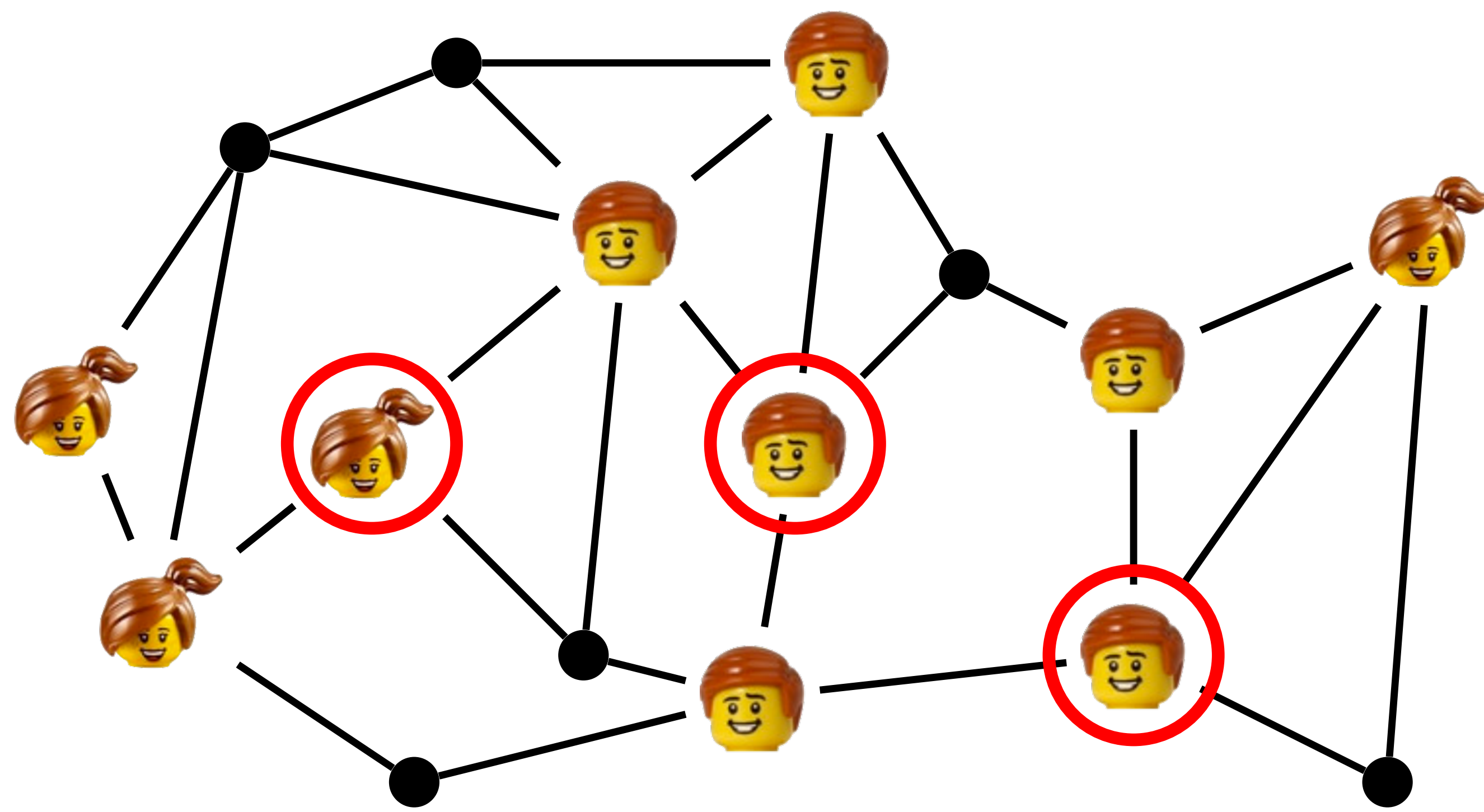


Fairness in Social Influence Maximization via Optimal Transport

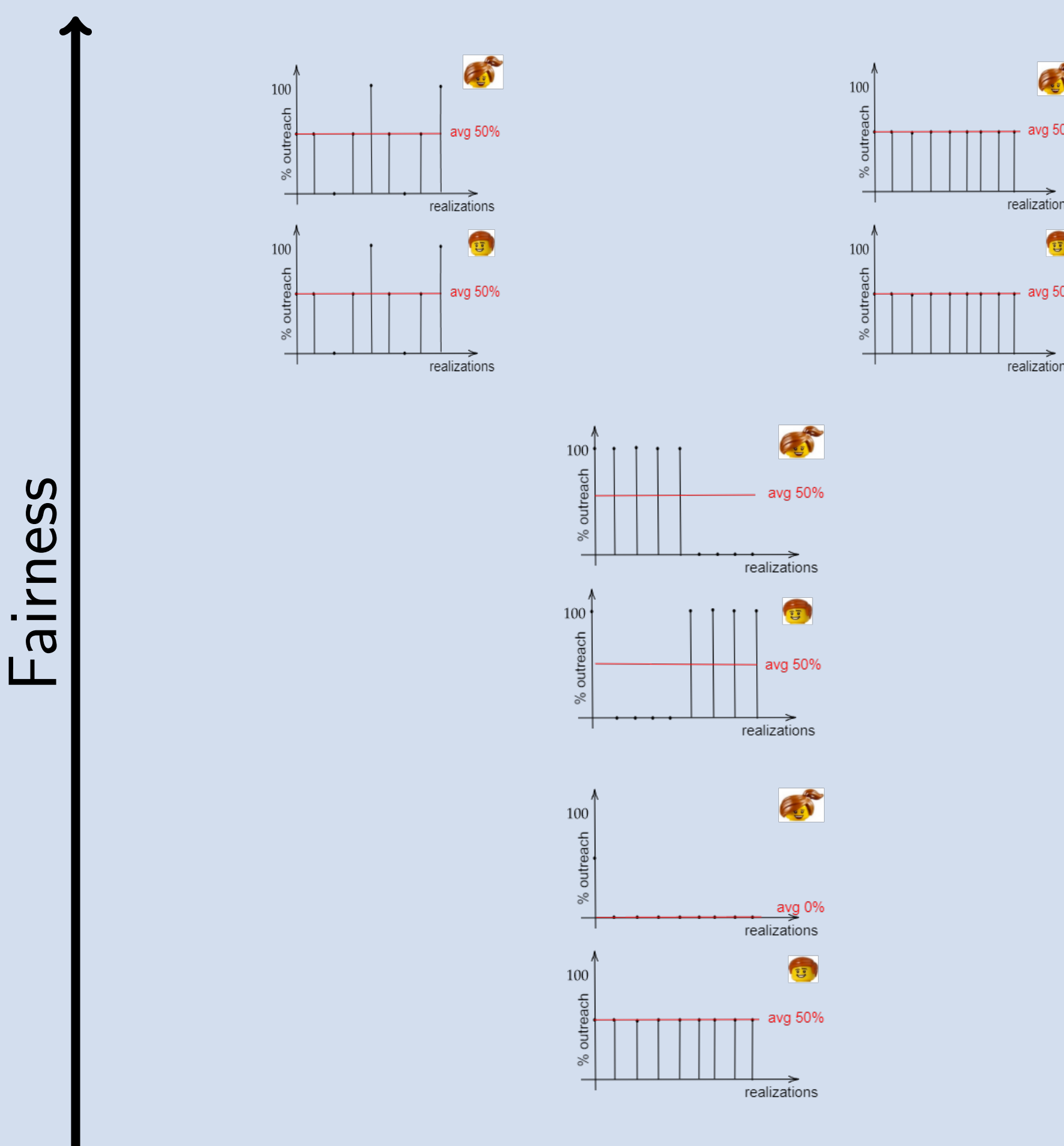
Shubham Chowdhary, Giulia De Pasquale, Nicolas Lanzetti, Ana-Andreea Stoica, Florian Dörfler



Social influence maximization

- Social influence maximization studies how to strategically select a pre-specified small proportion of nodes in the social network, **the early adopters or seeds**, so that the outreach generated by a diffusion process that starts at these early adopters is maximized.
- The problem of selecting early adopters is NP-hard, so various heuristics have been proposed. Most algorithms purely rely on the graph topology and are agnostic to users' demographics, which raises **significant fairness concerns**.
- For this reason, many definitions of fairness were proposed. However, all these definitions involve a marginal expected value of fairness in groups, without considering **the correlations** – or other higher-order moments – for the **joint** probability distribution of different groups adopting the information.

Your turn: Which outcome is fair?



Motivating example

Consider the following stochastic outcomes:

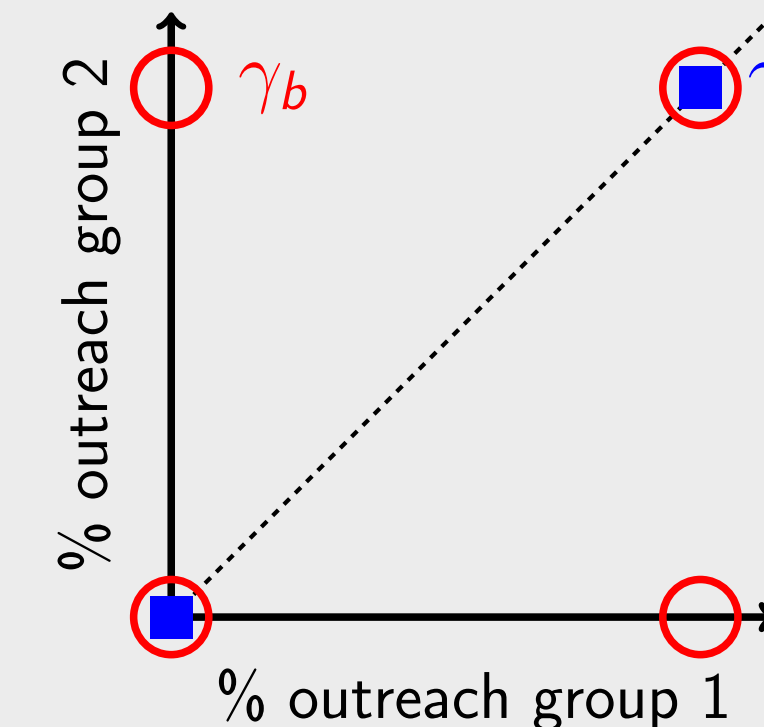
Outcome γ_a : in 50% of the cases everyone receives the information and in 50% no one.

Outcome γ_b : in 25% of the cases everyone receives the information, in 25% no one, in 25% only group 1, and in 25% only group 2.

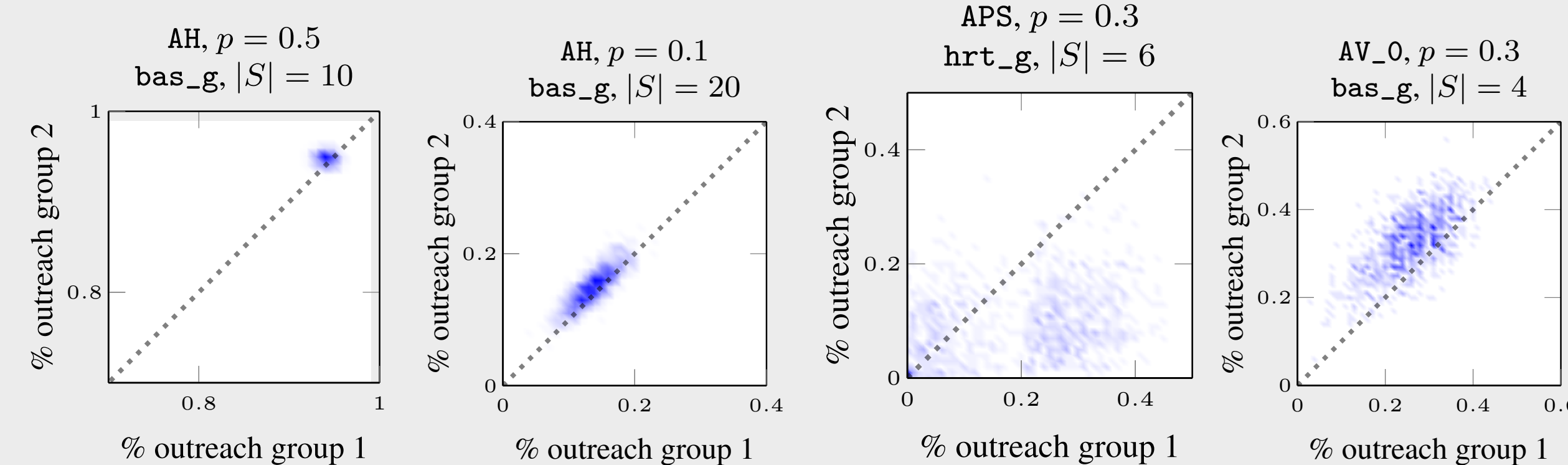
Which outcome is fair?

In γ_a , the percentage of members of group 1 who get the information *a*lways coincides with the percentage of people of group 2. In γ_b , this is not always true.

From a fairness perspective, γ_a and γ_b encode **very different outcomes** ... but γ_a and γ_b have the same marginals and so we need to look at correlations.



This happens in real datasets too



Fairness via optimal transport

use optimal transport to compare stochastic outcomes:
fairness is quantified via the optimal transport discrepancy from an ideal outcome

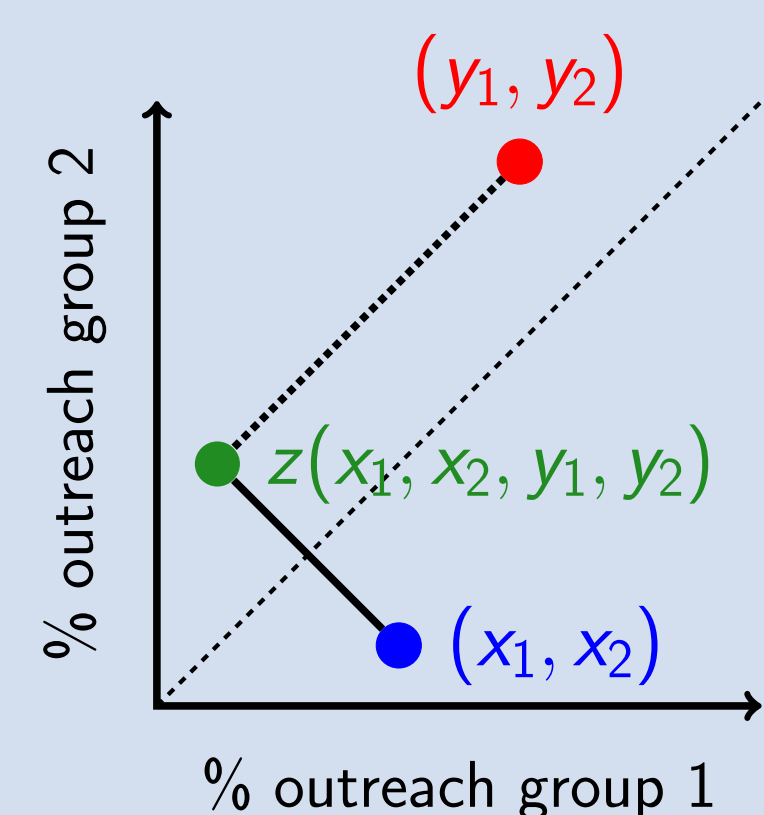
Ideal distribution: $\gamma^* = \delta_{(1,1)}$ (i.e., everyone receives the information)

Fairness-aware transportation cost:

Intuition: moving mass diagonally does not impact fairness, so it should not be penalized.

$$c((x_1, x_2), (y_1, y_2)) = \|z(x_1, x_2, y_1, y_2) - (x_1, x_2)\|$$

$$= \frac{\sqrt{2}}{2} |(x_2 - x_1) - (y_2 - y_1)|$$



Mutual Fairness of a stochastic outcome γ

$$\text{Fairness}(\gamma) = 1 - \sqrt{2} W_c(\gamma, \gamma^*) = \mathbb{E}_{(x_1, x_2) \sim \gamma} [1 - |x_1 - x_2|].$$

... which is just a “normalized” $W_c(\gamma, \gamma^*)$

Reminder: The optimal transport problem

For a transportation cost $c : ([0, 1] \times [0, 1]) \times ([0, 1] \times [0, 1]) \rightarrow \mathbb{R}_{\geq 0}$ the **optimal transport discrepancy** between $\gamma_a \in \mathcal{P}([0, 1] \times [0, 1])$ and $\gamma_b \in \mathcal{P}([0, 1] \times [0, 1])$ is

$$W_c(\gamma_a, \gamma_b) = \min_{\pi \in \Pi(\gamma_a, \gamma_b)} \mathbb{E}_{(x_1, x_2), (y_1, y_2) \sim \pi} [c((x_1, x_2), (y_1, y_2))]$$

where $\Pi(\gamma_a, \gamma_b)$ is the set of probability distributions over so that its first marginal is γ_a and its second marginal is γ_b .

S3D: our fairness metric to select seeds

β -fairness: revisit the transportation cost to tradeoff fairness and efficiency,

$c_\beta((x_1, x_2), (y_1, y_2)) = \beta \|z(x_1, x_2, y_1, y_2) - (x_1, x_2)\| + (1 - \beta) \|z(x_1, x_2, y_1, y_2) - (y_1, y_2)\|$,
to define the β -fairness metric

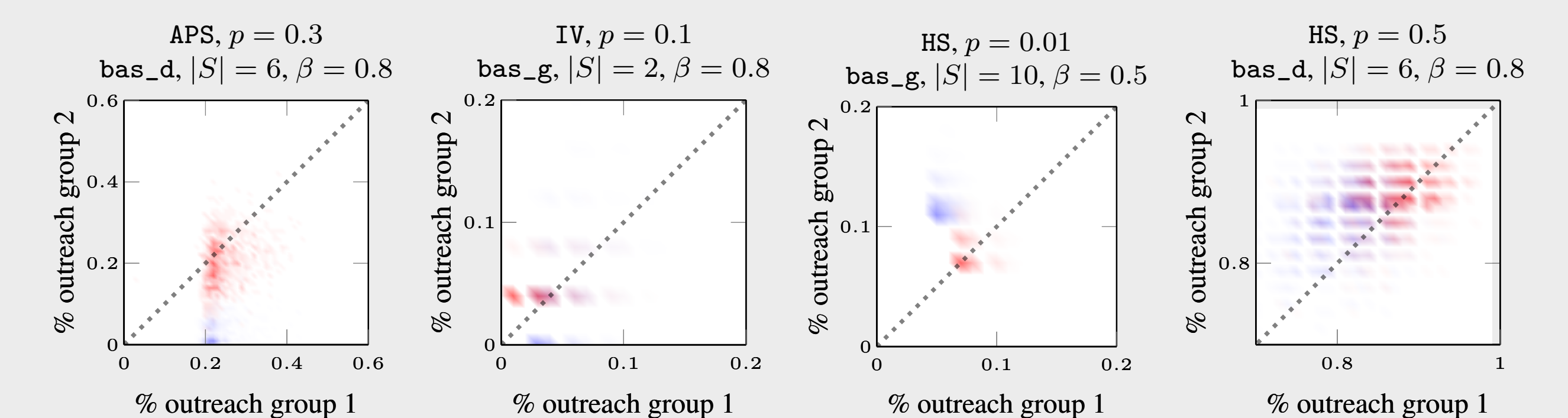
$$\beta - \text{Fairness}(\gamma) = 1 - \frac{\sqrt{2}}{\max\{1, 2 - 2\beta\}} W_{c_\beta}(\gamma, \gamma^*)$$

$$= \mathbb{E}_{(x_1, x_2) \sim \gamma} \left[1 - \frac{\beta |x_1 - x_2| + (1 - \beta) |x_1 + x_2 - 2|}{\max\{1, 2 - 2\beta\}} \right]$$

Our algorithm:

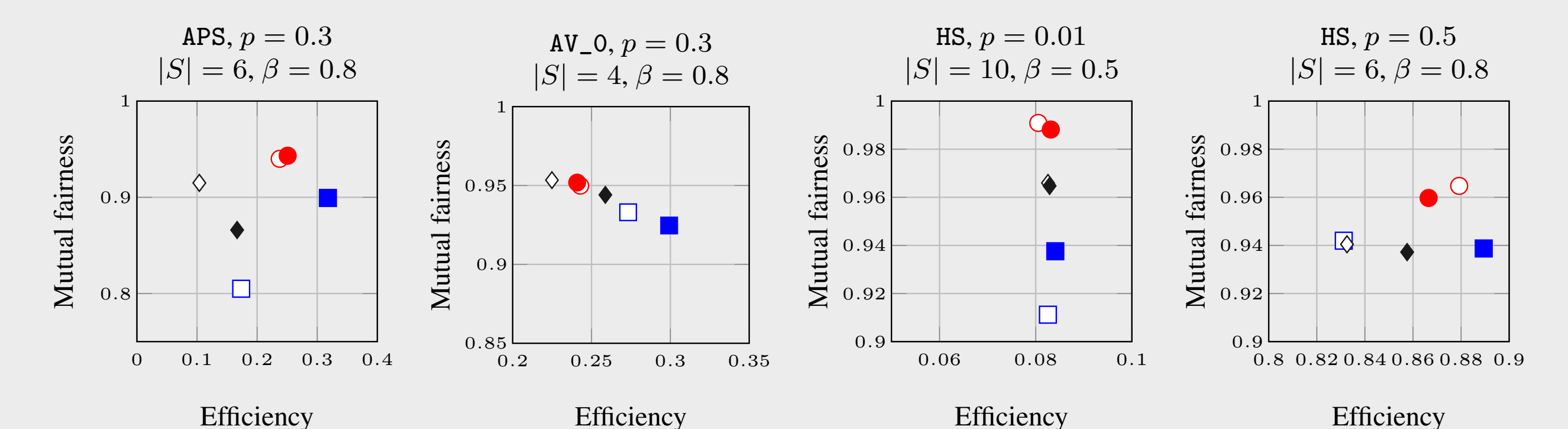
- 1: $\mathcal{S} \leftarrow \{\}$, $S \leftarrow S_0$ ▷ initial collection of candidates, running seedset
- 2: **for** k iterations **do** ▷ configurable k
- 3: $V_S \leftarrow$ nodes reachable from S via cascade, using SEEDSET_REACH routine
- 4: $S' \leftarrow \{\}$
- 5: **for** $|S|$ iterations **do** ▷ searching nearby states, V_S , to get S' (??)
- 6: $S' \leftarrow S' \cup \{v\} \mid v \sim V_S$
- 7: $V_{S'} \leftarrow$ nodes reachable from S' in a fixed horizon, using SEEDSET_REACH
- 8: $V_S \leftarrow V_S \setminus V_{S'}$
- 9: $E_S \leftarrow -\text{BETA_FAIRNESS}(S, \beta)$ ▷ expected potential energy defined on β -fairness
- 10: $E_{S'} \leftarrow -\text{BETA_FAIRNESS}(S', \beta)$
- 11: $p_{\text{accept}} \leftarrow \min\{1, e^{E_S - E_{S'}}\}$ ▷ S' acceptance on energy minimization
- 12: **if** $x \sim B(p_{\text{accept}})$ **then** ▷ Metropolis sampling
- 13: $S^+ \leftarrow S'$ ▷ get a better seedset
- 14: **else**
- 15: **if** $x \sim B(\epsilon)$ **then** ▷ for some small constant ϵ
- 16: $S^+ \leftarrow \{v_i\}_{i=1}^{|S|} \sim V_G$ ▷ random seedset
- 17: **else**
- 18: $S^+ \leftarrow S$ ▷ retain existing choice
- 19: $\mathcal{S} \leftarrow \mathcal{S} \cup \{S^+\}$
- 20: $S \leftarrow S^+$ ▷ for next iteration
- 21: $S^* \leftarrow S \in \mathcal{S} \mid \text{BETA_FAIRNESS}(S, \beta)$ is maximum ▷ via S3D_ITERATE
- 22: **return** S^*

Performance



blue: nominal outreach distribution **red:** outreach distribution with our algorithm

Comparison with other algorithms



Filled markers are greedy-based algorithms: **■** = bas_g, **●** = S3D_g, and **◆** = hrt_g.
Empty markers are degree-based algorithms: **□** = bas_d, **○** = S3D_d, and **◇** = hrt_d.