A MORE SCALABLE ALGORITHM FOR FAIR INFLUENCE MAXIMIZATION WITH CONCAVE FUNCTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

The Influence Maximization (IM) problem seeks a set of seed nodes that maximizes information spread in a social network. While IM has found applications in several domains, such as viral marketing, advertisement, public health and political campaigns, its solutions often concentrate influence among a subset of demographic groups, raising fairness concerns. Fair Influence Maximization (FIM) addresses this issue by requiring that influence is spread equitably across all groups. In particular, the Concave Fairness Framework (CFF), based on the poverty-reward principle of prioritizing under-reached groups, provides a general approach to FIM. However, greedy algorithms needed for its realization require many Monte Carlo simulations per step, making it impractical on large networks.

We here focus on a continuous relaxation of CFF, and propose a scalable algorithm for it, called FIM-SPACE. Our method works with any concave fairness function admissible in the CFF, and its runtime scales independently of the number of Monte Carlo samples required by greedy methods. We provide theoretical guarantees on the approximation quality of our approach and experimentally show that pipage rounding of its fractional solution produces a discrete seed set whose fairness and influence are competitive with those returned by the greedy baseline.

Our experiments show that FIM-SPACE achieves solution quality comparable to the greedy baseline on real-world social networks, while running up to $70\times$ faster. This means that on larger networks, where greedy methods become intractable, our approach is able to return high-quality fair seed sets within reasonable time.

1 INTRODUCTION

Identifying the most influential nodes in a network is one of the fundamental questions in network science and it has gained great attention in the last decades. This problem has been formalized in the *Influence Maximization* (IM) problem that, given a social network encoded as a graph, seeks a set of seed nodes from which to start an information diffusion campaign in order to maximize the spread of the information among the nodes in the network (Kempe et al., 2003). The IM problem found a considerable number of applications in several different domains such as viral marketing (Teng et al., 2024), advertising (Liang et al., 2023), HIV prevention and other public health challenges (Walsh et al., 2025; Auletta et al., 2025b), rumor control (Zhong et al., 2023), and the design of political campaigns (Jaouadi & Romdhane, 2024). The selection of the seeds strongly depends on the diffusion model adopted to describe how information spreads in the network. A largely adopted model is the Independent Cascade (IC) model (Kempe et al., 2003), a stochastic model where each activated node attempts to influence each inactive neighbor with a fixed probability, and the process unfolds in discrete time steps until no new activations occur.

While the IM problem is NP-hard (Kempe et al., 2003), the submodular nature of its objective function permits an efficient greedy approximation. Specifically, a greedy strategy achieves for every $\varepsilon > 0$ an $(1-1/e-\varepsilon)$ approximation guarantee relative to the optimal solution (Kempe et al., 2003). Although the aforementioned greedy framework has a good approximation ratio, it relies on Monte Carlo simulations (MCs) to estimate influence spread that makes this approach unfeasible in practice. To address this inefficiency issue several approximation algorithms and heuristics have been proposed for the IM problem (we refer the interested reader to the survey by Li et al. (2023)).

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.

However, since IM focuses on maximizing the spread of influence, it is naturally driven toward selecting seeds to disseminate information in the largest communities of the network, ignoring small communities and potentially creating large disparity among groups. In recent years, due to the increasing popularity of social networks as platforms to propagate information or opinions of social relevance that made them a significant influence source on people's lives, there is a growing concern about their potential discrimination against people, especially those belonging to small and marginalized groups. This has motivated research on the *Fair Influence Maximization* (FIM) problem, which extends IM to ensure an equitable distribution of influence across different social groups.

General fairness notions include *individual fairness*, which requires similar treatment for similar individuals (Dong et al., 2023), and *group fairness*, which requires that the algorithm does not yield discriminatory outcomes against individuals from any specific social group. In the context of FIM, research has focused primarily on fairness at the group level. In this paper, we adopt the framework proposed by Wang et al. (2024), which builds on the *poverty-reward* (PR) fairness notion. PR achieves fairness by prioritizing groups with low utility, and Wang et al. (2024) show that any concave function satisfying mild requirements realizes PR. This observation motivates the *Concave Fairness Framework* (CFF), which instantiates PR via concave functions and generalizes the approach of Rahmattalabi et al. (2021), that was restricted to isoelastic functions u^α/α for $\alpha < 1, \alpha \neq 0$, and $\log u$ for $\alpha = 0$, enabling flexible customization of fairness to different scenarios.

Based on PR, Wang et al. (2024) propose the Fair Influence Maximization within the Concave Fairness Framework (FIM-CFF) problem, and they show that it preserves submodularity and monotonicity, allowing to be solved with a greedy approach with $(1-1/e-\varepsilon)$-approximation guarantees for every $\varepsilon > 0$, regardless of the concave fairness function of interest. However, as for the classical IM problem, the greedy approach does not scale because the many per-step Monte Carlo simulations. Indeed, such greedy approaches are unaffordable not only for large networks (such as the one modeling social networks as Facebook or Twitter), but also for moderate-size networks, as the ones considered in the more recent applications of (F)IM for public health problems Walsh et al. (2025). It is then natural to ask whether the scalability limitations imposed by the execution of Monte Carlo simulations can be mitigated, and algorithms able to scale at least to these moderate size networks can be designed. For classical IM (Li et al., 2023), this has been achieved by Reverse Influence Sampling (RIS), which reformulates IM as a set coverage problem yielding $(1-1/e-\varepsilon)$-approximation guarantees without costly MCs (Borgs et al., 2014; Tang et al., 2015). An attempt to extend RIS to FIM was presented by Rui et al. (2023), who propose FIMM, an algorithm that achieves $(1-1/e-\epsilon)$-approximation guarantees, but *only* for the restricted class of concave fairness functions introduced by Rahmattalabi et al. (2021), failing to support the full generality of CFF[1].

Thus, the research question we address in this paper is the design of a solution method for FIM-CFF that supports any compliant concave fairness function, achieves solution quality comparable to the greedy baseline, and scales better than greedy being affordable at least on moderate-size networks.

Our Results. In this paper, we address the problem described above. To this end, we depart from its classical discrete formulation and formulate a continuous counterpart of the FIM-CFF problem, named cFIM-CFF, where, instead of looking for a set of seeds, we seek for a probability distribution over seeds guaranteeing that the expected size of the selected seed set does not exceed the budget and maximizes the fair objective function. We prove that cFIM-CFF preserves monotonicity and guarantees DR-submodularity (i.e., a natural generalization of submodularity to the continuous domain). To the best of our knowledge, we are the first to define and study such a continuous counterpart of the FIM problem. In this work we show that cFIM-CFF can be used as a proxy towards computing solutions for the original discrete problem, and we experimentally show its effectiveness in this role (see below). Note that continuous solutions may be of independent interest in applications where multiple diffusion campaigns are run (Auletta et al., 2025a).

We propose an algorithm, named FIM-SPACE, for cFIM-CFF that works for *all* the fairness functions consistent with the CFF framework and, for ϕ with bounded ϕ', ϕ'', returns a $(1-1/e)$-approximation of the optimum up to an additive ε. We show that the runtime of FIM-SPACE is $O\left(\frac{r^3 B^2}{\varepsilon^2}\,\lambda(\Lambda+\lambda^2)\,\Gamma\right)$, where B is the number of seeds to be selected, r is the number of groups, Γ is the average of the square of their sizes, while λ and Λ depend on how many nodes are able to influence a randomly selected node (more details below). On the worst case, $\Gamma = O(n^2/r)$, $\lambda = O(n)$,

[1] See Appendix A for more details about the obstacle behind the generalization of FIMM to other CFF.

and $\Lambda = O(n^2)$, where n is the number of nodes. Hence, in this worst case FIM-SPACE improves over the greedy algorithm for FIM-CFF by a factor $\frac{B \log n}{r^2}$. Despite this improvement over greedy can appear limited, we have evidences that the real improvement can be much larger.

The first of these evidences raises from the nature of these algorithms. Indeed, the huge complexity of greedy algorithms comes from Monte Carlo simulations: at each steps it estimates the marginal gain of all candidates, and each estimate needs many fresh cascade simulations. So greedy pays a per-candidate, per-step simulation cost that reappears every iteration. This cost component is unavoidable. FIM-SPACE instead never runs a Monte Carlo simulation. It draws one small batch of random realizations per iteration and uses it to update all coordinates at once, making the computational complexity of each iteration very limited. Clearly, this requires in principle many more iterations to provide the same theoretical guarantees as greedy. Anyway, for many choices of parameters the effective number of steps happens to be far smaller than that worst-case bound. Moreover, in many settings, many of the final steps of the algorithm can be avoided, as one can get sufficiently close to the desired outcome with just a few steps.

Furthermore, the regime that triggers the worst-case instance is precisely where fair influence maximization is largely unnecessary. Indeed, it requires that almost every node is able to infect a constant fraction of the graph. But then every seed set almost saturates every group, so there is no influence gap to correct and fairness is trivially achieved. In other words, in these cases, a near random algorithm would have been able to quickly compute a seed set guaranteeing a good approximation of the optimal fair influence. In principle, this can be merged with FIM-SPACE to achieve a better algorithm: a near-random selection when λ and Λ are large, and FIM-SPACE when they are small. Making this composition precise would require defining the thresholds on the value of λ and Λ for switching between regimes, and analyzing the cost of estimating them (which is not free). While this would have been an interesting theoretical direction, we defer it to future work and we here focus on the proposed optimization method used by FIM-SPACE and its analysis (see below). On the other hand, we experimentally observe that FIM-SPACE outperforms the greedy baseline over many realistic instances. Specifically, our experiments show that FIM-SPACE runs up to $70\times$ faster than the greedy baseline, and in almost all cases it reaches the desired guarantee after few steps, much less than how much would have been required by our theoretical bound. Experiments also highlight that FIM-SPACE serves as an effective proxy for FIM-CFF. Indeed, a randomized pipage rounding (Karimi et al., 2017) of its solution yields B seeds whose ex-ante guarantee is to return an $(1 - \frac{1}{e})$-approximation of the optimum up to an additive ε (App. B.3); experiments show that this guarantee tends to hold even ex-post, and the actual returned seed sets are of comparable quality to the greedy baseline. This means that on networks where greedy methods become intractable, FIM-SPACE still returns high-quality fair seeds within a reasonable time. Finally, despite its generality, FIM-SPACE achieves runtime comparable to or faster than FIMM (Rui et al., 2023) even on the fairness functions for which FIMM was designed, while producing solutions of comparable or superior quality. Results are also robust to variations of the parameters of FIM-SPACE.

Our Technical Contribution. The continuous nature of the problem considered in this work allows us to explore the use of continuous optimization tools. Several (stochastic)-gradient based methods do exist in literature for the maximization of DR-submodular functions (e.g., (Hassani et al., 2017; Mokhtari et al., 2020; Hassani et al., 2020; Lian et al., 2024; Zhang et al., 2020)) each of them coming with different guarantees. Most of these methods build on the Frank & Wolfe (1956) projection-free scheme, and try to achieve good approximation guarantees. Anyway, most of these methods have slow convergence rates (ε^{-3}), due to stochastic gradients and issues in facing variance. For this reason, latest works (Mokhtari et al., 2020) adopted variance-reduction methods like STORM (Cutkosky & Orabona, 2019). E.g., Hybrid SPIDER-CG algorithm (Lian et al., 2024) uses variance reduction and converges in ε^{-2}.

Unfortunately, this latter approach assumes that the objective function can be accessed through an unbiased stochastic gradient. However, our objective is a composition of two functions: an inner function g that estimates the spread of information within each group, and an outer concave function f that introduces non-linearity by rewarding groups with low utility more than those with high utility. Note that the inner function can be evaluated only averaging over samples, i.e., we are only able to compute an empirical estimate $\hat{g}(x)$ of $g(x) = \mathbb{E}[\hat{g}(x)]$. However, by the non-linearity of the outer function, the quantity that we are able to access, namely $f(\hat{g}(x))$, is a biased estimate of the right value $f(g(x)) = f(\mathbb{E}[\hat{g}(x)])$, since $\mathbb{E}[f(\hat{g}(x))] \neq f(\mathbb{E}[\hat{g}(x)])$. In other words, the natural stochastic

gradient is biased, and the approach of Lian et al. (2024) does not apply to our setting off the shelf. Hence, it is necessary to remove the bias introduced by the composition of these multiple functions.

Composed functions have been considered by Chen et al. (2020) and by Jiang et al. (2024). Specifically, Chen et al. (2020) considers the problem of maximizing an objective composed of multiple influence spread functions: however, in their case an unbiased estimation can be computed for all the composed functions. Hence, this approach fails to address our issue of biased estimation. Jiang et al. (2024) instead defined a projection-free, variance-reduced algorithm designed for *minimizing* a general smooth nested composition of functions, with no structure assumed beyond smoothness. Since our objective is instead to maximize a monotone and DR-submodular, and our target is a multiplicative approximation, their results do not transfer to our setting. Moreover, although the problems differ and the results are not directly comparable, their results either guarantee convergence at the weaker rate of $O(\varepsilon^{-3})$ or they require large, and thus computationally expensive, batches.

In FIM-SPACE we then combine the two approaches discussed above, namely a compositional bias control together with the DR-submodular maximization analysis, for our fair objective. More specifically, we reuse the potential argument of Lian et al. (2024), but the gradient-estimation part changes to handle the composition as done by Jiang et al. (2024). Actually, the nature of our problem, and specifically, the determinism of our outer function, also allows to achieve a convergence rate of $O(\varepsilon^{-2})$ with constant batch size, that is a faster and computationally cheaper rate than the one proved by Jiang et al. (2024) in their setting.

Roadmap. In section 2 we provide a formal definition of FIM-CFF and we define the continuous counterpart, cFIM-CFF. In section 3 we describe our algorithm FIM-SPACE, and bound its computational complexity and its approximation guarantees. Finally, in section 4 we provide experimental evidence of the scalability of our algorithm and of the quality of the returned outcomes.

2 PRELIMINARY DEFINITIONS

Consider an edge-weighted graph $H = (V, E, p)$, where V is the set of nodes, E is the set of edges, and $p\colon E \to [0, 1]$ assigns a probability to each edge. By sampling edges according to the edge probabilities p (i.e., each edge e is selected independently with probability $p(e)$), we obtain a *live graph* $G = (V, E')$, with $E' \subseteq E$. Let $\mathcal{G}_H$ be the distribution over the possible live graphs $G \sim \mathcal{G}_H$ that can be generated in this way. A subset $S \subseteq V$ of nodes is termed *seed set*. Note that we often find useful to represent S though its vector representation, that is, the vector $\mathbf{x}_S$ of $|V|$ entries such that $\mathbf{x}_S[i] = 1$ if $i \in S$ and $\mathbf{x}_S[i] = 0$ otherwise. Note that the size of the seed set is $|S| = |\{i\colon \mathbf{x}_S[i] = 1\}|$. Given a live graph G and a seed set S we denote with $g_G(S)$ the *influence* of S on G, where $g_G(S)$ is the number of nodes reachable in G from at least one node in S. The *expected influence* of a seed set S is then defined as $g(S) = \mathbb{E}_{G\sim\mathcal{G}_H}[g_G(S)]$. Given a live graph G and a node u, the *reverse reachable set* $R(G, u)$ denotes the set of nodes from which u is reachable in G. Note that u can be infected in G by a seed set S only if $S \cap R(G, u) \neq \emptyset$. Hence $g_G(S) = \sum_{u\in V} \mathbb{1}_{S\cap R(G,u)\neq\emptyset} = \sum_{u\in V} \Big(1 - \prod_{v\in R(G,u)} (1 - \mathbf{x}_S[v])\Big)$. Let us denote with $\mathcal{R}_H$ a probability distribution over reverse reachable sets given by choosing a live graph G according to $\mathcal{G}_H$ and a random node u. Then, it is not hard to see that $g(S) = |V| \cdot \mathbb{E}_{R\sim\mathcal{R}_H}\left[1 - \prod_{v\in R}(1 - \mathbf{x}_S[v])\right]$ (Chen et al. (2020)). It is immediate to check that this definition can be immediately extended from the discrete to the continuous domain as follows: let $\mathbf{x} \in [0, 1]^{|V|}$ be a fractional indicator vector, where each coordinate $\mathbf{x}[v]$ represents the independent probability of including node v in a random set S. Then we can extend $g(S)$ to $\gamma(\mathbf{x}) = |V| \left(1 - \mathbb{E}_{R\sim\mathcal{R}_H}\left[\prod_{v\in R}(1 - \mathbf{x}[v])\right]\right)$.

In the *Influence Maximization (IM)* problem, given $H = (V, E, p)$ and a budget $B \in \mathbb{N}^+$, the objective is to identify a seed set S^* of size at most B with maximum expected influence, i.e., $S^* \in \arg\max_{S\colon |S|\leq B} g(S)$. Kempe et al. (2003) established that the objective function $g(\mathbf{x}_S)$ is monotone and submodular, i.e., $g(\mathbf{x}_T) \leq g(\mathbf{x}_S)$ (*monotonicity*) and $g(\mathbf{x}_{S\cup\{u\}}) - g(\mathbf{x}_S) \leq g(\mathbf{x}_{T\cup\{u\}}) - g(\mathbf{x}_T)$ (*submodularity*) for any $T \subseteq S \subseteq V$ and $u \in V \setminus S$. Its continuous variants consist in identifying a probability distribution $\mathbf{x}^*$ over at most B seeds with maximum expected influence, i.e., $\mathbf{x}^* \in \arg\max_{\mathbf{x}\colon \sum_i \mathbf{x}[i]\leq B} \gamma(\mathbf{x})$.

Let $\mathcal{C} = \{C_1, \ldots, C_r\}$ be a partition of the set of nodes V into r disjoint groups. For a given seed set S, let $g_C(S)$ denote the expected number of influenced nodes in group $C \in \mathcal{C}$, i.e., $g_C(S) =$

$\mathbb{E}_{G\sim\mathcal{G}_H}[g_{G,C}(S)]$, where $g_{G,C}(S)$ is the number of nodes in C reachable from S in G. Similarly for a fractional indicator $\mathbf{x}$, we can define $\gamma_C(\mathbf{x}) = |C|\left(1 - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1-\mathbf{x}[v])\right]\right)$, where $\mathcal{R}_{H,C}$ is the same as $\mathcal{R}_H$ but restricted to choose a random node in C. The *utility* that group $C \in \mathcal{C}$ achieves from a seed set S is defined as the fraction of its members that are influenced, $u_C(S) = \frac{g_C(S)}{|C|}$. Similarly, the utility that C achieve from a distribution $\mathbf{x}$ is

$$\upsilon_C(\mathbf{x}) = \frac{\gamma_C(\mathbf{x})}{|C|} = 1 - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1-\mathbf{x}[v])\right]. \tag{1}$$

A *feasible fairness function* $\phi\colon [0,1] \to \mathbb{R}_{\geq 0}$ is a strictly increasing, twice-differentiable concave function. The concavity of ϕ ensures that "poorer" groups (those with lower utility) receive a higher marginal reward for influence gains than "richer" groups. We here set $\Omega_\phi = \max_{y\in[0,1]}\phi'(y)$, and $L_\phi = \max_{y_1,y_2\in[0,1]^2}\frac{|\phi'(y_1)-\phi'(y_2)|}{|y_1-y_2|}$. In the *Fair Influence Maximization problem under the Concave Fairness Framework (FIM-CFF)*, given $H = (V, E, p)$ and a budget $B \in \mathbb{N}^+$, the objective is to identify a seed set S^* of size at most B that maximizes the fair objective function $f(S) = \sum_{i=1}^{r}|C_i|\cdot\phi(u_{C_i}(S))$. Formally, $S^* \in \arg\max_{S\colon |S|\leq B} f(S)$.

Critically, Wang et al. (2024) proved that the fair objective function $f(S)$ inherits the fundamental properties of standard IM. Specifically, since the group influence $g_C(S)$ is non-decreasing and submodular, and ϕ is non-decreasing and concave, the composition $\phi(u_C(S))$ is also non-decreasing and submodular. Consequently, the aggregate fair objective $f(S)$ is both monotone and submodular. Hence, a simple greedy algorithm, named GREEDY-CFF, can be used to compute, for every $\varepsilon > 0$, an $\left(1-\frac{1}{e}-\varepsilon\right)$-approximation of $f(S^*)$, that choose the next seed as the one that maximizes the expected marginal contribution, where the expected marginal contribution of each seed can be estimated up to an error ε through Monte Carlo simulations. Unluckily, the need of a many Monte Carlo simulations makes this algorithm prohibitively slow beyond networks of a few thousand nodes. Specifically, the worst-case running time of the greedy algorithm is $O(\frac{B^3n^5L_\phi^2\log n}{\varepsilon^2})$ (Karimi et al., 2017).

Similarly, as done above for IM, we can here define even for FIM-CFF, its continuous counterpart, named cFIM-CFF, by simply replacing $u_C(S)$ with $\upsilon_C(\mathbf{x})$. That is, given $H = (V, E, p)$ and a budget $B \in \mathbb{N}^+$, the objective is to identify a distribution $\mathbf{x}$ over seed sets of size at most B (in expectation) that maximizes the fair objective function $\Phi(\mathbf{x}) = \sum_{i=1}^{r}|C_i|\cdot\phi(\upsilon_{C_i}(\mathbf{x}))$. Formally, $\mathbf{x}^* \in \arg\max_{\mathbf{x}\colon \sum_i \mathbf{x}[i]\leq B}\Phi(\mathbf{x})$. We now show that $\Phi(\mathbf{x})$ is monotone and DR-submodular on $[0,1]^{|V|}$, which is the structural property that enables the gradient-based methods we develop in the following sections. Note that $\Phi(\mathbf{x})$ is *not* the multilinear extension of $f(S)$: by Jensen's inequality, $\Phi(\mathbf{x}) = \sum_{C\in\mathcal{C}}|C|\cdot\phi(\mathbb{E}_{S\sim\mathbf{x}}[u_C(S)]) \geq \mathbb{E}_{S\sim\mathbf{x}}[f(S)]$, with equality only at the vertices of $[0,1]^{|V|}$. Hence the standard result that the multilinear extension of a monotone submodular function is DR-submodular (Calinescu et al. (2011)) does not apply to Φ directly, and we have to argue about Φ separately. Here, *monotonicity* ensures that increasing the selection probability of any node never decreases the total expected objective value, i.e., for any $\mathbf{x}, \mathbf{y} \in [0,1]^{|V|}$ with $\mathbf{x} \leq \mathbf{y}$ (componentwise), $\Phi(\mathbf{x}) \leq \Phi(\mathbf{y})$. *DR-submodularity* is the continuous analogue of diminishing returns: for all $\mathbf{x} \leq \mathbf{y}$ in $[0,1]^{|V|}$, all $i \in [|V|]$, and all $k \in \mathbb{R}_+$ such that $k\mathbf{e}_i + \mathbf{x}, k\mathbf{e}_i + \mathbf{y} \in [0,1]^{|V|}$, $\Phi(k\mathbf{e}_i + \mathbf{x}) - \Phi(\mathbf{x}) \geq \Phi(k\mathbf{e}_i + \mathbf{y}) - \Phi(\mathbf{y})$. Since Φ is twice differentiable, this is equivalent to $\frac{\partial^2\Phi}{\partial\mathbf{x}[i]\,\partial\mathbf{x}[j]} \leq 0$ for all $i, j \in V$ (Bian et al. (2017)).

Lemma 1. *$\Phi(\mathbf{x})$ is monotone non-decreasing and DR-submodular on $[0,1]^{|V|}$.*

The proof, deferred to Appendix B.1, follows by the chain rule applied to $\Phi(\mathbf{x}) = \sum_C |C|\cdot\phi(\upsilon_C(\mathbf{x}))$, using the fact that each υ_C is itself monotone DR-submodular (as the multilinear extension of a coverage function) and that ϕ is non-decreasing and concave.

By exploiting these properties, we here propose an algorithm, FIM-SPACE, that computes an $\left(1-\frac{1}{e}\right)$-approximation, up to an additive ε, of $\mathbf{x}^*$ in a time that improves by a factor $\Omega(\frac{B\log n}{r^2})$ over GREEDY-CFF in the worst case. FIM-SPACE leverages continuous optimization techniques by using gradient-based methods. To this aim, observe that

$$\frac{\partial\upsilon_C(\mathbf{x})}{\partial\mathbf{x}[i]} = \sum_R \Pr_{\mathcal{R}_{H,C}}(R)\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}[v]) \tag{2}$$

We set $m_{i,C}(\mathbf{x}) = \frac{\partial \upsilon_C(\mathbf{x})}{\partial \mathbf{x}[i]}$ and we call it the *marginal influence gain* of group C with respect to i in $\mathbf{x}$. We denote with $M(\mathbf{x})$ the matrix of all the $n \times r$ marginal gains. We set $w_C(\mathbf{x}) = \phi'(\upsilon_C(\mathbf{x}))$ and we call it the *marginal fairness weight*. Finally, we set $\mathbf{w}(\mathbf{x}) = (|C| w_C(\mathbf{x}))_{C \in \mathcal{C}}$. Hence,

$$\frac{\partial \Phi(\mathbf{x})}{\partial \mathbf{x}[i]} = \frac{\partial \sum_{j=1}^{r} |C_j| \phi(\upsilon_{C_j}(\mathbf{x}))}{\partial \mathbf{x}[i]} = \sum_{j=1}^{r} |C_j| \frac{\partial \phi(\upsilon_{C_j}(\mathbf{x}))}{\partial \mathbf{x}[i]}$$
$$= \sum_{j=1}^{r} |C_j| w_{C_j}(\mathbf{x}) \frac{\partial \upsilon_{C_j}(\mathbf{x})}{\partial \mathbf{x}[i]} = \sum_{j=1}^{r} |C_j| w_{C_j}(\mathbf{x}) m_{i,C_j}(\mathbf{x}),$$

and, consequently, $\nabla \Phi(\mathbf{x}) = M(\mathbf{x}) \mathbf{w}(\mathbf{x})$.

3 FIM-SPACE

In order to find an approximate solution to our problem, we propose a stochastic projection-free gradient-based approach building on the Frank & Wolfe (1956) scheme and its stochastic variants for DR-submodular maximization (Mokhtari et al., 2020). At each iteration t, we draw batches of β_t reverse reachable (RR) sets (line 3); use them to compute an estimate $\mathbf{v}_t$ of the gradient (lines 4-18); compute $\mathbf{z}_t$ by selecting the top B entries of $\mathbf{v}_t$, i.e., the maximizer of the linear approximation $\langle \mathbf{v}_t, \mathbf{s} \rangle$ of Φ at $\mathbf{x}_t$ over $\{\mathbf{s} \in \mathcal{X} : \sum_i \mathbf{s}[i] = B\}$, where $\mathcal{X} = \{\mathbf{x} \in [0,1]^{|V|} : \sum_i \mathbf{x}[i] \leq B\}$ (line 19); and finally move $\mathbf{x}_t$ by $\mathbf{z}_t / T$ (line 20), as in (Lian et al., 2024).

A naive instantiation of this scheme would be impractical due to the large variance of the RR-set-based gradient estimator. We thus adopt the STORM momentum-based variance reduction technique of Cutkosky & Orabona (2019) Concretely (line 18), rather than setting $\mathbf{v}_t$ to the fresh gradient estimate $M_t \mathbf{w}_t$, we take a convex combination, governed by a momentum parameter $\alpha_t \in [0,1]$, between this estimate and a corrected estimate $\tilde{\mathbf{v}}_t$. The latter is obtained (line 17) by adjusting $\mathbf{v}_{t-1}$ with the differential $M_t \mathbf{w}_t - \hat{M}_{t-1} \mathbf{w}_{t-1}$, where $\hat{M}_{t-1}$ is the marginal-influence-gain matrix evaluated at the previous iterate $\mathbf{x}_{t-1}$ on the *current* batch $\mathcal{S}^t$. Intuitively, $\tilde{\mathbf{v}}_t$ progressively cancels the per-batch noise so that in the long run $\mathbf{v}_t$ becomes increasingly insensitive to sampling variance.

This alone, however, is insufficient. Our objective Φ has a two-level compositional structure: an *inner* function mapping $\mathbf{x}$ to the vector of group utilities $(\upsilon_C(\mathbf{x}))_{C \in \mathcal{C}}$, accessed via stochastic RR-set samples, and an *outer* concave aggregation $\mathbf{u} \mapsto \sum_C |C| \phi(u_C)$ which is deterministic given the inner values. Standard variance-reduction techniques for stochastic DR-submodular maximization treat the gradient as a single stochastic estimator and do not account for the bias introduced by composing the outer concave layer with stochastic estimates of the inner utilities. Following Jiang et al. (2024), we handle this by tracking the inner utilities υ_C^t separately via the same momentum-based scheme (lines 4-10), and then composing them with the outer derivative ϕ' to form $\mathbf{w}_t$ (line 11) before plugging into the gradient estimator. We remark that, since our outer function is deterministic, the gradient tracker of Jiang et al. (2024) simplifies in our setting: the marginal fairness weight $\mathbf{w}_t$ factors out of the per-sample products, and the resulting update is the one shown in line 18.

We call our procedure *Fair Influence Maximization via Stochastic Projection-free Adaptive Compositional Estimation* (FIM-SPACE) and report the pseudocode in Algorithm 1.

Theorem 1. *Let $\lambda = \max_{C \in \mathcal{C}} \mathbb{E}_{R \sim \mathcal{R}_{H,C}}[|R|]$ and $\Lambda = \max_{C \in \mathcal{C}} \mathbb{E}_{R \sim \mathcal{R}_{H,C}}[|R|^2]$. For every feasible fairness function ϕ, there are choices of T, α_t, and β_t such that* FIM-SPACE *returns an outcome $\mathbf{x}^T$ such that $\mathbb{E}[\Phi(\mathbf{x}^T)] \geq \left(1 - \frac{1}{e}\right) \Phi(\mathbf{x}^*) - \varepsilon$ in time $O\left(\frac{1}{\varepsilon^2} r^2 \lambda B^2 \sum_{C \in \mathcal{C}} |C|^2 \left(\Lambda \Omega_\phi^2 + \lambda^2 L_\phi^2\right)\right)$.*

The proof of the theorem works by customizing the framework of Jiang et al. (2024) to prove that our algorithm successfully achieves variance reduction despite the non-linear fairness composition, and thus the estimated gradient $\mathbf{v}_t$ quickly approaches to the real gradient $\nabla \Phi(\mathbf{x}_t)$. Specifically, we show (cf. Lemma 9) that the composed gradient error decays as $O(1/t)$, by controlling the inner-utility error (c.f., Lemma 4) and the marginal-gain error (c.f., Lemma 7) separately and then composing them (c.f., Lemma 8). The convergence of the gradient enable us (c.f., Proposition 1) to estimate the approximation guarantee of FIM-SPACEwith respect to the number of iterations via a potential argument, similar to the one used by Lian et al. (2024). This allows bounds the number of iterations required by the algorithm to reach the desired approximation guarantee. Finally, to

Algorithm 1 FIM-SPACE

Require: Graph $H = (V, E, p)$, communities $\mathcal{C}$, budget B, iterations T, momentum $\{\alpha_t\}$, batch sizes $\{\beta_t\}$

Ensure: A fractional solution $\mathbf{x}^* \in \mathcal{X} = \{\mathbf{x} \in [0,1]^{|V|} : \sum_i \mathbf{x}[i] \leq B\}$

Initialize $\mathbf{x}_0 = \mathbf{0}$
for $t = 0$ **to** $T-1$ **do**
  Sample, for each $C \in \mathcal{C}$, independent batches $\mathbf{R}_C^t, \mathcal{S}_C^t$ of β_t RR sets from $\mathcal{R}_{H,C}$
  $\bar{v}_C^t \leftarrow 1 - \frac{1}{\beta_t} \sum_{R \in \mathbf{R}_C^t} \prod_{v \in R} (1 - \mathbf{x}_t[v])$, for each $C \in \mathcal{C}$
  if $t = 0$ **then**
    $v_C^t \leftarrow \bar{v}_C^t$, for each $C \in \mathcal{C}$
  else
    $\hat{v}_C^{t-1} \leftarrow 1 - \frac{1}{\beta_t} \sum_{R \in \mathbf{R}_C^t} \prod_{v \in R} (1 - \mathbf{x}_{t-1}[v])$, for each $C \in \mathcal{C}$
    $\tilde{v}_C^t \leftarrow v_C^{t-1} + \bar{v}_C^t - \hat{v}_C^{t-1}$, for each $C \in \mathcal{C}$
    $v_C^t \leftarrow (1 - \alpha_t)\, \tilde{v}_C^t + \alpha_t\, \bar{v}_C^t$, for each $C \in \mathcal{C}$
  $\mathbf{w}_t \leftarrow \left(|C| \cdot \phi'([v_C^t]_0^1)\right)_{C \in \mathcal{C}}$
  $M_t \leftarrow (m_{i,C}^t)_{i \in V,\, C \in \mathcal{C}}$ where $m_{i,C}^t = \frac{1}{\beta_t} \sum_{R \in \mathcal{S}_C^t} \mathbb{1}_{i \in R} \prod_{v \in R \setminus \{i\}} (1 - \mathbf{x}_t[v])$
  if $t = 0$ **then**
    $\mathbf{v}_t \leftarrow M_t \mathbf{w}_t$
  else
    $\hat{M}_{t-1} \leftarrow (\hat{m}_{i,C}^{t-1})_{i \in V,\, C \in \mathcal{C}}$ where $\hat{m}_{i,C}^{t-1} = \frac{1}{\beta_t} \sum_{R \in \mathcal{S}_C^t} \mathbb{1}_{i \in R} \prod_{v \in R \setminus \{i\}} (1 - \mathbf{x}_{t-1}[v])$
    $\tilde{\mathbf{v}}_t \leftarrow \mathbf{v}_{t-1} + M_t \mathbf{w}_t - \hat{M}_{t-1} \mathbf{w}_{t-1}$
    $\mathbf{v}_t \leftarrow (1 - \alpha_t)\, \tilde{\mathbf{v}}_t + \alpha_t\, (M_t \mathbf{w}_t)$
  $\mathbf{z}_t \leftarrow \arg\max_{\mathbf{s} \in \mathcal{X} : \sum_i \mathbf{s}[i] = B} \langle \mathbf{s}, \mathbf{v}_t \rangle$
  $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t + \mathbf{z}_t / T$
return $\mathbf{x}^* \leftarrow \mathbf{x}_T$

estimate the total running time of FIM-SPACE, we evaluate the complexity of each iteration. Due to the page limit we defer the full proof of this theorem to the appendix B.2.

Since $\lambda = O(n)$, $\Lambda = O(n^2)$, and $\sum_{C \in \mathcal{C}} |C|^2 = O(n^2)$, we have the following approximation.

Corollary 1 (Worst-case approximation)**.** *For every feasible fairness function ϕ, there are choices of T, α_t, and β_t such that* FIM-SPACE *returns an outcome $\mathbf{x}^T$ such that $\mathbb{E}[\Phi(\mathbf{x}^T)] \geq \left(1 - \frac{1}{e}\right) \Phi(\mathbf{x}^*) - \varepsilon$ in time $O(r^2 n^5 B^2 (\Omega_\phi^2 + L_\phi^2)/\varepsilon^2)$.*

For sparse graphs and few groups, $\lambda, \Lambda, r = O(1)$ (e.g., $\lambda \leq 8$, $\Lambda \leq 264$ in our large networks, Table 5), and FIM-SPACE requires only $O(n^2 B^2 (\Omega_\phi^2 + L_\phi^2)/\varepsilon^2)$ time.

4 EXPERIMENTAL EVALUATION

In this section we experimentally[2] evaluate FIM-SPACE with respect to the quality of the solution, and the scalability on moderate size networks. Other robustness results are reported in Appendix.

Datasets. To evaluate quality, we use 24 networks from Wilder et al. (2018), modeling a real-world intervention in the Antelope Valley region of California (AV networks). Each node represents a real individual; directed edges encode peer influence, and each network contains 500 nodes. Following Tsang et al. (2019) and Wang et al. (2024), nodes carry three demographic attributes partitioning the network into disjoint groups: *ethnicity* (5 groups), *age* (7 groups), and *gender* (2 groups). As in Wang et al. (2024), experiments on AV networks use a uniform influence probability 0.1 under the IC model, seed budget $B = 25$, and 15 independent runs per instance.

[2] All experiments are run on a server equipped with two AMD EPYC 9654 96-core processors. The code is implemented in Python 3.10.12 and worker-level parallelism is implemented, allowing for parallel independent executions. Complete source code, datasets, and scripts to reproduce the results are included in the supplementary material. The code will be made publicly available upon publication. Additionally, all stochastic components use fixed incremental seeds for the randomness for each of the independent runs, ensuring that results are exactly reproducible and that trials are independent draws.

Table 1: Solution quality on AV networks (24 graphs, 3 attributes). Each cell reports mean ± SD across all 72 graph–attribute pairs (n=24 graphs per attribute, trials averaged within each graph). Higher is better for Φ_φ, σ, RoF, F_2; lower is better for PoF and Time. FIMM is restricted to φ_{pow}.

φ	Algorithm	Φ_φ	σ	PoF	RoF	F_2	Time (s)
φ_{pow}	FIM-SPACE	400.04 ± 0.80	53.89 ± 1.07	0.005 ± 0.009	0.644 ± 0.314	0.883 ± 0.141	2.94 ± 0.49
	GREEDY-CFF	399.97 ± 0.83	53.81 ± 1.09	0.006 ± 0.009	0.634 ± 0.361	0.880 ± 0.158	209.28 ± 33.58
	FIMM	397.19 ± 1.20	51.23 ± 1.16	0.054 ± 0.016	0.082 ± 0.433	0.658 ± 0.202	0.81 ± 0.42
φ_{exp}	FIM-SPACE	329.25 ± 3.64	53.87 ± 1.05	0.005 ± 0.008	0.677 ± 0.299	0.895 ± 0.134	2.94 ± 0.51
	GREEDY-CFF	328.99 ± 3.89	53.79 ± 1.12	0.006 ± 0.010	0.713 ± 0.328	0.908 ± 0.138	209.52 ± 34.66
φ_{log}	FIM-SPACE	204.63 ± 1.77	53.92 ± 1.06	0.004 ± 0.008	0.585 ± 0.311	0.864 ± 0.141	2.94 ± 0.52
	GREEDY-CFF	204.46 ± 1.83	53.83 ± 1.09	0.006 ± 0.009	0.575 ± 0.381	0.863 ± 0.155	208.30 ± 36.44

To evaluate scalability, we consider four social networks of increasing scale from the SNAP repository (Yang & Leskovec, 2012): *email*, *amazon*, *youtube*, and *DBLP* (full descriptions and URLs in Appendix C). Following Rui et al. (2023), we set $p(u, v) = 1/d_{\text{in}}(v)$ under the weighted IC model, with $B \in \{10, 25, 50\}$ and 15 independent runs per instance. Communities with fewer than 10 nodes are pruned; preprocessing details are in Appendix C.

Baselines. We compare FIM-SPACE against two fair baselines and one standard IM reference. GREEDY-CFF (Wang et al., 2024) is the greedy realization of the CFF, iteratively selecting seeds by largest marginal gain in $\Phi_\varphi(\cdot)$ via Monte Carlo estimates. It supports any concave ϕ and achieves a $(1 - 1/e)$-approximation (Wang et al., 2024). FIMM (Rui et al., 2023) is a RIS-based algorithm with a $(1-1/e-\varepsilon)$-approximation guarantee, restricted to the power function $\phi(u) = u^\alpha$ for $\alpha > 0$. For computing fairness metrics, we also evaluate GREEDY-IM (Kempe et al., 2003), the standard greedy IM baseline (no fairness), replaced by the more scalable IMM(Tang et al., 2015) on larger graphs. Full baseline configurations are reported in Appendix C.

Metrics. For each returned seed set S we report the *fair objective* Φ_φ, *influence spread* σ, *Price of Fairness* (PoF), *Reward of Fairness* (RoF), and F_2-score, as defined by Wang et al. (2024) (formal definitions in Appendix C). Briefly: PoF measures the relative loss in influence spread incurred by enforcing fairness (lower is better); RoF quantifies the reduction in per-community utility variance relative to GREEDY-IM (higher is better); and the F_2-score summarizes the fairness–influence tradeoff as their weighted harmonic mean (higher is better). All metrics are estimated via 10,000 Monte Carlo simulations. Results are reported as mean ± standard deviation over 15 independent runs; on the AV collection, values are additionally averaged over all 24 networks. Reported running times include all per-algorithm steps: RR-set generation for FIMMand IMM, RR-set pool construction, gradient iterations and pipage rounding (required for returning a discrete solution comparable with other baselines) for FIM-SPACE, and Monte Carlo evaluations for GREEDY-CFF and GREEDY-IM.

Adopted Fairness Functions. Following Wang et al. (2024), we instantiate GREEDY-CFF and FIM-SPACE with three representative concave functions that cover qualitatively distinct shapes: *power*, $\varphi_{\text{pow}}(u) = u^{\zeta_1}$ with $\zeta_1 = 0.1$; *exponential*, $\varphi_{\text{exp}}(u) = 1 - e^{-\zeta_2 u}$ with $\zeta_2 = 10$; and *logarithmic*, $\varphi_{\text{log}}(u) = \log_2(u^{\zeta_3} + 1)$ with $\zeta_3 = 0.5$ (FIMM is only instantiated with the power one). The value $\zeta_1 = 0.1$ follows the default setting of FIMM (Rui et al., 2023). For ζ_2 and ζ_3 we select mid-curvature values from within the ranges swept by Wang et al. (2024).

Results: FIM Quality. We evaluate FIM-SPACE, GREEDY-CFF, FIMM, and GREEDY-IM on the 24 AV networks across the three demographic attributes and $\varphi \in \{\varphi_{\text{pow}}, \varphi_{\text{exp}}, \varphi_{\text{log}}\}$, using parameters $T = 500$, $\beta = 100$ (full configuration in App. C). Trials are averaged within each graph before aggregation; all comparisons are paired by graph ($n = 24$). Table 1 reports mean ± SD across the 72 graph–attribute pairs. Significance uses the two-sided Wilcoxon signed-rank test with Holm–Bonferroni correction within each (metric, comparison) family across the 9 (φ, attribute) cells (3 for FIMM, power-only); full per-cell results are in Appendix C.1.

FIM-SPACE and GREEDY-CFF achieve statistically indistinguishable F_2 and RoF across all fairness functions and demographic attributes, with both methods incurring a near-zero price of fairness (PoF ≤ 0.006). Both substantially outperform FIMM on every metric ($\Delta F_2 \simeq +0.22$, $\Delta\text{RoF} \simeq +0.56$), reflecting a regime effect where FIMM sacrifices quality for scalability. Crucially, FIM-SPACE achieves this matched quality in $\sim$3 s per instance versus GREEDY-CFF's $\sim$209 s, a $\sim$70$\times$ speedup. FIMM is fastest at $\sim$0.8 s but at the quality cost described above.

Table 2: Quality and scalability on SNAP datasets ($B = 25$); mean $\pm$ SD over 15 trials. ‡ algorithm stopped at the 1-hour deadline (runtime is a lower bound). * significant vs. GREEDY-CFF; † significant vs. FIMM (Wilcoxon signed-rank, $\alpha = 0.05$); runtime excluded from significance testing.

φ	Dataset	Algorithm	Φ_φ	σ	PoF	RoF	Time (s)
φ_{pow}	Email	FIM-SPACE	$826.46 \pm 0.37^{*\dagger}$	$330.57 \pm 1.59^{*\dagger}$	$-0.012 \pm 0.012^{*\dagger}$	$1.012 \pm 0.012^{*\dagger}$	19.19 ± 0.81
		GREEDY-CFF	826.04 ± 0.50	328.88 ± 1.98	-0.007 ± 0.009	1.007 ± 0.009	880.22 ± 18.95
		FIMM	825.28 ± 0.20	325.86 ± 0.68	0.003 ± 0.009	0.997 ± 0.009	80.63 ± 1.05
	Amazon	FIM-SPACE	$1059.99 \pm 132.47^{\dagger}$	$275.13 \pm 20.19^{\dagger}$	$0.473 \pm 0.040^{\dagger}$	$0.527 \pm 0.040^{\dagger}$	302.23 ± 28.01
		GREEDY-CFF‡	—	—	—	—	> 3600
		FIMM	1526.28 ± 23.30	526.09 ± 2.29	-0.007 ± 0.012	1.007 ± 0.012	325.52 ± 16.68
	YouTube	FIM-SPACE	$27904.98 \pm 5.37^{\dagger}$	$12359.83 \pm 24.04^{\dagger}$	$-0.002 \pm 0.004^{\dagger}$	$1.002 \pm 0.004^{\dagger}$	1044.52 ± 40.22
		FIMM	27688.26 ± 6.48	11248.61 ± 26.45	0.088 ± 0.004	0.912 ± 0.004	2168.74 ± 55.68
	DBLP	FIM-SPACE	$54310.97 \pm 56.80^{\dagger}$	$7296.36 \pm 108.15^{\dagger}$	$0.095 \pm 0.015^{\dagger}$	$0.905 \pm 0.015^{\dagger}$	2114.79 ± 99.04
		FIMM	53422.86 ± 66.92	6670.07 ± 61.43	0.173 ± 0.007	0.827 ± 0.007	1064.67 ± 17.17
φ_{exp}	Email	FIM-SPACE	$888.40 \pm 0.53^{*}$	329.29 ± 2.02	-0.008 ± 0.011	1.008 ± 0.011	18.94 ± 0.62
		GREEDY-CFF	887.75 ± 0.84	327.53 ± 2.98	-0.003 ± 0.015	1.003 ± 0.015	883.91 ± 24.53
	Amazon	FIM-SPACE	1410.61 ± 51.48	387.42 ± 18.69	0.258 ± 0.034	0.742 ± 0.034	295.66 ± 35.42
	YouTube	FIM-SPACE	29831.14 ± 5.63	11970.95 ± 97.76	0.029 ± 0.007	0.971 ± 0.007	1026.50 ± 53.93
	DBLP	FIM-SPACE	43936.23 ± 61.42	8038.83 ± 32.92	0.003 ± 0.008	0.997 ± 0.008	2102.27 ± 128.09
φ_{log}	Email	FIM-SPACE	$620.10 \pm 1.23^{*}$	330.90 ± 1.78	-0.013 ± 0.013	1.013 ± 0.013	18.90 ± 0.52
		GREEDY-CFF	618.40 ± 2.13	328.90 ± 2.96	-0.007 ± 0.010	1.007 ± 0.010	886.46 ± 20.93
	Amazon	FIM-SPACE	706.41 ± 48.57	305.76 ± 17.85	0.415 ± 0.036	0.585 ± 0.036	299.24 ± 36.73
	YouTube	FIM-SPACE	21542.33 ± 15.71	12362.41 ± 22.65	-0.002 ± 0.003	1.002 ± 0.003	1040.46 ± 60.74
	DBLP	FIM-SPACE	27494.10 ± 98.62	7772.70 ± 63.06	0.036 ± 0.011	0.964 ± 0.011	2098.94 ± 117.60

Results: Scalability. To evaluate scalability, we process datasets in increasing order of scale. For each (algorithm, dataset, budget) triple, if its first trial exceeds the one-hour deadline, the algorithm is marked intractable for that dataset and all larger ones, otherwise the following 14 trials are executed. The IMM reference is probed first: if intractable, relative metrics are unavailable for that dataset.

Table 2 reveals a clear scalability gap. GREEDY-CFF fails to complete beyond Email, exhausting the one-hour deadline on Amazon and timing out entirely on YouTube and DBLP. FIMM, while scalable, is restricted to φ_{pow} and cannot be applied when other fairness criteria are required. FIM-SPACE, by contrast, completes all 15 trials on every dataset and for all three fairness functions within the time limit, reaching runtimes of $\sim$19 s on Email, $\sim$300 s on Amazon, $\sim$1,040 s on YouTube, and $\sim$2,100 s on DBLP at $B = 25$. This represents the first demonstration of welfare-fair influence maximization beyond the power function on networks with tens of thousands of nodes.

Quality is preserved alongside scalability. On Email, FIM-SPACE matches GREEDY-CFF and FIMM to within 0.5% on Φ_φ. On YouTube and DBLP, FIM-SPACE consistently outperforms FIMM on both objective and influence spread (e.g., RoF $= 0.90$ vs. 0.83 on DBLP under φ_{pow}). The one exception is Amazon under φ_{pow}, where FIM-SPACE shows high variance ($\Phi_\varphi = 1{,}060 \pm 132$ vs. FIMM's $1{,}526 \pm 23$); this limitation is specific to the power function on this dataset, as FIM-SPACE recovers under φ_{exp} (RoF $= 0.74$) while GREEDY-CFF exceeds the time limit. Small negative PoF values and RoF slightly above 1.0 on Email are discussed in Appendix C.2.

5 LIMITATIONS AND FUTURE DIRECTIONS

Despite its shown scalability, FIM-SPACE still has a huge worst-case complexity. Hence, it would be interesting to evaluate if this complexity is unavoidable, or faster algorithms can be designed (e.g., by composing FIM-SPACE with a quasi-random algorithm, as discussed above, or by partitioning the network while accounting for the influence across parts). Extending FIM-SPACE to work even with other information diffusion models beyond IC would also be useful.

While we use the outcome of FIM-SPACE as a proxy for computing a set of seeds of size exactly B through a simple rounding approach, and experimental evaluation shows that these seeds have quality comparable with the one computed by GREEDY-CFF, we stress that our guarantee for them is *ex ante* (App. B.3): whether the expected fairness $\mathbb{E}[f(S)]$ of the actual returned set is still a $(1 - \frac{1}{e})$-approximation even *ex-post* remains open.

AI USE STATEMENT

In this work, we used generative AI tools to assist in the formulation of mathematical claims, to assist in the writing of proofs and code implementation, and to provide feedback on the research methodology. Additionally, we used generative AI tools to draft parts of the paper and to polish its writing. All AI-assisted content was reviewed by the authors, corrected and verified by hand; the authors take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

FIM-SPACE is specified in Algorithm 1, and complete proofs of all theoretical results are given in the appendix. All datasets are publicly available; their descriptions, the configuration of FIM-SPACE and of all baselines, and the evaluation metrics are reported in Appendix C. Source code, datasets, and scripts to reproduce all results are included in the supplementary material, and will be made publicly available upon acceptance.

REFERENCES

Vincenzo Auletta, Francesco Carbone, Diodato Ferraioli, and Cosimo Vinci. Adaptive multi-round influence maximization with limited information. In *International Conference on Principles and Practice of Multi-Agent Systems*, pp. 187–204. Springer, 2025a.

Vincenzo Auletta, Diodato Ferraioli, and Grazia Ferrara. Influence maximization in unknown social networks: A contextual bandit approach. In *International Conference on Principles and Practice of Multi-Agent Systems*, pp. 224–231. Springer, 2025b.

An Bian, Kfir Levy, Andreas Krause, and Joachim M Buhmann. Continuous dr-submodular maximization: Structure and algorithms. *Advances in Neural Information Processing Systems*, 30, 2017.

Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. Maximizing social influence in nearly optimal time. In *Proceedings of the 2014 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 946–957, 2014. doi: 10.1137/1.9781611973402.70.

Gruia Calinescu, Chandra Chekuri, Martin Pal, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.

Wei Chen, Weizhong Zhang, and Haoyu Zhao. Gradient method for continuous influence maximization with budget-saving considerations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 43–50, 2020.

Ashok Cutkosky and Francesco Orabona. Momentum-based variance reduction in non-convex sgd. *Advances in neural information processing systems*, 32, 2019.

Yushun Dong, Jing Ma, Song Wang, Chen Chen, and Jundong Li. Fairness in graph mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10583–10602, 2023.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

Hamed Hassani, Mahdi Soltanolkotabi, and Amin Karbasi. Gradient methods for submodular maximization. *Advances in Neural Information Processing Systems*, 30, 2017.

Hamed Hassani, Amin Karbasi, Aryan Mokhtari, and Zebang Shen. Stochastic conditional gradient++:(non) convex minimization and continuous submodular maximization. *SIAM Journal on Optimization*, 30(4):3315–3344, 2020.

Myriam Jaouadi and Lotfi Ben Romdhane. A survey on influence maximization models. *Expert Systems with Applications*, 248:123429, 2024.

Wei Jiang, Sifan Yang, Wenhao Yang, Yibo Wang, Yuanyu Wan, and Lijun Zhang. Projection-free variance reduction methods for stochastic constrained multi-level compositional optimization. In *International Conference on Machine Learning*, pp. 21962–21987. PMLR, 2024.

Mohammad Karimi, Mario Lucic, Hamed Hassani, and Andreas Krause. Stochastic submodular maximization: The case of coverage functions. *Advances in Neural Information Processing Systems*, 30, 2017.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146, 2003.

Yandi Li, Haobo Gao, Yunxuan Gao, Jianxiong Guo, and Weili Wu. A survey on influence maximization: From an ml-based combinatorial optimization. *ACM Transactions on Knowledge Discovery from Data*, 17(9):1–50, 2023.

Yuefang Lian, Donglei Du, Xiao Wang, Dachuan Xu, and Yang Zhou. Stochastic variance reduction for dr-submodular maximization. *Algorithmica*, 86(5):1335–1364, 2024.

Ziwei Liang, Qiang He, Hongwei Du, and Wen Xu. Targeted influence maximization in competitive social networks. *Information Sciences*, 619:390–405, 2023.

Aryan Mokhtari, Hamed Hassani, and Amin Karbasi. Stochastic conditional gradient methods: From convex minimization to submodular maximization. *Journal of machine learning research*, 21(105):1–49, 2020.

Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Max Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. Fair influence maximization: A welfare optimization approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11630–11638, 2021.

Xiaobin Rui, Zhixiao Wang, Jiayu Zhao, Lichao Sun, and Wei Chen. Scalable fair influence maximization. *Advances in Neural Information Processing Systems*, 36:66675–66691, 2023.

Youze Tang, Yanchen Shi, and Xiaokui Xiao. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pp. 1539–1554, 2015.

Ya-Wen Teng, Yishuo Shi, De-Nian Yang, Chih-Hua Tai, Philip S Yu, and Ming-Syan Chen. Multigrade revenue maximization for promotional and competitive viral marketing in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 37(3):1339–1353, 2024.

Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. Group-fairness in influence maximization. In Sarit Kraus (ed.), *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 5997–6005. ijcai.org, 2019.

Liam Walsh, Mengbin Ye, Brian DO Anderson, and Zhiyong Sun. Decentralised adaptive-gain control for eliminating epidemic spreading on networks. *Automatica*, 174:112143, 2025.

Zhixiao Wang, Jiayu Zhao, Chengcheng Sun, Xiaobin Rui, and Philip S Yu. A general concave fairness framework for influence maximization based on poverty reward. *ACM Transactions on Knowledge Discovery from Data*, 19(1):1–23, 2024.

Bryan Wilder, Han-Ching Ou, Kayla de la Haye, and Milind Tambe. Optimizing network structure for preventative health. In *AAMAS*, pp. 841–849, 2018.

Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD workshop on mining data semantics*, pp. 1–8, 2012.

Mingrui Zhang, Zebang Shen, Aryan Mokhtari, Hamed Hassani, and Amin Karbasi. One sample stochastic frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, pp. 4012–4023. PMLR, 2020.

Xiaojing Zhong, Yukun Yang, Feiqi Deng, and Guiyun Liu. Rumor propagation control with anti-rumor mechanism and intermittent control strategies. *IEEE Transactions on Computational Social Systems*, 11(2):2397–2409, 2023.

A On the limits of FIMM

FIMM needs an unbiased estimator of the fair objective, but the plug-in estimator is biased under a concave ϕ, as we discussed in the main paper. FIMM restores unbiasedness only for a specific objective that is the power function u^α: indeed, it rewrites u_C^α as the binomial series of $(1-u_C)^k$ and estimates each integer power unbiasedly. This approach works cleanly because the serial expansion of the power function converges on the whole utility range $[0, 1]$, so truncation is safe. A general ϕ has no such guarantee: one would need a new series per function, and its accuracy would depend on the specific function. E.g., For the $\log$ function we considered here, the expansion would be particularly bad where $u \to 0$, i.e. for the low-utility groups that fairness is meant to prioritize. As a result, the region of greatest importance to our analysis is also the most difficult to estimate accurately. Our approach sidesteps all of this: we never estimate the objective without bias. We drive the gradient bias to zero over iterations via compositional tracking, which works uniformly for any admissible ϕ with no per-function construction.

B Omitted Proofs

B.1 Proof of Lemma 1

Consider any $C \in \mathcal{C}$. The group utility

$$\upsilon_C(\mathbf{x}) = 1 - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1-\mathbf{x}[v])\right]$$

is the multilinear extension of the coverage function $S \mapsto u_C(S)$, hence monotone non-decreasing and DR-submodular on $[0,1]^{|V|}$ Calinescu et al. (2011). In particular, $\frac{\partial \upsilon_C}{\partial \mathbf{x}[i]} \geq 0$ and $\frac{\partial^2 \upsilon_C}{\partial \mathbf{x}[i]\,\partial \mathbf{x}[j]} \leq 0$ for all $i, j \in [|V|]$.

By the chain rule applied to $\Phi(\mathbf{x}) = \sum_{C\in\mathcal{C}} |C| \cdot \phi(\upsilon_C(\mathbf{x}))$,

$$\frac{\partial \Phi}{\partial \mathbf{x}[i]}(\mathbf{x}) = \sum_{C\in\mathcal{C}} |C| \cdot \phi'(\upsilon_C(\mathbf{x})) \cdot \frac{\partial \upsilon_C}{\partial \mathbf{x}[i]}(\mathbf{x}) \geq 0,$$

where the last inequality holds since $\phi' \geq 0$ (ϕ non-decreasing) and $\frac{\partial \upsilon_C}{\partial \mathbf{x}[i]} \geq 0$. This proves monotonicity.

Differentiating once more, for any $i, j \in [|V|]$,

$$\frac{\partial^2 \Phi}{\partial \mathbf{x}[i]\,\partial \mathbf{x}[j]}(\mathbf{x}) = \sum_{C\in\mathcal{C}} |C| \cdot \left[\phi''(\upsilon_C(\mathbf{x})) \cdot \frac{\partial \upsilon_C}{\partial \mathbf{x}[i]}(\mathbf{x}) \cdot \frac{\partial \upsilon_C}{\partial \mathbf{x}[j]}(\mathbf{x}) + \phi'(\upsilon_C(\mathbf{x})) \cdot \frac{\partial^2 \upsilon_C}{\partial \mathbf{x}[i]\,\partial \mathbf{x}[j]}(\mathbf{x})\right].$$

Both terms in the bracket are non-positive: the first because $\phi'' \leq 0$ (ϕ concave) and the two first-order partials of υ_C are non-negative; the second because $\phi' \geq 0$ and $\frac{\partial^2 \upsilon_C}{\partial \mathbf{x}[i]\,\partial \mathbf{x}[j]} \leq 0$. Hence $\frac{\partial^2 \Phi}{\partial \mathbf{x}[i]\,\partial \mathbf{x}[j]} \leq 0$ for all i, j, which gives DR-submodularity. □

B.2 Full Proof of Theorem 1

We first derive separately the bounds for estimation errors of the utility equation 1 (that is used to estimate the marginal influence weight) and of the marginal influence gain equation 2, separately. Because we decouple the tracking of the utility from the one of the marginal influence gain, we can bound their respective errors without introducing nested cross-variance terms. All unspecified expectations along the proof have to be considered with respect to the random choices of the algorithm.

Let us start by bounding the estimation error on the utility. To this aim, we first show the following two properties.

Lemma 2. *For every $t \geq 0$ and every community $C \in \mathcal{C}$, we have that*

$$\mathbb{E}\left[\left|\bar{v}_C^t - v_C(\mathbf{x}_t)\right|^2\right] \leq \frac{1}{\beta_t}.$$

Proof. By using that for each reachable set R we have that $\prod_{v\in R}(1-\mathbf{x}_t[v])^2 \leq \prod_{v\in R}(1-\mathbf{x}_t[v])$ since each factor lies in $[0,1]$, we achieve that for each group C and every $\mathbf{x}$

$$\begin{aligned}\mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\left(1-\prod_{v\in R}(1-\mathbf{x}[v])\right)^2\right] &= 1 - 2\mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1-\mathbf{x}[v])\right] \\ &\quad + \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1-\mathbf{x}[v])^2\right] \\ &\leq 1 - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1-\mathbf{x}[v])\right] = v_C(\mathbf{x}) \leq 1.\end{aligned}$$

Now observe that $\bar{v}_C^t$ is the average of β_t i.i.d. samples each with expectation $v_C(\mathbf{x}_t)$. Hence, we achieve

$$\begin{aligned}&\mathbb{E}\left[\left|\bar{v}_C^t - v_C(\mathbf{x}_t)\right|^2\right] \\ &= \frac{1}{\beta_t^2}\sum_{R\in\mathbf{R}_C^t}\mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\left(\left(1-\prod_{v\in R}(1-\mathbf{x}_t[v])\right) - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[1-\prod_{v\in R}(1-\mathbf{x}_t[v])\right]\right)^2\right] \\ &\leq \frac{1}{\beta_t}\mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\left(1-\prod_{v\in R}(1-\mathbf{x}_t[v])\right)^2\right] \leq \frac{1}{\beta_t}.\end{aligned}$$

□

Lemma 3. *For every $t \geq 1$ and every community $C \in \mathcal{C}$, we have that*

$$\mathbb{E}\left[\left|(v_C(\mathbf{x}_{t-1}) - v_C(\mathbf{x}_t)) - \left(\hat{v}_C^t - \bar{v}_C^t\right)\right|^2\right] \leq \frac{\lambda}{\beta_t}\mathbb{E}\left[\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2\right]$$

Proof. Recall that $\bar{v}_C^t = 1 - \frac{1}{\beta_t}\sum_{R\in\mathbf{R}_C^t}\prod_{v\in R}(1-\mathbf{x}_t[v])$ and $\hat{v}_C^t = 1 - \frac{1}{\beta_t}\sum_{R\in\mathbf{R}_C^t}\prod_{v\in R}(1-\mathbf{x}_{t-1}[v])$, and thus

$$\hat{v}_C^t - \bar{v}_C^t = \frac{1}{\beta_t}\sum_{R\in\mathbf{R}_C^t}\left(\prod_{v\in R}(1-\mathbf{x}_t[v]) - \prod_{v\in R}(1-\mathbf{x}_{t-1}[v])\right),$$

i.e., the average of β_t i.i.d. samples each with expectation exactly $v_C(\mathbf{x}_{t-1}) - v_C(\mathbf{x}_t)$. Let $X_R^t = \prod_{v\in R}(1-\mathbf{x}_t[v]) - \prod_{v\in R}(1-\mathbf{x}_{t-1}[v])$, then,

$$\begin{aligned}&\mathbb{E}\left[\left|(v_C(\mathbf{x}_{t-1}) - v_C(\mathbf{x}_t)) - \left(\hat{v}_C^t - \bar{v}_C^t\right)\right|^2\right] \\ &= \frac{1}{\beta_t^2}\sum_{R\in\mathbf{R}_C^t}\mathbb{E}_{R\sim\mathcal{R}_{\mathcal{H},\mathcal{C}}}\left[\left(X_R^t - \mathbb{E}_{R\sim\mathcal{R}_{\mathcal{H},\mathcal{C}}}\left[X_R^t\right]\right)^2\right] \\ &\leq \frac{1}{\beta_t}\mathbb{E}_{R\sim\mathcal{R}_{\mathcal{H},\mathcal{C}}}\left[\left(X_R^t\right)^2\right].\end{aligned} \tag{3}$$

By applying the telescoping sum property $|\prod a_i - \prod b_i| \leq \sum |a_i - b_i|$ (which holds since all factors $(1-\mathbf{x}_t[v]), (1-\mathbf{x}_{t-1}[v]) \in [0,1]$), followed by the Cauchy-Schwarz inequality, we obtain:

$$\left|X_R^t\right| \leq \sum_{v\in R}|\mathbf{x}_{t-1}[v] - \mathbf{x}_t[v]| \leq \sqrt{|R|}\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|. \tag{4}$$

The claim follows by plugging equation 4 in equation 3 and taking the expectation over $R \sim \mathcal{R}_{H,C}$, using $\mathbb{E}_{R\sim\mathcal{R}_{H,C}}[|R|] = \lambda$. □

Lemma 4 (Utility Error). *For every $t \geq 1$ and every community $C \in \mathcal{C}$, we have that:*

$$\mathbb{E}\left[\left|v_C^t - v_C(\mathbf{x}_t)\right|^2\right] \leq (1-\alpha_t)^2 \mathbb{E}\left[\left|v_C^{t-1} - v_C(\mathbf{x}_{t-1})\right|^2\right] + \frac{2\alpha_t^2}{\beta_t} + \frac{2(1-\alpha_t)^2\lambda}{\beta_t}\mathbb{E}\left[\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2\right].$$

Proof. By the update rule at line 10, we have that:

$$v_C^t - v_C(\mathbf{x}_t) = (1-\alpha_t)(v_C^{t-1} - v_C(\mathbf{x}_{t-1})) + \Delta_t,$$

where

$$\Delta_t = \bar{v}_C^t - v_C(\mathbf{x}_t) + (1-\alpha_t)\left(v_C(\mathbf{x}_{t-1}) - \hat{v}_C^t\right).$$

Recall that $\bar{v}_C^t$ is the estimate of $v_C(\mathbf{x}_t)$ according to samples in $\mathbf{R}_C^t$, and $\hat{v}_C^t$ is the estimate of $v_C(\mathbf{x}_{t-1})$ according the same set of samples. Hence, $\mathbb{E}\left[\bar{v}_C^t\right] = v_C(\mathbf{x}_t)$ and $\mathbb{E}\left[\hat{v}_C^t\right] = v_C(\mathbf{x}_{t-1})$, and thus $\mathbb{E}_{\mathbf{R}_C^t}\left[\Delta_t\right] = 0$. Using the variance property $\mathbb{E}\left[|A+B|^2\right] = \mathbb{E}\left[|A|^2\right] + \mathbb{E}\left[|B|^2\right]$ for any zero-mean independent increment B, we obtain:

$$\mathbb{E}\left[\left|v_C^t - v_C(\mathbf{x}_t)\right|^2\right] = (1-\alpha_t)^2\mathbb{E}\left[\left|v_C^{t-1} - v_C(\mathbf{x}_{t-1})\right|^2\right] + \mathbb{E}\left[\left|\Delta_t\right|^2\right].$$

To bound the second term, we rearrange Δ_t to isolate the variance of the current sample and the variance of the path integration:

$$\begin{aligned}
\mathbb{E}\left[\left|\Delta_t\right|^2\right] &= \mathbb{E}\left[\left|\alpha_t\left(\bar{v}_C^t - v_C(\mathbf{x}_t)\right) + (1-\alpha_t)\left((v_C(\mathbf{x}_{t-1}) - v_C(\mathbf{x}_t)) - \left(\hat{v}_C^t - \bar{v}_C^t\right)\right)\right|^2\right] \\
&\leq 2\alpha_t^2\mathbb{E}\left[\left|\bar{v}_C^t - v_C(\mathbf{x}_t)\right|^2\right] + 2(1-\alpha_t)^2\mathbb{E}\left[\left|(v_C(\mathbf{x}_{t-1}) - v_C(\mathbf{x}_t)) - \left(\hat{v}_C^t - \bar{v}_C^t\right)\right|^2\right] \\
&\leq \frac{2\alpha_t^2}{\beta_t} + \frac{2(1-\alpha_t)^2\lambda}{\beta_t}\mathbb{E}\left[\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2\right],
\end{aligned}$$

where the first inequality follows since $(a+b)^2 \leq 2a^2 + 2b^2$, and the second inequality follows from Lemma 2 and Lemma 3. □

We now consider the error on the marginal influence gain. In what follows, $[y]_0^1 = \min\{1, \max\{0, y\}\}$ denotes the clipping of y to $[0,1]$, and $\phi'(v_C^t)$ stands for $\phi'([v_C^t]_0^1)$ as in line 11. Since $[v_C^t]_0^1 \in [0,1]$, we have $0 \leq \phi'([v_C^t]_0^1) \leq \Omega_\phi$; moreover, since $v_C(\mathbf{x}_t) \in [0,1]$ and clipping is non-expansive, $|[v_C^t]_0^1 - v_C(\mathbf{x}_t)| \leq |v_C^t - v_C(\mathbf{x}_t)|$. As above, before to bound the error in the estimation of the inner gradient we prove some preliminary bounds.

Lemma 5. *For every $t \geq 0$ and every community $C \in \mathcal{C}$, we have that*

$$\mathbb{E}\left[\left\|M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] \leq \frac{r\Omega_\phi^2\lambda}{\beta_t}\sum_{C\in\mathcal{C}}|C|^2.$$

Proof. Let $\delta_t = M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t$. Note that $\mathbb{E}\left[\left\|M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] = \mathbb{E}\left[\sum_{i\in V}\delta_t[i]^2\right] = \sum_{i\in V}\mathbb{E}\left[\delta_t[i]^2\right]$. Moreover, we have that

$$\begin{aligned}
&\delta_t[i] = \\
&\sum_{C\in\mathcal{C}}|C|\phi'(v_C^t)\left(\frac{1}{\beta_t}\sum_{R\in\mathcal{S}_C^t}\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}_t[v]) - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}_t[v])\right]\right) \\
&\leq \Omega_\phi\sum_{C\in\mathcal{C}}|C|\left(\frac{1}{\beta_t}\sum_{R\in\mathcal{S}_C^t}\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}_t[v]) - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}_t[v])\right]\right).
\end{aligned}$$

Here the inequality is to be read in absolute value, i.e., $|\delta_t[i]| \le \Omega_\phi \sum_C |C|\,|\cdot|$, since $0 \le \phi'(\cdot) \le \Omega_\phi$. Since $\left(\sum_{i=1}^r a_i\right)^2 \le r \sum_{i=1}^r a_i^2$, then it follows that

$$
\begin{aligned}
&\mathbb{E}\left[\delta_t[i]^2\right] \\
&\le r\Omega_\phi^2 \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}\left[\left(\frac{1}{\beta_t}\sum_{R\in\mathcal{S}_C^t} \mathbb{1}_{i\in R} \prod_{v\in R\setminus\{i\}} (1-\mathbf{x}_t[v]) - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}_t[v])\right]\right)^2\right] \\
&= \frac{r\Omega_\phi^2}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\left(\mathbb{1}_{i\in R} \prod_{v\in R\setminus\{i\}} (1-\mathbf{x}_t[v]) - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\mathbb{1}_{i\in R}\prod_{v\in R\setminus\{i\}}(1-\mathbf{x}_t[v])\right]\right)^2\right] \\
&\le \frac{r\Omega_\phi^2}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\left(\mathbb{1}_{i\in R} \prod_{v\in R\setminus\{i\}} (1-\mathbf{x}_t[v])\right)^2\right] \\
&\le \frac{r\Omega_\phi^2}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\mathbb{1}_{i\in R}\right].
\end{aligned}
$$

Hence, we have that

$$
\begin{aligned}
\mathbb{E}\left[\left\|M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] &\le \frac{r\Omega_\phi^2}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\sum_{i\in V}\mathbb{1}_{i\in R}\right] \\
&= \frac{r\Omega_\phi^2}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[|R|\right].
\end{aligned}
$$

The lemma then follows since $\mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[|R|\right] \le \lambda$. □

Lemma 6. *For every $t \ge 1$, we have that*

$$
\begin{aligned}
&\mathbb{E}\left[\left\|(M(\mathbf{x}_{t-1})\mathbf{w}_{t-1} - M(\mathbf{x}_t)\mathbf{w}_t) - \left(\hat{M}_{t-1}\mathbf{w}_{t-1} - M_t\mathbf{w}_t\right)\right\|^2\right] \\
&\qquad\le \frac{2r}{\beta_t}\sum_{C\in\mathcal{C}} |C|^2 \left(\Omega_\phi^2\Lambda\,\mathbb{E}\left[\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2\right] + L_\phi^2\lambda\,\mathbb{E}\left[\left|v_C^t - v_C^{t-1}\right|^2\right]\right).
\end{aligned}
$$

Proof. Let $\mathcal{F}_t$ denote the history up to iteration $t-1$ together with the batches $\mathbf{R}_C^t$, so that $\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{w}_t, \mathbf{w}_{t-1}$ are $\mathcal{F}_t$-measurable, while the batches $\mathcal{S}_C^t$ are independent of $\mathcal{F}_t$. Conditional on $\mathcal{F}_t$, $\hat{M}_{t-1}\mathbf{w}_{t-1} - M_t\mathbf{w}_t$ is the average of β_t i.i.d. samples, each with expectation $M(\mathbf{x}_{t-1})\mathbf{w}_{t-1} - M(\mathbf{x}_t)\mathbf{w}_t$. Hence

$$
\mathbb{E}\left[\left\|(M(\mathbf{x}_{t-1})\mathbf{w}_{t-1} - M(\mathbf{x}_t)\mathbf{w}_t) - \left(\hat{M}_{t-1}\mathbf{w}_{t-1} - M_t\mathbf{w}_t\right)\right\|^2 \,\middle|\, \mathcal{F}_t\right] \le \frac{1}{\beta_t}\sum_{i\in V}\mathbb{E}\left[X_R[i]^2 \mid \mathcal{F}_t\right],
$$

where $R = (R_C)_{C\in\mathcal{C}}$ with independent $R_C \sim \mathcal{R}_{H,C}$ and, denoting $\pi_i^s(R_C) = \mathbb{1}_{i\in R_C}\prod_{v\in R_C\setminus\{i\}}(1-\mathbf{x}_s[v])$,

$$
X_R[i] = \sum_{C\in\mathcal{C}} |C|\left(\phi'(v_C^{t-1})\,\pi_i^{t-1}(R_C) - \phi'(v_C^t)\,\pi_i^t(R_C)\right) = Y_R[i] + Z_R[i],
$$

with $Y_R[i] = \sum_C |C|\,\phi'(v_C^t)\left(\pi_i^{t-1}(R_C) - \pi_i^t(R_C)\right)$ and $Z_R[i] = \sum_C |C|\left(\phi'(v_C^{t-1}) - \phi'(v_C^t)\right)\pi_i^{t-1}(R_C)$. Since $X_R[i]^2 \le 2Y_R[i]^2 + 2Z_R[i]^2$, we bound the two terms separately.

For Y_R, since $0 \le \phi'(v_C^t) \le \Omega_\phi$ and $(\sum_C a_C)^2 \le r\sum_C a_C^2$,

$$
Y_R[i]^2 \le r\Omega_\phi^2 \sum_{C\in\mathcal{C}} |C|^2\,\mathbb{1}_{i\in R_C}\left(\sum_{v\in R_C\setminus\{i\}} |\mathbf{x}_t[v] - \mathbf{x}_{t-1}[v]|\right)^2 \le r\Omega_\phi^2 \sum_{C\in\mathcal{C}} |C|^2\,\mathbb{1}_{i\in R_C}|R_C|\,\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2,
$$

where we used the telescoping product bound (valid since all factors are in $[0,1]$) and the Cauchy-Schwarz inequality. Summing over i and using $\sum_i \mathbb{1}_{i\in R_C} = |R_C|$ and $\mathbb{E}\left[|R_C|^2\right] \le \Lambda$, we get $\sum_i \mathbb{E}\left[Y_R[i]^2 \mid \mathcal{F}_t\right] \le r\Omega_\phi^2 \Lambda \sum_C |C|^2 \left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2$.

For Z_R, since clipping is non-expansive, $\left|\phi'(v_C^{t-1}) - \phi'(v_C^t)\right| \le L_\phi \left|v_C^{t-1} - v_C^t\right|$, and $0 \le \pi_i^{t-1}(R_C) \le \mathbb{1}_{i\in R_C}$. Hence $Z_R[i]^2 \le rL_\phi^2 \sum_C |C|^2 \mathbb{1}_{i\in R_C} \left|v_C^{t-1} - v_C^t\right|^2$, and summing over i and using $\mathbb{E}\left[|R_C|\right] \le \lambda$, we get $\sum_i \mathbb{E}\left[Z_R[i]^2 \mid \mathcal{F}_t\right] \le rL_\phi^2 \lambda \sum_C |C|^2 \left|v_C^t - v_C^{t-1}\right|^2$.

The claim follows by combining the two bounds and taking the expectation over $\mathcal{F}_t$. □

Lemma 7 (Marginal Influence Gain Error)**.** *For every $t \ge 1$, we have that:*

$$\mathbb{E}\left[\left\|\mathbf{v}^t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] \le (1-\alpha_t)^2 \mathbb{E}\left[\left\|\mathbf{v}^{t-1} - M(\mathbf{x}_{t-1})\mathbf{w}_{t-1}\right\|^2\right] + \frac{2\alpha_t^2 r\Omega_\phi^2 \lambda}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 + \frac{4(1-\alpha_t)^2 r}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \left(\Omega_\phi^2 \Lambda\, \mathbb{E}\left[\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2\right] + L_\phi^2 \lambda\, \mathbb{E}\left[\left|v_C^t - v_C^{t-1}\right|^2\right]\right).$$

Proof. By the update rule at line 18, we have that:

$$\mathbf{v}^t - M(\mathbf{x}_t)\mathbf{w}_t = (1-\alpha_t)(\mathbf{v}^{t-1} - M(\mathbf{x}_{t-1})\mathbf{w}_{t-1}) + \Delta_t,$$

where

$$\Delta_t = M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t + (1-\alpha_t)\left(M(\mathbf{x}_{t-1})\mathbf{w}_{t-1} - \hat{M}_{t-1}\mathbf{w}_{t-1}\right).$$

Let $\mathcal{F}_t$ denote the history up to iteration $t-1$ together with the batches $\mathbf{R}_C^t$. Then $\mathbf{x}_t$, $\mathbf{x}_{t-1}$, $\mathbf{w}_t$, $\mathbf{w}_{t-1}$ and $\mathbf{v}_{t-1}$ are $\mathcal{F}_t$-measurable, while M_t and $\hat{M}_{t-1}$ are computed on the batches $\mathcal{S}_C^t$, which are independent of $\mathcal{F}_t$. Hence, $\mathbb{E}\left[M_t\mathbf{w}_t \mid \mathcal{F}_t\right] = M(\mathbf{x}_t)\mathbf{w}_t$ and $\mathbb{E}\left[\hat{M}_{t-1}\mathbf{w}_{t-1} \mid \mathcal{F}_t\right] = M(\mathbf{x}_{t-1})\mathbf{w}_{t-1}$, and thus $\mathbb{E}\left[\Delta_t \mid \mathcal{F}_t\right] = 0$. Using the variance property $\mathbb{E}\left[\|A+B\|^2\right] = \mathbb{E}\left[\|A\|^2\right] + \mathbb{E}\left[\|B\|^2\right]$ for any zero-mean independent increment B, we obtain:

$$\mathbb{E}\left[\left\|\mathbf{v}^t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] = (1-\alpha_t)^2 \mathbb{E}\left[\left\|\mathbf{v}^{t-1} - M(\mathbf{x}_{t-1})\mathbf{w}_{t-1}\right\|^2\right] + \mathbb{E}\left[\left\|\Delta_t\right\|^2\right].$$

To bound the second term, we rearrange Δ_t to isolate the variance of the current sample and the variance of the path integration:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\Delta_t\right\|^2\right] = \\
&\mathbb{E}\left[\left\|\alpha_t\left(M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t\right) + (1-\alpha_t)\left(\left(M(\mathbf{x}_{t-1})\mathbf{w}_{t-1} - M(\mathbf{x}_t)\mathbf{w}_t\right) - \left(\hat{M}_{t-1}\mathbf{w}_{t-1} - M_t\mathbf{w}_t\right)\right)\right\|^2\right] \\
&\le 2\alpha_t^2 \mathbb{E}\left[\left\|M_t\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] \\
&\quad + 2(1-\alpha_t)^2 \mathbb{E}\left[\left\|\left(M(\mathbf{x}_{t-1})\mathbf{w}_{t-1} - M(\mathbf{x}_t)\mathbf{w}_t\right) - \left(\hat{M}_{t-1}\mathbf{w}_{t-1} - M_t\mathbf{w}_t\right)\right\|^2\right] \\
&\le \frac{2\alpha_t^2 r\Omega_\phi^2\lambda}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 + \frac{4(1-\alpha_t)^2 r}{\beta_t} \sum_{C\in\mathcal{C}} |C|^2 \left(\Omega_\phi^2\Lambda\, \mathbb{E}\left[\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2\right] + L_\phi^2\lambda\, \mathbb{E}\left[\left|v_C^t - v_C^{t-1}\right|^2\right]\right),
\end{aligned}$$

where the first inequality follows since $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$, and the second inequality follows from Lemma 5 and Lemma 6. □

We next compose errors.

Lemma 8 (Compositional Gradient Error)**.** *For every $t \ge 0$, we have that*

$$\mathbb{E}\left[\left\|\mathbf{v}_t - \nabla\Phi(\mathbf{x}_t)\right\|^2\right] \le 2\mathbb{E}\left[\left\|\mathbf{v}^t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] + 2r\lambda L_\phi^2 \sum_{C\in\mathcal{C}} |C|^2 \mathbb{E}\left[\left|v_C^t - v_C(\mathbf{x}_t)\right|^2\right].$$

Proof. From the definitions of $\mathbf{v}_t$ and $\nabla\Phi(\mathbf{x}_t)$, we have:

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{v}_t - \nabla\Phi(\mathbf{x}_t)\right\|^2\right] &= \mathbb{E}\left[\left\|\mathbf{v}_t - M(\mathbf{x}_t)\mathbf{w}(\mathbf{x}_t)\right\|^2\right] \\ &= \mathbb{E}\left[\left\|(\mathbf{v}_t - M(\mathbf{x}_t)\mathbf{w}_t) + (M(\mathbf{x}_t)\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}(\mathbf{x}_t))\right\|^2\right] \\ &\leq 2\mathbb{E}\left[\left\|\mathbf{v}_t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] + 2\mathbb{E}\left[\left\|M(\mathbf{x}_t)\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}(\mathbf{x}_t)\right\|^2\right].\end{aligned}$$

Now observe that

$$\begin{aligned}\mathbb{E}\left[\left\|M(\mathbf{x}_t)\mathbf{w}_t - M(\mathbf{x}_t)\mathbf{w}(\mathbf{x}_t)\right\|^2\right] &= \mathbb{E}\left[\left\|\sum_{C\in\mathcal{C}} |C|\nabla v_C(\mathbf{x}_t)\left(\phi'(v_C^t) - \phi'(v_C(\mathbf{x}_t))\right)\right\|^2\right] \\ &\leq r\sum_{C\in\mathcal{C}} |C|^2\mathbb{E}\left[\left\|\nabla v_C(\mathbf{x}_t)\right\|^2\left|\phi'(v_C^t) - \phi'(v_C(\mathbf{x}_t))\right|^2\right].\end{aligned} \tag{5}$$

Observe that $\left\|\nabla v_C(\mathbf{x}_t)\right\|^2 \leq \lambda$. Moreover $|\phi'(v_C^t) - \phi'(v_C(\mathbf{x}_t))| \leq L_\phi\left|v_C^t - v_C(\mathbf{x}_t)\right|$ by definition of L_ϕ. By plugging these values in equation 5, the claim follows. □

To prove that our algorithm successfully achieves variance reduction despite the non-linear fairness composition, we must bound the global gradient error over time.

Lemma 9 (Variance Convergence of the Composed Gradient Estimator)**.** *Suppose that Algorithm 1 is run with $\alpha_t = 1/t$. Then, for $t \geq 0$, the expected squared error of the composed gradient is bounded by:*

$$\mathbb{E}\left[\left\|\mathbf{v}_t - \nabla\Phi(\mathbf{x}_t)\right\|^2\right] \leq \frac{Q}{t+1},$$

where $Q_v = \frac{8\lambda B}{\beta_t}$, $Q_M = \frac{B\left(12\Lambda\Omega_\phi^2 + 496\lambda^2 L_\phi^2\right)}{\beta_t}$, and $Q = 2r\sum_{C\in\mathcal{C}}|C|^2\left(Q_M + L_\phi^2\lambda Q_v\right)$.

Proof. First of all observe that $\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|^2 = \frac{1}{T^2}\left\|\mathbf{z}_{t-1}\right\|^2 = \frac{B}{T^2}$, since $\mathbf{z}_{t-1}$ has exactly B entries equal to 1. Moreover, $\lambda \geq 1$ (every RR set contains its root), $\lambda \leq \Lambda$, and $B \geq 1$, so that $\lambda B \geq 1$.

Plugging these bounds in Lemma 4 and using that $1 - \alpha_t \leq 1$, we achieve that for every $t \geq 1$ and every $C \in \mathcal{C}$,

$$\begin{aligned}\mathbb{E}\left[\left|v_C^t - v_C(\mathbf{x}_t)\right|^2\right] &\leq \left(1 - \frac{1}{t}\right)^2\mathbb{E}\left[\left|v_C^{t-1} - v_C(\mathbf{x}_{t-1})\right|^2\right] + \frac{2}{\beta_t t^2} + \frac{2\lambda B}{\beta_t T^2} \\ &\leq \left(1 - \frac{1}{t}\right)^2\mathbb{E}\left[\left|v_C^{t-1} - v_C(\mathbf{x}_{t-1})\right|^2\right] + \frac{Q_v}{2t^2},\end{aligned}$$

where we used $T \geq t$ and $\lambda B \geq 1$. We prove by induction that $\mathbb{E}\left[\left|v_C^t - v_C(\mathbf{x}_t)\right|^2\right] \leq \frac{Q_v}{t+1}$. For $t = 0$, $v_C^0 = \bar{v}_C^0$ and the error is at most $\frac{1}{\beta_t} \leq Q_v$. For $t = 1$, since $\alpha_1 = 1$, the error is at most $\frac{Q_v}{2}$. For $t \geq 2$, assuming the bound for $t - 1$,

$$\mathbb{E}\left[\left|v_C^t - v_C(\mathbf{x}_t)\right|^2\right] \leq \frac{(t-1)^2}{t^2}\frac{Q_v}{t} + \frac{Q_v}{2t^2} = \frac{Q_v}{t^2}\left[t - \frac{3}{2} + \frac{1}{t}\right] \leq \frac{(t-1)Q_v}{t^2} \leq \frac{Q_v}{t+1},$$

where we used $t - \frac{3}{2} + 1/t \leq t - 1$ for $t \geq 2$, and $(t-1)/t^2 \leq 1/(t+1)$.

By line 10, $v_C^t - v_C^{t-1} = \alpha_t\left(\bar{v}_C^t - v_C^{t-1}\right) + (1 - \alpha_t)\left(\bar{v}_C^t - \hat{v}_C^{t-1}\right)$. For every RR set R, $\left|\prod_{v\in R}(1 - \mathbf{x}_t[v]) - \prod_{v\in R}(1 - \mathbf{x}_{t-1}[v])\right| \leq \sum_{v\in R}\left|\mathbf{x}_t[v] - \mathbf{x}_{t-1}[v]\right| \leq \sqrt{|R|}\left\|\mathbf{x}_t - \mathbf{x}_{t-1}\right\|$; hence, by Jensen's inequality, $\mathbb{E}\left[\left|\bar{v}_C^t - \hat{v}_C^{t-1}\right|^2\right] \leq \lambda\frac{B}{T^2}$ and $\left|v_C(\mathbf{x}_t) - v_C(\mathbf{x}_{t-1})\right|^2 \leq \lambda\frac{B}{T^2}$. Moreover, by Lemma 2, the bound above, and $\beta_t \geq 1$,

$$\mathbb{E}\left[\left|\bar{v}_C^t - v_C^{t-1}\right|^2\right] \leq 3\left(\frac{1}{\beta_t} + \frac{\lambda B}{T^2} + \frac{Q_v}{t}\right) \leq 30\lambda B.$$

Therefore, since $\alpha_t = 1/t$ and $T \geq t$, $\mathbb{E}\left[\left|v_C^t - v_C^{t-1}\right|^2\right] \leq \frac{60\lambda B}{t^2} + \frac{2\lambda B}{t^2} = \frac{62\lambda B}{t^2}$.

Plugging these bounds in Lemma 7, and using $\lambda \leq \Lambda$ and $B \geq 1$, we get for every $t \geq 1$

$$\mathbb{E}\left[\left\|\mathbf{v}^t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] \leq \left(1 - \frac{1}{t}\right)^2 \mathbb{E}\left[\left\|\mathbf{v}^{t-1} - M(\mathbf{x}_{t-1})\mathbf{w}_{t-1}\right\|^2\right] + \frac{r\sum_C |C|^2 B\left(6\Omega_\phi^2\Lambda + 248L_\phi^2\lambda^2\right)}{\beta_t t^2},$$

where the last term equals $\frac{r\sum_C |C|^2 Q_M}{2t^2}$. For $t = 0$, by Lemma 5, the error is at most $\frac{r\Omega_\phi^2\lambda}{\beta_t}\sum_C |C|^2 \leq r\sum_C |C|^2 Q_M$. Hence, the same inductive argument as above yields $\mathbb{E}\left[\left\|\mathbf{v}^t - M(\mathbf{x}_t)\mathbf{w}_t\right\|^2\right] \leq \frac{r\sum_C |C|^2 Q_M}{t+1}$ for every $t \geq 0$.

Composed error. Finally, applying Lemma 8:

$$\mathbb{E}\left[\left\|\mathbf{v}_t - \nabla\Phi(\mathbf{x}_t)\right\|^2\right] \leq 2r\sum_{C\in\mathcal{C}} |C|^2 \left(\frac{Q_M}{t+1} + \lambda L_\phi^2 \frac{Q_v}{t+1}\right) = \frac{Q}{t+1}.$$

□

We now use the convergence of the gradient to estimate the approximation guarantee of FIM-SPACE.

Proposition 1 (Approximation Guarantee of FIM-SPACE)**.** *Suppose that Algorithm 1 is run with* $T = \max\left\{\frac{64BQ}{\varepsilon^2}, \frac{LB}{\varepsilon}\right\}$ *and* $\alpha_t = \frac{1}{t}$*, where* $L = \sqrt{2r\sum_{C\in\mathcal{C}} |C|^2 \left(\Omega_\phi^2\Lambda + L_\phi^2\lambda^2\right)}$*. Then*

$$\mathbb{E}\left[\Phi(\mathbf{x}_T)\right] \geq \left(1 - \frac{1}{e}\right)\Phi(\mathbf{x}^*) - \varepsilon,$$

where $\mathbf{x}^* \in \arg\max_{\mathbf{x}} \Phi(\mathbf{x})$ *and* Q *is as defined in Lemma 9.*

Proof. We analyze the convergence trajectory using a potential function framework. Specifically, let $p_t = t/T$ for $t \in \{0, \ldots, T\}$. We then define the potential function $\mathcal{E}$ as follows:

$$\mathcal{E}(p_t) = e^{p_t}\left(\mathbb{E}\left[\Phi(\mathbf{x}_t)\right] - \Phi(\mathbf{x}^*)\right).$$

Let L be such that for every $\mathbf{x}, \mathbf{x}'$ it holds that $\|\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{x}')\|^2 \leq L^2\|\mathbf{x} - \mathbf{x}'\|^2$. Then, by (Lian et al., 2024, Lemma 3.3), we have that

$$\mathcal{E}(p_{t+1}) - \mathcal{E}(p_t) \geq -\frac{LBe^{p_{t+1}}}{2T^2} - e^{p_{t+1}}\left(1 - e^{-1/T}\right)\mathbb{E}\left[\langle\nabla\Phi(\mathbf{x}_t) - \mathbf{v}_t, \mathbf{x}^* - \mathbf{z}_t\rangle\right].$$

Here we use that $\langle\mathbf{v}_t, \mathbf{z}_t\rangle \geq \langle\mathbf{v}_t, \mathbf{x}^*\rangle$: since Φ is monotone, we can assume without loss of generality that $\sum_i \mathbf{x}^*[i] = B$, so $\mathbf{x}^*$ is feasible for the maximization at line 19. By the Cauchy-Schwarz inequality, and noting that $\|\mathbf{x}^* - \mathbf{z}_t\|^2 \leq 4B$ and $1 - e^{-1/T} \leq 1/T$:

$$\mathcal{E}(p_{t+1}) - \mathcal{E}(p_t) \geq -e^{p_{t+1}}\left(\frac{LB}{2T^2} + \frac{2\sqrt{B}}{T}\mathbb{E}\left[\|\nabla\Phi(\mathbf{x}_t) - \mathbf{v}_t\|\right]\right).$$

From Lemma 9, we have that $\mathbb{E}\left[\|\nabla\Phi(\mathbf{x}_t) - \mathbf{v}_t\|\right] \leq \sqrt{Q/(t+1)}$, and thus

$$\mathcal{E}(p_{t+1}) - \mathcal{E}(p_t) \geq -e^{p_{t+1}}\left(\frac{LB}{2T^2} + \frac{2\sqrt{BQ}}{T\sqrt{t+1}}\right).$$

Telescoping over $t = 0$ to $T - 1$ and bounding $e^{p_{t+1}} \leq e$:

$$\mathcal{E}(p_T) - \mathcal{E}(p_0) \geq -e\left(\frac{LB}{2T} + \frac{2\sqrt{BQ}}{T}\sum_{t=0}^{T-1}\frac{1}{\sqrt{t+1}}\right). \tag{6}$$

Using $\sum_{j=1}^{T} j^{-1/2} \leq 2\sqrt{T}$, the discrete sum is bounded by $2\sqrt{T}$. Expanding the potential function, we have that

$$e\left(\mathbb{E}\left[\Phi(\mathbf{x}_T)\right] - \Phi(\mathbf{x}^*)\right) - \left(\mathbb{E}\left[\Phi(\mathbf{x}_0)\right] - \Phi(\mathbf{x}^*)\right) \geq -e\left(\frac{LB}{2T} + \frac{4\sqrt{BQ}}{\sqrt{T}}\right).$$

By dividing by e, observing that $\mathbb{E}[\Phi(\mathbf{x}_0)] \geq 0$, and rearranging, we achieve that:

$$\mathbb{E}\left[\Phi(\mathbf{x}_T)\right] \geq \left(1 - \frac{1}{e}\right)\Phi(\mathbf{x}^*) - \frac{LB}{2T} - \frac{4\sqrt{BQ}}{\sqrt{T}}.$$

This is well defined since Φ is L-smooth with L as in the statement: indeed, $\nabla\Phi(\mathbf{x}) = \sum_C |C|\,\phi'(v_C(\mathbf{x}))\,\nabla v_C(\mathbf{x})$, and $\|\nabla v_C(\mathbf{x})\|^2 \leq \lambda$, $|v_C(\mathbf{x}) - v_C(\mathbf{x}')|^2 \leq \lambda\|\mathbf{x} - \mathbf{x}'\|^2$, and $\|\nabla v_C(\mathbf{x}) - \nabla v_C(\mathbf{x}')\|^2 \leq \Lambda\|\mathbf{x} - \mathbf{x}'\|^2$ (as in the proof of Lemma 6), so that $\|\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{x}')\|^2 \leq 2r\sum_C |C|^2\left(L_\phi^2\lambda^2 + \Omega_\phi^2\Lambda\right)\|\mathbf{x} - \mathbf{x}'\|^2$. Finally, since $T \geq \frac{64BQ}{\varepsilon^2}$, we have $\frac{4\sqrt{BQ}}{\sqrt{T}} \leq \frac{\varepsilon}{2}$, and since $T \geq \frac{LB}{\varepsilon}$, we have $\frac{LB}{2T} \leq \frac{\varepsilon}{2}$, from which the claim follows. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. It follows from Proposition 1, that if Algorithm 1 is run for $O\left(\frac{BQ}{\varepsilon^2} + \frac{LB}{\varepsilon}\right)$ steps, then it returns an outcome with the desired approximation.

Anyway, to estimate the total running time of FIM-SPACE, we need to evaluate the complexity of each iteration. To this aim observe that at each time step t we generate $2\beta_t \cdot r$ reverse reachable sets, with each reverse reachable set requiring time $O(|R|)$ to be generated. Hence, the expected time required for this operation is $O(\beta_t r\lambda)$. Moreover, it is immediate to check that each remaining operation requires at most this same cost except for at most operation at line 19. Anyway, it is not hard to check that this operation can be performed by choosing the top B entries according to $\mathbf{v}_t$, requiring in this way at most time $O(n\log n)$.

Hence, the total time complexity is $O\left(\frac{1}{\varepsilon^2}\max\{\beta_t r\lambda, n\log n\} \cdot \frac{rB^2\sum_{C\in\mathcal{C}}|C|^2}{\beta_t}\left(\Lambda\Omega_\phi^2 + \lambda^2 L_\phi^2\right)\right)$. In particular, by choosing β_t sufficiently large, one may have that $\max\{\beta_t r\lambda, n\log n\} = \beta_t r\lambda$, and thus the total time complexity is $O\left(\frac{1}{\varepsilon^2}r^2\lambda B^2\sum_{C\in\mathcal{C}}|C|^2\left(\Lambda\Omega_\phi^2 + \lambda^2 L_\phi^2\right)\right)$ (the additional term $O\left(\frac{LB}{\varepsilon}\beta_t r\lambda\right)$ due to $\frac{LB}{\varepsilon}$ is of lower order in $\frac{1}{\varepsilon}$, and thus dominated for ε small enough), as desired. □

Sanity check: standard IM. For $\phi(u) = u$ we have $\Omega_\phi = 1$ and $L_\phi = 0$, so the weights $\mathbf{w}_t$ are constant, the utility trackers play no role, and $\Phi(\mathbf{x}) = \sum_C |C|\,v_C(\mathbf{x})$ is the multilinear extension of the (unfair) influence function. In this case the gradient estimator of FIM-SPACE coincides with the variance-reduced estimator of Hybrid SPIDER-CG (Lian et al., 2024), and Theorem 1 recovers, up to constant factors, their $(1 - 1/e)$ guarantee for this objective. Hence, fairness affects our bound only through Ω_ϕ (the steepness of ϕ) and L_ϕ (its curvature), while the dependence on the network, through λ and Λ, is the same as for standard IM.

B.3 EX-ANTE GUARANTEE FOR THE ROUNDED SEED SET

To obtain a seed set of exactly B nodes, we apply randomized pipage rounding (Karimi et al., 2017) to the output $\mathbf{x}^T$ of FIM-SPACE. This rounding does not preserve Φ in expectation, but it preserves the utility of every group, which yields the following guarantee.

Lemma 10 (Ex-ante guarantee)**.** *Let S be obtained by randomized pipage rounding of the output $\mathbf{x}^T$ of* FIM-SPACE*, run as in Theorem 1. Then $|S| = B$ and*

$$\sum_{C\in\mathcal{C}} |C|\,\phi\big(\mathbb{E}\left[u_C(S)\right]\big) \;\geq\; \mathbb{E}\left[\Phi(\mathbf{x}^T)\right] \;\geq\; \left(1 - \tfrac{1}{e}\right) f(S^*) - \varepsilon,$$

where expectations are over the randomness of both FIM-SPACE *and the rounding.*

Proof. Since each $\mathbf{z}_t$ has exactly B entries equal to 1, $\sum_i \mathbf{x}^T[i] = B$. Fix a group C. If S contains each node v independently with probability $\mathbf{x}[v]$, then $\mathbb{E}_S\left[u_C(S)\right] = 1 - \mathbb{E}_{R\sim\mathcal{R}_{H,C}}\left[\prod_{v\in R}(1 - \mathbf{x}[v])\right] = v_C(\mathbf{x})$, i.e., v_C is the multilinear extension of u_C, and $v_C(\mathbf{x}_S) = u_C(S)$. Each step of randomized pipage rounding (Algorithm 3 of Karimi et al. (2017)) changes only two fractional coordinates i, j and preserves their sum, moving the current point $\mathbf{y}$ to $\mathbf{y} + t(\mathbf{e}_i - \mathbf{e}_j)$,

where t takes one of two values with probabilities such that $\mathbb{E}[t] = 0$. Let $h(t) = \upsilon_C(\mathbf{y}+t(\mathbf{e}_i-\mathbf{e}_j))$. Since υ_C is multilinear, $\frac{\partial^2 \upsilon_C}{\partial \mathbf{x}[i]^2} = \frac{\partial^2 \upsilon_C}{\partial \mathbf{x}[j]^2} = 0$, and hence

$$h''(t) = -2\,\frac{\partial^2 \upsilon_C}{\partial \mathbf{x}[i]\,\partial \mathbf{x}[j]} = 2\,\mathbb{E}_R\left[\mathbb{1}_{\{i,j\in R\}} \prod_{v\in R\setminus\{i,j\}} (1-\mathbf{y}[v])\right] \geq 0.$$

Thus h is convex and, by Jensen's inequality, $\mathbb{E}[h(t)] \geq h(0) = \upsilon_C(\mathbf{y})$. The rounding steps do not depend on C, preserve $\sum_i \mathbf{y}[i]$, and terminate after at most n steps at an integral point $\mathbf{x}_S$. Therefore $|S| = B$ and $\mathbb{E}\left[u_C(S) \mid \mathbf{x}^T\right] \geq \upsilon_C(\mathbf{x}^T)$ for every $C \in \mathcal{C}$ (see also Theorem 7 of Karimi et al. (2017)). Taking expectations also over FIM-SPACE, and using that ϕ is non-decreasing and concave,

$$\phi\big(\mathbb{E}\left[u_C(S)\right]\big) \geq \phi\big(\mathbb{E}\left[\upsilon_C(\mathbf{x}^T)\right]\big) \geq \mathbb{E}\left[\phi(\upsilon_C(\mathbf{x}^T))\right].$$

Multiplying by $|C|$ and summing over $C \in \mathcal{C}$ gives the first inequality. For the second, $\mathbf{x}_{S^*}$ is feasible and $\Phi(\mathbf{x}_{S^*}) = f(S^*)$, so $\Phi(\mathbf{x}^*) \geq f(S^*)$, and the claim follows from Theorem 1. □

By Lemma 10 we have that, before the seeds are drawn, every group is reached in expectation at least as much as promised by the fractional solution, and the fairness of these expected reaches is within a $(1 - 1/e)$ factor of the optimal seed set, up to ε. Notice that this is the same ratio as GREEDY-CFF, against the same benchmark. Clearly, this is only an *ex-ante* guarantee, and it does not bound the expected fairness $\mathbb{E}[f(S)]$ of the returned seed set, which, by concavity of ϕ, can only be smaller. For standard IM the two values coincide, since the objective is linear in the group utilities, and rounding is lossless (Karimi et al., 2017); the concave fairness function is precisely what separates them.

C EXTENDED EXPERIMENTAL RESULTS

Dataset descriptions. The SNAP datasets are: (i) *email*[3], where each node is a department member and a directed edge (u, v) means u sent v at least one email; (ii) *amazon*[4], where nodes are products and an undirected edge (u, v) means the two products were frequently co-purchased; (iii) *youtube*[5], where nodes are users and edges encode friendship; and (iv) *DBLP*[6], where nodes are authors and edges encode co-authorship. Additional statistics for the AV networks[7] are also reported in this appendix.

Table 3: Statistics of the SNAP networks after preprocessing (communities with fewer than 10 nodes removed; nodes belonging to no surviving community discarded).

Dataset	$\lvert V\rvert$	$\lvert E\rvert$	#Comm.	Mean size	Median size	Max size
Email	916	21,510	26	35.2	27.5	109
Amazon	12,698	40,096	642	19.8	15.0	223
YouTube	30,696	198,867	1,130	27.2	11.0	2,133
DBLP	72,875	268,346	1,353	53.9	13.0	6,030

These networks also highlight on realistic instances the parameters driving the complexity are usually small. Indeed, under the weighted IC model used in our SNAP experiments, no single seed dominates the cascade, so RR sets stay small. In particular, we measured λ and Λ directly on all four SNAP graphs: As n grows, λ stays in single digits and λ/n falls by two orders of magnitude, while Λ stays in the tens-to-hundreds rather than growing as n^2. The worst-case values $\lambda = \Theta(n)$ and $\Lambda = \Theta(n^2)$ thus overshoot the measured ones by several orders of magnitude.

[3] https://snap.stanford.edu/data/email-Eu-core.html

[4] https://snap.stanford.edu/data/com-Amazon.html

[5] https://snap.stanford.edu/data/com-Youtube.html

[6] https://snap.stanford.edu/data/com-DBLP.html

[7] https://github.com/bwilder0/fair_influmax_code_release/tree/master/networks

Table 4: Statistics of the Antelope Valley (AV) networks. All 24 networks contain exactly 500 nodes and a mean of 1,642.6 edges. Group sizes are reported as percentage of nodes, averaged over the 24 networks (mean $\pm$ standard deviation).

Attribute	Group	Size (%)
Ethnicity	Asian	3.53 ± 0.86
	Black	12.89 ± 1.32
	Latino	29.72 ± 2.26
	Other	4.16 ± 0.94
	White	49.70 ± 2.57
Age	18–24	14.46 ± 1.73
	25–29	9.41 ± 1.08
	30–39	21.95 ± 2.44
	40–49	26.03 ± 1.42
	50–59	11.75 ± 1.33
	60–64	6.05 ± 0.83
	65+	10.36 ± 1.26
Gender	Female	48.71 ± 2.17
	Male	51.29 ± 2.17

Table 5: Values of λ and Λ for the SNAP networks under the weighted IC model used in our experiments.

Dataset	n	λ	Λ
Email	916	8.0	264
Amazon	12,698	2.5	9.7
YouTube	30,696	6.4	99
DBLP	72,875	4.0	33

Baseline Implementation Details.

- FIM-SPACE: number of gradient iterations $T = 500$ and batch size $\beta = 100$. The same parameters are used for both AV and SNAP experiments. After optimization, the fractional solution is rounded to a discrete seed set via Randomized Pipage Rounding (Karimi et al., 2017).
- GREEDY-CFF: the greedy algorithm runs for exactly B steps, selecting at each step the node with the largest marginal gain in $\Phi_\varphi(\cdot)$ estimated via Monte Carlo simulation. We use 100 Monte Carlo samples per step for AV networks and 50 for SNAP networks; the lower value for SNAP is enforced by timing tractability, following the settings of Wang et al. (2024).
- FIMM: approximation parameter $\varepsilon = 0.5$, confidence parameter $\ell = 1$, Taylor expansion parameter $Q = 3$. The number of RR sets per community θ_c is computed according to the theoretical formula of Rui et al. (2023). FIMM is restricted to the power fairness function $\phi(u) = u^\alpha$ with $\alpha = 0.1$, following the default of Rui et al. (2023).
- GREEDY-IM and IMM: GREEDY-IM behaves analogously to GREEDY-CFF, selecting seeds by largest marginal gain in $\sigma(\cdot)$ via Monte Carlo simulation with 100 samples per step. On SNAP networks where GREEDY-IM is intractable, we substitute IMM (Tang et al., 2015) with $\varepsilon = 0.5$ and $\ell = 1$ as the standard IM reference for computing PoF and RoF, applied uniformly across all datasets for consistency.

Preprocessing details. For the SNAP networks, undirected edgelists are loaded as directed graphs preserving the original edge orientation from the file; each undirected edge therefore contributes a single directed arc. Influence probabilities follow the weighted IC model, $p(u, v) = 1/d_{\text{in}}(v)$, for all four SNAP datasets. Communities are drawn from the top-5000 ground-truth community files provided by SNAP; communities with fewer than 10 nodes are then removed following Rui et al. (2023), and nodes belonging to no surviving community are discarded. Each node in the

Table 6: Per-attribute breakdown of solution quality on AV networks (mean ± SD, n=24 graphs). Significance markers on FIM-SPACE rows only: * significant vs. GREEDY-CFF; † significant vs. FIMM. Tests use a two-sided Wilcoxon signed-rank test with Holm–Bonferroni correction within each (metric, comparison) family across the 9 (φ, attribute) cells (3 cells for FIMM, power-only). Absence of a marker indicates no significant difference.

φ	Attribute	Algorithm	Φ_φ	σ	PoF	RoF	F_2	Time (s)
φ_{pow}	Ethnicity	FIM-SPACE	$399.99 \pm 0.79^\dagger$	$53.83 \pm 1.06^\dagger$	$0.005 \pm 0.008^\dagger$	$0.640 \pm 0.244^\dagger$	$0.891 \pm 0.081^\dagger$	$3.01 \pm 0.30^{*\dagger}$
		GREEDY-CFF	399.87 ± 0.83	53.68 ± 1.10	0.008 ± 0.007	0.654 ± 0.237	0.896 ± 0.076	214.35 ± 31.46
		FIMM	397.06 ± 1.15	51.21 ± 1.11	0.054 ± 0.016	0.130 ± 0.282	0.696 ± 0.125	0.85 ± 0.12
	Age	FIM-SPACE	$399.90 \pm 0.82^\dagger$	$53.79 \pm 1.10^\dagger$	$0.005 \pm 0.010^\dagger$	$0.735 \pm 0.119^\dagger$	$0.924 \pm 0.037^\dagger$	$3.37 \pm 0.36^{*\dagger}$
		GREEDY-CFF	399.79 ± 0.88	53.69 ± 1.17	0.007 ± 0.011	0.698 ± 0.137	0.914 ± 0.041	216.22 ± 32.35
		FIMM	396.70 ± 1.20	51.39 ± 1.18	0.049 ± 0.015	0.086 ± 0.315	0.676 ± 0.134	1.28 ± 0.18
	Gender	FIM-SPACE	$400.23 \pm 0.80^\dagger$	$54.04 \pm 1.07^\dagger$	$0.004 \pm 0.008^\dagger$	$0.558 \pm 0.463^\dagger$	$0.834 \pm 0.222^\dagger$	$2.43 \pm 0.25^{*\dagger}$
		GREEDY-CFF	400.24 ± 0.75	54.05 ± 1.01	0.003 ± 0.009	0.550 ± 0.561	0.832 ± 0.256	197.28 ± 34.88
		FIMM	397.83 ± 0.98	51.09 ± 1.22	0.058 ± 0.016	0.030 ± 0.629	0.603 ± 0.295	0.32 ± 0.06
φ_{exp}	Ethnicity	FIM-SPACE	329.02 ± 3.61	53.82 ± 1.05	0.006 ± 0.008	0.651 ± 0.241	0.894 ± 0.079	$3.02 \pm 0.30^*$
		GREEDY-CFF	328.78 ± 3.48	53.73 ± 1.02	0.007 ± 0.008	0.701 ± 0.304	0.907 ± 0.105	215.67 ± 31.79
	Age	FIM-SPACE	328.55 ± 3.67	53.76 ± 1.08	0.005 ± 0.010	0.760 ± 0.108	0.932 ± 0.032	$3.38 \pm 0.40^*$
		GREEDY-CFF	328.07 ± 4.14	53.63 ± 1.20	0.008 ± 0.012	0.778 ± 0.102	0.937 ± 0.029	216.20 ± 34.26
	Gender	FIM-SPACE	330.17 ± 3.59	54.03 ± 1.06	0.004 ± 0.008	0.621 ± 0.442	0.859 ± 0.214	$2.43 \pm 0.25^*$
		GREEDY-CFF	330.12 ± 3.89	54.01 ± 1.14	0.004 ± 0.010	0.659 ± 0.470	0.880 ± 0.212	196.68 ± 35.59
φ_{log}	Ethnicity	FIM-SPACE	204.52 ± 1.77	53.88 ± 1.07	0.004 ± 0.008	0.579 ± 0.257	0.870 ± 0.088	$3.01 \pm 0.33^*$
		GREEDY-CFF	204.32 ± 1.66	53.76 ± 1.02	0.007 ± 0.008	0.523 ± 0.321	0.852 ± 0.113	213.71 ± 33.96
	Age	FIM-SPACE	204.37 ± 1.79	53.85 ± 1.08	0.004 ± 0.009	0.674 ± 0.136	0.907 ± 0.044	$3.38 \pm 0.42^*$
		GREEDY-CFF	204.10 ± 1.98	53.70 ± 1.18	0.006 ± 0.010	0.655 ± 0.180	0.900 ± 0.060	214.67 ± 36.83
	Gender	FIM-SPACE	204.99 ± 1.77	54.05 ± 1.07	0.003 ± 0.008	0.502 ± 0.446	0.815 ± 0.217	$2.42 \pm 0.25^*$
		GREEDY-CFF	204.95 ± 1.80	54.03 ± 1.09	0.004 ± 0.010	0.546 ± 0.550	0.837 ± 0.237	196.51 ± 37.02

AV networks carries exactly one value per demographic attribute, so community memberships are non-overlapping by construction.

Metric Definitions. Let $S_{\text{GREEDY-IM}}$ denote the seed set returned by the standard IM baseline and S_{FIM} the seed set returned by a fair algorithm. The Price of Fairness and Reward of Fairness are defined as:

$$\text{PoF} = \frac{g(S_{\text{GREEDY-IM}}) - g(S_{\text{FIM}})}{g(S_{\text{GREEDY-IM}})}, \qquad \text{RoF} = \frac{\text{Var}(\mathbf{u}(S_{\text{GREEDY-IM}})) - \text{Var}(\mathbf{u}(S_{\text{FIM}}))}{\text{Var}(\mathbf{u}(S_{\text{GREEDY-IM}})) + \text{Var}(\mathbf{u}(S_{\text{FIM}}))}, \tag{7}$$

where $g(S)$ is the expected number of activated nodes and $\mathbf{u}(S) = (u_{c_1}(S), \ldots, u_{c_{|\mathcal{C}|}}(S))$ is the vector of per-community utilities. The F_ω-score jointly captures both dimensions as their weighted harmonic mean:

$$F_\omega\text{-score} = (1+\omega^2)\,\frac{\frac{1+\text{RoF}}{2} \cdot (1 - \text{PoF})}{\omega^2 \cdot \frac{1+\text{RoF}}{2} + (1 - \text{PoF})}, \tag{8}$$

where ω controls the relative weight of fairness vs. influence; following Wang et al. (2024) we set $\omega = 2$.

C.1 Extended Solution Quality Results

Table 6 reports the per-attribute breakdown of all quality metrics on the AV networks, complementing the aggregated results in Table 1. Significance is assessed via two-sided Wilcoxon signed-rank test with Holm–Bonferroni correction ($\alpha = 0.05$) within each (metric, comparison) family across the 9 (φ, attribute) cells (3 for FIMM, power-only). The per-attribute results are consistent with the aggregate: no statistically significant difference was detected between FIM-SPACE and GREEDY-CFF on F_2 and RoF across all attributes and fairness functions, while both methods significantly outperform FIMM on every metric ($\Delta F_2 \simeq +0.22$, $\Delta\text{RoF} \simeq +0.56$). Figure 1 further illustrates the per-graph fairness–influence tradeoff for FIM-SPACE and the baselines.

C.2 Extended Scalability Results

This section provides a detailed breakdown of the scalability experiment summarized in Section 4 (table 2), covering solution quality (Figures 2–4), runtime scaling (Figure 5), and a discussion of

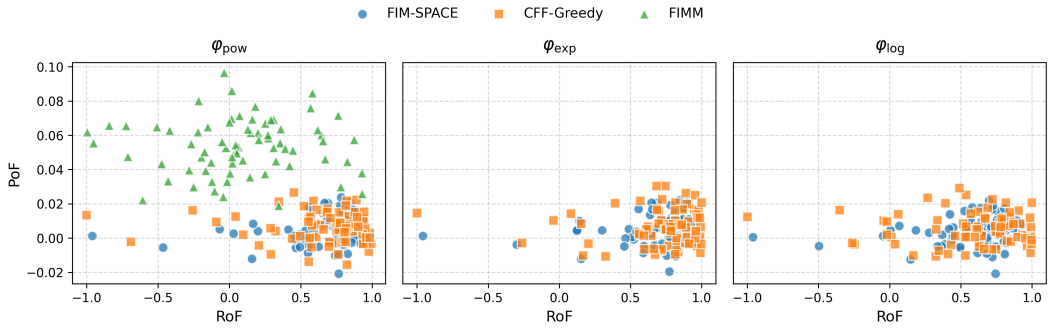


Figure 1: PoF vs. RoF per graph (n=24) and demographic attribute, faceted by φ. Each point is the mean over 15 trials. FIM-SPACE and GREEDY-CFF cluster in the high-RoF, zero-PoF corner across all three fairness functions; FIMM sits clearly off-frontier with higher PoF and lower RoF.

estimation artifacts. Significance markers in Table 2 are based on two-sided Wilcoxon signed-rank tests paired by trial index within each (algorithm, dataset, φ) combination ($\alpha = 0.05$); no multiple testing correction is applied as comparisons are few and pre-specified. Results are consistent across all tested budgets $B \in \{10, 25, 50\}$; figures and tables report $B = 25$ unless otherwise noted.

Tractability summary. We process datasets in increasing order of scale. For each (algorithm, dataset, budget) triple, if the first trial exceeds the one-hour deadline, the algorithm is marked intractable for that dataset and all larger ones; otherwise the remaining 14 trials are executed. The IMM reference is probed first: if intractable, relative metrics (PoF, RoF) are unavailable for that dataset. The resulting tractability picture is as follows:

- FIM-SPACE completes all 15 trials on every dataset and for all three fairness functions within the time limit, across all budgets $B \in \{10, 25, 50\}$.
- GREEDY-CFF completes on Email but exceeds the one-hour deadline on Amazon (one timed-out trial, no approximation guarantee) and is intractable on YouTube and DBLP.
- FIMM completes on all datasets but is restricted to φ_{pow} by design and cannot be applied when other fairness criteria are required.

Solution quality. Figure 2 shows Φ_φ across datasets and fairness functions. On Email, where all algorithms complete, FIM-SPACE matches GREEDY-CFF and FIMM to within 0.5% on Φ_φ. On YouTube and DBLP, FIM-SPACE equals or exceeds FIMM on Φ_φ across all budgets, with statistically significant advantages on both objective and influence spread (Table 2, † markers). The one exception is Amazon under φ_{pow}, where FIM-SPACE shows high variance ($\Phi_\varphi = 1{,}060 \pm 132$ vs. FIMM's $1{,}526 \pm 23$)[8]; this limitation is specific to the power function on this dataset, as FIM-SPACE recovers under φ_{exp} (RoF$= 0.74$) while GREEDY-CFF exceeds the time limit.

Figure 3 shows influence spread σ across datasets. FIM-SPACE tracks the IMM reference closely on YouTube and DBLP under φ_{exp} and φ_{log} (RoF≥ 0.96), confirming that the fairness objective is achieved at negligible influence cost at scale. Under φ_{pow}, both FIM-SPACE and FIMM trade some influence for fairness, with FIM-SPACE trading less on YouTube and DBLP.

[8]The effect comes from a very specific interaction between the power function and the Amazon graph.

First, the power function $\phi_{pow}(u) = u^{0.1}$ has a derivative $\phi'(u) = 0.1u^{-0.9}$ that blows up as $u \to 0$. The marginal fairness weight in our gradient is exactly $w_C = \phi'_{pow}(v_C)$, so for any group with near-zero utility this weight is huge and extremely sensitive: a small change in the estimated utility v_C produces a large swing in $\phi'(v_C)$, and hence in the gradient.

Second, Amazon is our sparsest graph by a wide margin: its cascades are tiny (see, e.g., values of λ and Λ in Table 5), so under the weighted IC model many groups end up with utility very close to zero. This is precisely the region where ϕ'_{pow} is most unstable, which is why the gradient variance is high there.

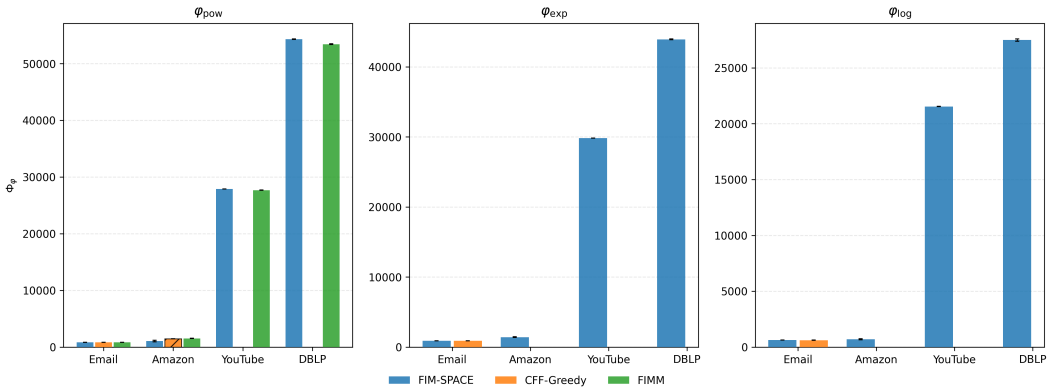


Figure 2: Fair objective Φ_φ across the four SNAP datasets ($B = 25$). Absent bars indicate intractability (timeout) for GREEDY-CFF on larger datasets, or architectural restriction for FIMM on non-power fairness functions.

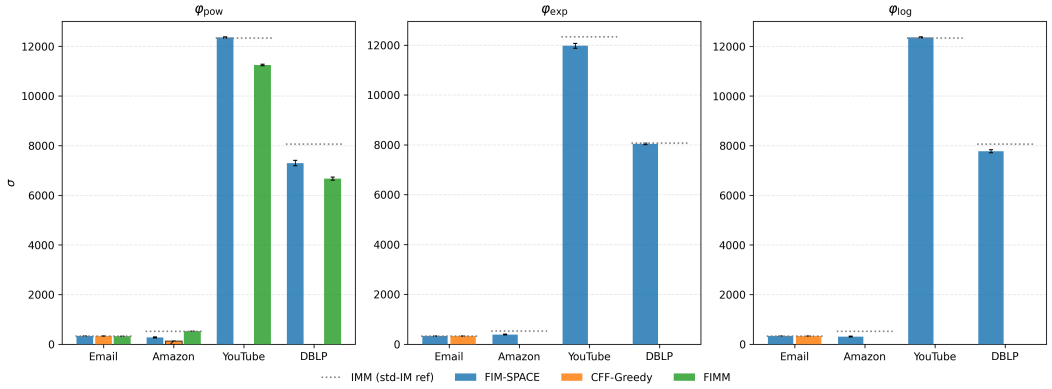


Figure 3: Influence spread σ across the SNAP datasets ($B = 25$). The dotted grey line is the IMM reference, illustrating the influence cost of enforcing fairness.

Fairness metrics. Figure 4 shows PoF and RoF across all datasets and fairness functions, with dashed reference lines at PoF$= 0$ and RoF$= 1$ marking the IMM baseline. FIM-SPACE stays at or near the reference on YouTube and DBLP under $\varphi_{\exp}$ and $\varphi_{\log}$, confirming that fairness is achieved at minimal influence cost. Under φ_{pow}, FIMM falls below RoF$= 1$ on YouTube and DBLP, while FIM-SPACE achieves higher RoF at comparable or lower PoF on both datasets. The Amazon anomaly under φ_{pow} is visible here as well: FIM-SPACE shows lower RoF than FIMM for this specific combination, consistent with the high variance discussed above.

Runtime scaling. Figure 5 shows execution time on a log scale. FIM-SPACE grows consistently with network size across all three fairness functions, from $\sim$19 s on Email to $\sim$2,100 s on DBLP at $B = 25$. FIMM is faster than FIM-SPACE on DBLP ($\sim$1,065 s vs. $\sim$2,115 s) due to its more targeted RR-set construction, but this advantage is limited to φ_{pow} only. GREEDY-CFF is the slowest by a wide margin even on Email ($\sim$880 s) and is intractable beyond that; quality metrics for the Amazon row in Table 2 are based on a single timed-out trial and carry no approximation guarantee. The runtime ordering of the algorithms is consistent across all tested budgets $B \in \{10, 25, 50\}$.

Estimation artifacts. Small negative PoF values and RoF slightly above 1.0 visible on Email arise because the IMM reference is itself a $(1 - 1/e - \varepsilon)$ approximation whose Monte Carlo influence estimate is subject to sampling noise. When a fair algorithm's seed set achieves marginally

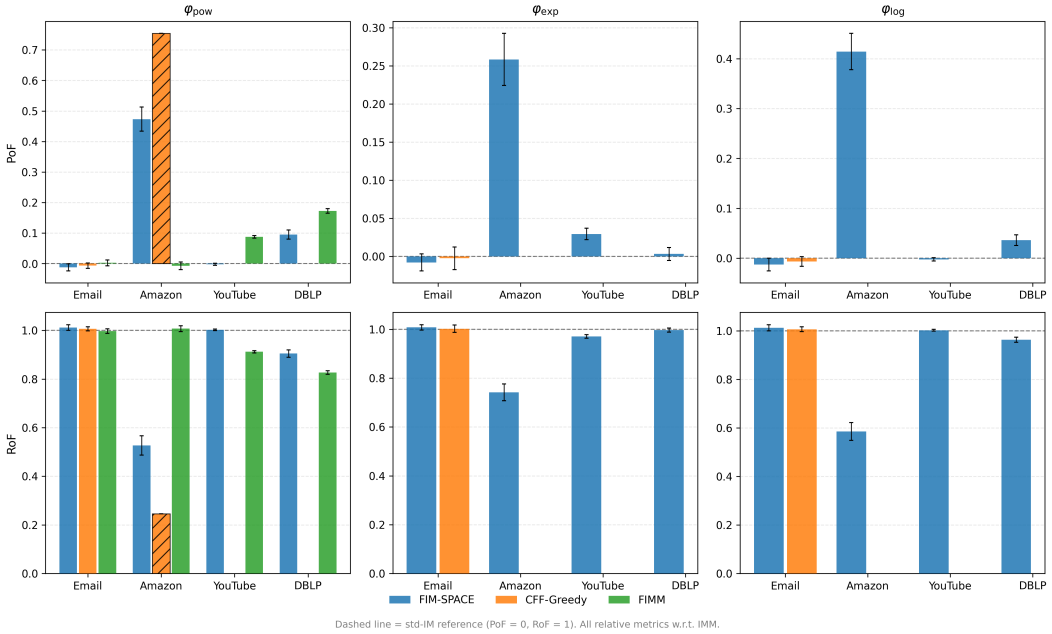


Figure 4: PoF (top row) and RoF (bottom row) across SNAP datasets ($B = 25$), faceted by φ. Dashed lines at PoF $= 0$ and RoF $= 1$ mark the IMM baseline.

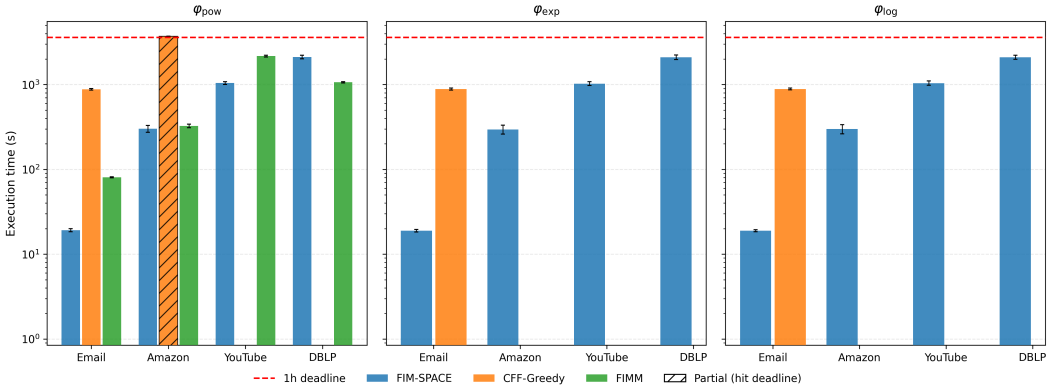


Figure 5: Execution time on a log scale ($B = 25$). The red dashed line marks the one-hour deadline. Hatched bars indicate timed-out runs.

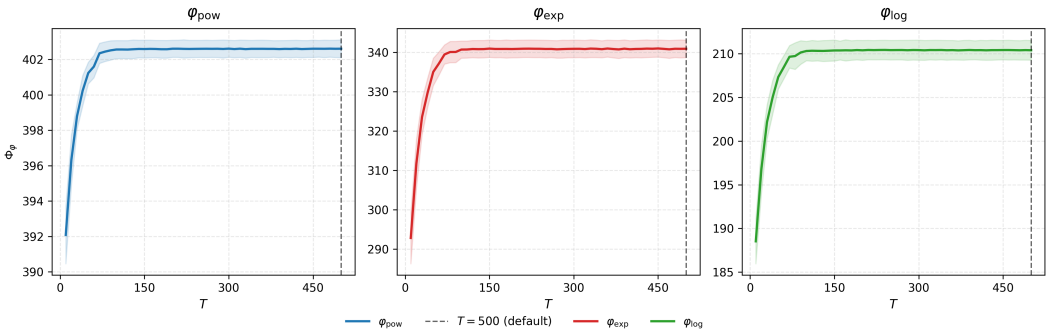


Figure 6: Objective value Φ_φ as a function of the number of iterations T, shown separately for the three fairness functions (left: φ_{pow}, centre: φ_{exp}, right: $\varphi_{\log}$). Each curve shows the mean over 15 trials $\times$ 3 attributes. The vertical dashed line marks the default budget $T = 500$.

higher sampled influence than IMM under the same evaluation budget, PoF becomes negative and RoF exceeds 1. The deviations are consistently below 1% in magnitude and carry no substantive interpretation.

C.3 Extended Analysis of Hyperparameter Sensitivity and Convergence

This appendix provides a detailed breakdown of the empirical convergence and hyperparameters' sensitivity of the FIM-SPACE algorithm.

To assess the robustness and efficiency of the FIM-SPACE algorithm, we analyze its convergence behavior and sensitivity to key hyperparameters. We evaluate the objective function Φ_ϕ (using the same fairness functions considered in the previous sections) and the total execution time. Results are averaged over 15 independent runs across the three attributes (ethnicity, age and gender) on one of the AV networks. Default parameters are the same used in section 4. To isolate the effect of each parameter, we vary just one of them each time while others remain fixed at their default values.

Empirical Convergence and Number of Iterations T. To analyze the convergence rate of the continuous optimization, we recorded the objective value at intermediate milestones up to the default budget of $T = 500$ iterations. Across all three fairness functions, FIM-SPACE shows steep initial learning curve. For instance, when instantiated with the power fairness function, the objective reaches value ~ 392 within just 10 iterations and climbs to ~ 402.5 by iteration 100. After $T = 100$, the objective plateaus, maxing out at 402.6 at $T = 500$. This trend is mirrored also in the exponential and logarithmic variants, proving that FIM-SPACE captures the vast majority of the optimal influence spread in the earliest stages of optimization, making it highly efficient for time-sensitive applications. Results are reported in Figure 6.

Varying the total number of iterations $T \in \{100, 200, 500, 1000\}$ demonstrates the computational cost of over-optimizing. Since the algorithm converges around $T = 100$, increasing the budget to $T = 1000$ yields no measurable improvement in the objective function's value. However, the execution time scales perfectly linearly with T, increasing from 0.55 seconds at $T = 100$ to 4.85 seconds at $T = 1000$. This confirms that smaller iteration budgets can be safely deployed without sacrificing solution quality. This can be observed in figure 7.

Batch Size β. The batch size β controls the number of Reverse Reachable (RR) sets sampled per step to estimate the gradient. Our sweep over sizes $\{50, 100, 200, 500\}$ reveals that increasing the batch size monotonically improves the objective value. However, we pay this quality in time penalty: a batch size of 500 takes roughly 13.6 seconds, compared to just 2.5 seconds for a batch size of 100. Consequently, a batch size of 100 for example, serves as a good tradeoff between high-quality gradient and reasonable execution time. Results are shown in Figure 8.

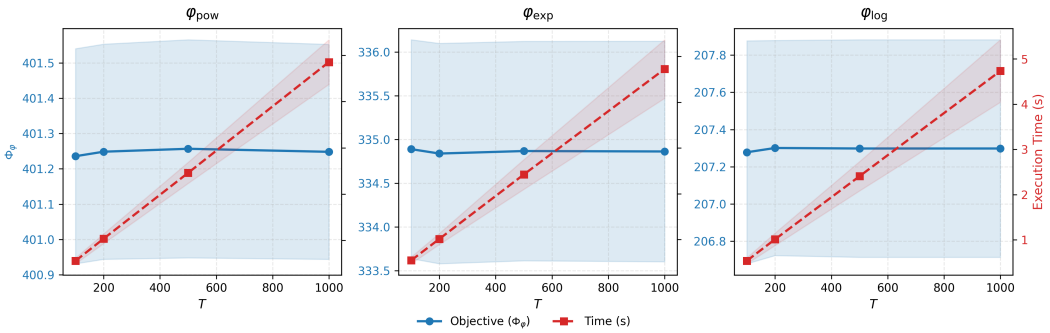


Figure 7: Final objective Φ_φ (blue, left y-axis) and execution time in seconds (red, right y-axis) as a function of the total iteration budget $T \in \{100, 200, 500, 1000\}$, for each fairness function. Filled circles and solid lines correspond to the objective; squares and dashed lines correspond to time.

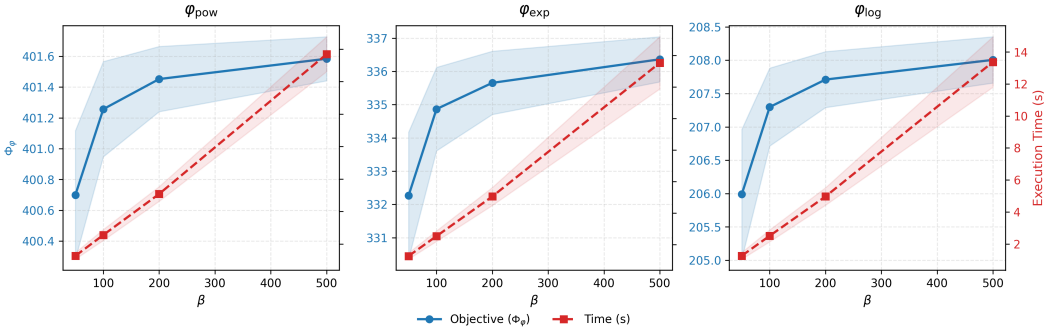


Figure 8: Final objective Φ_φ (blue, left y-axis) and execution time in seconds (red, right y-axis) as a function of the RR-set batch size $\beta \in \{50, 100, 200, 500\}$, for each fairness function. A larger batch reduces gradient variance and monotonically improves solution quality, but at a proportionally higher computational cost.

Overall, we can conclude that FIM-SPACE shows high stability to hyperparameter variations. Increasing the number of iterations T beyond 100 strictly increases execution time linearly (from roughly 0.5 s at $T = 100$ to 4.8 s at $T = 1000$) without yielding further improvements to the objective. Similarly, while increasing the batch size marginally improves the objective function, it incurs a severe linear time penalty, confirming that a batch size of 100 provides an optimal trade-off between gradient accuracy and speed.

C.4 EXTENDED ANALYSIS OF BUDGET SENSITIVITY

To verify the robustness of our approach under varying resource constraints, in this appendix we evaluate the algorithms across a sweep of seed budgets ($B \in \{5, 10, 15, 20, 25\}$). Figures 9, 10, 11, and 12 illustrate the evolution of the proxy objective (Φ_φ), Price of Fairness (PoF), Reward of Fairness (RoF), and F_2-score, respectively, as the available seed budget B increases from 5 to 25.

Objective Function and PoF. As expected, Figure 9 shows that the proxy objective Φ_φ grows monotonically with the budget across all algorithms and fairness families. FIM-SPACE consistently matches or slightly exceeds the discrete GREEDY-CFF baseline at every budget step. The Price of Fairness plot (Figure 10) provides critical context to this growth: at $B = 5$, enforcing fairness carries a severe penalty, resulting in a PoF of approximately 72% across all methods. This occurs because the standard utility-maximizing baseline can easily reach the dense core of the network with 5 seeds, whereas fair algorithms are forced to select sub-optimal peripheral seeds to satisfy

demographic constraints. However, as B increases, the network saturates. FIM-SPACE efficiently exploits this saturation, dropping the PoF to a statistically negligible $0.1\% - 0.2\%$ at $B = 25$. FIMM, by contrast, retains a structurally higher PoF of 5.1%.

Fairness and F_2-score. Figures 11 and 12 highlight the contrast of GREEDY-CFF and FIM-SPACE with respect to FIMM. As the budget grows, both GREEDY-CFF and FIM-SPACE successfully maintain high fairness, with their performance curves appearing closely matched. The two methods achieve comparable fairness at every budget. In stark contrast, the most striking failure belongs to FIMM under φ_{pow}: its RoF drastically collapses from 0.423 at $B = 5$ down to just 0.103 at $B = 25$. Because FIMM relies on rigid, exact combinatorial constraints, it fails to balance the marginal gains of different communities as the combinatorial search space expands. This translates directly to Figure 12, where both FIM-SPACE and GREEDY-CFF achieve highly efficient F_2-scores (≈ 0.90). This demonstrates that our continuous relaxation sacrifices zero solution quality compared to the best discrete greedy heuristics.

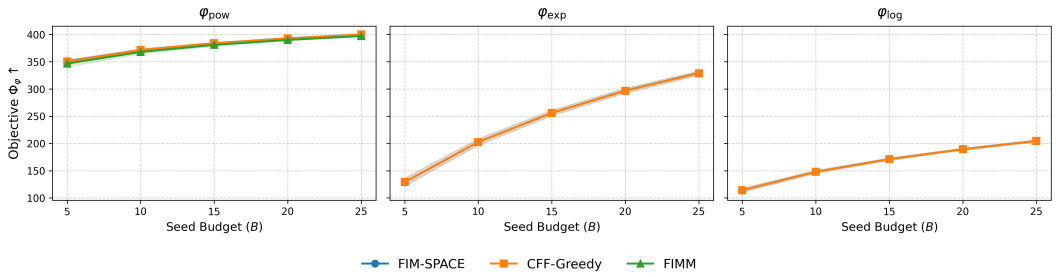


Figure 9: Proxy objective (Φ_φ) as a function of the seed budget (B).

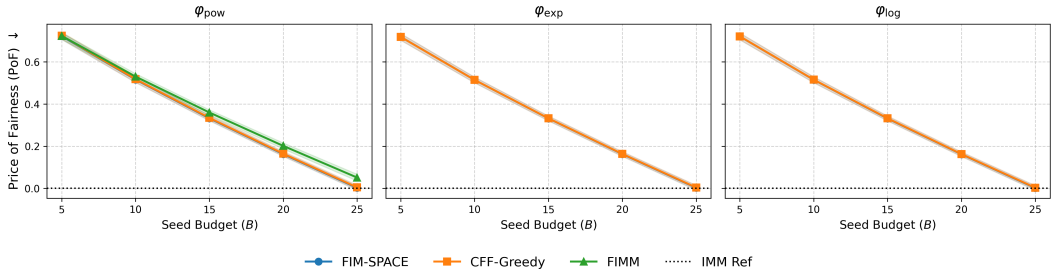


Figure 10: Price of Fairness (PoF) as a function of the seed budget (B). The dashed line represents the utility-maximizing IMM reference (PoF $= 0$).

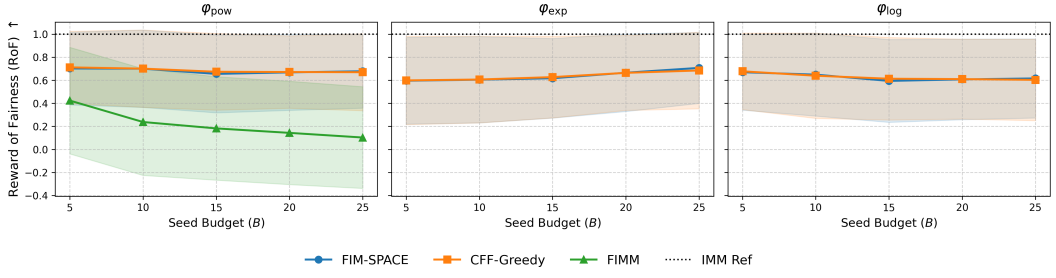


Figure 11: Reward of Fairness (RoF) as a function of the seed budget (B). The dashed line represents the IMM reference (RoF $= 1$).

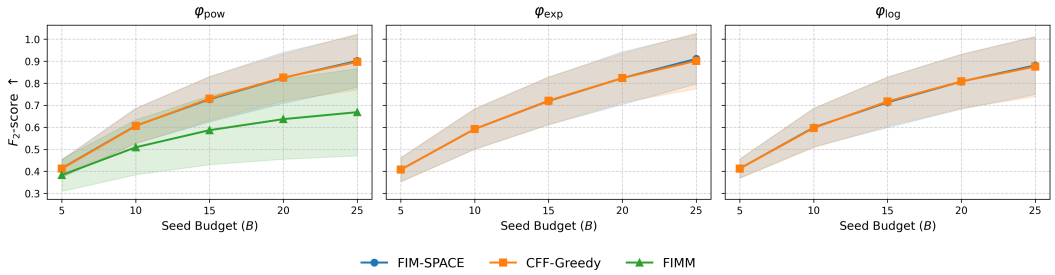


Figure 12: F_2-score across seed budgets.

Overall, the results demonstrate that FIM-SPACE is highly efficient at utilizing additional seeds to simultaneously maximize utility and equity. At highly restricted budgets ($B = 5$), all algorithms inherently incur a massive Price of Fairness (PoF $\approx 72\%$) because a tiny seed set cannot adequately cover multiple communities. However, as the budget expands to $B = 25$, FIM-SPACE perfectly allocates the new seeds, driving the PoF to near-zero (0.1%–0.2%) while securing the highest Reward of Fairness (RoF) and overall F_2-score (> 0.88 across all fairness functions). Conversely, methods like FIMM fail to scale their fairness guarantees; as B increases, FIMM's RoF severely collapses to 0.10, indicating it wastes the expanded budget on redundant or inequitable nodes.

C.5 Extended Analysis of Phi Parameter Sensitivity

In this appendix, we evaluate the algorithms' sensitivity to the parameters governing the fairness-influence tradeoff within each concave function family: ζ_1 for φ_{pow}, ζ_2 for φ_{exp}, and ζ_3 for φ_{log}. Tuning these parameters allows us to smoothly interpolate between strict community-level fairness and pure utility maximization.

We sweep each phi family over its free parameter ($\zeta_1 \in \{0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0\}$ for φ_{pow}, $\zeta_2 \in \{2, 5, 10, 20, 40, 80\}$ for φ_{exp}, and $\zeta_3 \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for φ_{log}) and compare FIM-SPACE against GREEDY-CFF over 24 AV graphs $\times$ 3 demographic attributes $\times$ 15 trials at $B = 25$.

Figures 13, 14, 15, and 16 illustrate the algorithms' behavior across varying levels of fairness stringency for the objective value Φ_φ, Price of Fairness (PoF), Reward of Fairness (RoF), and F_2-score, respectively. Note that we omit the plot for raw influence spread (σ), as it is perfectly inversely proportional to the PoF metric.

The x-axes in all plots are oriented such that moving from left to right corresponds to a transition from a strict fairness requirement to a relaxed, pure-influence objective. Consequently, the axis for the exponential parameter ζ_2 is reversed.

Objective Function's Value. Figure 13 demonstrates that FIM-SPACE and the GREEDY-CFF baseline achieve nearly identical objective values (Φ_φ) across the entire parameter spectrum, resulting in visually overlapping curves. Under φ_{pow}, for example, as ζ_1 increases from 0.05 (heavy fairness penalty) to 1.0 (standard linearity), both algorithms successfully maximize the proxy submodular function to the same peak values. However, the true advantage of the continuous relaxation becomes apparent when decomposing this objective: as shown in the subsequent metrics, FIM-SPACE attains these matching objective values while discovering seed sets that yield slightly better actual fairness (higher RoF) and F2-score (F_2). This suggests that FIM-SPACE finds solutions that are equally good in Φ_φ and comparable or slightly better in how equitably influence is distributed.

The Fairness-Influence Tradeoff. Figures 14 and 15 deconstruct this objective into its relative components: PoF and RoF. As the parameters enforce stricter fairness (moving left), RoF naturally climbs. However, FIM-SPACE achieves a systematically higher RoF than GREEDY-CFF across all three function families. For instance, at $\zeta_2 = 40$ under φ_{exp}, FIM-SPACE reaches a mean RoF of 0.76 (compared to GREEDY-CFF's 0.70). Crucially, Figure 14 shows that FIM-SPACE achieves these fairness gains with a strictly lower PoF. Even at maximum fairness stringency, the PoF for

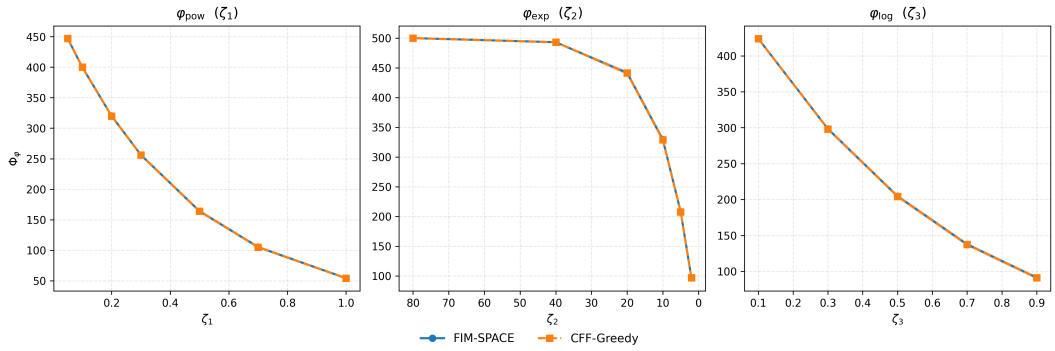


Figure 13: Mean objective value (Φ_φ) across the parameter sweeps for the three fairness functions. The x-axis is oriented such that moving from left to right transitions from strict fairness to pure utility.

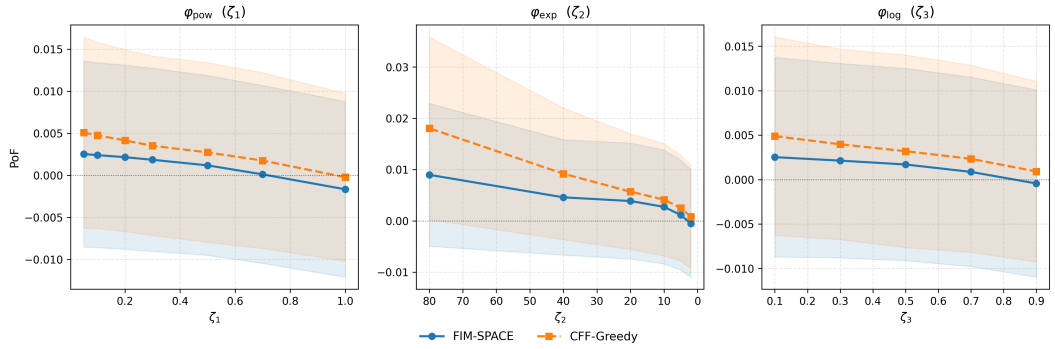


Figure 14: Price of Fairness (PoF) across the parameter sweeps. The horizontal dashed line at PoF $= 0$ represents the IMM utility baseline.

FIM-SPACE rarely exceeds 0.5%, meaning it redistributes influence equitably among communities without shrinking the overall size of the cascade.

From Figure 16 we can see that the trend for the F2-score is comparable to the RoF's one.

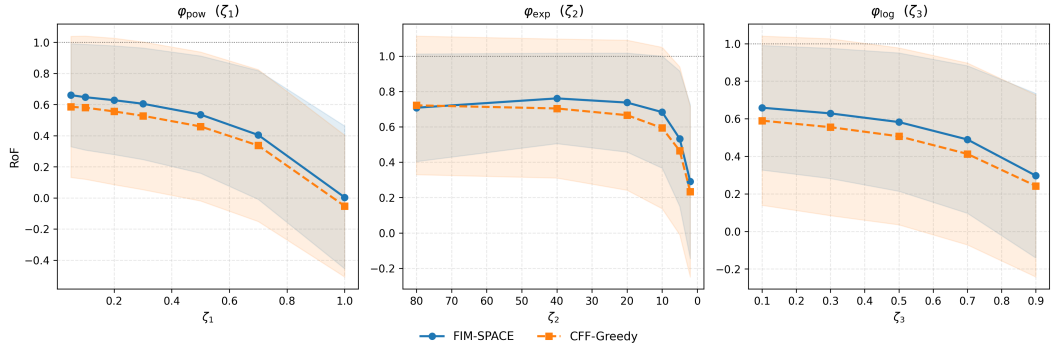


Figure 15: Reward of Fairness (RoF) across the parameter sweeps. The horizontal dashed line at RoF $= 1$ represents the maximum achievable value.

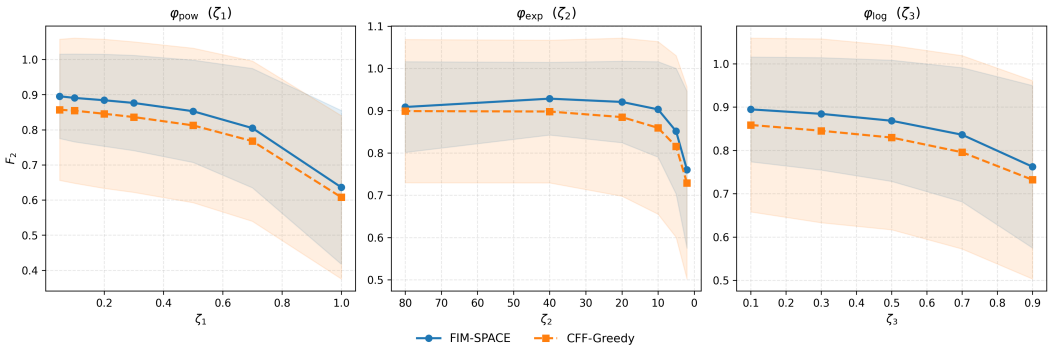


Figure 16: F_2-score across the parameter sweeps.

We can conclude that, across the entire parameter spectrum, FIM-SPACE performs on par with the combinatorial GREEDY-CFF baseline. While both algorithms successfully maximize the proxy submodular function Φ_φ to nearly identical peak values (fig. 13), FIM-SPACE discovers seed sets that capture a superior F2-score in practice (fig. 16). Under the strictest fairness settings (e.g., $\zeta_1 = 0.05$ or $\zeta_2 = 40$), FIM-SPACE incurs a negligible PoF (under 0.5%) while securing substantial RoF (e.g., reaching RoF > 0.76 under $\varphi_{\exp}$).

As the parameters relax to emulate standard influence maximization (e.g., $\zeta_1 \to 1.0$), FIM-SPACE gracefully converges toward the pure utility optimum, safely driving both RoF and PoF toward the baseline limits.