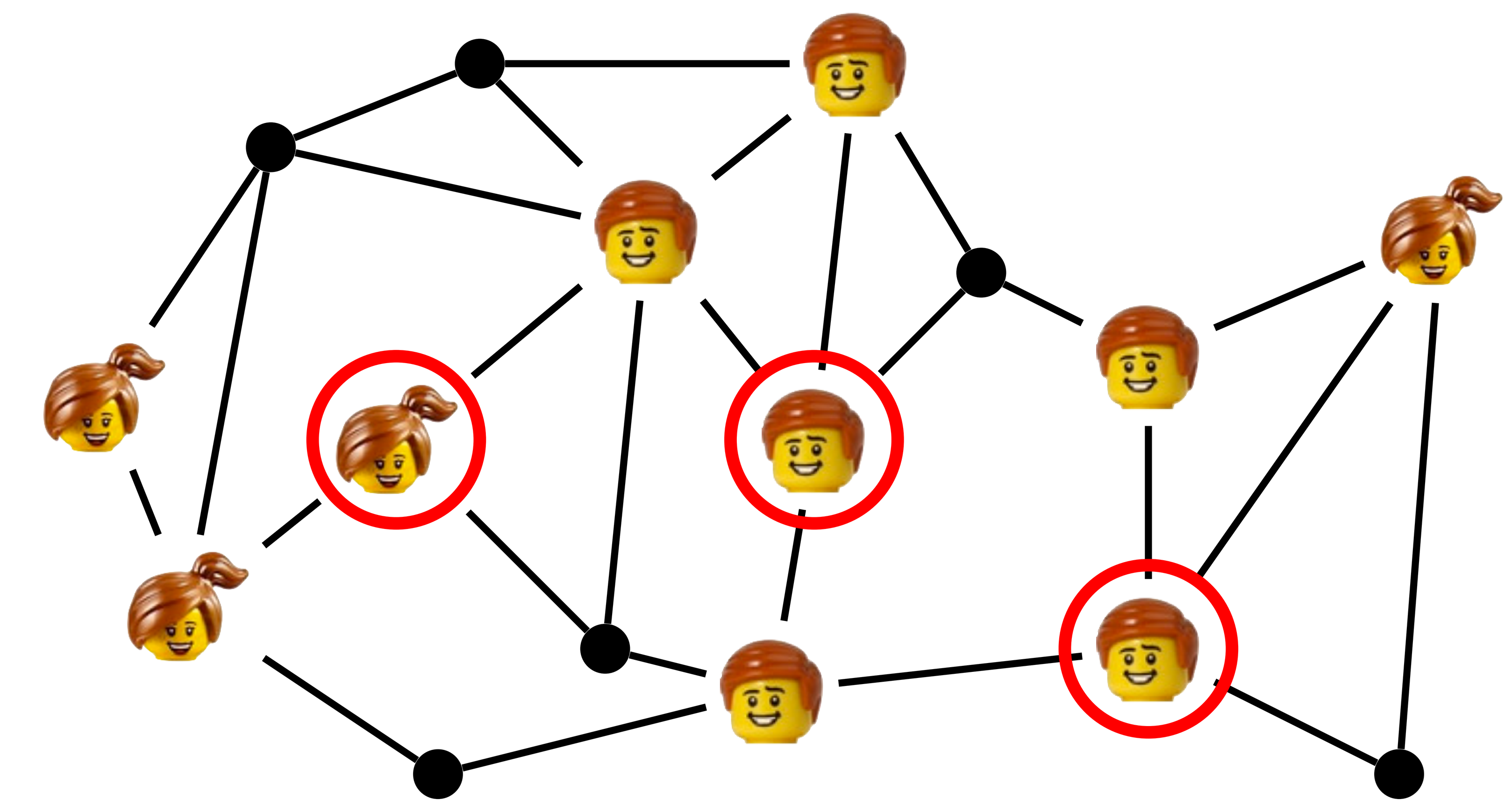
# Fairness in Social Influence Maximization via Optimal Transport

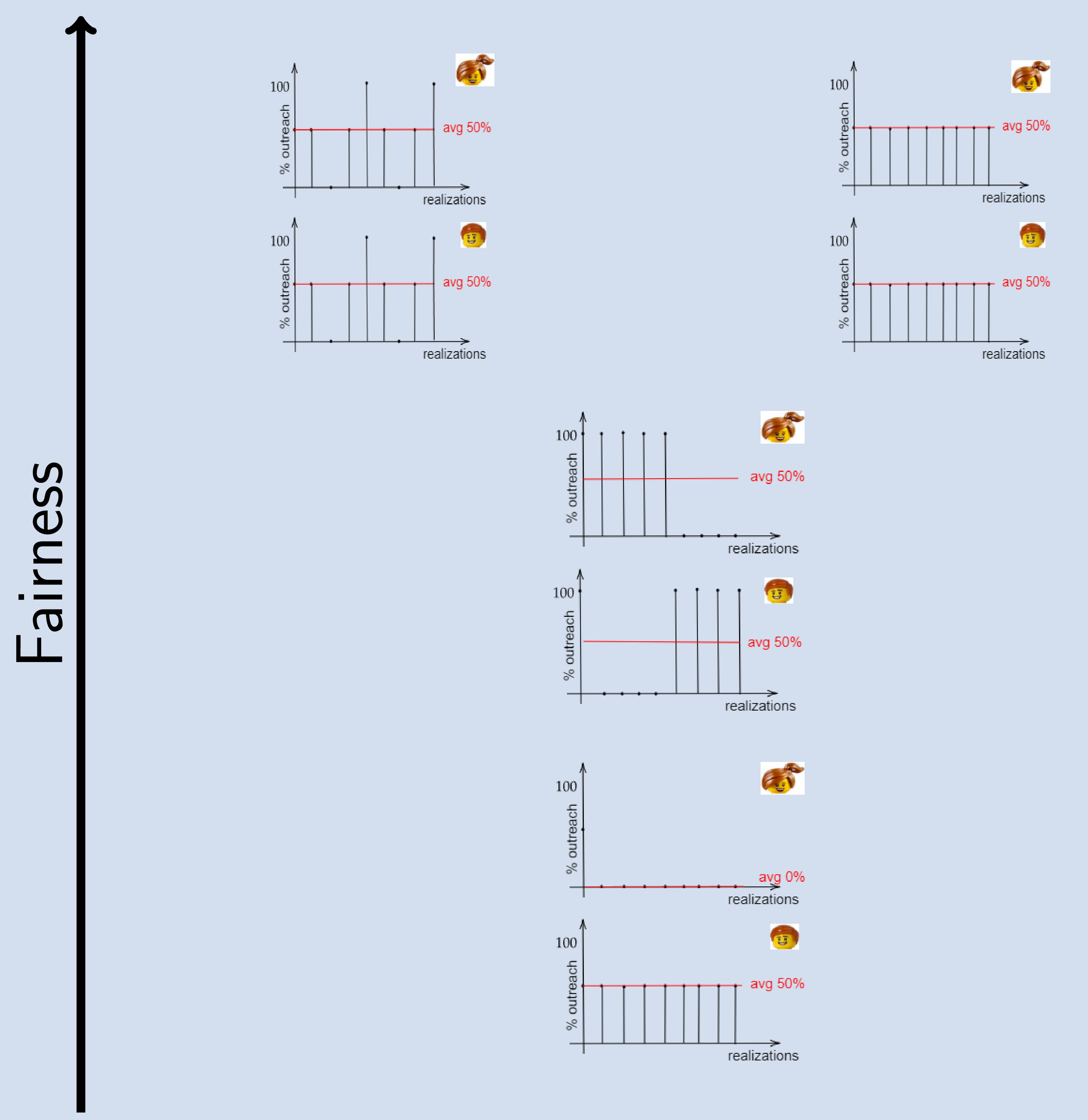
Shubham Chowdhary, Giulia De Pasquale, Nicolas Lanzetti, Ana-Andreea Stoica, Florian Dörfler



## 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?

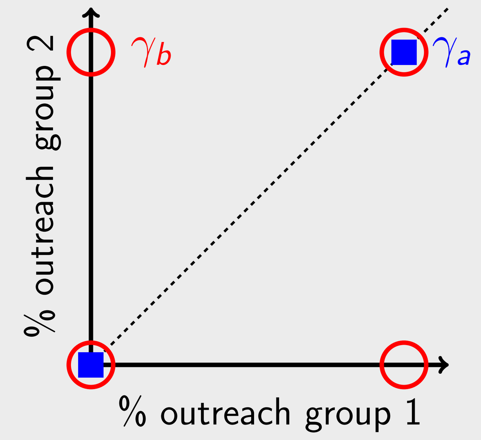



## Motivating example

Consider the following stochastic outcomes:

Outcome $\gamma_a$: in 50% of the cases everyone receives the information and in 50% no one.

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

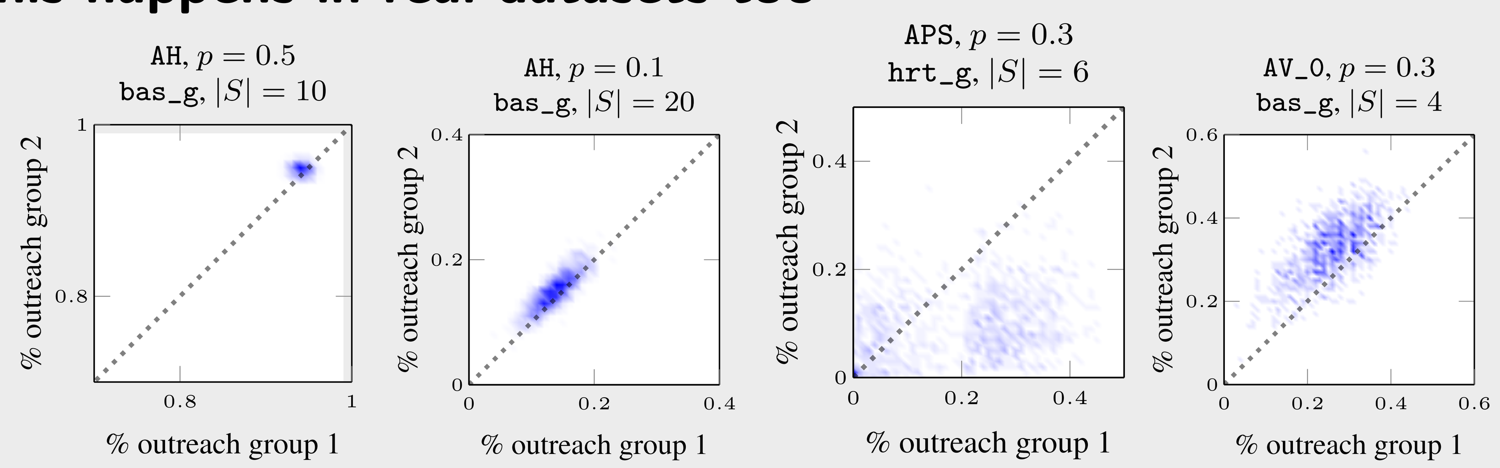



**Which outcome is fair?**

In $\gamma_a$, the percentage of members of group 1 who get the information *always* coincides with the percentage of people of group 2. In $\gamma_b$, this is not always true.

From a fairness perspective, $\gamma_a$ and $\gamma_b$ encode **very different outcomes**
…but $\gamma_a$ and $\gamma_b$ have the same marginals
and so we need to look at correlations.

## This happens in real datasets too



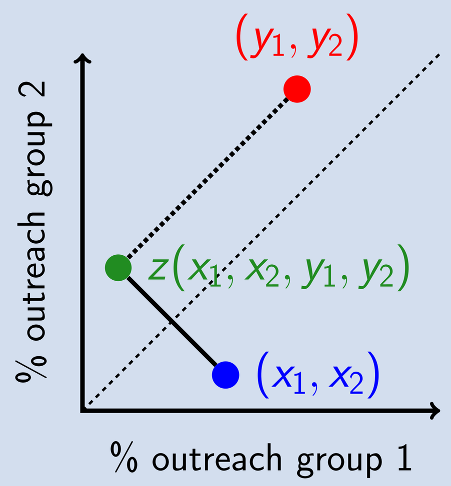

## Fairness via optimal transport

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** $\gamma$

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

…which is just a "normalized" $W_c(\gamma,\gamma^*)$

## Reminder: The optimal transport problem

For a transportation cost $c : ([0,1]\times[0,1]) \times ([0,1]\times[0,1]) \to \mathbb{R}_{\geq 0}$ 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

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

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

## S3D: our fairness metric to select seeds

**$\beta$-fairness:** revisit the transportation cost to tradeoff fairness and efficiency,

$$c_\beta((x_1,x_2),(y_1,y_2)) = \beta\|z(x_1,x_2,y_1,y_2) - (x_1,x_2)\| + (1-\beta)\|z(x_1,x_2,y_1,y_2) - (y_1,y_2)\|,$$

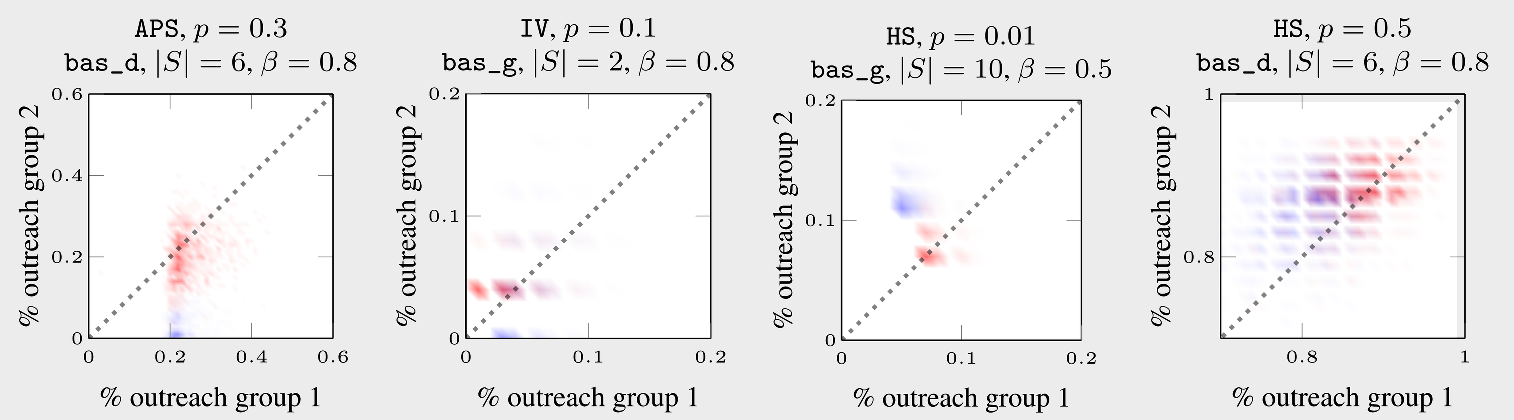
to define the $\beta$-fairness metric

$$\beta-\text{Fairness}(\gamma) = 1 - \frac{\sqrt{2}}{\max\{1,2-2\beta\}}W_{c_\beta}(\gamma,\gamma^*) = \mathbb{E}_{(x_1,x_2)\sim\gamma}\left[1 - \frac{\beta|x_1-x_2| + (1-\beta)|x_1+x_2-2|}{\max\{1,2-2\beta\}}\right]$$

**Our algorithm:**

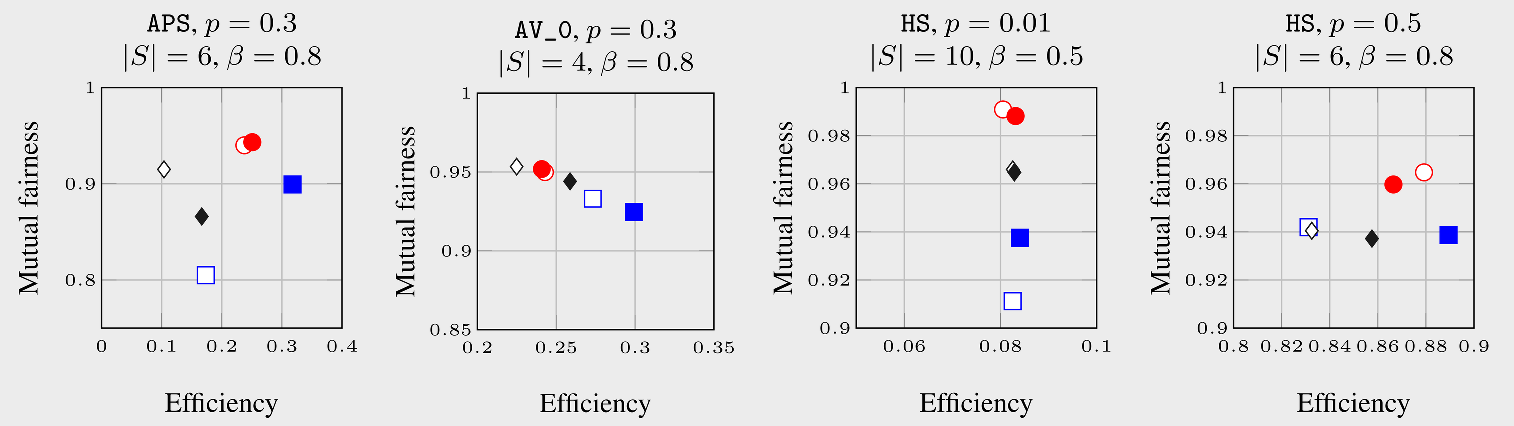
$S \leftarrow \{\}$, $S \leftarrow S_0$ ▷ initial collection of candidates, running seedset
**for** $k$ iterations **do** ▷ configurable $k$
  $V_S \leftarrow$ nodes reachable from $S$ via cascade, using SEEDSET_REACH routine
  $S' \leftarrow \{\}$
  **for** $|S|$ iterations **do** ▷ searching nearby states, $V_{S'}$, to get $S'$ (??)
    $S' \leftarrow S' \cup \{v\} \mid v \sim V_S$
    $V_{S'} \leftarrow$ nodes reachable from $S'$ in a fixed horizon, using SEEDSET_REACH
    $V_S \leftarrow V_S \setminus V_{S'}$
  $E_S \leftarrow -$BETA_FAIRNESS$(S,\beta)$ ▷ expected potential energy defined on $\beta$-fairness
  $E_{S'} \leftarrow -$BETA_FAIRNESS$(S',\beta)$
  $p_{\text{accept}} \leftarrow \min\{1, e^{E_S - E_{S'}}\}$ ▷ $S'$ acceptance on energy minimization
  **if** $x \sim \mathcal{B}(p_{\text{accept}})$ **then** ▷ Metropolis sampling
    $S^+ \leftarrow S'$ ▷ get a better seedset
  **else**
    **if** $x \sim \mathcal{B}(\epsilon)$ **then** ▷ for some small constant $\epsilon$
      $S^+ \leftarrow \{v_i\}_{i=1}^{|S|} \overset{|S|}{\sim} V_G$ ▷ random seedset
    **else**
      $S^+ \leftarrow S$ ▷ retain existing choice
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{S^+\}$
  $S \leftarrow S^+$ ▷ for next iteration
$S^* \leftarrow S \in \mathcal{S} \mid$ BETA_FAIRNESS$(S,\beta)$ is maximum ▷ via S3D_ITERATE
**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.