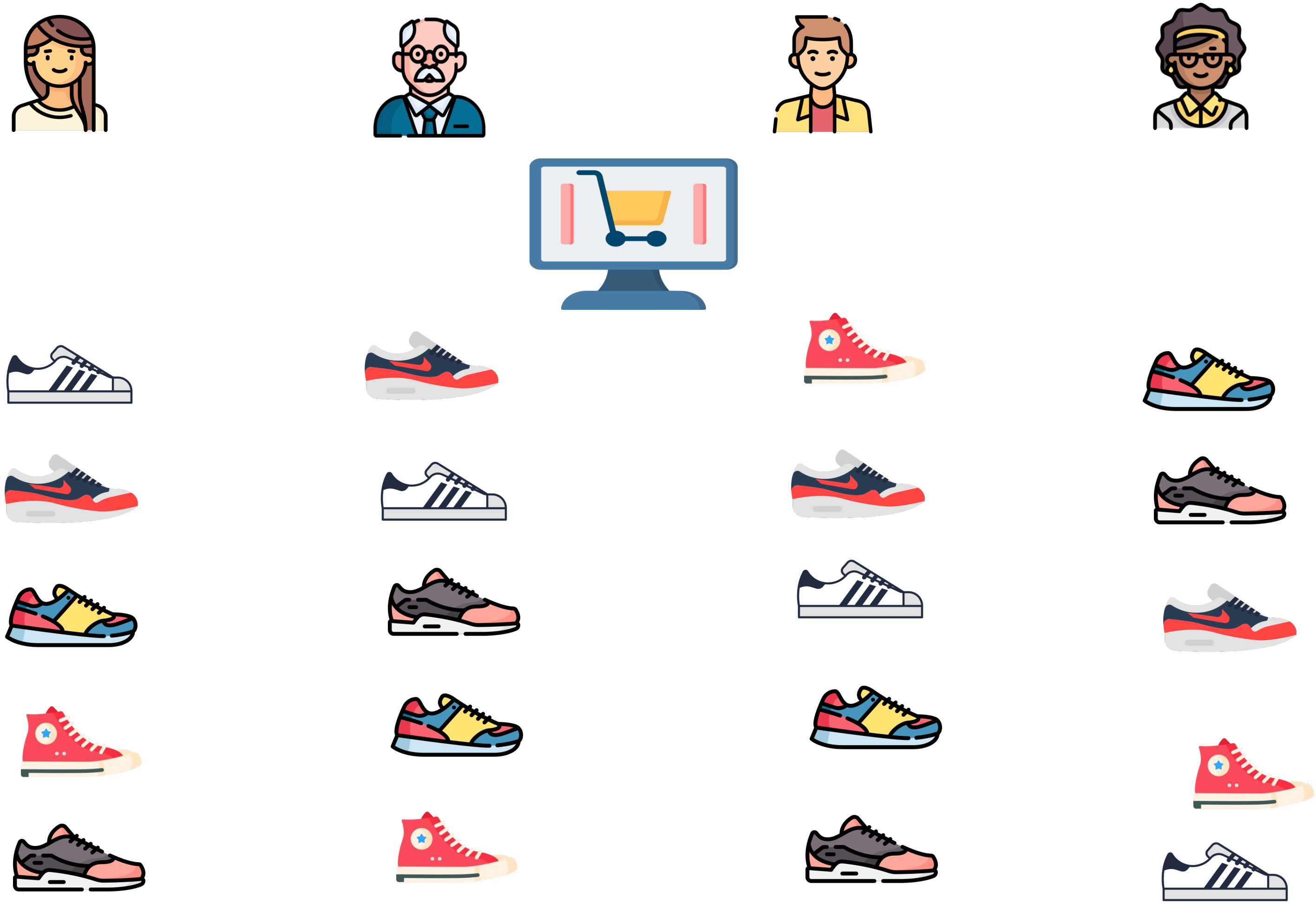
- **Utility of an individual:** $u_i(f) = \mathbb{I}[f(x_i) = \ y_i]$
- **Utility of a group:** $u_S(f) = \frac{1}{|S|} \sum_{i \in S} u_i\ (f)$
- **Optimal Classifier for a group:** $f_S^* = argmax_{f \in F} u_S(f)$
- **Best-effort Guarantees**
  - Return f such that $u_S(f) \geq \alpha \cdot u_S(f_S^*)$, with $\alpha \leq 1$, for each $S \subseteq N$
- **Observation:** No imperfect classifier f provides any reasonable guarantee to best-effort
  - Let $S = \{i \in N \colon f(x_i) \neq y_i\}$ and $u_s(f_S^*) = 1$
- **Randomized Classifiers:** Let D_f be a distribution over F
  - $u_i\big(D_f\big) = \mathbb{E}_{f \sim D_f}[u_i(f)]$
  - $u_S\big(D_f\big) = \frac{1}{|S|} \sum_{i \in S} \mathbb{E}_{f \sim D_f}[u_i(f)]$

Classification

[Krishnaswamy, Jiang, Wang, Cheng, Munagala, 2021]

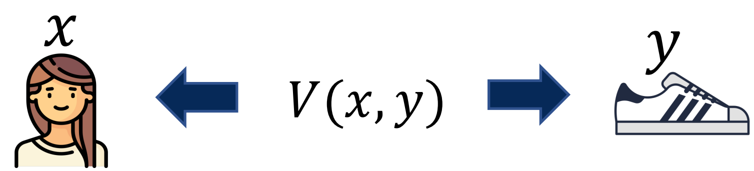
- **Theorem:** There is an instance in which there is no distribution D_f over classifiers such that for all $S \subseteq N$ with $u_s(f_S^*) = 1$, $u_s(D_f) > \frac{|S|}{|N|}$
- $\boldsymbol{D_f^{NSW} = argmax_{D_f \in \Delta(F)} \prod_{i \in N} u_i(D_f)}$
- **Theorem:**
  1. For every group $S \subseteq N$ that admits a perfect classifier, $u_S(D_f^{NSW}) \geq \frac{|S|}{|N|}$
  2. For every group $S \subseteq N$, $u_S(D_f^{NSW}) \geq \frac{|S|}{|N|}[u_S(f_S^*)]^2$

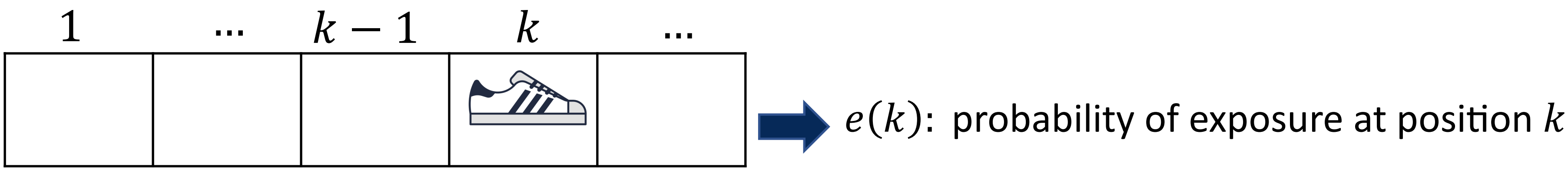
Rankings



Fair Rankings for Products

[Saito, Joachims, 2022]

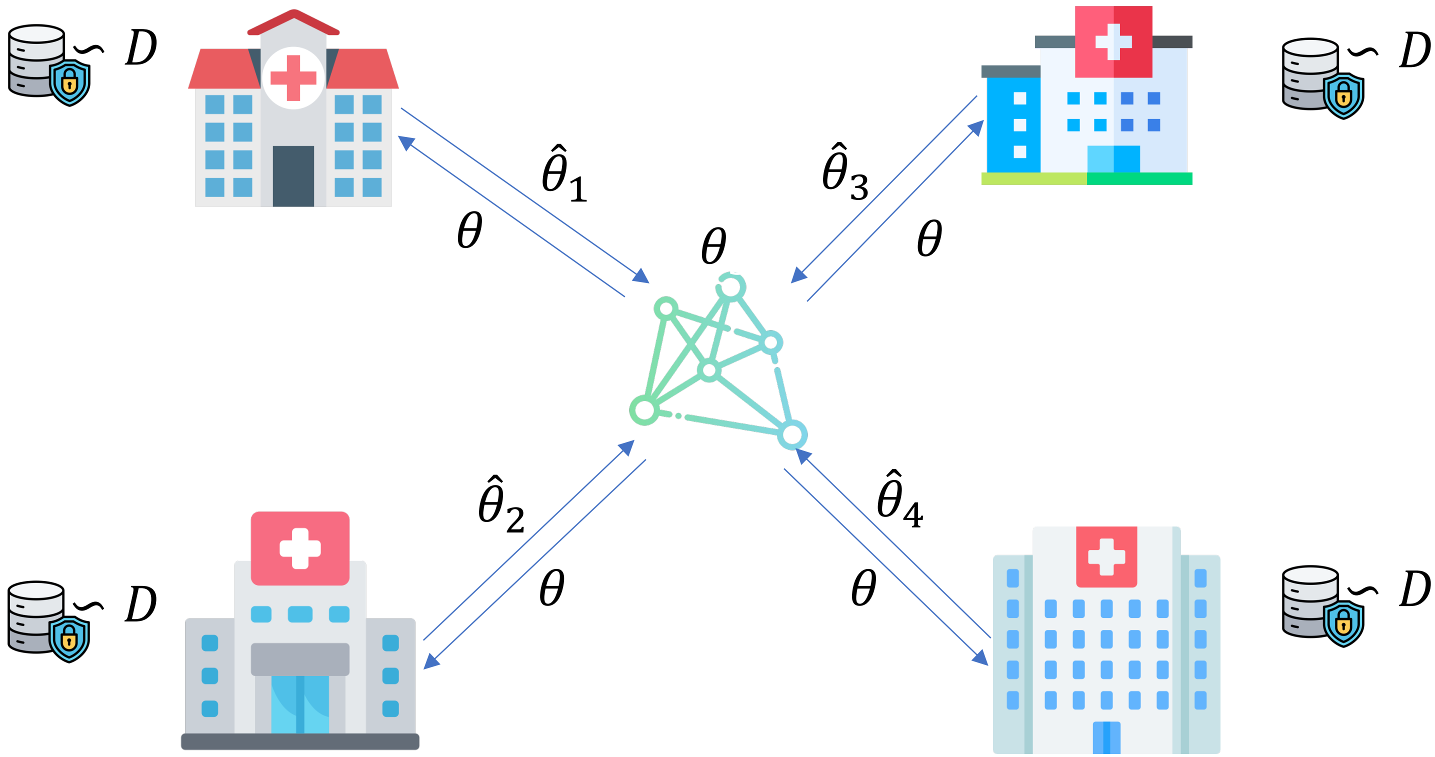




- **Recommendation Policy:** $\pi(y, x, k)$ probability y to exposed at position k in x's ranking
- **Utility of item for a policy:** $u_y(\pi) = \sum_{x \in X} \sum_{k=1}^{|Y|} V(x, y) \cdot e(k) \cdot \pi(y, x, k)$
- **NSW:** $arxgmax_{\pi} \prod_{y \in Y} u_y(\pi\ )$ s. t.
  - $\sum_{y \in Y} \pi(y, x, k) = 1\,, \forall x, k;$
  - $\sum_{k=1}^{|Y|} \pi(y, x, k) = 1, \forall y, k;$
- **Theorem (informal):** NSW achieves Pareto optimality and approximates envy-freeness

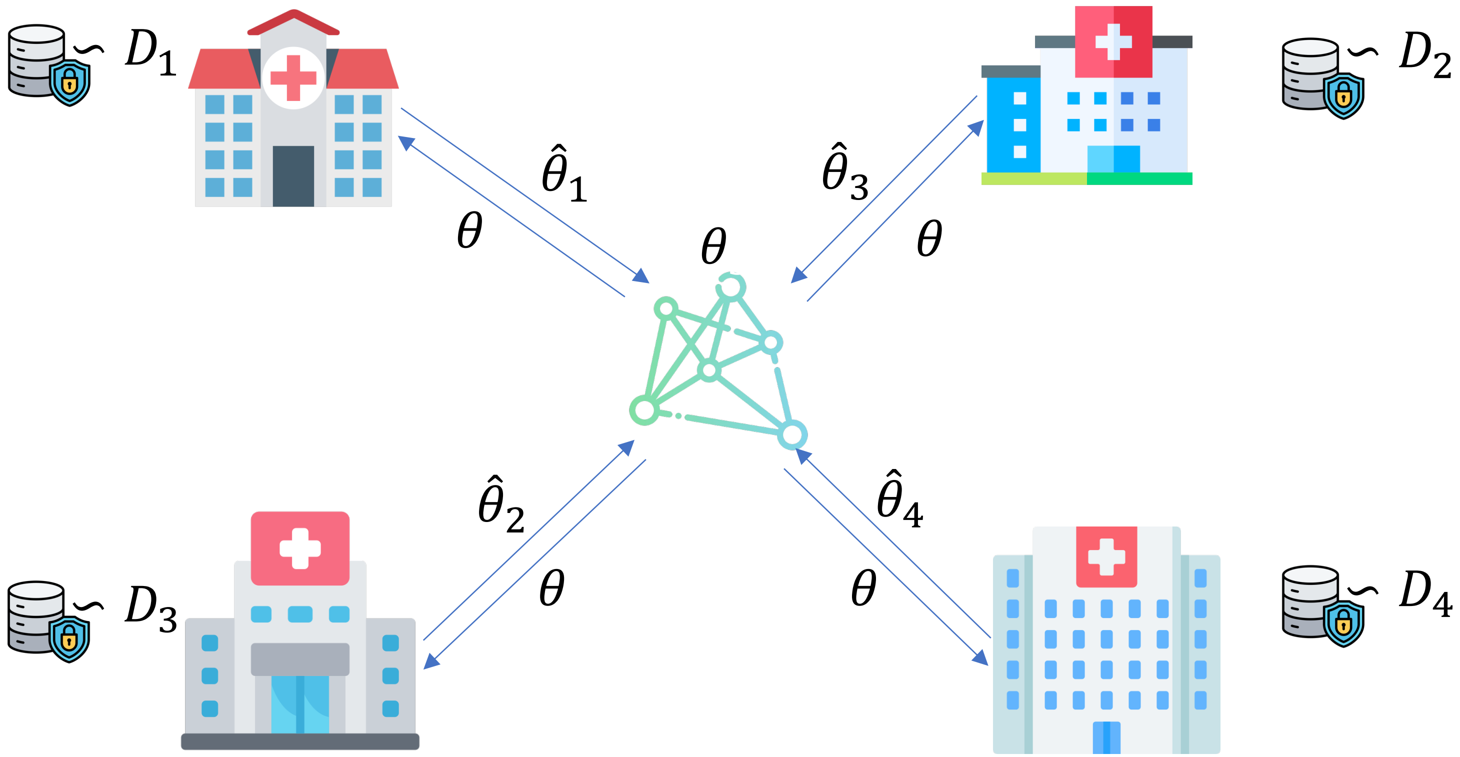
Core in ML

Federated Learning



- Goal: Choose f_θ: $\mathbb{R}^d \rightarrow \mathbb{R}$ from $F = \{f_\theta\text{: } \theta \in P \subseteq \mathbb{R}^d\}$

Federated Learning



- Goal: Choose $f_\theta\colon \mathbb{R}^d \to \mathbb{R}$ from $F = \{f_\theta\colon \theta \in P \subseteq \mathbb{R}^d\}$

Federated Learning

[Chaudhury, Li, Kang, Li, Mehta, 2022]

- **Utility of each agent:**
  - $u_i(\theta) = M - \mathbb{E}_{(x,y)\sim D_i}\left[\ell_i(f_\theta(x), y)\right]$
- **Goal:** Choose θ that is fair for all agents
- **Core:** A parameter vector $\theta \in P$ is in the core if for all $\theta' \in P$ and $S \subseteq N$, it holds

  $u_i(\theta) \geq \frac{|S|}{|N|} u_i(\theta')$ for all $i \in S$, with at lost one strict inequality
- **Pareto Optimality:** A parameter vector $\theta \in P$ is Pareto Optimal if there exists no $\theta' \in P$ such that $u_i(\theta') \geq u_i(\theta)$ for all $i \in N$, with at lost one strict inequality
- **Proportionality:** A parameter vector $\theta \in P$ is proportionally fair if for all $\theta' \in P$, it holds

  $u_i(\theta) \geq \frac{u_i(\theta')}{|N|}$ for all $i \in N$

Federated Learning

[Chaudhury, Li, Kang, Li, Mehta, 2022]

- **Theorem:** When the agents' utilities are continuous and the set of maximizers of any conical combination of the agents' utilities is convex, a parameter vector $\theta \in P$ in the core always exists

- **Theorem:** When the agents' utilities are concave, then the parameter vector $\theta \in P$ that maximizes the NSW is in the core

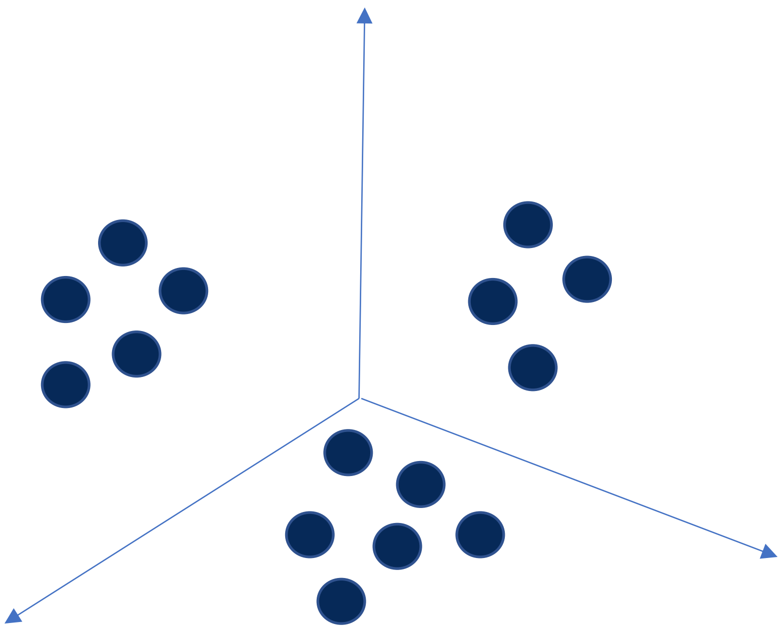
maximize $\prod_{i \in N} u_i(\theta)$

subject to $\theta \in P$

maximize $\sum_{i \in N} \log(u_i(\theta))$

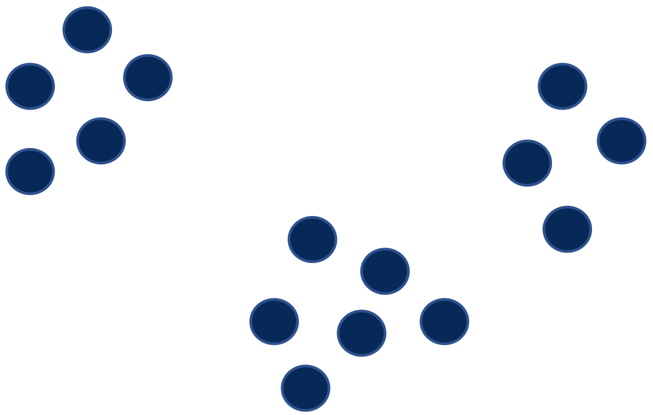
subject to $\theta \in P$

Clustering



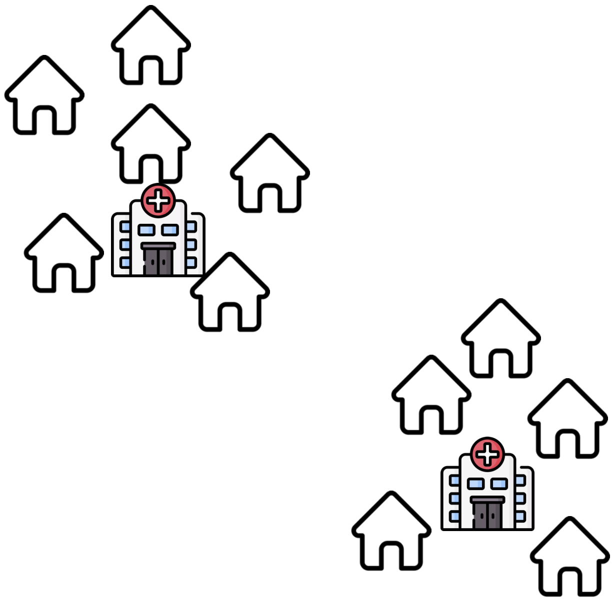
Clustering in ML

- Goal:
  - Analyze data sets to summarize their characteristics
  - Objects in the same group are similar



Clustering in Economics

- Goal:
  - Allocate a set of facilities that serve a set of agents (e.g. hospitals)



Centroid Clustering

- Input:
  - Set N of n data points
  - Set M of m feasible cluster centers
  - $\forall i, j \in N \cup M$: we have $d(i,j)$ (which forms a ***Metric Space***)
    - $\boldsymbol{d(i,i) = 0}$, $\forall\, i \in N \cup M$
    - $\boldsymbol{d(i,j) = d(j,i)}$, $\forall\, i, j \ \in N \cup M$
    - $\boldsymbol{d(i,j) \leq d(i,\ell) + d(\ell,j)}$, $\forall\, i, j, \ell \in N \cup M$, (**Triangle Inequality**)
- Output:
  - A set $C \subseteq M$ of k centers, i.e. $C = \{c_1, \dots, c_k\}$
  - Each data point is assigned to its closest cluster center
    - $C(i) = argmin_{c \in C}\, d(i,c)$

Famous-Objective Functions

- k-median: Minimizes the sum of the distances
  - $\min_{\substack{C \subseteq M: \\ |C| \leq k}} \sum_{i \in N} d(i, C(i))$
- k-means: Minimizes the sum of the square of the distances
  - $\min_{\substack{C \subseteq M: \\ |C| \leq k}} \sum_{i \in N} d^2(i, C(i))$
- k-center: Minimizes the maximum distance
  - $\min_{\substack{C \subseteq M: \\ |C| \leq k}} \max_{i \in N} d(i, C(i))$

Fairness in Clustering

[Chen, Fain, Lyu, Munagala, 2019]

- *Fair Clustering through Proportional Entitlement:*
  - *Every group of n/k agents "deserves" its own cluster center*

- Definition in Committee Selection: W is in the core if
  - For all $S \subseteq N$ and $T \subseteq M$
  - If $|S| \geq |T| \cdot n/k$ (**large**)
  - Then, $|A_i \cap W| \geq |A_i \cap T|$ for some $i \in S$
  - "If a group can afford T, then T should not be a (strict) Pareto improvement for the group"

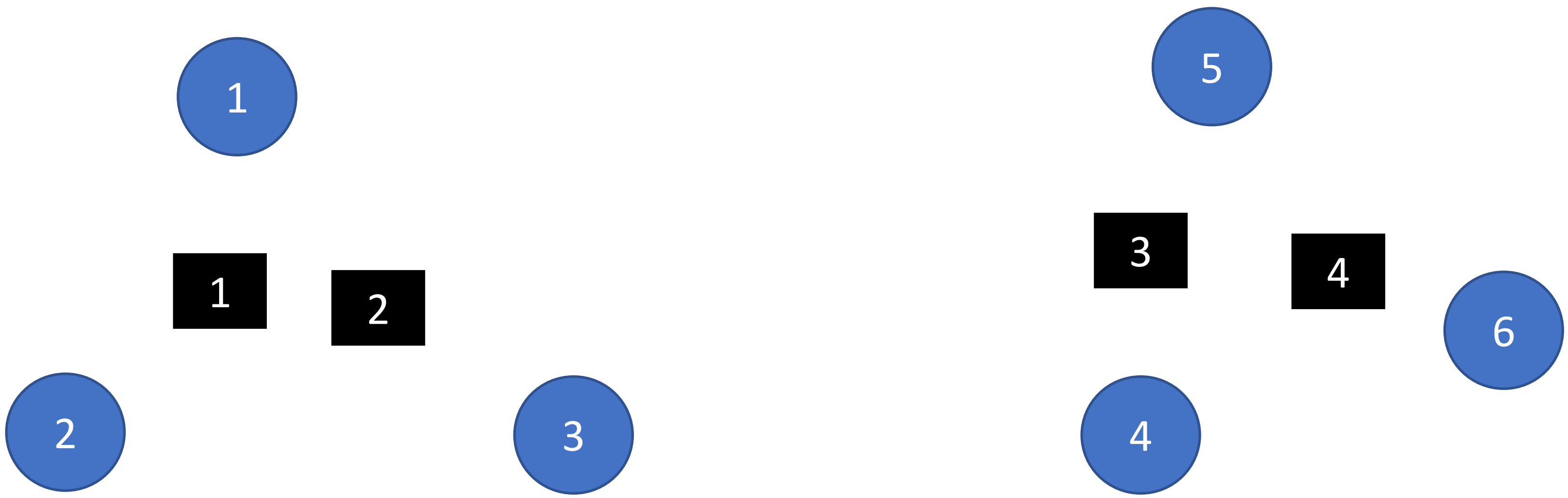
- Definition in Clustering: C is in the core if
  - For all $S \subseteq N$ and $y \subseteq M$
  - If $|S| \geq n/k$ (**large**)
  - Then, $d(i, C(i)) \leq d(i, y)$ for some $i \in S$
  - "If a group can afford a center y, then y should not be a (strict) Pareto improvement for the group"

Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

Example

k=2

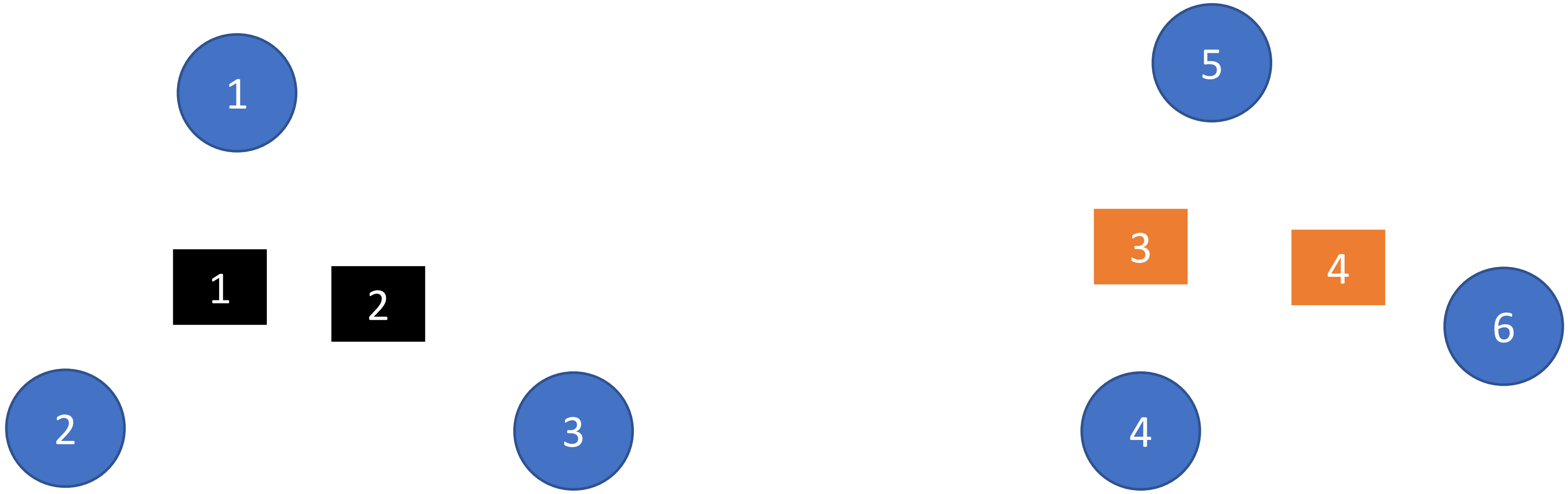


Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

<u>Example</u>

k=2

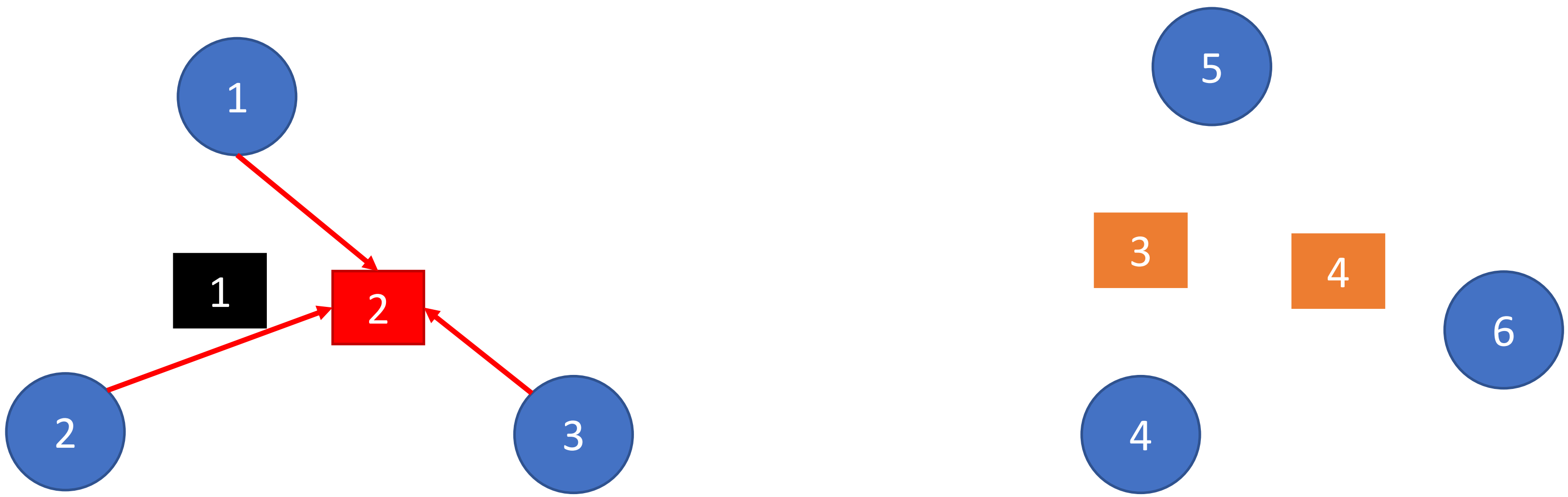


Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

Example

k=2

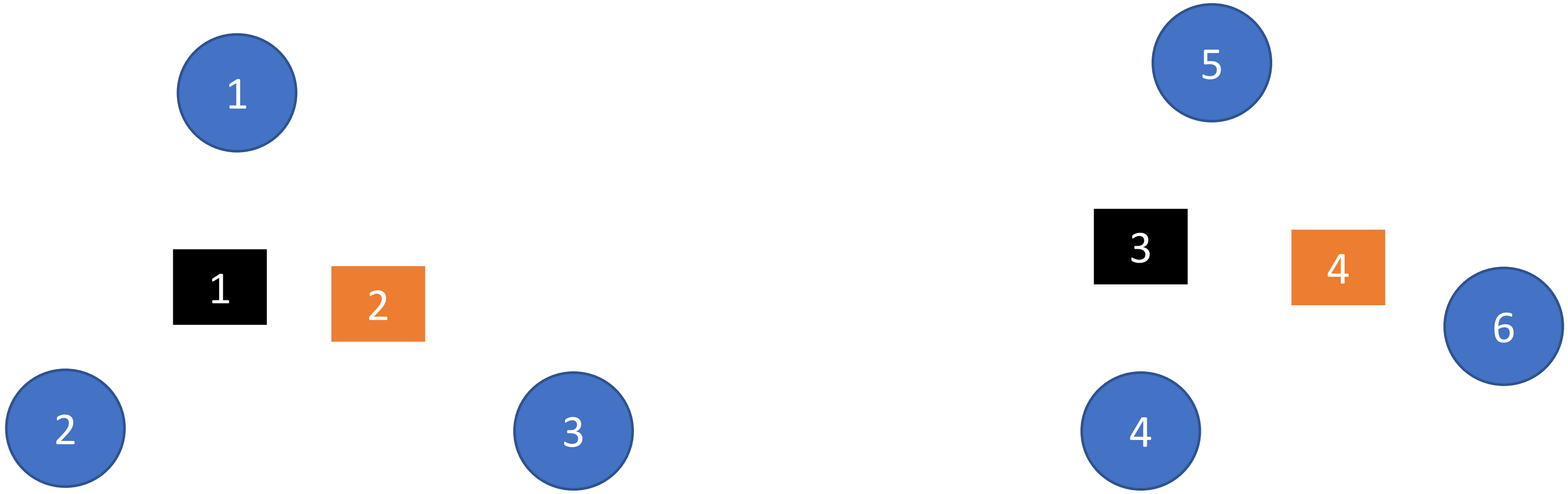


Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

<u>Example</u>

k=2

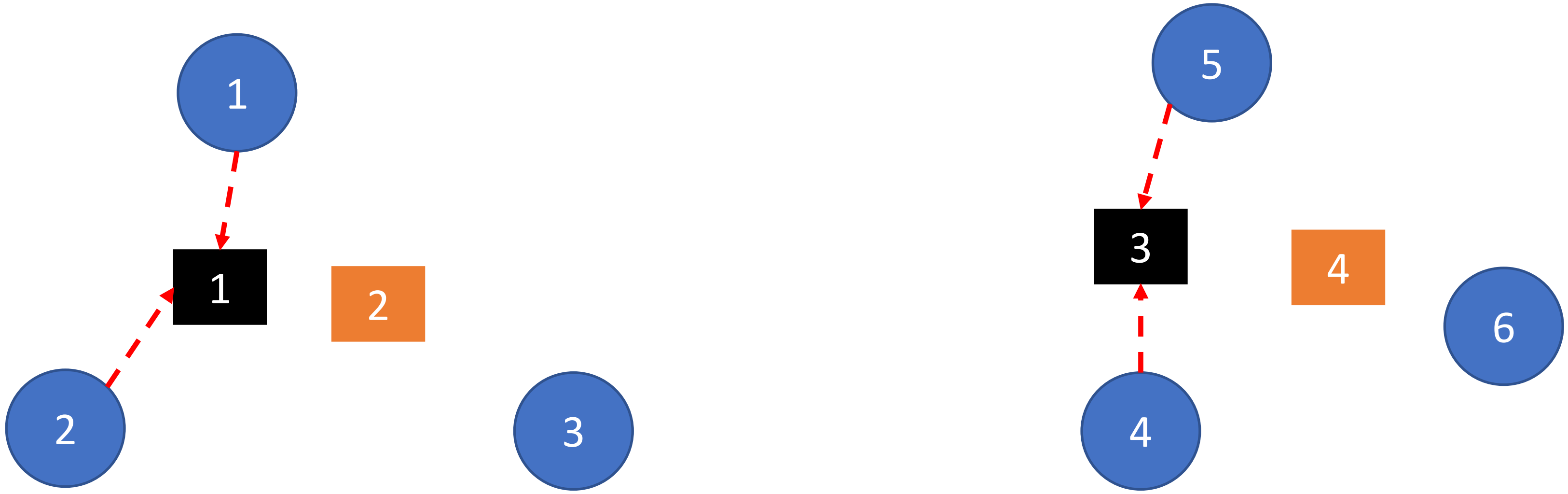


Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

Example

k=2

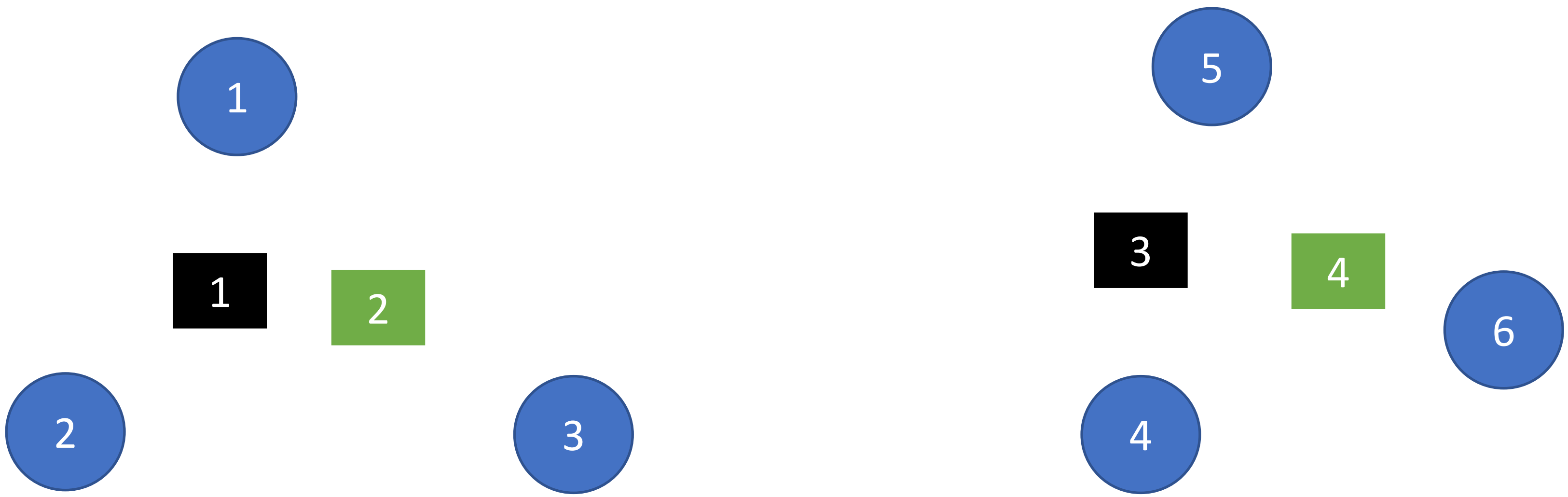


Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

Example

k=2



Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

- A clustering solution in the core does not always exist

α-Core:
A solution C is in the α-core, with $\alpha \geq 1$ if there is ***no*** *group of points $S \subseteq N$ with $|S| \geq n/k$ and $y \in M$ such that:*

$$\forall i \in S, \alpha \cdot d(i, y) < d(i, C(i))$$

Core in Centroid Clustering

[Chen, Fain, Lyu, Munagala, 2019]

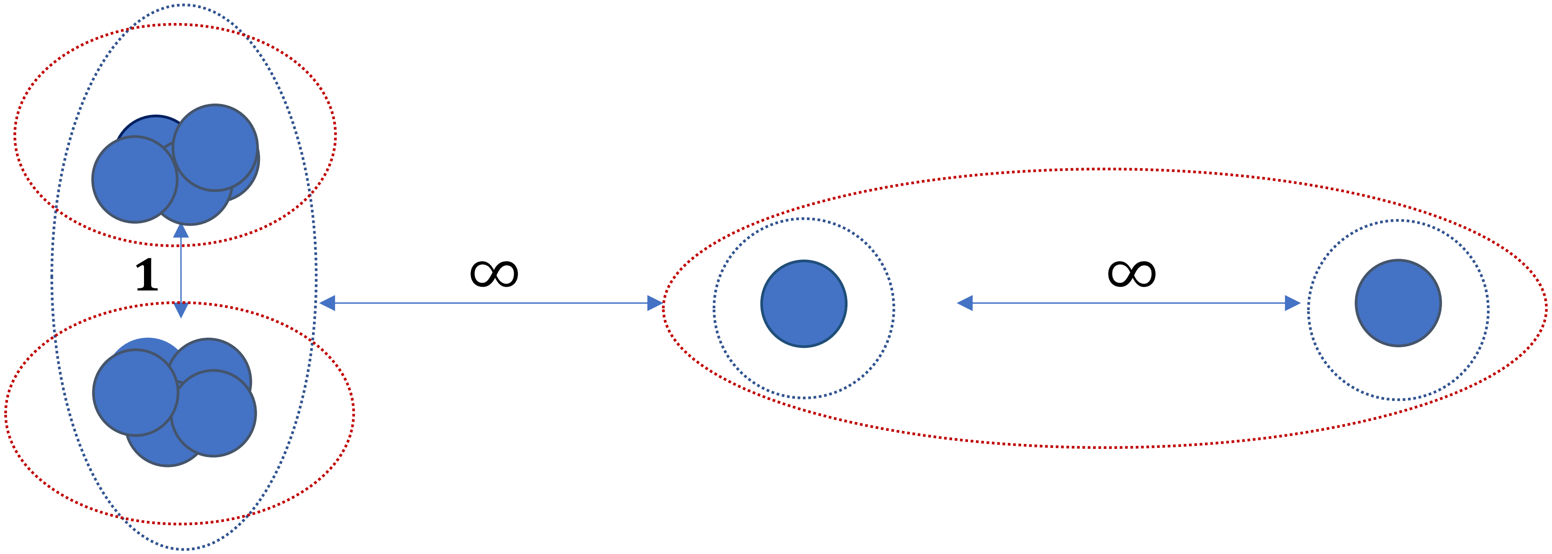
[M, S, 2020]

- **Theorem [Chen et al.]:**
  - There exists an algorithm called, Greedy Capture, that returns a clustering solution in the $(1+\sqrt{2})$-core under any metric space
  - For arbitrary metric spaces and $\alpha < 2$, a clustering solution in the α-core is not guaranteed to exist

- **Theorem [M and S]:**
  - For L_2, Greedy Capture returns a clustering solution in the 2-core
  - For L_1 and L_∞, Greedy Capture does not always return a clustering solution in the α-core, with $\alpha < 1+\sqrt{2}$
  - For L_2 and $\alpha < 1.154$, a clustering solution in the α-core is not guaranteed to exist
  - For L_1 and L_∞, and $\alpha < 1.4$, a clustering solution in the α-core is not guaranteed to exist

Core vs Classic Objectives

$k = 3$



Non-Centroid Clustering

- Input:
  - ➢ Set N of n data points
  - ➢ $\forall i, j \in N \cup M$: we have $d(i, j)$ (which forms a ***Metric Space***)
    - $\boldsymbol{d(i,i) = 0}$, $\forall\ i \in N \cup M$
    - $\boldsymbol{d(i,j) = d(j,i)}$, $\forall\ i, j\ \in N \cup M$
    - $\boldsymbol{d(i,j) \leq d(i,\ell) + d(\ell,j)}$, $\forall\ i, j, \ell \in N \cup M$, **(Triangle Inequality)**
- Output:
  - ➢ Partition the individuals into k clusters, i.e. $C = \{C_1, \dots, C_k\}$
- Loss for Each Cluster:
  - ➢ For $S \subseteq N$ and $i \in$ S, $\ell_i(S) \geq 0$

Core in Non-Centroid Clustering

[Caragiannis, M, S, 2023]

- ***α-Core***: A solution C is in the *α-core*, with $\alpha \geq 1$, if there is **no** group of points $S \subseteq N$ with $|S| \geq n/k$ such that:

$$\forall i \in S, \ell_i(S) < \ell_i(C(i))$$

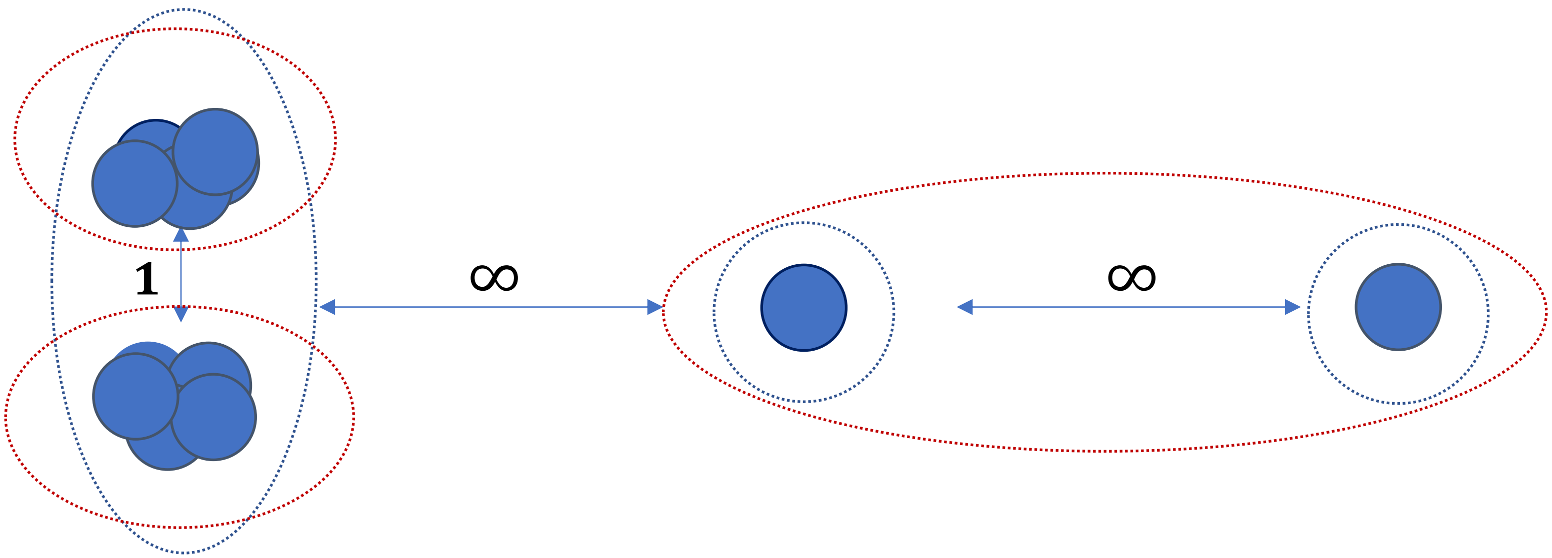
- **Average Loss:** For each $S \subseteq N$, $\ell_i(S) = \frac{1}{|S|} \sum_{i' \in S} d(i, i')$

- **Theorem:**
  - Greedy Capture returns a clustering solution in the (n/k)-core under any metric space
  - For arbitrary metric spaces and $\alpha < 1.366$, a clustering solution in the α-core is not guaranteed to exist

- **Open Question:** Does a clustering solution in the $O(1)$-core always exist?

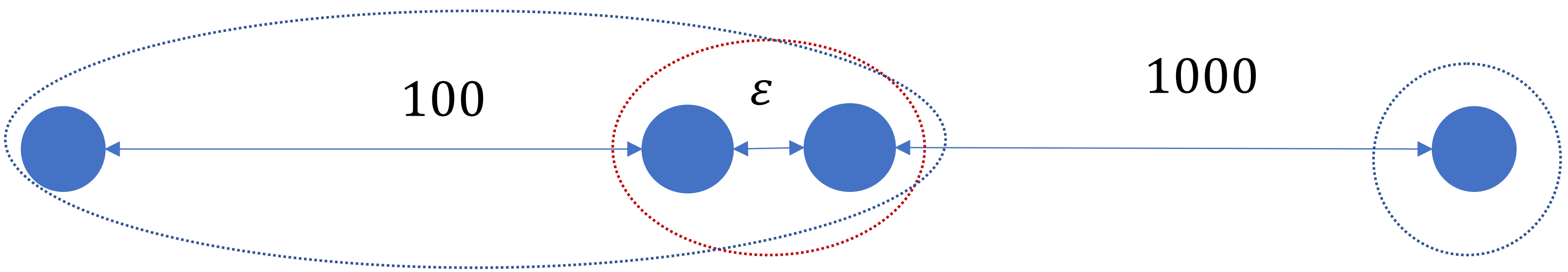
Core vs Classic Objectives

$k = 3$



Core vs Envy-Freeness

$k = 2$



References

- A. Aamand, J. Chen, A. Liu, S. Silwal, P. Sukprasert, A. Vakilian and F. Zhang. Constant Approximation for Individual Preference Stable Clustering. NeurIPS, pp. 24005-24017, 2023.
- S. Ahmadi, P. Awasthi, S. Khuller, M. Kleindessner, J. Morgenstern, P. Sukprasert, A. Vakilian. Individual Preference Stability for Clustering. ICML, pp. 197-246, 2022.
- M.F. Balcan, T. Dick, R.Noothigattu, and A. D. Procaccia. Envy-Free Classification. NeurIPS, pp. 1238-1248, 2019.
- A. Biswas, G.K.Patro, N.Ganguly, K.Gummadi, A.Chakraborty. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. WWW, pp. 1194-1204 2020.
- I. Caragiannis, E. Micha and N. Shah. Proportional Fairness in Non-Centroid Clustering. Manuscript.
- B. R. Chaudhury, L. Li, M. Kang, B. Li, R. Mehta. Fairness in Federated Learning via Core-Stability. NeurIPS, pp. 5738--5750, 2022.

References

- X. Chen, B. Fain, C. Lyu and K. Munagala. Proportionally fair clustering. ICML pp. 1032-1041, 2019.
- V. Do, S. Corbett-Davies, J. Atif and N. Usunier. Online certification of preference-based fairness for personalized recommender systems. AAAI, pp. 6532-6540, 2022.
- C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel. Fairness through Awareness. ITCS, pp. 214-226, 2012.
- R. Freeman, E. Micha and N. Shah. Two-sided Matching Meets Fair Division. IJCAI, pp. 203-209 , 2021.
- S. Hossain, A. Mladenovic, and N. Shah. Designing Fairly Fair Classifiers Via Economic Fairness Notions. WWW, pp. 1559-1569, 2020.
- S. Hossain, E. Micha and N. Shah. Fair Algorithms for Multi-Agent Multi-Armed Bandits. NeurIPS, pp. 24005-24017, 2021.
- M. Jones, H. L. Nguyen and T. Nguyen. An Efficient Algorithm for Fair Multi-Agent Multi-Armed Bandit with Low Regret. AAAI, pp. 8159-8167. 2023.

References

- P. Kim, A. Korolova, G. N. Rothblum and G. Yona. Preference-Informed Fairness. FAccT, pp. 546, 2020.
- A. Krishnaswamy, Z. Jiang., K. Wang, Y. Chen and K. Munagala. Fair for all: Best-effort fairness guarantees for classification AISTATS, pp. 3259-3267, 2021.
- L. Li, E. Micha, A. Nikolov and N. Shah. Partitioning Friends Fairly. AAAI, pp. 5747-5754, 2023.
- E. Micha and N. Shah. Proportionally Fair Clustering Revisited. ICALP, pp. 85:1-85:16, 2020.
- J. B. Orlin. Improved algorithms for computing Fisher's market clearing prices. STOC, pp. 291-300, 2010.
- Y. Saito and T. Joachims. Fair Ranking as Fair Division: Impact-Based Individual Fairness in Ranking. KDD, pp. 1514-1524, 2022.
- B. Ustun, Y. Liu, D. C. Parkes. Fairness without Harm: Decoupled Classifiers with Preference Guarantees. ICML, pp. 6373-6382, 2019.
- H. R. Varian. Equity, envy and efficiency. Journal of Economic Theory, 9:63–91, 1974.

Thank you!

Questions?