# Fairness in Social Influence Maximization via Optimal Transport

SHUBHAM CHOWDHARY, ETH Zürich, Switzerland
GIULIA DE PASQUALE*, Eindhoven University of Technology, The Netherlands
NICOLAS LANZETTI*, ETH Zürich, Switzerland
ANA-ANDREEA STOICA, Max Plank Institute, Germany
FLORIAN DÖRFLER, ETH Zürich, Switzerland

We examine fairness in social influence maximization, where the goal is to strategically select seeds that spread information through a network while ensuring balanced outreach among different communities. Existing fairness metrics focus on marginalized expected outreach within groups but fail to account for a joint stochasticity in information diffusion. This oversight can lead to misleading fairness evaluations, as highly unequal outcomes may still be classified as fair. To address this issue, we introduce a new fairness metric, *mutual fairness*, which considers outreach variability using optimal transport theory. We propose a novel seed-selection algorithm that optimizes both outreach and mutual fairness. Experiments on real datasets show that the proposed approach enhances fairness with minimal or even no loss in efficiency.

Keywords: Social Influence Maximization, Fairness, Optimal Transport

**Reference Format:**
Shubham Chowdhary, Giulia De Pasquale, Nicolas Lanzetti, Ana-Andreea Stoica, and Florian Dörfler. 2025. Fairness in Social Influence Maximization via Optimal Transport. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 7 pages.

## 1 Introduction

Social Influence Maximization (SIM) is the problem of selecting a small group of early adopters in a social network to maximize the spread of information, such as product endorsements [16], job advertisements [3], or public health awareness [25]. Since finding the optimal selection is NP-hard [10], heuristic strategies like greedy algorithms and centrality measures are commonly used. However, these approaches focus solely on network topology and ignore user demographics, raising fairness concerns. Real-world social networks consist of diverse social groups with varying sizes and connectivity patterns. Prioritizing highly connected nodes often excludes less-connected minorities, leading to inequitable information distribution and reinforcing biases [9, 23], particularly in critical domains like health, education, and employment.

*Both authors contributed equally to this research.

Authors' Contact Information: Shubham Chowdhary, ETH Zürich, Zürich, Switzerland, schowdhary@ethz.ch; Giulia De Pasquale, Eindhoven University of Technology, Eindhoven, The Netherlands, g.de.pasquale@tue.nl; Nicolas Lanzetti, ETH Zürich, Zürich, Switzerland, lnicolas@ethz.ch; Ana-Andreea Stoica, Max Plank Institute, Tübingen, Germany, ana-andreea.stoica@tuebingen.mpg.de; Florian Dörfler, ETH Zürich, Zürich, Switzerland, dorfler@ethz.ch.


*EWAF'25, June 30–July 02, 2025, Eindhoven, NL*

Several multiple group-level fairness metrics have been proposed over the years [6]. They fall under the notions of *equity* [8, 9, 20], *equality* [6], *max-min fairness* [7, 28], *welfare* [15], and *diversity* [23]: All of them quantify the fair distribution of influence across groups. In particular, Stoica et al. [20] propose a new SIM algorithm that operates under the constraint that, in expectation, the same percentage of users in each category is reached.

Many models of diffusion processes in the SIM problem are inherently stochastic, meaning that *who* gets the information transmitted can vary greatly from one run to another. Consider, as an example, the case in which 50% of realizations over a diffusion process, no one in group 1 receives the information and everyone in group 2 does, whereas in the other 50% it is the opposite. This circumstance would be classified as fair in expectation, even though it is commonly not perceived as fair. We show how this phenomenon is common in real-world data and how our proposed framework can detect such undesired scenarios. This prompts us to propose a novel fairness metric. Our work [5] presents two key contributions: mutual fairness, a new fairness metric based on optimal transport, and a novel seeding algorithm that balances fairness and efficiency in social influence maximization. Mutual fairness addresses limitations in existing metrics by considering all groups simultaneously and measuring deviations from an ideal equitable distribution. We use this metric to evaluate popular influence-spreading algorithms, revealing that network topology significantly impacts fairness—group-label-blind seed selection can lead to inequality in moderately homophilic networks, while only highly integrated or segregated networks tend to achieve a fairer outreach. The proposed seed selection algorithm optimizes for mutual fairness, enhancing fairness with minimal or even no loss in efficiency. This approach offers a comprehensive framework for designing and evaluating fair influence maximization strategies. Our work [5] is inspired by a recent line of work that draws on optimal transport theory [26] for fairness guarantees [2, 4, 18, 19, 22, 27]. To our knowledge, this is the first work to develop novel metrics and seeding algorithms that leverage optimal transport for the SIM problem.

## 2 Mutual Fairness via Optimal Transport

### 2.1 Motivating Example

Consider the SIM problem with network nodes belonging to two groups, group 1 and group 2, with each group having the outreach probability distribution $\mu_g = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, g \in \{1,2\}$, with $\delta_k$ representing the delta distribution at $k \in [0,1]$. That is, in 50% of the cases all members in group $g$ receive the information, and in 50% of the cases no one in group $g$ receives the information. It is tempting to say that this setting is fair since $\mu_1$ and $\mu_2$ coincide and therefore share the same expected value. We argue that this information does not suffice to claim fairness. Indeed, consider the two following probability distributions over the final configurations:

$$\gamma_a = 0.5 \cdot \delta_{(0,0)} + 0.5 \cdot \delta_{(1,1)}, \qquad \gamma_b = 0.25 \cdot \delta_{(0,0)} + 0.25 \cdot \delta_{(1,1)} + 0.25 \cdot \delta_{(0,1)} + 0.25 \cdot \delta_{(1,0)},$$

with $\delta_{(i,j)}$, representing the delta distribution at $(i,j) \in [0,1] \times [0,1]$. Interestingly, both $\gamma_a$ and $\gamma_b$ are "compatible" with $\mu_1$ and $\mu_2$: If we compute their marginals, we obtain $\mu_1$ and $\mu_2$. However, $\gamma_a$ and $\gamma_b$ encode two fundamentally different final configurations. In $\gamma_a$, the percentage of members of group $a$ who get the information *always* coincides with the percentage of people of group $b$. Conversely, in $\gamma_b$, more outcomes are possible; in particular, there is a probability of $0.25 + 0.25 = 0.5$ that all members of one group receive the information and no member of the other group receives it (see Fig. 1). Thus, from a fairness perspective, $\gamma_a$ and $\gamma_b$ encode very different outcomes. We therefore argue that a fairness metric should be expressed in terms of *joint* probability distribution $\gamma$, and not solely based on its marginals $\mu_1$ and $\mu_2$, as commonly done in the literature [8, 20].

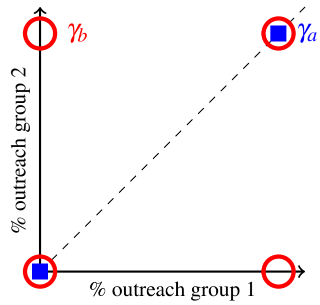


Fig. 1. Illustration of the $(\gamma_a, \gamma_b)$ example. The outcome of $\gamma_a$ is always fair: In all realizations the share of agents of group 1 receiving the information coincides with the share of agents of group 2. The outcome of $\gamma_b$ is instead largely unfair in 50% of the cases, as only one group receives the information. However, $\gamma_a$ and $\gamma_b$ share the same marginal distributions: In 50% of the cases everyone in the group receives the information and in 50% of the cases no one receives it.

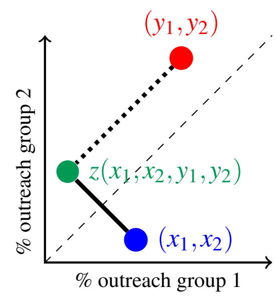


Fig. 2. The transportation cost $c$ measures the length of the solid segment, denoted by the perpendicular distance between the red and blue points w.r.t. the ideal fairness diagonal; shifts along the diagonal (dotted) are not considered for fairness and are only relevant for efficiency.

### 2.2 Mutual Fairness

Our motivating example prompts us to reason about fairness in terms of the joint probability measure $\gamma$, instead of its marginal distributions $\mu_1$ and $\mu_2$. Since $\gamma$ is a probability distribution (over all possible final configurations), we can quantify fairness by computing its "distance" from an "ideal" reference distribution $\gamma^*$ along the diagonal, capturing the ideal situation in which both groups receive the information in the same proportion. We do so by using tools from optimal transport.

*Background in optimal transport.* For a given (continuous) transportation cost $c : ([0,1] \times [0,1]) \times ([0,1] \times [0,1]) \to \mathbb{R}_{\geq 0}$, the optimal transport discrepancy between two probability distributions $\gamma_a \in \mathscr{P}([0,1] \times [0,1])$ and $\gamma_b \in \mathscr{P}([0,1] \times [0,1])$ is defined as

$$W_c(\gamma_a, \gamma_b) = \min_{\pi \in \Pi(\gamma_a, \gamma_b)} \mathbb{E}_{(x_1,x_2),(y_1,y_2) \sim \pi}, [c((x_1,x_2),(y_1,y_2))], \tag{1}$$

where $\Pi(\gamma_a, \gamma_b)$ is the set of probability distributions over $([0,1] \times [0,1]) \times ([0,1] \times [0,1])$ so that the first marginal is $\gamma_a$ and the second marginal is $\gamma_b$. Intuitively, the optimal transport problem quantifies the minimum transportation cost to morph $\gamma_a$ into $\gamma_b$ when transporting a unit of mass from $(x_1,x_2)$ to $(y_1,y_2)$ costs $c((x_1,x_2),(y_1,y_2))$. The optimization variable $\pi$ is called transportation plan and $\pi((x_1,x_2),(y_1,y_2))$ indicates the amount of mass at $(x_1,x_2)$ displaced to $(y_1,y_2)$. When the probability distributions are discrete (or the space $[0,1]$ is discretized), the transportation problem (1) is a finite-dimensional linear program and can therefore be solved efficiently [14].

*Optimal transport to measure fairness.* To address the identified limitations in evaluating fairness, we define fairness using the joint outreach probability distribution $\gamma \in P([0,1] \times [0,1])$, where $\gamma(x_1,x_2)$ represents the probability of $x_1$ and $x_2$ respective fractions of group 1 and 2 receiving the information. We introduce an optimal transport-based fairness metric, computing the minimal transformation cost required to shift $\gamma$ to an ideal fair

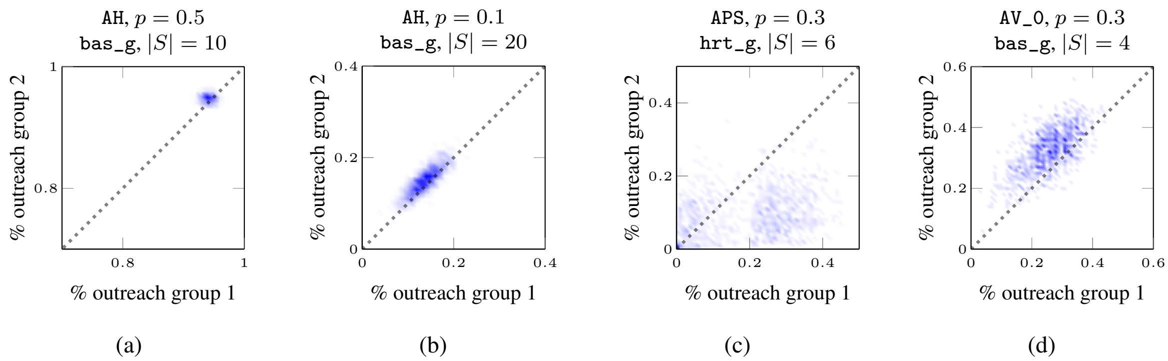


Fig. 3. Joint outreach probability distribution for different datasets, different propagation probabilities $p$, and seedsets cardinalities $|S|$.

distribution $\gamma^*$–wherein both groups receive the same proportion. Our *mutual fairness* metric is:

$$\text{FAIRNESS}(\gamma) = 1 - \sqrt{2}W_c(\gamma, \gamma^*), \tag{2}$$

where $W_c$ represents the optimal transport discrepancy between $\gamma$ and $\gamma^*$, and $c$ is the transportation cost defined as the Euclidean distance from the fairness axis, which is represented by the diagonal line in Fig. 2, and $\gamma^*$ is the reference distribution located on the top right corner of the plane (i.e., everyone receives the transmitted information). Finally, $\sqrt{2}$ is a normalization factor that ensures that the metric lies in the interval $[0, 1]$. This approach ensures that fairness is accounted for through each realization, and not just in expectation. Note that mutual fairness can be combined with other known metrics that account for efficiency using a re-weighting parameter $\beta \in [0, 1]$ that balances the trade-off.

## 3 Experiments and Results

We evaluate our approach on several real-world datasets: Add Health (`AH`), Antelope Valley variants 0 to 23 (`AV_{0-23}`) [24], APS Physics (`APS`) [12], Deezer (`DZ`) [17], High School Gender (`HS`) [13], Indian Villages (`IV`) [1], and Instagram (`INS`) [21].

Our first case study, Fig. 3, examines the joint outreach probability distribution across different datasets, revealing four distinct outcomes. In dense graphs with high cross-group connectivity and extreme propagation probabilities ($p \geq 0.5$ or $p \to 0$), outreach is nearly deterministic and highly fair (Fig. 3a), making fairness metrics redundant. For moderate $p$ (e.g., $p = 0.1$), outreach aligns along the diagonal (Fig. 3b), ensuring fairness in both expected and realized outreach. However, in cases such as APS with $p = 0.3$ (Fig. 3c), the outreach becomes highly stochastic, sometimes excluding one group entirely despite having a similar expected value across each group. Lastly, `AV_0` (Fig. 3d) shows a stochastic outreach with slight bias towards one group, where mutual fairness provides deeper insights than equity metrics alone.

The study in Figure 4 evaluates our proposed seed-selection algorithm, called Stochastic Seedset Selection Descent (`S3D`, detailed in Appendix A), across various datasets against baseline methods like the degree centrality (`bas_d`) and greedy (`bas_g`) based algorithms [11] and other heuristic based algorithms [20] (`hrt_d`, `hrt_g`). By

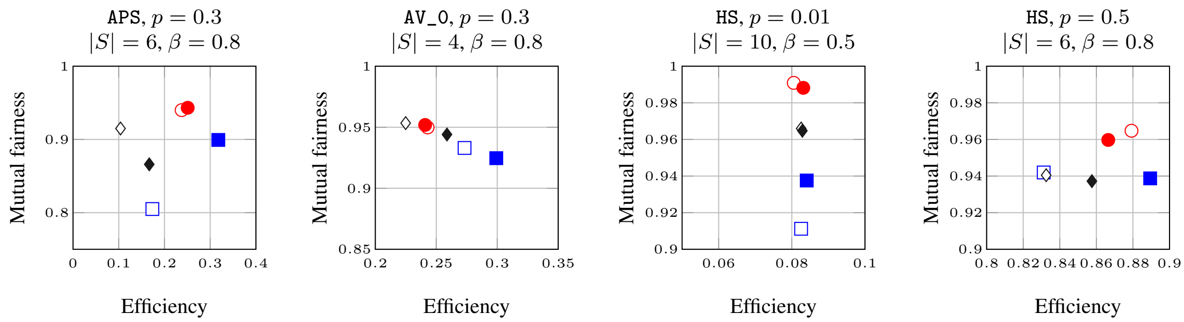


Fig. 4. S3D trade-off and improvement against other label-aware and label-blind algorithms for several datasets, propagation probabilities $p$, seed set cardinalities $|S|$ and fairness-efficiency tradeoffs $\beta$. Filled markers refer to greedy-based algorithms: ■ = bas_g, ● = S3D_g, and ♦ = hrt_g. Empty markers refer to degree-based algorithms: □ = bas_d, ○ = S3D_d, and ◇ = hrt_d. Along with S3D's performance one can also see difference between standard greedy, degree, and other heuristic-baseed algorithms.

initializing S3D with these baselines (S3D_d, S3D_g), our results show that S3D shifts the joint outreach probability distribution toward the diagonal, enhancing fairness while maintaining or slightly reducing efficiency (Fig. 4). In datasets with high cross-group connections (AH, DZ, INS), label-blind seed selection already achieves moderate fairness. For low cross-group connectivity datasets (APS), label-blind strategies naturally select diverse seeds, functioning similarly to S3D. However, in moderate cases (AV, HS, IV), S3D significantly outperforms label-blind strategies in terms of fairness.

## References

[1] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2013. The diffusion of microfinance. *Science* 341, 6144 (2013), 1236498.

[2] Emily Black, Samuel Yeom, and Matt Fredrikson. 2020. FlipTest: fairness testing via optimal transport. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT'20)* (2020), 111–121.

[3] Wei Chen, Wei Lu, and Ning Zhang. 2012. Time-critical influence maximization in social networks with time-delayed diffusion process. *In Proceedings of AAAI Conference of Artificial Intelligence* 26, 1 (2012), 1–7.

[4] Silvia Chiappa, Ray Jiang, Tom Stepleton, Aldo Pacchiano, Heinrich Jiang, and John Aslanides. 2020. A General Approach to Fairness with Optimal Transport. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT'20)* (2020), 3633–3640.

[5] Shubham Chowdhary, Giulia De Pasquale, Nicolas Lanzetti, Ana Andreea Stoica, and Florian Dörfler. 2025. Fairness in Social Influence Maximization via Optimal Trasport. *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)* (2025), 1–34.

[6] Golnoosh Farnadi, Behrouz Babaki, and Michel Gendreau. 2020. A Unifying Framework for Fairness-Aware Influence Maximization. *International World Wide Web Conference 2020* (2020), 714–722.

[7] Benjamin Fish, Ashkan Bashardoust, Danah Boyd, Sorelle Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2020. Gaps in Information Access in Social Networks? *International World Wide Web Conference 2019, San Francisco, USA* (2020), 480–490.

[8] Ali Junaid, Babaei Mahmoudreza, Abhijnan Chakraborty, Baharan Mirzasoleiman, Krishna P. Gummadi, and Adish Singla. 2023. On the Fairness of Time-Critical Influence Maximization in Social Network. *IEEE Transaction on knowledge and data engineering* 35, 3 (2023), 480–490.

[9] Fariba Karimi, Mathieu Génois, Claudia Wagner, Philipp Singer, and Markus Strohmaier. 2018. Homophily influences ranking of minorities in social networks. *Scientific reports* 8, 1 (2018), 11077.

[10] David Kempe, Jon Kleinberg, and Eva Tardos. 2003. Maximizing the Spread of Influence through a Social Network. *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2003), 137–146.

[11] David Kempe, Jon Kleinberg, and Éva Tardos. 2005. Influential nodes in a diffusion model for social networks. In *Automata, Languages and Programming: 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005. Proceedings 32*. Springer, 1127–1138.

[12] Eun Lee, Fariba Karimi, Claudia Wagner, Hang-Hyun Jo, Markus Strohmaier, and Mirta Galesic. 2019. Homophily and minority-group size explain perception biases in social networks. *Nature human behaviour* 3, 10 (2019), 1078–1087.

[13] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. 2015. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one* 10, 9 (2015), e0136497.

[14] Gabriel Peyré, Marco Cuturi, et al. 2019. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning* 11, 5-6 (2019), 355–607.

[15] Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Max Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. 2021. Fair influence maximization: A welfare optimization approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11630–11638.

[16] Matthew Richardson and Pedro Domingos. 2002. Mining knowledge-sharing sites for viral marketing. *In Proceedings of 8th International Conference on Knowledge, Discovery and Data Mining* (2002), 61–70.

[17] Benedek Rozemberczki and Rik Sarkar. 2020. Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 1325–1334.

[18] Yves Rychener, Bahar Taskesen, and Daniel Kuhn. 2022. Metrizing fairness. *arXiv preprint arXiv:2205.15049* (2022).

[19] Nian Si, Karthyek Murthy, Jose Blanchet, and Viet Anh Nguyen. 2021. Testing Group Fairness via Optimal Transport Projections. *Proceedings of the 38th International Conference on Machine Learning* (2021), 9649–9659.

[20] Ana-Andreea Stoica, Jessy Xinyi Han, and Augustin Chaintreau. 2020. Seeding network influence in biased networks and the benefits of diversity. *Proceedings of The Web Conference 2020* (2020), 2089–2098.

[21] Ana-Andreea Stoica, Christopher Riederer, and Augustin Chaintreau. 2018. Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity. In *Proceedings of the 2018 World Wide Web Conference*. 923–932.

[22] Bahar Taskesen, Jose Blanchet, Daniel Kuhn, and Viet Anh Nguyen. 2021. A statistical test for probabilistic fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 648–665.

[23] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. 2018. Group-Fairness in Influence Maximization. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)* (2018), 5997–6005.

[24] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. 2019. Group-fairness in influence maximization. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)* (2019), 5997–6005.

[25] Thomas W. Valente and Patchareeya Pumpuang. 2007. Identifying opinion leaders to promote behaviour change. *Health, Education & Behaviour* 34, 6 (2007), 881–896.

[26] Cédric Villani. 2009. *Optimal transport: old and new*. Vol. 338. Springer.

[27] Meike Zehlike, Alex Loosley, Håkan Jonsson, Emil Wiedemann, and Philipp Hacker. 2022. Beyond incompatibility: Trade-offs between mutually exclusive fairness criteria in machine learning and law. *arXiv preprint arXiv:2212.00469* (2022).

[28] Jianming Zhu, Smita Ghosh, and Weili Wu. 2019. Group Influence Maximization Problem in Social Networks. *IEEE Transactions on Computational Social Sciences* 6, 6 (2019), 1156–1164.

## A Sochastic Seed Selection Descent (S3D)

**Algorithm 1** Stochastic Seedset Selection Descent

**Input**: Social Graph $G(V_G, E_G)$, initial seed set $S_0$, $\beta$ fairness weight, $\varepsilon$-tolerance
**Output**: Optimal seedset $S^*$

$\mathscr{S} \leftarrow \{\}$, $S \leftarrow S_0$ ▷ 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
    $V_S \leftarrow V_S \setminus V_{S'}$
  $E_S \leftarrow -\text{BETA\_FAIRNESS}(S, \beta)$ ▷ expected potential energy defined on $\beta$-fairness
  $E_{S'} \leftarrow -\text{BETA\_FAIRNESS}(S', \beta)$
  $p_{\text{accept}} \leftarrow \min\{1, \mathrm{e}^{E_S - E_{S'}}\}$ ▷ $S'$ acceptance on energy minimization
  **if** $x \sim \mathscr{B}(p_{\text{accept}})$ **then** ▷ Metropolis sampling
    $S^+ \leftarrow S'$ ▷ get a better seedset
  **else**
    **if** $x \sim \mathscr{B}(\varepsilon)$ **then** ▷ for some small constant $\varepsilon$
      $S^+ \leftarrow \{v_i\}_{i=1}^{|S|} \overset{|S|}{\sim} V_G$ ▷ random seedset
    **else**
      $S^+ \leftarrow S$ ▷ retain existing choice
  $\mathscr{S} \leftarrow \mathscr{S} \cup \{S^+\}$
  $S \leftarrow S^+$ ▷ for next iteration
$S^* \leftarrow S \in \mathscr{S} \mid \text{BETA\_FAIRNESS}(S, \beta)$ is maximum ▷ via S3D_ITERATE
**return** $S^*$