FAIR INFLUENCE MAXIMIZATION BEYOND PREDEFINED GROUPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Fair influence maximization aims to spread information widely while avoiding large disparities across different parts of a network. Most existing methods assume that protected groups or community labels are known. This assumption is often unrealistic when demographic information is missing, incomplete, sensitive, or defined by multiple overlapping attributes. In addition, raw reach alone may not provide a fair comparison because some nodes are structurally harder to reach than others. We introduce AURA-IM, a framework for fair influence maximization without predefined groups. AURA-IM does not use observed group memberships or the true numbers of attributes/groups during seed selection. Instead, it uses target-specific reverse-reachable samples to estimate how much service each node can realistically attain under a given seed budget. This attainable service is used to define a normalized opportunity signal. A structural multi-view encoder then learns node representations, and a latent auditor adaptively identifies underserved marginal and intersectional groups. Their risk is modeled using KL-robust utility and upper-tail aggregation, and is jointly optimized with influence spread through an adversarial extragradient–Frank-Wolfe procedure followed by discrete refinement. Experiments on eight real-world networks show that AURA-IM achieves strong influence spread while maintaining competitive observed-group fairness. These results suggest that fairness-aware influence maximization is possible even when protected groups are not available during optimization.

1 INTRODUCTION

Influence maximization (IM) asks how to select a small seed set whose activation yields large expected diffusion over a network (Kempe et al., 2003). Classical IM is deliberately population-agnostic: it optimizes aggregate spread, even when the resulting reach is systematically uneven across structurally different parts of the network. Fair IM (FIM) augments the spread objective with distributional requirements intended to protect populations that would otherwise receive disproportionately little service (Tsang et al., 2019; Khajehnejad et al., 2020).

A recurring limitation of group-fair FIM is that the protected partition is typically available before optimization. This assumption is restrictive when demographic memberships are missing, incomplete, sensitive, or simply not part of the operational data. It is also limiting in multi-attribute settings, where vulnerability can arise simultaneously along several axes and through their intersections. A second difficulty is structural: raw reach probabilities are not directly comparable across nodes because topology imposes different ceilings on how much influence each node can realistically receive under a finite seed budget. A node can appear poorly served simply because it is intrinsically difficult to reach.

We introduce Adaptive Unobserved-Group Attainability for Influence Maximization (AURA-IM), a static-graph framework for fair influence maximization beyond predefined groups. AURA-IM separates seed optimization from observed group-label values. It first builds a global reverse-reachable (RR) (Borgs et al., 2014) oracle for spread together with target-specific RR banks that quantify how much service each target can attain under a budget K. These target-specific samples define a budget-conditioned attainability frontier. AURA-IM then constructs stochastic structural views of the same fixed topology and learns an opportunity-aware representation using only graph structure and the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

graph-derived attainability target. Realized target reach is normalized by its attainable frontier to produce a node-level opportunity signal.

AURA-IM next learns an overcomplete factorized latent auditor. The auditor contains multiple latent axes and group slots with learnable continuous activity gates, but it is never told the observed demographic attributes or their group counts. It searches for low-opportunity marginal and intersectional subpopulations and evaluates their service through KL-distributionally robust lower utility and a finite-sample CVaR upper-tail risk (Rockafellar et al., 2000). Structural regularizers prevent the adversary from manufacturing arbitrarily small, incoherent, redundant, or degenerate latent groups.

Seed selection is formulated as an adversarial game. For a fixed relaxed seed vector, the auditor searches for structurally plausible vulnerable groups by extragradient ascent (Korpelevich, 1976). For the updated auditor state, the seed player maximizes differentiable RR influence minus robust latent fairness risk using Frank-Wolfe (Frank et al., 1956) ascent on the capped simplex, which preserves the exact relaxed budget. The relaxed solution is then converted to a hard Top-K seed set and refined by a bounded one-for-one swap search under the same hard RR/fairness objective. Observed multi-attribute label values enter only the post-hoc evaluation after the seed sets have been fixed. Main contributions are as follows:

- Attainability-normalized opportunity. We define a node-wise, budget-conditioned attainable service frontier from target-specific RR maximum coverage, and normalize realized target reach by this frontier. The resulting opportunity variable distinguishes insufficient service from structural non-attainability.
- Adaptive latent multi-axis fairness without observed group memberships. A factorized auditor learns soft memberships on multiple latent axes together with continuous axis/group activities, while size, coherence, crispness, independence, complexity, and minimum-capacity penalties regularize the discovered structure.
- Robust marginal and intersectional protection. KL-DRO provides a conservative lower bound on subgroup opportunity, a CVaR top-tail operator concentrates on the most vulnerable latent risks, and a mass-constrained beam search identifies low-opportunity intersections without enumerating the full Cartesian product.
- Adversarial Frank-Wolfe seed optimization. A genuine two-stage extragradient update trains the latent auditor, while the seed player uses a cardinality-preserving Frank-Wolfe step on the capped simplex. Hard Top-K extraction and RR-aware fairness swaps complete the discrete solution.

2 Related Work

Early work on FIM formalized group constraints and multi-objective submodular formulations (Tsang et al. 2019), unified multiple fairness criteria through mixed-integer optimization (Farnadi, Babaki, and Gendreau 2020), and adopted social-welfare objectives to control the fairness-efficiency trade-off (Rahmattalabi et al. 2021). Later variants considered randomization for maximin fairness (Becker et al. 2022), time-critical disparities in diffusion (Ali et al. 2021), and balance under homophily (Anwar, Saveski, and Roy 2021). These formulations established fairness as a first-class optimization target, but they typically assume that the relevant demographic or community partition is specified before optimization.

Subsequent work has diversified both representations and optimizers. Khajehnejad et al. (2020) learn adversarial graph embeddings that suppress observed sensitive-group information while retaining influence-relevant structure. Multi-objective approaches search for spread-fairness trade-offs (Gong and Guo 2023), while FIM-GERL couples neural optimization with multigroup welfare objectives (Zhang et al. 2025). Community-aware methods such as SFIMCO exploit overlapping structure for scalable welfare-oriented selection (Chen and Rezaeipanah 2025). More recent methods increasingly use learned sequential policies and stronger robustness machinery: HOLF and DQ4FairIM formulate group-aware seed selection with deep reinforcement learning (Guo, Li, and Mo 2026; Saxena et al. 2026); Gao and Ito (2026) optimize worst-case performance across multiple plausible community partitions; Zhang et al. (2026) develop randomized maximin FIM via stochastic Frank-Wolfe; and FIMMOGA++ combines reverse-influence sampling with multi-objective evolutionary search (Srinivas et al. 2026). These advances improve expressiveness and scalability, but

the fairness signal is still usually defined with respect to explicit groups, community partitions, or fixed group-aware rewards.

AURA-IM addresses a different perspective; during optimization it receives only graph structure and graph-derived attainability signals, while observed attributes, group memberships, and the true numbers of attributes and groups are withheld. Rather than suppressing known sensitive labels or optimizing a known partition, it learns an overcomplete latent multi-axis auditor that searches for underserved marginal and intersectional populations, then couples their robust risk to RR-based seed optimization. This makes the fairness target adaptive to latent structural disadvantage while preserving observed groups solely for post-hoc evaluation. The distinction is complementary to prior group-aware FIM: AURA-IM is designed for settings where the protected partition is unavailable, incomplete, sensitive, or inappropriate to hard-code before seed selection.

3 Proposed Method

We propose AURA-IM as an adversarial auditor–seed procedure: sample global and target-specific RR banks; derive target attainability; train a structural opportunity encoder; initialize a latent multi-axis auditor; alternate an auditor update that seeks underserved latent groups with a seed update that balances influence and their risk; and round and refine each seed set. The two players are the auditor and seed selector, rather than the observed groups. Figure 1 gives the data flow. Throughout selection, attribute-file node identifiers align the node universe, while their observed attribute values remain unused.

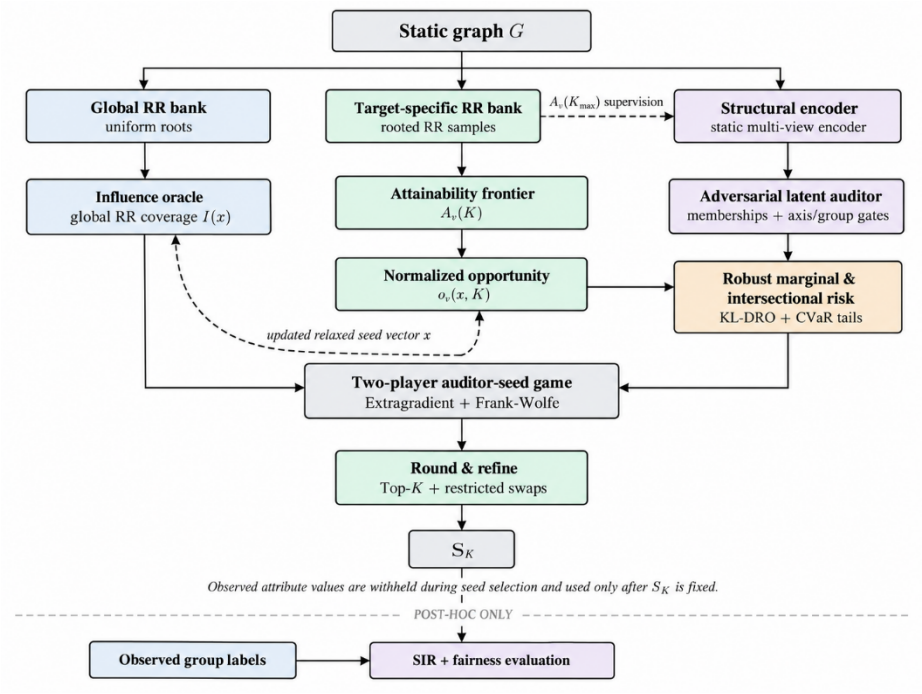


Figure 1: The implemented static AURA-IM pipeline. Target RR samples supply service and a greedy attainable reference, with the latter supervising the encoder. The global RR bank estimates influence. The adversarial auditor–seed game alternates risk-seeking extragradient and influence–fairness Frank–Wolfe updates before rounding and hard swaps; observed group values are revealed only for evaluation.

Graph construction and sampled influence. Let $G = (V, E)$ have $N = |V|$ nodes and $\mathcal{K}$ be the requested seed budgets. The node universe is the union of edge endpoints and identifiers in the attribute file; the latter's attribute values are sealed during selection. The implementation maps identifiers to contiguous indices, removes self-loops, and merges repeated *ordered* edge pairs. The active configuration is static. For infection probability β and recovery probability γ, the reverse traversal probability is $T = \beta/max\{\beta + \gamma - \beta\gamma, \epsilon_T\}$, where ϵ_T is a denominator safeguard. An

RR sample starts at its root and traverses incoming diffusion neighbors to a depth limit, accepting a previously unseen neighbor independently with probability T. Let $\mathcal{R}^G = \{R_i^G\}_{i=1}^M$ be M samples with uniformly drawn roots. The normalized sampled influence of a discrete seed set S is

$$I(S) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}\{S \cap R_i^G \neq \varnothing\}. \quad (1)$$

Target-specific service and attainable reference. Separately from the global bank, R target samples R_{vj}^T are drawn at each $v \in V$, indexed by $j \in \{1, \ldots, R\}$. Excluding their root, $\bar{R}_{vj}^T = R_{vj}^T \backslash \{v\}$, defines sampled service from other nodes:

$$p_v(S) = R^{-1} \sum_{j=1}^{R} \mathbf{1}\{S \cap \bar{R}_{vj}^T \neq \varnothing\}. \quad (2)$$

For each candidate source $u \neq v$, its mask $B_{vu} = \{j : u \in \bar{R}_{vj}^T\}$ contains the target samples it can cover. Starting with $C_v(0) = \varnothing$, the proposed method repeatedly adds a distinct mask with maximum positive gain $|B_{vu} \backslash C_v(k-1)|$ and stops when no positive gain remains. Its union $C_v(k)$ after at most k selections defines the sampled attainability frontier

$$A_v(k) = |C_v(k)| / R, \qquad k \leq K_{\max} = max\mathcal{K}. \quad (3)$$

It stays fixed if no positive gain remains and equals one after all samples become covered. The encoder predicts $A_v(K_{\max})$. For either a hard set S or relaxed vector x, budget-specific opportunity is

$$o_v(\cdot, K) = \begin{cases} min\{1, p_v(\cdot) / (A_v(K) + \epsilon_o)\}, & A_v(K) > \epsilon_o, \\ 1, & A_v(K) \leq \epsilon_o, \end{cases} \quad (4)$$

where ϵ_o is a denominator floor and a zero-frontier target is assigned full opportunity by convention. With $x_v \in [0, 1]$ and logarithm floor ϵ_x, the relaxed RR-set coverage and its two bank averages are

$$\begin{aligned} c_B(x) &= 1 - exp\left(\sum_{u \in B} \log clip\left(1 - x_u, \epsilon_x, 1\right)\right), \\ I(x) &= M^{-1} \sum_i c_{R_i^G}(x), \qquad p_v(x) = R^{-1} \sum_j c_{\bar{R}_{vj}^T}(x). \end{aligned} \quad (5)$$

The expression computes a product of nonselection factors over an RR set with a numerical logarithm floor. The proposed method stores both banks as sparse incidence matrices, so the inner sums in Eq. (5) are sparse matrix–vector products. For a hard set, Eqs. (1)–(4) are recomputed with Boolean RR hits rather than the clipped relaxation.

Proposition 1 (Fixed-bank coverage). *For fixed sampled RR banks, $I(S)$ and every $p_v(S)$ are normalized, monotone, and submodular; $p_v(\{v\}) = 0$. Let $p_v^\star(k) = \max_{|S| \leq k} p_v(S)$. The greedy frontier is nondecreasing and satisfies $A_v(k) \geq \left[1 - (1 - 1/k)^k\right] p_v^\star(k)$ for $1 \leq k \leq K_{\max}$.*

Proofs appear in Appendix B.5.

Opportunity-aware multi-view representation. Independent edge drops of the static oriented structural adjacency form two persistent views $r \in \{1, 2\}$ and a more strongly dropped context view $r = c$. For each view let $A^{(r)}$ be its adjacency, $d_v^{(r)} = \sum_u A_{vu}^{(r)}$ its row degree, $t_v^{(r)} = 2d_v^{(r)}$, and $\widetilde{A}^{(r)} = D_r^{-1}\left(A^{(r)} + I_N\right)$ its row-normalized self-loop adjacency, with $(D_r)_{vv} = max\{\sum_u \left(A^{(r)} + I_N\right)_{vu}, 1\}$. A stationary structural signal $\pi^{(r)}$ is obtained by a restart iteration on $\left(\widetilde{A}^{(r)}\right)^\mathsf{T}$, initialized uniformly and normalized at the end (Appendix B.1). The proposed method constructs five structural coordinates and concatenates their zero-, one-, and two-hop versions:

$$\begin{aligned} f_v^{(r)} = \; & [log\left(1 + d_v^{(r)}\right), log\left(1 + d_v^{(r)}\right), log\left(1 + t_v^{(r)}\right), \\ & \left(\widetilde{A}^{(r)} \log\left(1 + t^{(r)}\right)\right)_v, log\left(1 + N\pi_v^{(r)}\right)], \\ H_{raw}^{(r)} = \; & \left[f^{(r)}, \widetilde{A}^{(r)} f^{(r)}, \left(\widetilde{A}^{(r)}\right)^2 f^{(r)}\right], \quad H_{vj}^{(r)} = \frac{\left(H_{raw}^{(r)}\right)_{vj} - \mu_j}{\sigma_j + \epsilon_H}. \end{aligned} \quad (6)$$

Here μ_j and σ_j are the mean and population standard deviation of raw coordinate j jointly over nodes and all three views, and ϵ_H is a standardization floor. The repeated first two coordinates reflect the static-view implementation, which uses the same row degree for its nominal out- and in-

degree features. A shared two-layer MLP e gives $z_v^{(1)} = e\left(H_v^{(1)}\right)$ and $z_v^{(2)} = e\left(H_v^{(2)}\right)$; a separate context MLP gives $z_v^c = e_c\left(H_v^{(c)}\right)$. The persistent embedding is $z_v^p = \left(z_v^{(1)} + z_v^{(2)}\right)/2$. The representation is trained once before the auditor–seed game and then detached. Training minimizes

$$L_{enc} = \lambda_{view}L_{view} + \lambda_{NCE}L_{NCE} + \lambda_{ctx}L_{ctx} + \lambda_{orth}L_{orth} + \lambda_{opp}L_{opp} + \lambda_{var}\{L_{var}\left(Z^p\right) + L_{var}\left(Z^c\right)\}. \quad (7)$$

For row-normalized embeddings $\widehat{z} = z/max\left(\parallel z \parallel_2, \epsilon_n\right)$, a sampled contrastive node subset $\mathcal{B} \subseteq V$, and $s_{ij} = \langle \widehat{z}_i^{(1)}, \widehat{z}_j^{(2)} \rangle / T_{NCE}$, the implemented terms are

$$\begin{aligned}
L_{view} &= MSE\left(\widehat{Z}^{(1)}, \widehat{Z}^{(2)}\right), \\
L_{NCE} &= -\tfrac{1}{2|\mathcal{B}|}\textstyle\sum_{i\in\mathcal{B}}\left[\log\frac{e^{s_{ii}}}{\sum_{j\in\mathcal{B}}e^{s_{ij}}} + log\frac{e^{s_{ii}}}{\sum_{j\in\mathcal{B}}e^{s_{ji}}}\right], \\
L_{ctx} &= MSE\left(d_c\left(Z^c\right), H^{(c)} - \tfrac{1}{2}\left(H^{(1)} + H^{(2)}\right)\right), \\
L_{orth} &= N^{-1}\textstyle\sum_v \langle \widehat{z}_v^p, \widehat{z}_v^c \rangle^2, \\
L_{opp} &= N^{-1}\textstyle\sum_v \left[\sigma\left(h_o\left(z_v^p\right)\right) - A_v\left(K_{\max}\right)\right]^2, \\
L_{var}(Z) &= d_z^{-1}\textstyle\sum_{j=1}^{d_z}\left[s_{\min} - \sqrt{Var_v\left(Z_{vj}\right) + \epsilon_{var}}\right]_+ .
\end{aligned} \quad (8)$$

Here capital Z stacks node embeddings, MSE averages squared entries, d_c is a context decoder, h_o an opportunity-prediction MLP, σ the logistic function, T_{NCE} the contrastive temperature, d_z the embedding dimension, $s_{\min}$ the variance target, $\epsilon_n, \epsilon_{var}$ numerical floors, and $[y]_+ = max(y, 0)$. The sample variance Var_v is over nodes. View agreement and symmetric contrastive alignment train the persistent views; context-difference reconstruction and orthogonality retain a separate context component; the opportunity head predicts $A_v\left(K_{\max}\right)$; and the variance terms deter collapse. The six coefficients in Eq. (7) weight the corresponding losses. Adam (Kingma et al., 2014) trains the encoder with weight decay and gradient norm clipping; the stationary iteration and numerical settings are in Appendix B.1 and Section D. The auditor uses $Z^p = \left(z_v^p\right)_{v\in V}$ after training.

Adversarial latent-group auditor. Let A_* and G_* be capacity limits on latent axes a and group slots g, not observed attribute or group counts. A two-layer GELU MLP applied to z_v^p produces A_*G_* logits $\updownarrow_{vag}$, reshaped by axis and slot. Learned axis and slot logits ξ_a, ζ_{ag} yield

$$\begin{aligned}
\alpha_a &= \sigma\left(\xi_a/T_g\right), \qquad \delta_{ag} = \sigma\left(\zeta_{ag}/T_g\right), \\
q_{vag} &= softmax_g\left(\updownarrow_{vag} + logmax\{\delta_{ag}, \epsilon_g\}\right), \qquad w_{ag} = \alpha_a\delta_{ag}.
\end{aligned} \quad (9)$$

Here T_g is gate temperature, ϵ_g a log floor, q_{vag} a node's soft membership within axis a, and w_{ag} the slot activity. The axis gate α_a does not modify the within-axis softmax; the slot gate δ_{ag} both shifts its logit and contributes to w_{ag}. For intersection search and hard refinement the active axes are $\mathcal{A}_+ = \{a : \alpha_a \geq t_a\}$ and their active slots are $\mathcal{G}_a^+ = \{g : \delta_{ag} \geq t_g\}$, with top-gate fallback to the minimum capacities n_a, n_g when either threshold leaves too few. The auditor network and gate logits are optimized together.

The auditor objective subtracts penalties that discourage undersized or oversized slots, fragmented membership, redundant axes, diffuse membership, and degenerate gate activity:

$$\begin{aligned}
P(\theta) = \quad & \lambda_s P_{size} + \lambda_e P_{smooth} + \lambda_i P_{ind} + \lambda_h P_{ent} \\
& + \lambda_a P_{axis} + \lambda_g P_{group} + \lambda_b P_{binary} + \lambda_c P_{capacity}.
\end{aligned} \quad (10)$$

Let $\bar{q}_{ag} = N^{-1}\sum_v q_{vag}$, $Q_a = \left(q_{vag} - \bar{q}_{ag}\right)_{vg}$, and $d_P(z) = max\left(z, \epsilon_P\right)$, where ϵ_P is a denominator floor. The fixed sample $\mathcal{E}_s$ consists of retained input edge rows; l, u bound mean slot membership. The eight penalties used in Eq. (10) are

$$
\begin{aligned}
P_{size} &= \frac{\sum_{ag} w_{ag}\left([l-\bar{q}_{ag}]_+^2+[\bar{q}_{ag}-u]_+^2\right)}{d_P\left(\sum_{ag} w_{ag}\right)},\\
P_{smooth} &= \frac{\sum_{(v,v')\in\mathcal{E}_s}\sum_{ag} w_{ag}\left(q_{vag}-q_{v'ag}\right)^2}{|\mathcal{E}_s| d_P\left(\sum_{ag} w_{ag}\right)},\\
P_{ind} &= \frac{\sum_{a<b}\alpha_a\alpha_b mean\left[\left(Q_a^{\mathsf{T}}Q_b/N\right)^{\odot 2}\right]}{d_P\left(\sum_{a<b}\alpha_a\alpha_b\right)},\\
P_{ent} &= \frac{\sum_a \alpha_a\left[-N^{-1}\sum_{vg} q_{vag}\log(q_{vag}+\epsilon_P)\right]}{d_P\left(\sum_a \alpha_a\right)},\\
P_{axis} &= A_*^{-1}\textstyle\sum_a \alpha_a, \qquad P_{group} = \left(A_*G_*\right)^{-1}\sum_{ag}\alpha_a\delta_{ag},\\
P_{binary} &= A_*^{-1}\textstyle\sum_a \alpha_a\left(1-\alpha_a\right) + \left(A_*G_*\right)^{-1}\sum_{ag}\delta_{ag}\left(1-\delta_{ag}\right),\\
P_{capacity} &= \left[n_a-\textstyle\sum_a \alpha_a\right]_+^2 + \frac{\sum_a \alpha_a\left[n_g-\sum_g \delta_{ag}\right]_+^2}{d_P\left(\sum_a \alpha_a\right)}.
\end{aligned}
\tag{11}
$$

Here $[y]_+ = max(y, 0)$, $mean$ averages entries, and $\odot 2$ squares entries elementwise; n_a, n_g are minimum axis and within-axis slot capacities. Set $P_{smooth} = 0$ when $\mathcal{E}_s$ is empty. The first four penalties address size, graph variation, redundancy across axes, and uncertain membership; the remaining four control structure and discourage all gates from closing. Each coefficient λ weighs the indicated term. Sampling and initialization details are in Appendix B.2.

KL-robust marginal and intersectional risk. For a nonnegative membership vector m of positive mass, define $P_m(v) = m_v/\sum_u m_u$. On a finite positive temperature grid $\mathcal{H}$, the implemented KL-robust lower utility is

$$U(m, o) = \max_{\eta\in\mathcal{H}}\left\{-\eta\rho - \eta\log\textstyle\sum_{v\in V} P_m(v)e^{-o_v/\eta}\right\}, \tag{12}$$

where ρ is the KL radius and $o = (o_v)_{v\in V}$ is current opportunity. This is a finite-grid evaluation of the dual, not an asserted exact continuous optimization. Marginal risks are $r_{ag} = w_{ag}\left[1 - U\left(q_{\cdot ag}, o\right)\right]$. For a vector of m risks sorted as $r_{(1)}^{\downarrow} \geq \cdots \geq r_{(m)}^{\downarrow}$, the empirical upper-tail operator is

$$Tail_\tau(r) = h^{-1}\textstyle\sum_{i=1}^{h} r_{(i)}^{\downarrow}, \qquad h = max\{1, \lceil(1-\tau)m\rceil\}, \tag{13}$$

with tail level τ. Applied to all A_*G_* gated marginal risks, it gives R_{attr}; threshold activity is not used to discard marginal terms in the differentiable game. To find intersection candidates, the proposed method instead searches the threshold-active slots. For a soft membership vector m define its mass $M(m) = \sum_v m_v$ and nominal mean service $\bar{o}(m) = \sum_v m_v o_v/\left(M(m) + \epsilon_{beam}\right)$, where ϵ_{beam} is a denominator floor. Active slots with $M\left(q_{\cdot ag}\right) \geq \mu_{\min} N$ are ranked by increasing $\bar{o}$ and the lowest B form a beam. A beam slot can be extended only with a slot from another axis; a product membership below the same mass threshold is rejected. The lowest-$\bar{o}$ retained *pairs* comprise $\mathcal{C}$. For $c = ((a, g), (b, h)) \in \mathcal{C}$, $m_v^c = q_{vag}q_{vbh}$ and $r_c = w_{ag}w_{bh}\left[1 - U\left(m^c, o\right)\right]$; their upper tail gives R_{inter}. Thus nominal service ranks candidates, whereas the KL-robust utility determines their final risk. For nonempty evaluable $\mathcal{C}$, the mixture weights $(\omega_a, \omega_i) = softmax\left(\left(R_{attr}, R_{inter}\right)/T_r\right)$ at temperature T_r define $R = \omega_a R_{attr} + \omega_i R_{inter}$; otherwise $R = R_{attr}$. Here $\mu_{\min}$ is the minimum membership fraction, B the beam width, and T_r the mixture temperature.

Adversarial two-player game and discrete refinement. For budget K, the relaxed feasible region is $\mathcal{D}_K = \{x \in [0, 1]^N : \sum_v x_v = K\}$. The auditor and seed selector have opposing incentives with respect to the risk R:

$$
\begin{aligned}
F(\theta; x, K) &= R\left(o(x, K); \theta\right) - P(\theta), && \text{auditor ascent,}\\
J(x; \theta, K) &= I(x) - \lambda_{fair} R\left(o(x, K); \theta\right), && x \in \mathcal{D}_K, \quad \text{seed ascent.}
\end{aligned}
\tag{14}
$$

Here θ comprises the auditor network and gate parameters, $P(\theta)$ is the structural and capacity penalty defined above, and λ_{fair} weighs the learned latent-group risk. The auditor searches for high-risk memberships while paying P; the seed player increases influence and reduces the risk induced by those memberships. For fixed detached opportunity in one auditor step, two gradients perform extragradient ascent: $\theta' = \theta_t + \eta_a\nabla_\theta F\left(\theta_t; x_t, K\right)$, followed by $\theta_{t+1} = \theta_t + \eta_a\nabla_\theta F\left(\theta'; x_t, K\right)$ (15), with step size η_a. The seed gradient $g_t = \nabla_x J\left(x_t; \theta_{t+1}, K\right)$ holds auditor outputs fixed. If s_t selects the K largest coordinates of g_t, the Frank–Wolfe step is

$$x_{t+1} = (1 - \gamma_t)\, x_t + \gamma_t s_t, \qquad \gamma_t = clip\{2/\left(t + c_{FW}\right), \gamma_{\min}, 1\}. \tag{16}$$

The offset c_{FW} and floor $\gamma_{\min}$ set the schedule. At each game iteration the relaxed RR state is computed once; intersection keys are refreshed at a fixed cadence before the auditor update and, on that cadence, after it. The auditor receives detached current opportunity, and its memberships and gates are detached during the seed gradient. Auditor parameters carry over across increasing budgets, whereas each budget starts from its own feasible x. Let $n_v = \left|\{i : v \in R_i^G\}\right|$ be the global RR occurrence count. The first budget starts at the indicator of the K largest n_v; a later budget projects $y = x^{prev} + \mathbf{1}/N$ onto $\mathcal{D}_K$ by $x_v = clip\,(y_v - \nu, 0, 1)$, using bisection on ν to approximate $\sum_v x_v = K$.

Proposition 2 (Linear seed oracle). *The top-K indicator s_t maximizes $\langle g_t, s\rangle$ over $s \in \mathcal{D}_K$. If $x_t \in \mathcal{D}_K$ and $\gamma_t \in [0,1]$, Eq. (16) remains in $\mathcal{D}_K$ in exact arithmetic.*

The proof is in Appendix B.5.

Discrete rounding and restricted swaps. The final x is rounded to its K largest coordinates. Hard RR influence, target service, opportunity, and intersection keys are recomputed from this set. In the hard risk, only threshold-active slots and selected pairs enter the tails, and their risks have no continuous gate multipliers; the same marginal/intersection mixture applies (Appendix B.4). The hard objective is $J_{hard}(S,K) = I(S) - \lambda_{fair} R_{hard}(S,K)$. To form a bounded one-for-one neighborhood, removal candidates are selected from S by smallest x_v, while nonseed addition candidates are ranked by

$$b_v = x_v + \kappa_d \frac{n_v}{\max_u n_u + \epsilon_d}, \quad (17)$$

where n_v is the global RR occurrence count, κ_d weights it, and ϵ_d is a denominator floor. Within a pass, candidate swaps are scored using the currently cached intersection keys. The best candidate is accepted only if this cached-key objective exceeds the current objective by the swap tolerance. After acceptance the hard objective and intersection keys are recomputed, so the procedure does not guarantee monotone improvement of the newly refreshed objective. Its candidate caps and number of passes are specified in Appendix C.

Algorithm 1: AURA-IM, implemented static branch
Input: G, requested budgets $\mathcal{K}$, diffusion and model settings.
Output: $\{S_K\}_{K\in\mathcal{K}}$.

1	Build diffusion/structural adjacencies, global and root-excluded target RR banks.
2	Greedily compute $A_v(k)$; train the view encoder with $A_v\,(K_{\max})$ and retain Z^p.
3	Initialize the gated auditor and sampled-edge regularizer.
4	**for** increasing $K\in\mathcal{K}$: initialize $x \in \mathcal{D}_K$ from RR counts or project the prior x.
5	**for** each game iteration: compute soft RR influence and opportunity once; refresh active pair keys when scheduled.
6	Adversarial auditor: ascend $F = R - P$ by extragradient at fixed o; refresh keys after the update when scheduled.
7	Seed player: detach auditor outputs and ascend $J = I - \lambda_{fair} R$ by the top-K Frank–Wolfe oracle.
8	**end for**; round top-K, recompute hard RR state, run cached-key restricted swaps, and save S_K.
9	**end for**; return $\{S_K\}$.

4 EVALUATION

Processed datasets and protocol. We use eight processed networks (Table 1); node attribute values are withheld during AURA-IM selection and revealed for evaluation. High-School-2013 aggregates recorded proximity contacts; Od is an OULAD-derived student similarity graph; Ego-Facebook has binary profile indicators; MovieLens is a user preference-similarity graph; DBLP, Pokec, AH, and APS supply one observed group attribute each. Dataset construction, exact group counts, and the distinction between raw and processed graph statistics are documented in Appendix D. Our primary evaluation uses SIR (Kermack et al., 1927) infection probability $\beta = .01$ for the across-budget trade-off plots and component analyses. At higher propagation probabilities, many

methods approach the ceiling of the entropy-based fairness measure, so the lower rate leaves more resolution for comparing both fairness and spread. To examine behavior under stronger diffusion, we additionally report the supplied baseline-comparison tables at $\beta \in \{.03, .05\}$ for $K \in \{20, 40\}$. The protocol values are recorded in Appendix C.

Table 1: Processed graph sizes and observed attributes used only in evaluation. Group counts follow attribute order.

Dataset	N	$\lvert E\rvert$	**Attributes**	**Groups per attribute**
High-School-2013	327	5,818	2	9, 3
Od	2,302	18,484	9	2, 13, 5, 11, 3, 2, 4, 4, 2
Ego-Facebook	4,039	88,234	8	eight binary attributes
MovieLens	6,040	152,225	3	2, 7, 21
DBLP	5,000	27,322	1	15
Pokec	5,000	48,454	1	2
AH	1,997	8,523	1	2
APS	1,281	3,064	1	2

Observed-group measures. For the qth SIR world let $\mathcal{A}_q(S)$ be the final activated set. If Q worlds are evaluated, normalized spread is $\widehat{I}_{SIR}(S) = (QN)^{-1} \sum_{q=1}^{Q} |\mathcal{A}_q(S)|$. For observed attribute a with groups V_{ag}, group coverage is $C_{ag}(S) = Q^{-1} \sum_q |\mathcal{A}_q(S) \cap V_{ag}| \,/\, |V_{ag}|$. With m_a groups and positive total coverage, put $\pi_{ag} = C_{ag} / \sum_h C_{ah}$ and $H_a = -\sum_{g:\pi_{ag}>0} \pi_{ag} \log \pi_{ag} / log m_a$. Macro entropy is $H = A^{-1} \sum_{a=1}^{A} H_a$ over A observed attributes. The trade-off plots also use macro attribute-wise worst-group coverage $W = A^{-1} \sum_a \min_g C_{ag}$ and macro normalized L1 dispersion $D = A^{-1} \sum_a \left[\left(m_a^{-1} \sum_g \left| C_{ag} - \bar{C}_a \right| \right) / \{2 \lfloor m_a/2 \rfloor \lceil m_a/2 \rceil / m_a^2\} \right]$, where $\bar{C}_a = m_a^{-1} \sum_g C_{ag}$. Lower D indicates closer coverages. The standardized evaluator also exports Worst-Group Coverage as $\min_{a,g} C_{ag}$, which differs from plotted W. High entropy alone need not imply high service; these *observed* measures differ from the *latent* opportunity risk optimized in Section 3.

Across-budget behavior. The primary $\beta = .01$ trade-off plots trace seed budgets from $K = 4$ to 40 and show spread against macro entropy, macro worst-group coverage, and macro L1 dispersion. Figure 2 shows Od: as spread rises across budgets, entropy generally increases with local reversals, and worst-group coverage rises. High-School-2013 shows a similar broad increase in entropy, but its dispersion panel does not establish the lowest dispersion for AURA-IM at every operating point. Ego-Facebook's entropy varies across budgets rather than rising steadily. On MovieLens, spread and worst-group coverage increase while entropy declines over the larger budgets and dispersion rises. On DBLP, entropy can be high even when absolute worst-group coverage remains small. Pokec and AH have several methods clustered near the entropy ceiling, limiting that metric's power to separate them. APS displays a clear spread–entropy tension: AURA-IM reaches relatively high spread while several alternatives have higher plotted entropy and worst-group coverage. Appendix F gives every three-panel plot for each dataset. These are separately optimized budget-specific solutions and need not form a monotone fairness trajectory. Their lower propagation probability is chosen to retain visible fairness variation; the table checks two higher-probability operating points.

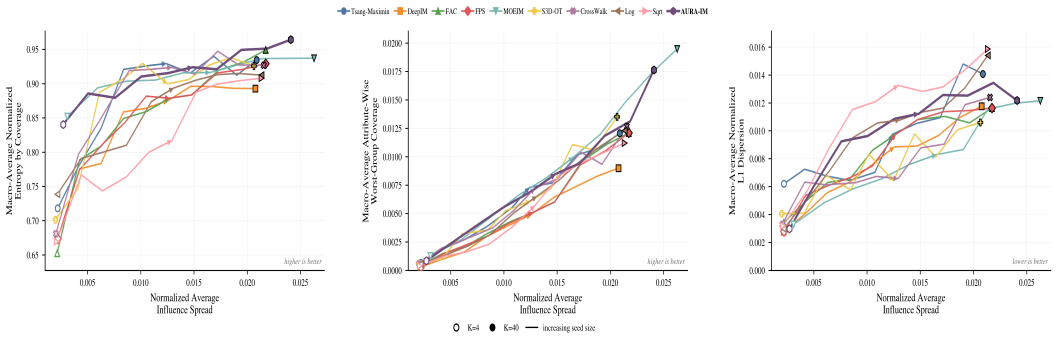


Figure 2: Two panels of the primary Od trade-off plot at $\beta = .01$, across budgets $K = 4, 8, \ldots, 40$: macro entropy (left) and macro attribute-wise worst-group coverage (right) against normalized spread. The full plot, including L1 dispersion, and seven other graphs appear in Appendix F.

Component study and attribute diagnostics. The two primary $\beta = .01$ four-dataset ablation figures in Appendix G compare full AURA-IM with six variants. Removing the broad fairness configuration markedly lowers entropy at small budgets on High-School-2013 and Od; its spread is also lower on Od and MovieLens, so the visual evidence does not show a simple spread gain from removing fairness. The separate component curves are more mixed: the variant without attainability sometimes has higher High-School entropy, whereas the full configuration tends to have higher Od spread. On MovieLens all entropy traces occupy a narrow band near one even while spread separates. The full-method per-attribute heatmap (Appendix H) exposes differences concealed by macro entropy, notably among Od and Ego-Facebook attributes.

5 Conclusion

AURA-IM presents a static, demographic-free framework for fair influence maximization beyond predefined groups. Rather than relying on observed demographic memberships during seed selection, the framework constructs budget-conditioned attainable opportunity from target-specific reverse-reachable samples and learns structural representations to identify potentially underserved populations. A factorized latent auditor then searches for vulnerable marginal and intersectional groups, while KL-robust utility, upper-tail risk aggregation, and structural regularization guide fairness-aware optimization. The resulting latent fairness risk is jointly optimized with RR-based influence spread through an adversarial auditor–seed procedure, followed by Frank–Wolfe optimization, discrete rounding, and restricted swap refinement. Observed group attributes are used only after the seed sets are fixed for post-hoc evaluation. Experiments across heterogeneous real-world networks show that AURA-IM can maintain strong diffusion performance while achieving competitive fairness without using protected-group labels during optimization. In the additional-probability comparisons, AURA-IM achieves higher normalized influence spread than the strongest reported baseline in 28 of 32 configurations, while improvements in macro entropy are less uniform. The across-budget results further demonstrate that entropy, worst-group coverage, and L1 dispersion capture complementary aspects of fairness and do not necessarily vary monotonically with influence spread. Results on APS, Ego-Facebook, and MovieLens particularly illustrate that higher spread does not always imply better observed-group balance. These findings support the feasibility of fairness-aware influence maximization when demographic group definitions are unavailable or intentionally withheld. Nevertheless, AURA-IM does not guarantee fairness for every possible observed partition, nor do the finite-sample propositions establish convergence of the full adversarial game or global optimality. The framework therefore provides a practical foundation for demographic-free fair influence maximization while motivating future work on stronger theoretical guarantees, statistical validation, and broader robustness across graph and diffusion regimes.

AI use statement

Generative AI used for grammar check and proofreading.

Reproducibility statement

The appendix specifies the implemented sampled oracles, encoder and auditor penalties, optimization steps, parameter values, dataset preprocessing, complete tabulated comparisons, and proofs.

References

Ali, J., Babaei, M., Chakraborty, A., Mirzasoleiman, B., Gummadi, K. P., & Singla, A. (2021). "On the fairness of time-critical influence maximization in social networks." IEEE Transactions on Knowledge and Data Engineering, 35(3), 2875-2886.

Anwar, Muhammad Shahzad, Martin Saveski, and Deb Roy. 2021. “Balanced Influence Maximization in the Presence of Homophily.” In Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 175–83. https://doi.org/10.1145/3437963.3441787.

Bearman, Peter S., Jo Jones, and J. Richard Udry. 1997. “The National Longitudinal Study of Adolescent Health: Research Design.” Carolina Population Center, University of North Carolina at Chapel Hill.

Becker, Ruben, Gianlorenzo D’Angelo, Sajjad Ghobadi, and Hugo Gilbert. 2022. “Fairness in Influence Maximization Through Randomization.” Journal of Artificial Intelligence Research 73: 1251–83. https://doi.org/10.1613/jair.1.13367.

Borgs, C., Brautbar, M., Chayes, J., & Lucier, B. (2014, January). Maximizing social influence in nearly optimal time. In Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms (pp. 946-957). Society for Industrial and Applied Mathematics.

Chen, L., and A. Rezaeipanah. 2025. “SFIMCO: Scalable Fair Influence Maximization Based on Overlapping Communities and Optimization Algorithms.” Neurocomputing 629: 129687. https://doi.org/10.1016/j.neucom.2025.129687.

Chowdhary, S., G. De Pasquale, N. Lanzetti, A.-A. Stoica, and F. D
"orfler. 2024. “Fairness in Social Influence Maximization via Optimal Transport.” In Advances in Neural Information Processing Systems 37.

Cunegatti, E., Custode, L., & Iacca, G. (2024, July). Many-objective evolutionary influence maximization: Balancing spread, budget, fairness, and time. In Proceedings of the genetic and evolutionary computation conference companion (pp. 655-658).

Farnadi, Golnoosh, Behrouz Babaki, and Michel Gendreau. 2020. “A Unifying Framework for Fairness-Aware Influence Maximization.” In Companion Proceedings of the Web Conference 2020, 714–22. https://doi.org/10.1145/3366424.3383555.

Feng, Y., Patel, A., Cautis, B., & Vahabi, H. (2023, August). Influence maximization with fairness at scale. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 4046-4055).

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. Naval research logistics quarterly, 3(1-2), 95-110.

Gao, T., and T. Ito. 2026. “Robust Fair Influence Maximization under Multiple Community Partitions.” Proceedings of the ACM on Management of Data 4 (1): Article 78, 1-25. https://doi.org/10.1145/3786692.

Gong, H., and C. Guo. 2023. “Influence Maximization Considering Fairness: A Multi-Objective Optimization Approach with Prior Knowledge.” Expert Systems with Applications 214: 119138. https://doi.org/10.1016/j.eswa.2022.119138.

Guo, P., H. Li, and X. Mo. 2026. “Higher-Order Learning for Fairness: A Reinforcement Learning Framework for Equitable and Efficient Influence Maximization.” Journal of King Saud University - Computer and Information Sciences 38: Article 403. https://doi.org/10.1007/s44443-026-00798-0.

Harper, F. Maxwell, and Joseph A. Konstan. 2015. “The MovieLens Datasets: History and Context.” ACM Transactions on Interactive Intelligent Systems 5 (4): 19. https://doi.org/10.1145/2827872.

Kempe, David, Jon Kleinberg, and
Éva Tardos. 2003. “Maximizing the Spread of Influence Through a Social Network.” In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 137–46. https://doi.org/10.1145/956750.956769.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772), 700-721.

Khajehnejad, Moein, Ahmad Asgharian Rezaei, Mahmoudreza Babaei, Jessica Hoffmann, Mahdi Jalili, and Adrian Weller. 2020. “Adversarial Graph Embeddings for Fair Influence Maximization

over Social Networks." In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 4306–12. https://doi.org/10.24963/ijcai.2020/594.

Khajehnejad, A., Khajehnejad, M., Babaei, M., Gummadi, K. P., Weller, A., & Mirzasoleiman, B. (2022, June). Crosswalk: Fairness-enhanced node representation learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 11, pp. 11963-11970).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Korpelevich, G. M. (1976). The extragradient method for finding saddle points and other problems. Matecon, 12, 747-756.

Kuzilek, Jakub, Martin Hlosta, and Zdenek Zdrahal. 2017. "Open University Learning Analytics Dataset." Scientific Data 4: 170171. https://doi.org/10.1038/sdata.2017.171.

Lee, Eun, 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: 1078–87. https://doi.org/10.1038/s41562-019-0677-4.

Ling, C., Jiang, J., Wang, J., Thai, M. T., Xue, R., Song, J., ... & Zhao, L. (2023, July). Deep graph representation learning and optimization for influence maximization. In International conference on machine learning (pp. 21350-21361). PMLR.

Mastrandrea, Rossana, 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): e0136497. https://doi.org/10.1371/journal.pone.0136497.

Leskovec, J., & Mcauley, J. (2012). "Learning to discover social circles in ego networks." Advances in neural information processing systems, 25.

Oord, A. V. D., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.

Rahmattalabi, Aida, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Mark Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. 2021. "Fair Influence Maximization: A Welfare Optimization Approach." Proceedings of the AAAI Conference on Artificial Intelligence 35 (13): 11630–38. https://doi.org/10.1609/aaai.v35i13.17383.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of risk, 2, 21-42.

Saxena, A., H. K. Yadav, B. Rutten, and S. S. Jha. 2026. "DQ4FairIM: Fairness-Aware Influence Maximization Using Deep Reinforcement Learning." IEEE Transactions on Computational Social Systems. Advance online publication. https://doi.org/10.1109/TCSS.2026.3705520.

Srinivas, A. J., P. Potikas, W. Andreopoulos, and K. Potika. 2026. "Fair Influence Maximization with Reverse Influence Sampling Boosted Multi-Objective Genetic Algorithm." In 2026 IEEE 12th International Conference on Big Data Computing Service and Machine Learning Applications (BigDataService), 138-145. IEEE. https://doi.org/10.1109/BigDataService70481.2026.00027.

Takac, Lubos, and Michal Zabovsky. 2012. "Data Analysis in Public Social Networks." In Present Day Trends of Innovations.

Tsang, Alan, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. 2019. "Group-Fairness in Influence Maximization." In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 5997–6005. https://doi.org/10.24963/ijcai.2019/831.

Yang, Jaewon, and Jure Leskovec. 2012. "Defining and Evaluating Network Communities Based on Ground-Truth." In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, 1–8. https://doi.org/10.1145/2350190.2350193.

Zhang, Y., S. Chen, L. Du, Z. Zhang, and W. Yang. 2026. "Fairness-Aware Influence Maximization with Randomized Strategies: A Stochastic Frank-Wolfe Framework." ACM Transactions on Knowledge Discovery from Data. Advance online publication. https://doi.org/10.1145/3827606.

Zhang, Z., D. Li, Y. Wang, W. Chen, and Y. Zhu. 2025. "Fairness-Constrained Multi-group Influence Maximization." Knowledge and Information Systems 67 (4): 3487-3511. https://doi.org/10.1007/s10115-024-02314-0.

APPENDIX/SUPPLEMENTARY MATERIAL

APPENDIX A ADDITIONAL-PROBABILITY COMPARISON AGAINST SUPPLIED BASELINES

At the two additional probabilities and two selected budgets, the supplied tables compare AURA-IM with Log (Ali et al., 2021), Sqrt (Ali et al., 2021), MOEIM (Cunegatti et al., 2024), S3D-OT (Chowdhary et al., 2024), Tsang (Tsang et al., 2019), DeepIM (Ling et al., 2023), FAC (Feng et al., 2023), FPS (Feng et al., 2023), and CrossWalk (Khajehnejad et al., 2022). Table 2 shows every dataset and configuration alongside the best of those nine baselines in each cell.

Table 2: Additional-probability checks: AURA-IM compared with the best of nine baselines at $\beta = .03, .05$ and $K = 20, 40$. Entries are proposed/best baseline for normalized spread (I) and macro normalized entropy (H).

		$K = 20$		$K = 40$	
Dataset	**Metric**	$p = .03$	$p = .05$	$p = .03$	$p = .05$
High-School-2013	I	**0.392/0.391**	**0.741/0.738**	**0.485/0.483**	**0.774/0.768**
	H	**0.997/0.995**	**0.999/0.998**	**0.997/0.995**	**0.999/0.998**
Od	I	**0.032/0.031**	**0.097/0.091**	**0.055/0.054**	**0.138/0.127**
	H	**0.989/0.983**	**0.994/0.993**	**0.994/0.990**	**0.997/0.996**
Ego	I	**0.377/0.376**	**0.556/0.553**	**0.383/0.382**	0.538/0.561
	H	**0.954/0.951**	**0.980/0.978**	**0.958/0.953**	0.977/0.979
Mov	I	**0.577/0.576**	**0.843/0.842**	0.581/0.583	**0.844/0.843**
	H	0.998/0.999	0.999/0.999	0.999/0.999	0.999/0.999
DBLP	I	**0.041/0.040**	**0.137/0.130**	**0.058/0.051**	**0.150/0.143**
	H	**0.931/0.926**	**0.964/0.963**	**0.960/0.950**	**0.977/0.971**
Pok	I	**0.114/0.112**	**0.334/0.331**	**0.132/0.123**	**0.344/0.338**
	H	0.996/0.998	0.998/0.998	0.996/0.998	0.998/0.998
AH	I	0.019/0.020	**0.033/0.029**	**0.035/0.030**	**0.056/0.055**
	H	**0.998/0.995**	0.996/0.997	**0.997/0.989**	0.999/0.999
APS	I	0.029/0.030	**0.043/0.037**	**0.051/0.047**	**0.069/0.061**
	H	0.933/0.982	0.897/0.980	**0.995/0.988**	0.975/0.999

Spread exceeds the strongest baseline in 28 of 32 configurations. The largest displayed advantage is $+.011$ on Od at $K = 40, p = .05$ ($.138$ versus $.127$); the largest deficit is $-.023$ on Ego-Facebook at the same K, p ($.538$ versus $.561$). The other spread deficits occur on MovieLens at $K = 40, p = .03$, AH at $K = 20, p = .03$, and APS at $K = 20, p = .03$. Macro entropy is higher in 18 configurations, equal in six, and lower in eight. The largest gain is $+.010$ on DBLP at $K = 40, p = .03$ ($.960$ versus $.950$), while APS at $K = 20, p = .05$ has the largest deficit, $-.083$ ($.897$ versus $.980$). In MovieLens, the entropy values cluster near one, creating ties at displayed precision.

High-School-2013, Od, and DBLP each exceed the strongest reported baseline on both displayed measures in all four tabulated configurations. The High-School entropy gains are small ($.001$–$.002$), while Od's spread gain is most visible at the larger propagation probability and budget. DBLP has larger entropy gains at $K = 40$ ($.010$ and $.006$), but entropy measures *balance of coverage*, not its absolute minimum. The spread advantage on Pokec is accompanied by two entropy deficits and two displayed ties. On MovieLens, three of four entropy entries tie at three decimals and the remaining one is below the strongest baseline. APS gains spread in three configurations but loses entropy in three; at $K = 20, p = .05$, its spread is $.043$ versus $.037$ while its entropy is $.897$ versus $.980$. Ego-Facebook at $K = 40, p = .05$ is a joint deficit. The training risk protects learned latent groups, while these observed groups are revealed only after selection.

APPENDIX B SUPPLEMENTARY DETAILS OF THE PROPOSED METHOD

5.1 STRUCTURAL INPUTS AND ENCODER LOSSES

For a sampled structural view $A^{(r)}$, let $d^{(r)} = A^{(r)}\mathbf{1}$, $t^{(r)} = 2d^{(r)}$, and

$$\widetilde{A}^{(r)} = D_r^{-1}\left(A^{(r)} + I_N\right), \qquad (D_r)_{vv} = max\left\{\sum_u \left(A^{(r)} + I_N\right)_{vu}, 1\right\}. \tag{18}$$

For these static views, both of the first two degree coordinates use the same row degree. With $\pi_0^{(r)} = \mathbf{1}/N$, the stationary structural signal is iterated and then normalized as

$$\pi_{j+1}^{(r)} = (1-d_\pi)\,\mathbf{1}/N + d_\pi\left(\widetilde{A}^{(r)}\right)^{\mathsf{T}}\pi_j^{(r)}, \qquad \pi^{(r)} = \frac{\pi_{n_\pi}^{(r)}}{max\{\mathbf{1}^{\mathsf{T}}\pi_{n_\pi}^{(r)}, \epsilon_\pi\}}, \tag{19}$$

where n_π is the iteration count, d_π the retention factor, and ϵ_π a numerical floor. The descriptor and propagated input are

$$\begin{aligned} f_v^{(r)} = \; & [log\left(1+d_v^{(r)}\right), log\left(1+d_v^{(r)}\right), log\left(1+t_v^{(r)}\right), \\ & \left(\widetilde{A}^{(r)}\log\left(1+t^{(r)}\right)\right)_v, log\left(1+N\pi_v^{(r)}\right)], \\ H_{raw}^{(r)} = \; & \left[f^{(r)}, \widetilde{A}^{(r)}f^{(r)}, \left(\widetilde{A}^{(r)}\right)^2 f^{(r)}\right], \qquad H_{vj}^{(r)} = \frac{\left(H_{raw}^{(r)}\right)_{vj} - \mu_j}{\sigma_j + \epsilon_H}. \end{aligned} \tag{20}$$

Thus every $H^{(r)}$ has the feature dimension reported in Table 3. The mean μ_j and population standard deviation σ_j of each raw coordinate j are computed jointly across all nodes and all three views; ϵ_H is the denominator floor. The edge drops are independent draws for the retained oriented entries of the structural adjacency.

To specify Eq. (7), let $\widehat{z} = z/max\left(\| z \|_2, \epsilon_n\right)$ denote row-wise $\updownarrow_2$ normalization with numerical floor ϵ_n, and let MSE average squared differences over all entries. The embedding dimension is d_z, and Var_v below is sample variance across nodes. The view and context terms are

$$L_{view} = MSE\left(\widehat{Z}^{(1)}, \widehat{Z}^{(2)}\right), \qquad L_{ctx} = MSE\left(d_c\left(Z^c\right), H^{(c)} - \tfrac{H^{(1)}+H^{(2)}}{2}\right), \tag{21}$$

where d_c is the context decoder. Given a uniformly sampled subset $\mathcal{B} \subseteq V$ capped at $B_{\max}$, put $s_{ij} = \langle \widehat{z}_i^{(1)}, \widehat{z}_j^{(2)} \rangle / T_{NCE}$. Symmetric InfoNCE (Oord et al., 2018) is

$$L_{NCE} = -\tfrac{1}{2|\mathcal{B}|}\sum_{i\in\mathcal{B}}\left[\log\tfrac{e^{s_{ii}}}{\sum_{j\in\mathcal{B}}e^{s_{ij}}} + log\tfrac{e^{s_{ii}}}{\sum_{j\in\mathcal{B}}e^{s_{ji}}}\right]. \tag{22}$$

Let h_o be the opportunity head and $[z]_+ = max(z, 0)$. The remaining terms are

$$\begin{aligned} L_{orth} &= \tfrac{1}{N}\sum_v \langle \widehat{z}_v^p, \widehat{z}_v^c \rangle^2, \\ L_{opp} &= \tfrac{1}{N}\sum_v \left[\sigma\left(h_o\left(z_v^p\right)\right) - A_v\left(K_{\max}\right)\right]^2, \\ L_{var}(Z) &= \tfrac{1}{d_z}\sum_{j=1}^{d_z}\left[s_{\min} - \sqrt{Var_v\left(Z_{vj}\right) + \epsilon_{var}}\right]_+. \end{aligned} \tag{23}$$

The variance in this last expression is the column-wise sample variance. The view and context encoders are separate multilayer perceptrons with a GELU hidden layer. The opportunity head applies a sigmoid after its final linear layer. The encoder is optimized by Adam with the learning schedule and clipping shown in Table 3.

5.2 AUDITOR PENALTY DEFINITIONS

Let $\bar{q}_{ag} = N^{-1}\sum_v q_{vag}$, $w_{ag} = \alpha_a\delta_{ag}$, and $\overline{z} = max\left(z, \epsilon_P\right)$. Let $\mathcal{E}_s$ be the fixed uniform sample of retained input edge rows used for graph smoothness, and let l, u bound allowed mean slot membership. For each axis a, define the centered matrix $Q_a \in \mathbb{R}^{N\times G_*}$ by $(Q_a)_{vg} = q_{vag} - \bar{q}_{ag}$. The size, smoothness, and cross-axis independence penalties are

$$
\begin{aligned}
P_{size} &= \frac{\sum_{ag} w_{ag}\{[l-\bar{q}_{ag}]_+^2+[\bar{q}_{ag}-u]_+^2\}}{\sum_{ag} w_{ag}},\\
P_{smooth} &= \frac{\sum_{(v,v')\in\mathcal{E}_s}\sum_{ag} w_{ag}\left(q_{vag}-q_{v'ag}\right)^2}{|\mathcal{E}_s|\sum_{ag} w_{ag}},\\
P_{ind} &= \frac{\sum_{a<b}\alpha_a\alpha_b\, mean\left[\left(Q_a^{\mathsf{T}}Q_b/N\right)^{\odot 2}\right]}{\sum_{a<b}\alpha_a\alpha_b}.
\end{aligned}
\tag{24}
$$

Here $\odot 2$ squares entries and $mean$ averages the resulting matrix. When the retained graph has no edges, P_{smooth} is set to zero.

The entropy, complexity, gate-binarization, and capacity terms are

$$
\begin{aligned}
P_{ent} &= \frac{\sum_a \alpha_a\left[-N^{-1}\sum_{vg} q_{vag}\log(q_{vag}+\epsilon_P)\right]}{\sum_a \alpha_a},\\
P_{axis} &= A_*^{-1}\textstyle\sum_a \alpha_a, \qquad P_{group} = (A_*G_*)^{-1}\sum_{ag}\alpha_a\delta_{ag},\\
P_{binary} &= A_*^{-1}\textstyle\sum_a \alpha_a\,(1-\alpha_a) + (A_*G_*)^{-1}\sum_{ag}\delta_{ag}\,(1-\delta_{ag}),\\
P_{capacity} &= [n_a-\textstyle\sum_a \alpha_a]_+^2 + \frac{\sum_a \alpha_a\left[n_g-\sum_g \delta_{ag}\right]_+^2}{\sum_a \alpha_a}.
\end{aligned}
\tag{25}
$$

The capacity targets n_a, n_g also define the fallback for threshold selection. With the penalty coefficients in Table 4, the auditor objective uses

$$
\begin{aligned}
P(\theta) = \quad & \lambda_s P_{size} + \lambda_e P_{smooth} + \lambda_i P_{ind} + \lambda_h P_{ent}\\
& + \lambda_a P_{axis} + \lambda_g P_{group} + \lambda_b P_{binary} + \lambda_c P_{capacity}.
\end{aligned}
\tag{26}
$$

The axis logits start at evenly spaced values between their stated endpoints. Group logits start at evenly spaced values between their endpoints on every axis plus independent Gaussian perturbations with the reported standard deviation. Their parameters and the auditor network are optimized together.

5.3 Sparse coverage, active pairs, and update timing

The proposed method stores global incidence $\mathsf{G}_{iu} = \mathbf{1}\{u \in R_i^G\}$ and root-excluded target incidence $\mathsf{T}_{(v,j),u} = \mathbf{1}\{u \in \bar{R}_{vj}^T\}$ as sparse matrices. If $b_u(x) = logclip\,(1-x_u, \epsilon_x, 1)$, then the rows of $\mathsf{G}b(x)$ and $\mathsf{T}b(x)$ are the exponents in Eq. (5); this gives all global and target relaxed covers without separately looping over RR sets.

For an active axis a and slot g, let $m^{(a,g)} = q_{\cdot ag}$, $M(m) = \sum_v m_v$, and $\bar{o}(m) = \sum_v m_v o_v / \left(M(m) + \epsilon_{beam}\right)$. The search retains the B active slots with smallest $\bar{o}(m)$ among those satisfying $M(m) \geq \mu_{\min} N$. Each retained slot is multiplied by an active slot from a distinct axis. Duplicate pairs and products with mass below $\mu_{\min} N$ are removed; the B lowest-mean surviving products become the pair keys. The active structure is produced by gate thresholds with minimum-capacity top-gate fallback. Pair discovery uses detached membership and opportunity and is discrete; it is not differentiated through the beam. During the relaxed game every axis and slot still contributes to R_{attr} through its continuous gate, while only retained pair keys contribute to R_{inter}.

At an iteration the proposed method computes $I(x)$ and $o(x, K)$ once, optionally refreshes pair keys, performs auditor extragradient on detached o, optionally refreshes pair keys again, detaches the updated memberships and gates, and finally differentiates J with respect to x. The pre- and post-auditor key refreshes use the same configured cadence; keys are otherwise reused. The next budget reuses the trained auditor parameters, but the relaxed seed vector is warm-started through the capped-simplex projection.

5.4 Discrete risk and implementation boundary

At hard refinement, let $\mathcal{G}_+$ be the threshold-active slots and $\mathcal{C}_+$ the selected pair keys. For a set S, its two component risks are

$$
\begin{aligned}
\widetilde{R}_{attr}(S) &= Tail_\tau\left(\{1-U\left(q_{\cdot ag}, o(S,K)\right) : (a,g)\in\mathcal{G}_+\}\right),\\
\widetilde{R}_{inter}(S) &= Tail_\tau\left(\{1-U\left(m^c, o(S,K)\right) : c\in\mathcal{C}_+\}\right).
\end{aligned}
\tag{27}
$$

The hard risk is the softmax-temperature mixture of these two scores when $\mathcal{C}_+$ is nonempty and equals the marginal score otherwise. The hard KL routine assigns unit utility if its membership vector has negligible total mass; it uses a separate small logarithm floor in the finite-grid calculation. Intersection keys are first obtained from the current hard set. A swap pass considers only seeds with the smallest final relaxed coordinates and nonseeds with the largest rank in Eq. (17), subject to the candidate caps in Table 4. Candidate sets are evaluated against cached keys; a candidate is accepted only if this cached-key score exceeds the current stored objective by the swap tolerance. After acceptance the hard objective is recomputed with refreshed keys. Consequently a pass does not certify monotone improvement of the newly refreshed objective. The resulting seed sets are evaluated separately after selection, using the observed group values at that stage.

5.5 Proofs of properties of the implemented operators

Proof of Proposition 1, sampled coverage. Fix a set $B \subseteq V$ and define $h_B(S) = \mathbf{1}\{S \cap B \neq \varnothing\}$. For $S \subseteq T \subseteq V$ and $u \notin T$, the marginal $h_B(S \cup \{u\}) - h_B(S)$ equals one precisely when $u \in B$ and $S \cap B = \varnothing$. Replacing S by T can only change this marginal from one to zero. Thus h_B is monotone and submodular, with $h_B(\varnothing) = 0$. Equations (1) and (2) are nonnegative averages of such functions and inherit these properties. Every R_{vj}^T excludes v, hence $p_v(\{v\}) = 0$. □

Proof of Proposition 1, greedy frontier. For a fixed target, each source u covers the sample-index mask B_{vu}. Removing duplicate masks does not alter the maximum coverage attainable with at most k sources. Let C_i be the union after i greedy choices and let O be a union of at most k masks attaining $p_v^\star(k) = |O|/R$. By submodularity of sample coverage, the sum of the marginal uncovered gains of those at most k optimal masks relative to C_i is at least $|O \backslash C_i|$, which is at least $|O| - |C_i|$. Therefore, if a positive gain remains, the mask chosen by greedy gains at least $(|O| - |C_i|)/k$ samples. Writing $a_i = |C_i|/R$ gives $p_v^\star(k) - a_{i+1} \leq (1 - 1/k)(p_v^\star(k) - a_i)$, with $a_0 = 0$. Iterating to $i = k$ proves the frontier bound in Proposition 1. If no positive gain remains, all target samples reachable by any source are already covered, so the inequality still holds. Greedy adds a mask only for positive gain; hence $A_v(k)$ is nondecreasing. □

Proof of Proposition 2. Order the coordinates so that $g_{t,1} \geq \cdots \geq g_{t,N}$, breaking ties arbitrarily. For any $s \in \mathcal{D}_K$, if some $i \leq K$ has $s_i < 1$ and some $j > K$ has $s_j > 0$, transferring $\delta = min\{1 - s_i, s_j\}$ from j to i preserves the box constraints and coordinate sum. The linear objective changes by $\delta(g_{t,i} - g_{t,j}) \geq 0$. Repeated transfers produce the top-K indicator, so it attains the maximum in Proposition 2. Finally, $\mathcal{D}_K$ is convex; a convex combination of $x_t, s_t \in \mathcal{D}_K$ with $\gamma_t \in [0, 1]$ remains in $\mathcal{D}_K$. The bisection projection used for a later budget approximates its target sum to numerical precision. □

Appendix C Parameter Settings

Tables 3 and 4 record the values used for proposed method. The latent capacities and thresholds refer to the auditor's learned groups. Symbolic notation is defined at the point of use in the main method and Appendix B.1.

Table 3: Graph, RR, and representation settings.

Quantity	**Value**
Quantity	Value
budgets $\mathcal{K}$	$4, 8, \ldots, 40$
Primary-plot infection probability β; recovery probability γ	0.01; 1
RR transmissibility denominator floor ϵ_T	10^{-12}
Global RR bank size M; target samples per root R	8000; 8
Persistent-view edge-drop rates r_1, r_2	$0.05, 0.10$
Context-view edge-drop rate r_c	0.25
Stationary-signal iterations n_π; retention d_π	15; 0.85
Stationary-signal restart $1 - d_\pi$; mass floor ϵ_π	0.15; 10^{-12}
Joint feature-standardization floor ϵ_H	10^{-6}
Embedding normalization floor ϵ_n (framework default)	10^{-12}

Quantity	Value
Structural input dimension; encoder hidden width h_e; embedding dimension d_z	15; 64; 32
Context decoder widths	$32 \to 64 \to 15$
Opportunity head widths	$32 \to 32 \to 1$
Encoder epochs; Adam learning rate	80; 0.002
Encoder weight decay; gradient norm clip	10^{-5}; 5
Contrastive sample cap $B_{\max}$; temperature T_{NCE}	1024; 0.20
View and NCE weights $\lambda_{view}, \lambda_{NCE}$	1; 0.8
Context and orthogonality weights $\lambda_{ctx}, \lambda_{orth}$	0.5; 0.10
Opportunity and variance weights $\lambda_{opp}, \lambda_{var}$	1.2; 0.08
Variance target $s_{\min}$; variance stabilizer ϵ_{var}	0.5; 10^{-4}
Relaxed RR logarithm floor ϵ_x; opportunity denominator floor ϵ_o	10^{-6}; 10^{-8}

Table 4: Auditor, risk, game, refinement, and evaluation settings.

Quantity	Value
Quantity	Value
Latent axes A_*; slots per axis G_*; auditor hidden width h_a	6; 8; 64
Gate temperature T_g; log-gate floor ϵ_g	0.75; 10^{-6}
Active axis threshold t_a; active group threshold t_g	0.45; 0.40
Fallback minimum active axes n_a; groups per active axis n_g	1; 2
Axis-logit initialization endpoints	$1, -0.25$
Group-logit initialization endpoints; Gaussian perturbation standard deviation	$0.9, -0.35$; 0.05
Mean-membership lower and upper bounds l, u	0.06; 0.80
Fixed edge-row sample cap; its seed offset from the main seed	5000; 999
Size and smoothness penalty weights λ_s, λ_e	8; 0.30
Independence and entropy penalty weights λ_i, λ_h	0.30; 0.03
Axis and group complexity weights λ_a, λ_g	0.020; 0.010
Gate-binary and minimum-capacity weights λ_b, λ_c	0.020; 2
KL radius ρ; geometric dual-grid interval $\mathcal{H}$	0.05; $\left[10^{-3}, 20\right]$
Number of grid temperatures; differentiable log-membership floor ϵ_{KL}	12; 10^{-30}
Differentiable membership-mass floor; hard zero-mass threshold	10^{-12}; 10^{-15}
Hard-risk log-membership floor	10^{-300}
Empirical upper-tail level τ; mixture temperature T_r	0.90; 0.08
Largest intersection order; beam width B	2; 6
Minimum intersection membership fraction $\mu_{\min}$	0.015
Intersection beam nominal-mean denominator floor ϵ_{beam}	10^{-12}
Fairness coefficient λ_{fair}; game iterations T_{game}	3; 16
Extragradient steps per game iteration; ascent rate η_a	1; 0.001
Intersection refresh interval	4
Frank–Wolfe offset c_{FW}; step floor $\gamma_{\min}$	6; 0.01
Projection bisection iterations; hard swap passes	80; 1
Removal candidate cap; addition candidate cap	3; 16
Addition rank's RR-degree weight κ_d; denominator floor ϵ_d	0.05; 10^{-9}
Swap acceptance tolerance ϵ_{swap}; penalty denominator floor ϵ_P	10^{-8}; 10^{-8}
Additional tabulated infection probabilities (reported results; separate runs)	$0.03, 0.05$

Table 5: Additional experimental settings.

Quantity	Code value
Ablation repeats; encoder epochs per repeat	10; 80
Structural-only auditor pretraining steps; learning rate	80; 0.001
Ablation budget grid; game iterations per budget	$4, 8, \ldots, 40$; 16
Sensitivity budget	20
Sensitivity grid for fairness weight	$.5, 1.5, 3, 4.5, 6$
Sensitivity grid for tail level	.70, .80, .90, .95, .99
Sensitivity grid for KL radius	0, .02, .05, .10, .20
Sensitivity grid for gate temperature	.40, .55, .75, 1, 1.25

APPENDIX D PROCESSED DATASETS AND GROUP DISTRIBUTIONS

High-School-2013 aggregates proximity contacts into a static graph; class has nine values and processed gender has three (Mastrandrea, Fournet, and Barrat 2015). Od starts from an Open University Learning Analytics Dataset cohort; log-scaled virtual-learning-environment clicks form a cosine-similarity graph, with at most ten neighbors per student above similarity .05, retaining the largest connected component (Kuzilek, Hlosta, and Zdrahal 2017). Ego-Facebook uses eight binary profile indicators (McAuley and Leskovec 2012). MovieLens constructs a user similarity network from binary positive ratings (ratings at least four), retains at most thirty neighbors above cosine similarity .05, and evaluates gender, age, and occupation (Harper and Konstan 2015). DBLP uses a single processed 15-way community label rather than the full overlapping raw community set (Yang and Leskovec 2012). Pokec uses one gender-derived binary field (Takac and Zabovsky 2012). AH uses an Add Health school friendship graph and binary white/non-white race partition (Bearman, Jones, and Udry 1997; Chowdhary et al. 2024); APS uses two research-topic categories on the largest component of a citation graph (Lee et al. 2019).

Table 6: Exact processed group sizes and percentages. A_i indexes attributes in the order stated below.

Dataset	Attribute	Groups	Sizes (percentage)
Dataset	Attribute	Groups	Sizes (percentage)
High-School-2013	A1	9	44 (13.46%), 40 (12.23%), 39 (11.93%), 38 (11.62%), 36 (11.01%), 34 (10.40%), 34 (10.40%), 33 (10.09%), 29 (8.87%)
High-School-2013	A2	3	175 (53.52%), 145 (44.34%), 7 (2.14%)
Od	A1	2	1,718 (74.63%), 584 (25.37%)
Od	A2	13	336 (14.60%), 255 (11.08%), 201 (8.73%), 194 (8.43%), 185 (8.04%), 183 (7.95%), 162 (7.04%), 159 (6.91%), 158 (6.86%), 146 (6.34%), 121 (5.26%), 116 (5.04%), 86 (3.74%)
Od	A3	5	1,081 (46.96%), 638 (27.72%), 500 (21.72%), 63 (2.74%), 20 (0.87%)
Od	A4	11	239 (10.38%), 233 (10.12%), 226 (9.82%), 226 (9.82%), 221 (9.60%), 220 (9.56%), 217 (9.43%), 199 (8.64%), 199 (8.64%), 182 (7.91%), 140 (6.08%)
Od	A5	3	1,583 (68.77%), 685 (29.76%), 34 (1.48%)
Od	A6	2	2,116 (91.92%), 186 (8.08%)
Od	A7	4	890 (38.66%), 708 (30.76%), 398 (17.29%), 306 (13.29%)
Od	A8	4	603 (26.19%), 597 (25.93%), 575 (24.98%), 527 (22.89%)
Od	A9	2	2,087 (90.66%), 215 (9.34%)
Ego-Facebook	A1	2	3,217 (79.65%), 822 (20.35%)
Ego-Facebook	A2	2	3,292 (81.51%), 747 (18.49%)

Dataset	Attribute	Groups	Sizes (percentage)
Ego-Facebook	A3	2	3,301 (81.73%), 738 (18.27%)
Ego-Facebook	A4	2	3,302 (81.75%), 737 (18.25%)
Ego-Facebook	A5	2	3,363 (83.26%), 676 (16.74%)
Ego-Facebook	A6	2	3,420 (84.67%), 619 (15.33%)
Ego-Facebook	A7	2	3,448 (85.37%), 591 (14.63%)
Ego-Facebook	A8	2	3,453 (85.49%), 586 (14.51%)
MovieLens	A1	2	4,331 (71.71%), 1,709 (28.29%)
MovieLens	A2	7	2,096 (34.70%), 1,193 (19.75%), 1,103 (18.26%), 550 (9.11%), 496 (8.21%), 380 (6.29%), 222 (3.68%)
MovieLens	A3	21	759 (12.57%), 711 (11.77%), 679 (11.24%), 528 (8.74%), 502 (8.31%), 388 (6.42%), 302 (5.00%), 281 (4.65%), 267 (4.42%), 241 (3.99%), 236 (3.91%), 195 (3.23%), 173 (2.86%), 144 (2.38%), 142 (2.35%), 129 (2.14%), 112 (1.85%), 92 (1.52%), 72 (1.19%), 70 (1.16%), 17 (0.28%)
DBLP	A1	15	3,804 (76.08%), 333 (6.66%), 160 (3.20%), 136 (2.72%), 120 (2.40%), 83 (1.66%), 66 (1.32%), 63 (1.26%), 53 (1.06%), 48 (0.96%), 33 (0.66%), 31 (0.62%), 27 (0.54%), 24 (0.48%), 19 (0.38%)
Pokec	A1	2	2,733 (54.66%), 2,267 (45.34%)
AH	A1	2	1,306 (65.40%), 691 (34.60%)
APS	A1	2	874 (68.23%), 407 (31.77%)

APPENDIX E COMPLETE SUPPLIED COMPARISON TABLES

These are the additional $\beta = .03, .05$ comparisons at $K = 20, 40$, distinct from the primary $\beta = .01$ plots.

Table 7: Normalized average influence spread: additional β=0.03,0.05 comparisons at K=20,40. Greedy and multi-objective methods.

Dataset	K, β	Log	Sqrt	MOEIM	S3D-OT	Tsang
High-School-2013	20, 0.03	0.391	0.391	0.390	0.381	0.389
High-School-2013	20, 0.05	0.738	0.738	0.713	0.732	0.720
High-School-2013	40, 0.03	0.482	0.483	0.470	0.483	0.470
High-School-2013	40, 0.05	0.768	0.768	0.729	0.764	0.747
Od	20, 0.03	0.025	0.025	0.031	0.016	0.018
Od	20, 0.05	0.085	0.083	0.091	0.050	0.058
Od	40, 0.03	0.044	0.045	0.054	0.034	0.035
Od	40, 0.05	0.124	0.123	0.120	0.087	0.127
Ego	20, 0.03	0.376	0.376	0.363	0.355	0.368
Ego	20, 0.05	0.553	0.553	0.534	0.527	0.534
Ego	40, 0.03	0.381	0.382	0.363	0.374	0.364
Ego	40, 0.05	**0.561**	0.560	0.547	0.554	0.550
Mov	20, 0.03	0.570	0.570	0.576	0.575	0.575
Mov	20, 0.05	0.841	0.841	0.842	0.841	0.842
Mov	40, 0.03	**0.583**	**0.583**	0.576	0.582	0.581
Mov	40, 0.05	0.841	0.841	0.842	0.841	0.842
DBLP	20, 0.03	0.032	0.031	0.040	0.013	0.013
DBLP	20, 0.05	0.128	0.127	0.126	0.117	0.113

Dataset	K, β	Log	Sqrt	MOEIM	S3D-OT	Tsang
DBLP	40, 0.03	0.046	0.041	0.051	0.016	0.029
DBLP	40, 0.05	0.137	0.136	0.143	0.132	0.134
Pok	20, 0.03	0.109	0.109	0.112	0.100	0.100
Pok	20, 0.05	0.331	0.331	0.316	0.320	0.315
Pok	40, 0.03	0.123	0.123	0.121	0.110	0.116
Pok	40, 0.05	0.338	0.338	0.321	0.329	0.323
AH	20, 0.03	0.014	0.014	**0.020**	0.014	0.013
AH	20, 0.05	0.020	0.019	0.029	0.020	0.019
AH	40, 0.03	0.026	0.027	0.030	0.026	0.028
AH	40, 0.05	0.036	0.036	0.055	0.040	0.049
APS	20, 0.03	0.021	0.021	**0.030**	0.019	0.018
APS	20, 0.05	0.033	0.032	0.037	0.019	0.023
APS	40, 0.03	0.040	0.040	0.047	0.036	0.045
APS	40, 0.05	0.055	0.054	0.060	0.046	0.061

Table 8: Normalized average influence spread: additional β=0.03,0.05 comparisons at K=20,40. Deep methods and AURA-IM.

Dataset	K, β	DeepIM	FAC	FPS	CrossWalk	AURA-IM
High-School-2013	20, 0.03	0.383	0.388	0.388	0.353	**0.392**
High-School-2013	20, 0.05	0.722	0.724	0.723	0.736	**0.741**
High-School-2013	40, 0.03	0.477	0.471	0.471	0.476	**0.485**
High-School-2013	40, 0.05	0.743	0.744	0.744	0.765	**0.774**
Od	20, 0.03	0.018	0.019	0.020	0.018	**0.032**
Od	20, 0.05	0.054	0.049	0.049	0.059	**0.097**
Od	40, 0.03	0.035	0.038	0.038	0.036	**0.055**
Od	40, 0.05	0.093	0.088	0.088	0.098	**0.138**
Ego	20, 0.03	0.351	0.358	0.357	0.351	**0.377**
Ego	20, 0.05	0.521	0.498	0.524	0.523	**0.556**
Ego	40, 0.03	0.361	0.358	0.358	0.368	**0.383**
Ego	40, 0.05	0.537	0.534	0.534	0.537	0.538
Mov	20, 0.03	0.576	0.573	0.573	0.576	**0.577**
Mov	20, 0.05	0.841	0.842	0.842	0.841	**0.843**
Mov	40, 0.03	0.581	0.579	0.579	0.582	0.581
Mov	40, 0.05	0.843	0.843	0.843	0.840	**0.844**
DBLP	20, 0.03	0.008	0.031	0.030	0.011	**0.041**
DBLP	20, 0.05	0.101	0.124	0.130	0.067	**0.137**
DBLP	40, 0.03	0.021	0.048	0.044	0.018	**0.058**
DBLP	40, 0.05	0.130	0.132	0.138	0.122	**0.150**
Pok	20, 0.03	0.093	0.105	0.106	0.030	**0.114**
Pok	20, 0.05	0.317	0.312	0.316	0.263	**0.334**
Pok	40, 0.03	0.112	0.119	0.119	0.060	**0.132**
Pok	40, 0.05	0.329	0.321	0.322	0.318	**0.344**
AH	20, 0.03	0.014	0.015	0.015	0.011	0.019
AH	20, 0.05	0.019	0.023	0.023	0.014	**0.033**
AH	40, 0.03	0.027	0.030	0.030	0.024	**0.035**
AH	40, 0.05	0.037	0.043	0.043	0.031	**0.056**
APS	20, 0.03	0.023	0.018	0.018	0.017	0.029
APS	20, 0.05	0.032	0.020	0.020	0.022	**0.043**
APS	40, 0.03	0.044	0.036	0.036	0.036	**0.051**
APS	40, 0.05	0.057	0.038	0.040	0.044	**0.069**

Table 9: Macro normalized entropy by coverage: additional β=0.03,0.05 comparisons at K=20,40. Greedy and multiobjective methods.

Dataset	K, β	Log	Sqrt	MOEIM	S3D-OT	Tsang
High-School-2013	20, 0.03	0.991	0.991	0.993	0.994	0.990
High-School-2013	20, 0.05	0.993	0.995	0.998	0.998	0.998
High-School-2013	40, 0.03	0.990	0.991	0.994	0.991	0.991
High-School-2013	40, 0.05	0.990	0.990	0.998	0.991	0.992
Od	20, 0.03	0.983	0.982	0.978	0.954	0.956
Od	20, 0.05	0.993	0.991	0.991	0.993	0.993
Od	40, 0.03	0.985	0.982	0.990	0.972	0.971
Od	40, 0.05	0.994	0.994	0.991	0.996	0.993
Ego	20, 0.03	0.951	0.951	0.946	0.940	0.949
Ego	20, 0.05	0.978	0.978	0.974	0.972	0.974
Ego	40, 0.03	0.953	0.951	0.946	0.950	0.946
Ego	40, 0.05	**0.979**	**0.979**	0.977	**0.979**	0.978
Mov	20, 0.03	**0.999**	**0.999**	0.998	**0.999**	**0.999**
Mov	20, 0.05	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
Mov	40, 0.03	**0.999**	**0.999**	0.998	**0.999**	**0.999**
Mov	40, 0.05	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
DBLP	20, 0.03	0.890	0.872	0.904	0.913	0.694
DBLP	20, 0.05	0.963	0.958	0.932	0.928	0.935
DBLP	40, 0.03	0.950	0.892	0.947	0.910	0.839
DBLP	40, 0.05	0.971	0.958	0.959	0.941	0.953
Pok	20, 0.03	0.996	0.996	0.997	0.997	0.997
Pok	20, 0.05	**0.998**	**0.998**	**0.998**	**0.998**	**0.998**
Pok	40, 0.03	0.996	0.996	0.997	0.995	0.996
Pok	40, 0.05	**0.998**	**0.998**	**0.998**	**0.998**	**0.998**
AH	20, 0.03	0.964	0.939	0.990	0.989	0.995
AH	20, 0.05	0.982	0.970	0.995	0.992	**0.997**
AH	40, 0.03	0.971	0.951	0.989	0.979	0.927
AH	40, 0.05	0.984	0.974	0.997	**0.999**	**0.999**
APS	20, 0.03	0.926	0.784	0.782	0.979	0.889
APS	20, 0.05	**0.980**	0.928	0.757	0.970	0.919
APS	40, 0.03	0.918	0.854	0.832	0.981	0.988
APS	40, 0.05	0.963	0.907	0.788	0.992	0.938

Table 10: Macro normalized entropy by coverage: additional β=0.03,0.05 comparisons at K=20,40. Deep methods and AURA-IM.

Dataset	K, β	DeepIM	FAC	FPS	CrossWalk	AURA-IM
High-School-2013	20, 0.03	0.994	0.995	0.995	0.995	**0.997**
High-School-2013	20, 0.05	0.998	0.998	0.998	0.998	**0.999**
High-School-2013	40, 0.03	0.990	0.995	0.995	0.995	**0.997**
High-School-2013	40, 0.05	0.998	0.998	0.998	0.998	**0.999**
Od	20, 0.03	0.958	0.961	0.964	0.983	**0.989**
Od	20, 0.05	0.993	0.982	0.982	0.987	**0.994**
Od	40, 0.03	0.984	0.981	0.980	0.975	**0.994**

Dataset	K, β	DeepIM	FAC	FPS	CrossWalk	AURA-IM
Od	40, 0.05	0.994	0.991	0.990	0.991	**0.997**
Ego	20, 0.03	0.936	0.942	0.942	0.936	**0.954**
Ego	20, 0.05	0.969	0.961	0.970	0.970	**0.980**
Ego	40, 0.03	0.942	0.943	0.942	0.945	**0.958**
Ego	40, 0.05	0.974	0.974	0.974	0.973	0.977
Mov	20, 0.03	0.998	0.998	0.998	**0.999**	0.998
Mov	20, 0.05	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
Mov	40, 0.03	0.998	**0.999**	**0.999**	**0.999**	**0.999**
Mov	40, 0.05	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
DBLP	20, 0.03	0.790	0.926	0.889	0.880	**0.931**
DBLP	20, 0.05	0.927	0.955	0.953	0.940	**0.964**
DBLP	40, 0.03	0.877	0.921	0.895	0.868	**0.960**
DBLP	40, 0.05	0.942	0.954	0.954	0.961	**0.977**
Pok	20, 0.03	**0.998**	0.997	0.996	0.994	0.996
Pok	20, 0.05	**0.998**	**0.998**	**0.998**	**0.998**	**0.998**
Pok	40, 0.03	**0.998**	0.996	0.996	0.996	0.996
Pok	40, 0.05	**0.998**	**0.998**	**0.998**	**0.998**	**0.998**
AH	20, 0.03	0.981	0.979	0.936	0.917	**0.998**
AH	20, 0.05	0.990	0.994	0.993	0.988	0.996
AH	40, 0.03	0.940	0.987	0.987	0.977	**0.997**
AH	40, 0.05	**0.999**	**0.999**	**0.999**	0.973	**0.999**
APS	20, 0.03	0.950	0.937	0.937	**0.982**	0.933
APS	20, 0.05	0.918	0.849	0.849	0.976	0.897
APS	40, 0.03	0.954	0.983	0.988	0.975	**0.995**
APS	40, 0.05	0.931	0.998	**0.999**	**0.999**	0.975

APPENDIX F ACROSS-BUDGET TRADE-OFFS

The eight primary $\beta = .01$ plots display normalized spread horizontally and macro observed-group entropy, macro attribute-wise worst-group coverage, and macro normalized L1 dispersion vertically, respectively. The additional $\beta = .03, .05$ results at two budgets test stronger diffusion separately.

High-School-2013 and Od. On High-School-2013, AURA-IM moves to high spread while keeping macro entropy high as the budget rises. Entropy has local dips, and several alternative methods show comparable or lower L1 dispersion in parts of the plotted range. On Od, the AURA-IM entropy and macro worst-group coverage traces generally rise with spread, with smaller budget-to-budget reversals.

Ego-Facebook and MovieLens. The Ego-Facebook AURA-IM trace gains spread but exhibits nonmonotone entropy and dispersion, so neighboring budgets can alter observed balance even as spread grows. On MovieLens, entropy starts near its ceiling for the displayed methods; for AURA-IM it declines at larger budgets while macro worst-group coverage increases and L1 dispersion rises. These metrics answer different questions: increasing the least covered group does not require the normalized coverages to become more uniform.

DBLP and Pokec. On DBLP, AURA-IM combines comparatively high spread and entropy in the plotted range, yet the worst-group coverage remains small in absolute units. Thus the entropy comparison alone should not be read as adequate protection of every group. On Pokec, entropy traces for several methods lie close to one; the worst-group and dispersion panels offer more resolution, but the relative ordering changes with budget. A visually higher spread is not a proof of a fairness improvement.

AH and APS. On AH, AURA-IM reaches high spread and a high-entropy region, with macro worst-group coverage increasing as budget grows. APS gives the strongest counterpoint to a uniform fairness claim: AURA-IM reaches relatively high spread but several alternatives exhibit higher entropy and worst-group coverage at plotted operating points.

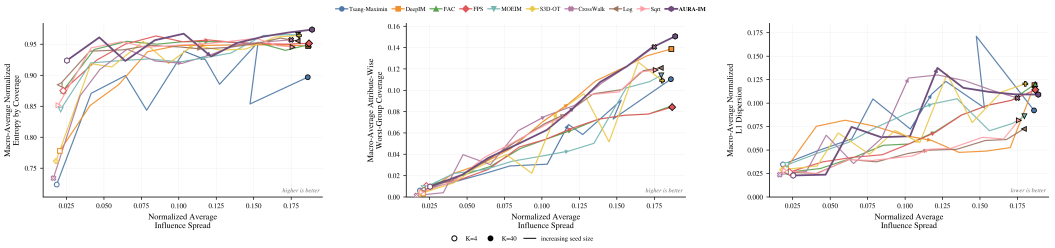


Figure 3: High-School-2013: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

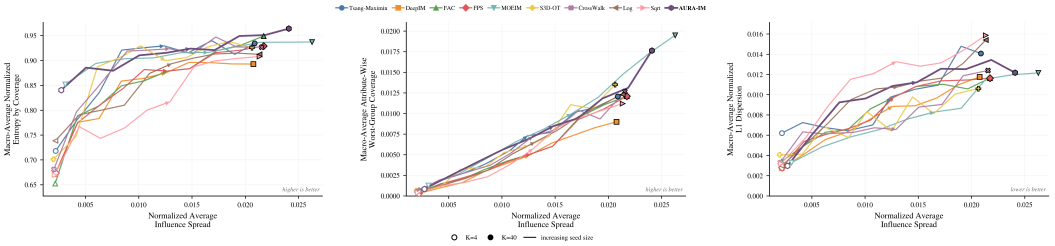


Figure 4: Od: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

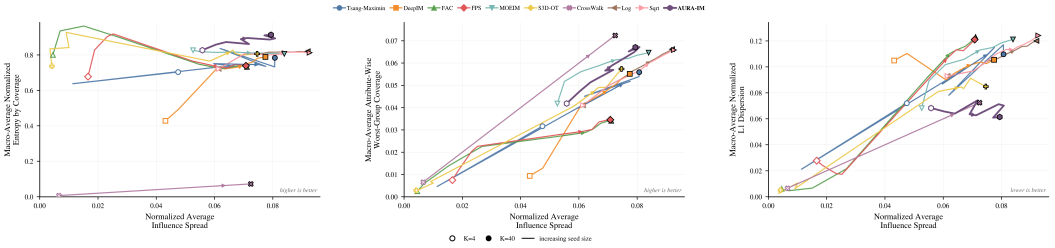


Figure 5: Ego-Facebook: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

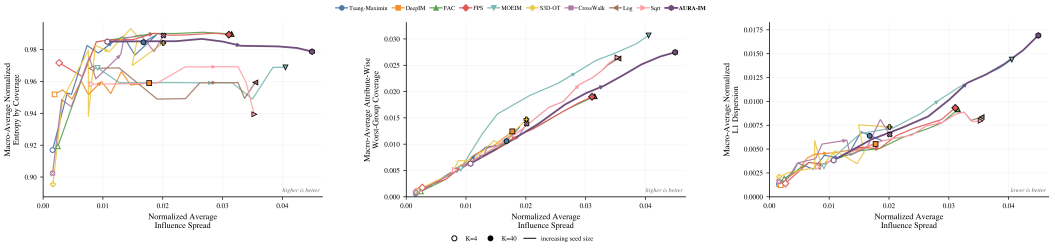


Figure 6: MovieLens: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

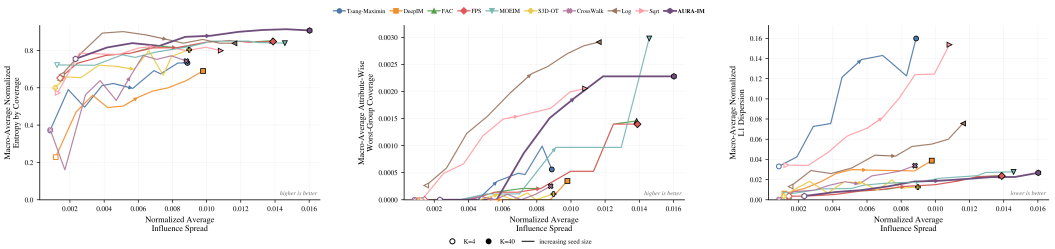


Figure 7: DBLP: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

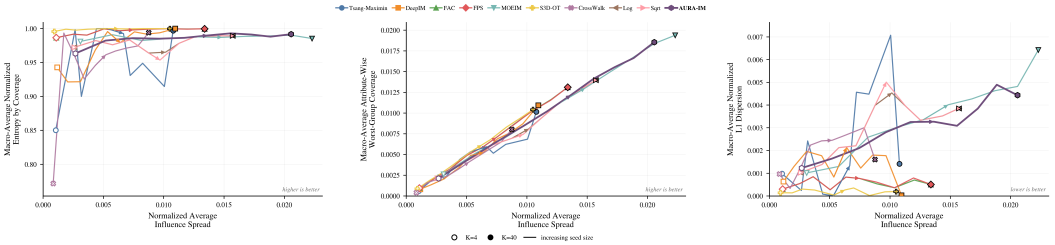


Figure 8: Pokec: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

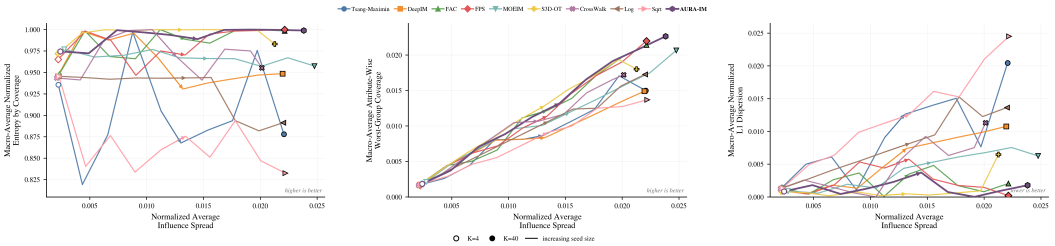


Figure 9: AH: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

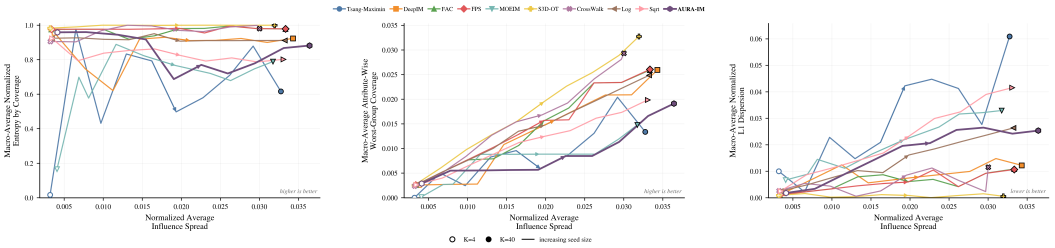


Figure 10: APS: primary $\beta = .01$ across-budget normalized spread versus macro entropy, macro attribute-wise worst-group coverage, and macro normalized L1 coverage dispersion, from left to right. Each point corresponds to one seed budget.

APPENDIX G COMPONENT ABLATIONS

The two primary $\beta = .01$ four-panel plots compare the full method with six variants on High-School-2013, Od, Ego-Facebook, and MovieLens. No-Fairness sets the fairness coefficient to zero, disables the adversary, KL robustness, upper-tail risk, pairwise intersections, and attainable-opportunity normalization, and retrains the representation without opportunity supervision. No-Adversarial Auditor trains a structural-penalty-only auditor without worst-served-group discovery. No-Attainability omits opportunity supervision and replaces the selector's service/frontier ratio with raw sampled target service. No-Robust-Tail replaces KL robustness and upper-tail aggregation with nominal mean risk. No-Intersectional Modeling removes pair search and pair risk. No-Fairness-Aware Representation omits opportunity supervision from the encoder while retaining attainability normalization during selection.

In the entropy plot, No-Fairness is markedly below Full at small budgets on High-School-2013 and Od, and generally below Full on Ego-Facebook. Its MovieLens entropy fluctuates in a narrow near-ceiling range, illustrating why the small vertical differences there should not be overread. No-Attainability exceeds Full at several High-School budgets and briefly on Ego-Facebook; No-Robust-

Tail also exceeds Full at some Od budgets. In the spread plot, Full is visibly above No-Fairness on Od and MovieLens and close to most component variants on High-School-2013; Ego-Facebook traces cross.

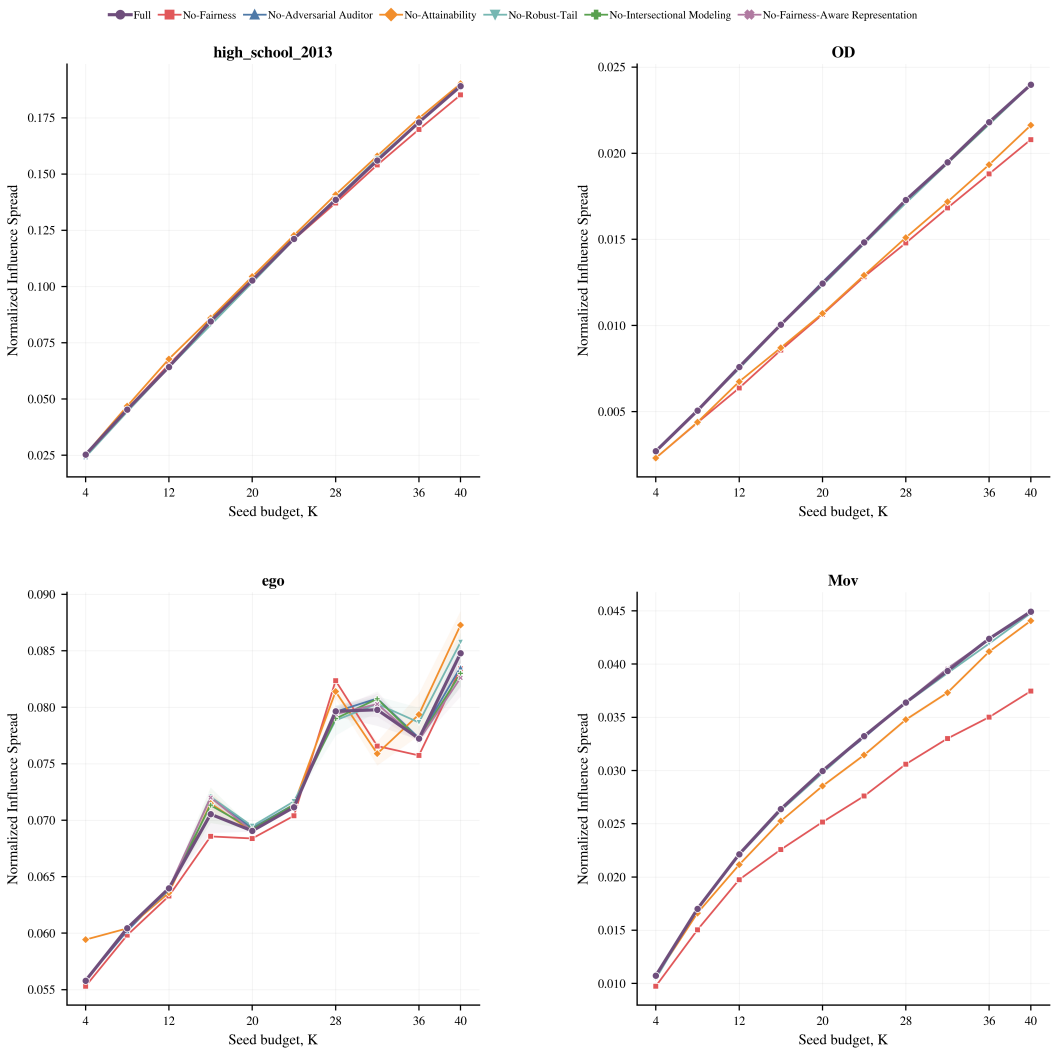


Figure 11: Supplied four-dataset ablation plots: Normalized average influence spread at the primary $\beta = .01$ setting. Panels show High-School-2013, Od, Ego-Facebook, and MovieLens.

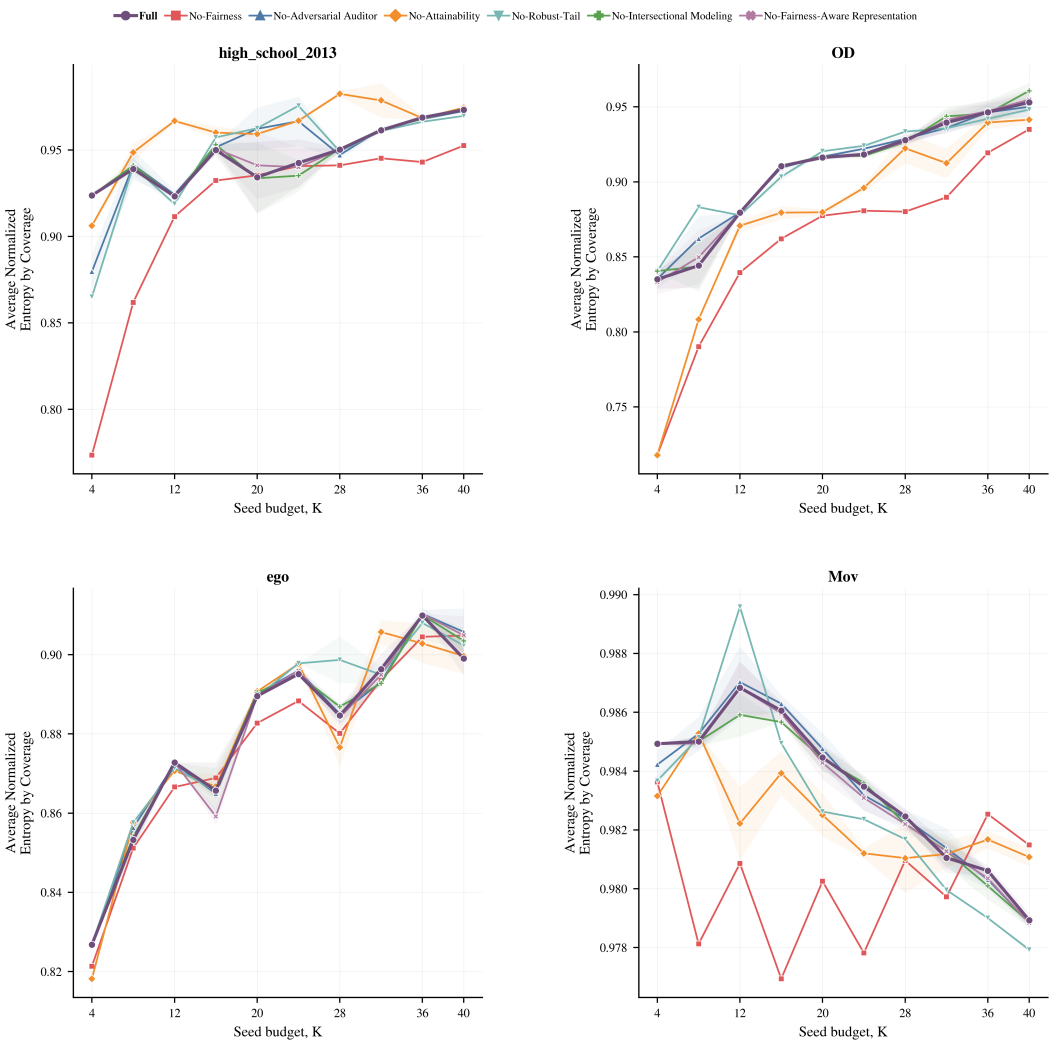


Figure 12: Supplied four-dataset ablation plots: Macro-average normalized entropy by coverage at the primary $\beta = .01$ setting. Panels show High-School-2013, Od, Ego-Facebook, and MovieLens.

APPENDIX H FULL-METHOD ATTRIBUTE DIAGNOSTICS

The supplied per-attribute heatmap plots entropy for the *Full* variant averaged over the paired repeats at each budget, not a comparison of ablation variants. The High-School-2013 rows are comparatively high, whereas the Od rows differ substantially: some attributes remain relatively low across several budgets despite the higher macro mean. In Ego-Facebook, the attributes indexed A7 and A8 stay markedly below several other attributes throughout the displayed range. MovieLens is near the entropy ceiling for all three plotted attributes.

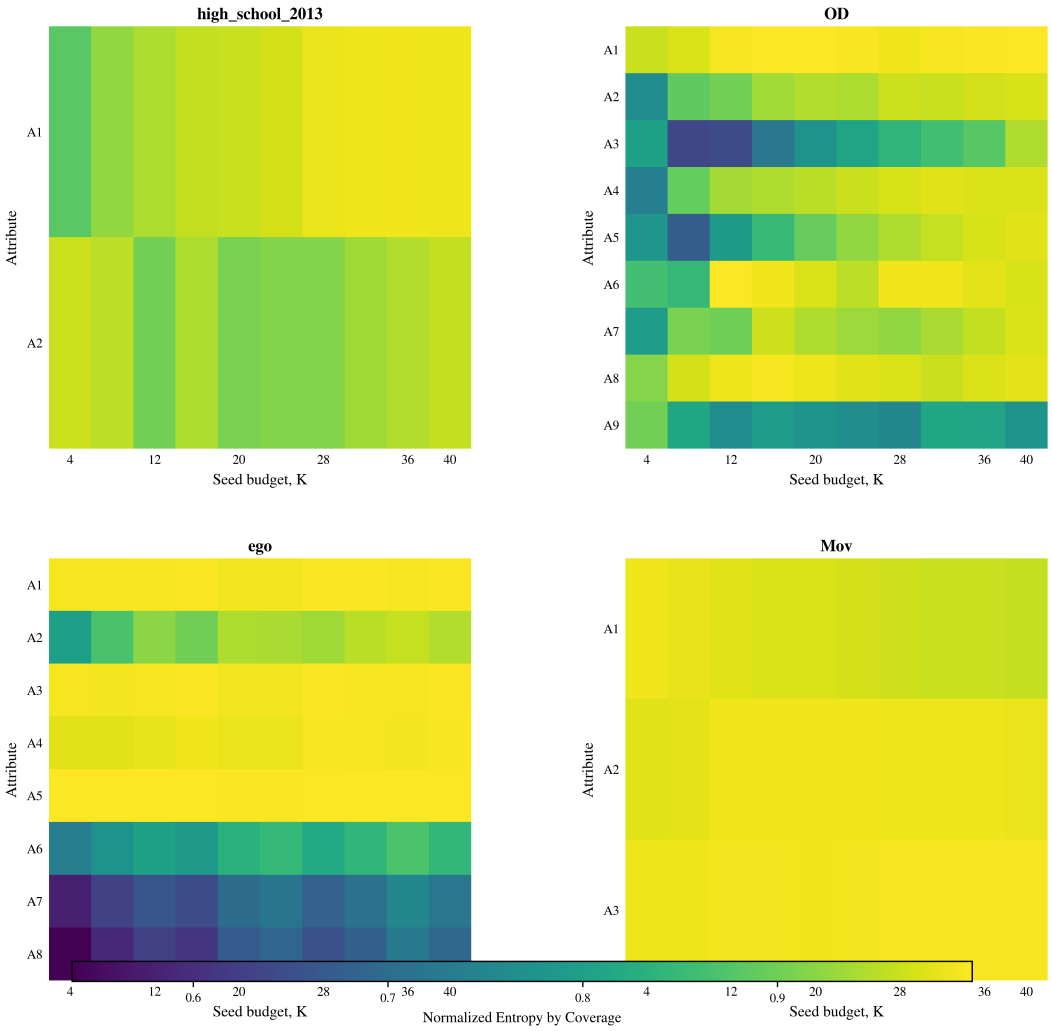


Figure 13: Full-method per-attribute normalized entropy across budgets at $\beta = .01$ on four datasets. Rows index observed evaluation attributes; colors encode mean entropy over the paired repeats.

The separate four-panel plot below follows only AURA-IM. It makes local reversals explicit on High-School-2013 and Ego-Facebook, the rising Od trajectory, and the decline in MovieLens entropy toward the largest budget.

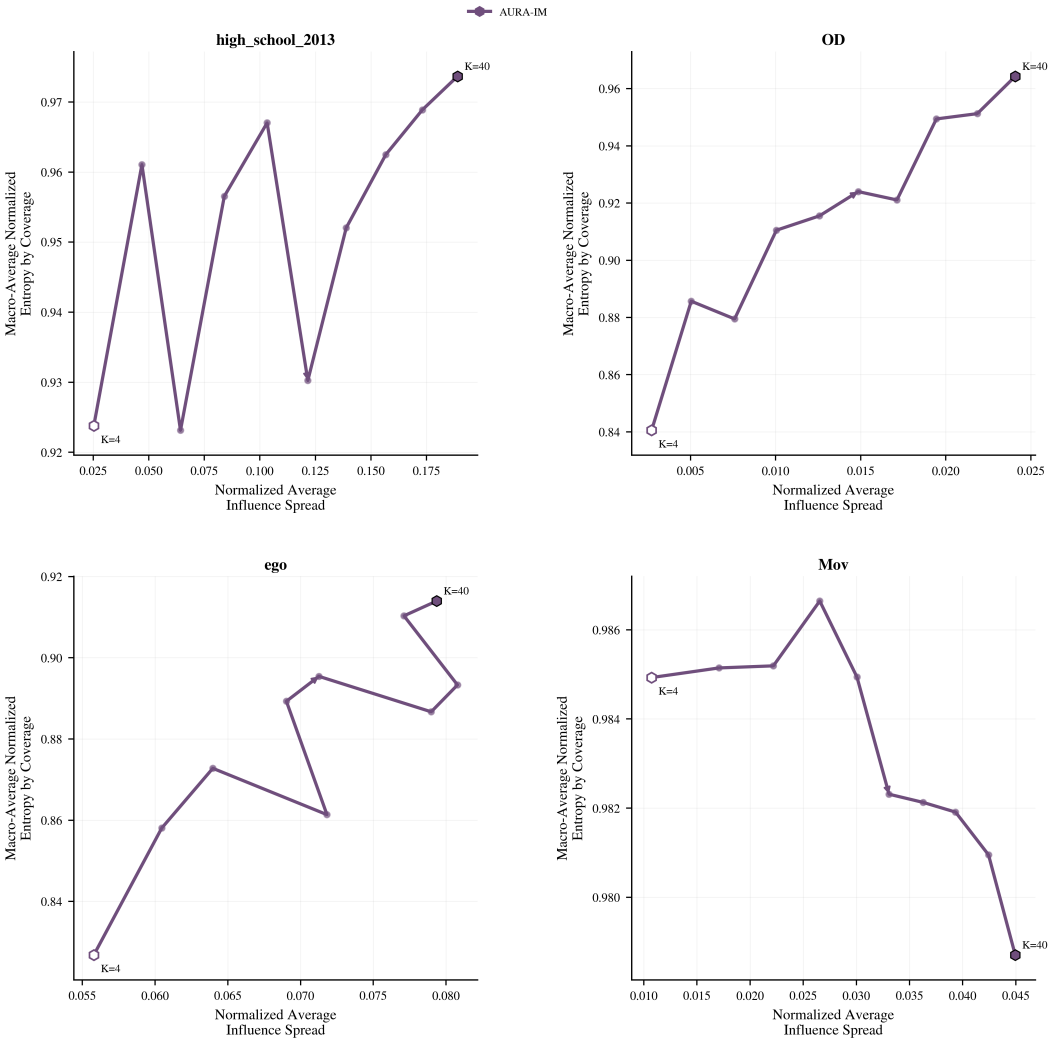


Figure 14: Full-method normalized spread versus macro entropy across budgets at $\beta = .01$ for High-School-2013, Od, Ego-Facebook, and MovieLens. Points join distinct budget-specific seed sets.