

Group-Fairness in Influence Maximization

Alan Tsang^{1*}, Bryan Wilder^{2*}, Eric Rice², Milind Tambe², Yair Zick¹

¹ Department of Computer Science, National University of Singapore,
akhtsang@nus.edu.sg, zick@comp.nus.edu.sg

² Center for Artificial Intelligence in Society, University of Southern California, {bwilder,
ericr, tambe}@usc.edu

*Equal contribution

Abstract. Influence maximization is a widely used model for information dissemination in social networks. Recent work has employed such interventions across a wide range of social problems, spanning public health, substance abuse, and international development (to name a few examples). A critical but understudied question is whether the benefits of such interventions are *fairly distributed* across different groups in the population; e.g., avoiding discrimination with respect to sensitive attributes such as race or gender. Drawing on legal and game-theoretic concepts, we introduce formal definitions of fairness in influence maximization. We provide an algorithmic framework to find solutions which satisfy fairness constraints, and in the process improve the state of the art for general multi-objective submodular maximization problems. Experimental results on real data from an HIV prevention intervention for homeless youth show that standard influence maximization techniques oftentimes neglect smaller groups which contribute less to overall utility, resulting in a disparity which our proposed algorithms substantially reduce.

Keywords: Influence Maximization · Fair Allocation.

1 Introduction

Influence maximization in social networks is a well-studied problem with applications in a broad range of domains. Consider, for example, a group of at-risk youth; outreach programs try to provide as many people as possible with useful information (e.g., HIV safety, or available health services). Since resources (e.g., social workers) are limited, it is not possible to personally reach every at-risk individual. It is thus important to target *key community figures* who are likely to spread vital information to others. Formally, individuals are nodes V in a social network, and we would like to influence or *activate* as many of them as possible. This can be done by initially *seeding* k nodes (where $k \ll |V|$). The seed nodes activate their neighbors with some probability, who activate their neighbors and so forth. Our goal is to identify k seeds such that the maximal number of nodes is activated. This is the classic *influence maximization problem* [13], that has received much attention in the literature.

In recent years, the influence maximization framework has seen application to many social problems, such as HIV prevention for homeless youth [28,25], public health awareness [23], financial inclusion [2], and more. Frequently, small and marginalized

groups within a larger community are those who benefit the most from attention and assistance. It is important, then, to ensure that the allocation of resources reflects and respects the diverse composition of our communities, and that each group receives a fair allocation of the community’s resources. For instance, in the HIV prevention domain we may wish to ensure that members of racial minorities or of LGBTQ identity are not disproportionately excluded; this is where our work comes in.

Our Contributions: This paper introduces the problem of fair resource allocation in influence maximization. Our *first contribution* is to propose fairness concepts for influence maximization. We start with a *maximin* concept inspired by the legal notion of disparate impact; formally it requires us to maximize the minimum fraction of nodes within each group that are influenced. While intuitive and well-motivated, this definition suffers from shortcomings that lead us to introduce a second concept, *diversity constraints*. Roughly, diversity constraints guarantee that every group receives influence commensurate with its “demand”, i.e., what it could have generated on its own, based on a number of seeds proportional to its size. Here, to compute a group’s demand, we allow it a number of seeds proportional to its size, but require that it spreads influence using only nodes in the group. Hence, a small but well connected group may have a better claim for influence than a large but sparsely connected group.

Our *second contribution* is an algorithmic framework for finding solutions that satisfy either fairness concept. While the classical influence maximization problem is submodular (and hence easily solved with a greedy algorithm), fairness considerations produce strongly non-submodular objectives. This renders standard techniques inapplicable. We show that both fairness concepts can be reduced to *multi-objective* submodular optimization problems, which are substantially more complex. Our key algorithmic contribution is a new method for general multi-objective submodular optimization which has substantially better approximation guarantee than the current best algorithm [22], and often better runtime as well. This result may be of independent interest.

Our *third contribution* is an analytical exploration of the *price of group fairness* in influence maximization, i.e., the reduction in social welfare with respect to the unconstrained influence maximization problem due to imposing a fairness concept. We show that the price of diversity can be high in general for both concepts and under a range of settings.

Our *fourth contribution* is an empirical study on real-world social networks that have been used for a socially critical application: HIV prevention for homeless youth. Our results show that standard influence maximization techniques often cause substantial fairness violations by neglecting small groups. Our proposed algorithm substantially reduces such violations at relatively small cost to overall utility.

Related Work: Kempe et al. [13] introduced influence maximization and proved that since the objective is submodular, greedily selecting nodes gives a $(1 - \frac{1}{e})$ -optimal solution. There has since been substantial interest among the AI community both in developing more scalable algorithms (see [15] for a recent survey), as well as in addressing the challenges of deployment in public health settings [27,24]. Recently, such algorithms have been used in real-world pilot tests for HIV prevention amongst homeless youth [28,25], driving home the need to consider fairness as influence maximization

is applied in socially sensitive domains. To our knowledge, no previous work considers fairness specifically for influence maximization. The techniques we introduce to optimize fairness metrics are related to research on multi-objective submodular maximization (outside the context of fairness), and we improve existing theoretical guarantees for this general problem [7,22].

Outside of influence maximization, the general idea of diversity as an optimization constraint has received considerable attention in recent years; it has been studied in multiwinner elections (see [6,10] for an overview), resource allocation [5], and matching problems [1,11]. We note that some of the above works (e.g. [1] and [18]) use a submodular objective function as a means of achieving diversity; interestingly, while the classic influence maximization target function is submodular, it is no longer so under diversity constraints. Group fairness has been studied extensively in the voting theory literature, where the objective is to identify a committee of k candidates that will satisfy subsets of voters (see a comprehensive overview in [10]). There have also been several works on group fairness in fair division, defining notions of group envy-freeness [8,9,19,21], and a group maximin share guarantee [3,20].

2 Model

Agents are embedded in a social network $G = (V, E)$. An edge $(i, j) \in E$ represents the ability for agent v_i to influence or *activate* v_j . G may be undirected or directed.

Diversity: Each agent in our network may identify with one or more groups within the larger population. These represent different ethnicities, genders, sexual orientations, or other groups for which fair treatment is important. Our goal is to maximize influence in a way such that each group receives at least a “fair” share of influence (more on this below). Let us designate these groups as $\mathcal{C} = \{C_1, \dots, C_m\}$. Each **group** C_i represents a non-empty subset of V , $\emptyset \neq C_i \subseteq V$. Each agent must belong to at least one group, but may belong to multiple groups; i.e. $C_1 \cup C_2 \cup \dots \cup C_m = V$. In particular, this allows for the expression of intersectionality, where an individual may be part of several minority groups.

Influence maximization: We model influence using the *independent cascade model* [13], the most common model in the literature. All nodes begin in the inactive state. The decision maker then selects k *seed nodes* to activate. Each node that is activated makes one attempt to activate each of its inactive neighbors; each attempt succeeds independently with probability p . Newly activated nodes attempt to activate their neighbors and so on, with the process terminating once there are no new activations.

We define the **influence** of nodes $A \subseteq V$, denoted $\mathcal{I}_G(A)$, as the expected number of nodes activated by seeding A . Of these, let $\mathcal{I}_{G,C_i}(A)$ be the expected number of activated vertices from C_i . Traditional influence maximization seeks a set A , $|A| \leq k$, maximizing $\mathcal{I}_G(A)$. Using a slight abuse of notation, let $\mathcal{I}_G(k)$ be the maximum influence that can be achieved by selected k seed nodes. That is, $\mathcal{I}_G(k) = \max_{|A|=k} \mathcal{I}_G(A)$. Analogously, we define $\mathcal{I}_{G,C_i}(k)$ as the maximum expected number of vertices from C_i that can be activated by k seeds. We now propose two means of capturing group fairness in influence maximization.

Maximin Fairness: Maximin Fairness captures the straightforward goal of improving the conditions for the *least well-off* groups. That is, we want to maximize the minimum influence received by any of the groups, as proportional to their population. This leads to the following utility function:

$$U^{\text{Maximin}}(A) = \min_i \frac{\mathcal{I}_{G,C_i}(A)}{|C_i|}$$

Subject to this maximin constraint, we seek to maximize overall influence. Thus, we define $\mathcal{I}_G^{\text{Maximin}} = \mathcal{I}_G(B)$ with $B = \operatorname{argmax}_{A \subseteq V, |A|=k} U^{\text{Maximin}}(A)$. That is, $\mathcal{I}_G^{\text{Maximin}}$ is the expected number of nodes activated by a seed configuration that maximizes the minimum proportional influence received by any group. This corresponds to the legal concept of *disparate impact*, which roughly states that a group has been unfairly treated if their “success rate” under a policy is substantially worse than other groups (see [4] for an overview). Therefore, maximin fairness may be significant to governmental or community organizations which are constrained to avoid this form of disparity. However, optimizing for equality of outcomes may be undesirable when some groups are simply much better suited than others to a network intervention. For instance, if one group is very poorly connected, maximin fairness would require that large number of nodes be spent trying to reach this group, even though additional seeds have relatively small impact.

Diversity Constraints: We now propose an alternate fairness concept by extending the notion of individual rationality to **Group Rationality**. The key idea is that no group should be better off by leaving the (influence maximization) game with their proportional allocation of resources and allocating them internally. For each group C_i , let $k_i = \lceil k|C_i|/|V| \rceil$ be the number of seeds that would be fairly allocated to the group C_i based on the group’s size within the larger population, rounded up to remove any doubt that this group receives a fair share. k_i is the **fair allocation** of seeds to the group.

Let $G[C_i]$ be the subgraph induced from G by the nodes C_i . This represents the network formed by group C_i if they were to separate from the original network. Now, we define the **group rational influence** that each group C_i can expect to receive as the number of nodes they expect to activate if they left the network, with their fair allocation of k_i seeds. We denote this group rational influence for C_i as $\mathcal{I}_{G[C_i]}(k_i)$. Then, we devise a set of **diversity constraints** that any group rational seeding configuration A with k seeds must satisfy: $\mathcal{I}_{G,C_i}(A) \geq \mathcal{I}_{G[C_i]}(k_i), \forall i$. That is, the influence received by each group is at least equal to what each group may accomplish on its own when given its fair share of k_i seed nodes.

The diversity constraint objective function is to maximize the expected number of nodes activated, subject to the above diversity constraint. The utility for selecting seed nodes A is:

$$U^{\text{Rational}}(A) = \begin{cases} \mathcal{I}_G(A), & \text{if } \mathcal{I}_{G,C_i}(A) \geq \mathcal{I}_{G[C_i]}(k_i), \forall i. \\ 0, & \text{otherwise.} \end{cases}$$

The maximum expected influence obtained via a group rational seeding configuration A is called the **rational influence** $\mathcal{I}_G^{\text{Rational}} = \mathcal{I}_G(B)$, where $B = \operatorname{argmax}_{A \subseteq V, |A|=k} U^{\text{Rational}}(A)$.

Price of Fairness: To measure the cost of ensuring a fair outcome for the diverse population, we will measure the Price of Fairness, the ratio of optimal influence to the best achievable influence under our two fairness criteria. Here **optimal influence** $\mathcal{I}_G^{\text{OPT}} = \mathcal{I}_G(k)$, which is the maximum amount of expected influence that can be obtained using any choice of k seed nodes. We omit the subscript where the context is clear.

$$PoF^{\text{Rational}} = \frac{\mathcal{I}^{\text{OPT}}}{\mathcal{I}^{\text{Rational}}} \quad PoF^{\text{Maximin}} = \frac{\mathcal{I}^{\text{OPT}}}{\mathcal{I}^{\text{Maximin}}}$$

3 Optimization

The standard approach to influence maximization is based on *submodularity*. Formally, a set function f on ground set V is submodular if for every $A \subseteq B \subseteq V$ and $x \in V \setminus B$, $f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$. This captures the intuition that additional seeds provide diminishing returns. However, both of our fairness concepts are easily shown to violate this property (proofs are deferred to the appendix)³:

Theorem 1. U^{Maximin} and U^{Rational} are not submodular.

Hence, we cannot apply the greedy heuristic to group-fair influence maximization. However, we now show that optimizing either utility function reduces to *multiobjective submodular maximization*, for which we give an improved algorithm below. Consider the following generic problem: given monotone submodular functions $f_1 \dots f_m$ and corresponding target values $W_1 \dots W_m$, find a set S satisfying $|S| \leq k$ with $f_i(S) \geq W_i$ for all i (under the promise that such an S exists). Roughly, f_i will be group i 's utility, and W_i will be the utility that we want to guarantee for i . Suppose that we have an algorithm for the above multiobjective problem. Then, we can optimize the maximin objective by letting $f_i = \frac{\mathcal{I}_{G, C_i}}{|C_i|}$ and binary searching for the largest W such that $f_i \geq W$ is feasible for all groups i . For diversity constraints, we let $f_i = \mathcal{I}_{G, C_i}$ and set the target $W_i = \mathcal{I}_{G[C_i]}(k_i)$. We then add another objective function $f_{\text{total}} = \mathcal{I}_G$ representing the combined utility and binary search for the highest value W_{total} such that the targets $W_1 \dots W_m, W_{\text{total}}$ are feasible. This represents the largest achievable total utility, subject to diversity constraints. Having reduced both fairness concepts to multiobjective submodular maximization, we now give an improved algorithm for this core problem.

The multiobjective submodular problem was introduced by Chekuri et al. [7], who gave an algorithm which guarantees $f_i \geq (1 - \frac{1}{e})W_i$ for all i provided that the number of objectives m is smaller than the budget k (when $m = \Omega(k)$, the problem is provably inapproximable [14]). Unfortunately, this algorithm is of mostly theoretical interest since it runs in time $O(n^8)$. Udvani [22] recently introduced a practically efficient algorithm; however it obtains an asymptotic $(1 - \frac{1}{e})^2$ -approximation instead of the optimal $(1 - \frac{1}{e})$. We remedy this gap by providing a practical algorithm obtaining an asymptotic $(1 - \frac{1}{e})$ -approximation (Algorithm 1). Its runtime is comparable to, and under many conditions faster than, the algorithm of [22]. We present the high-level idea behind the algorithm here, with additional details present in the appendix.

³ https://www.dropbox.com/s/5egb2wpl1gce7h5/appendix_fairness.pdf?dl=0

Previous algorithms [7,22] start from a common template in submodular optimization, which we also build on. The main idea is to relax the discrete problem to a continuous space. For a given submodular function f , its *multilinear extension* F is defined on n -dimensional vectors x where $0 \leq x_j \leq 1$ for all j . x_j represents the probability that item j is included in the set. Formally, let $S \sim x$ denote a set which includes each j independently with probability x_j . Then, we define $F(x) = \mathbb{E}_{S \sim x}[f(S)]$, which can be evaluated using random samples.

The main challenge is to solve the continuous optimization problem, which is where our technical contribution lies. Algorithm 1 describes the high-level procedure, which runs our continuous optimization subroutine (line 2) and then rounds the output to a discrete set (line 3). Line 1, which ensures that all items with value above a threshold τ are included in the solution, is a technical detail needed to ensure the rounding succeeds. The rounding process captured in lines 1 and 3 is fairly standard and used by both previous algorithms [7,22]. Our main novelty lies in an improved algorithm for the continuous problem, MULTIFW.

Algorithm 1 Multiobjective Optimization($\gamma, \tau, T, T', \eta$)

- 1: $S_1 = \{j : f_i(\{j\}) \geq \tau \text{ for some } i\}$
 - 2: $x = \text{MULTIFW}(k - |S_1|, \{\gamma(W_i - f_i(S))\}_{i=1}^m)$
 - 3: $S_2 = \text{SWAPROUND}(x_{\text{ini}})$ //see [7]
 - 4: **return** $S_1 \cup S_2$
-

Algorithm 2 Multiobjective Frank-Wolfe($k, \{W_i\}$)

- 1: $x^0 = 0$
 - 2: **for** $t = 1 \dots T$ **do**
 - 3: $v^t = \text{S-SP-MD}(x, \{i : W_i - F_i(x^{t-1}) \geq \epsilon\})$
 - 4: $x^t = x^{t-1} + \frac{1}{T} v^t$
 - 5: **return** $\text{APPROXDECOMPOSITION}(x^T)$ //see [16]
 - 6: **function** $\text{S-SP-MD}(x, \mathcal{I})$
 - 7: Initialize v s.t. $\|v\|_1 = k$ and $y \in \Delta(\mathcal{I})$ arbitrarily
 - 8: **for** $\ell = 1 \dots T'$ **do**
 - 9: Sample $i \sim y$; set $\hat{\nabla}_v = \frac{1}{W_i - F_i(x)} \mathcal{A}_{\text{grad}}^i(x)$
 - 10: Sample $j \sim v$; $\hat{\nabla}_y = k \cdot \text{diag}\left(\frac{1}{W - F(x)}\right) \mathcal{A}_{\text{item}}^j(x)$
 - 11: $y = \frac{y e^{-\eta \hat{\nabla}_y}}{\|y e^{-\eta \hat{\nabla}_y}\|_1}$
 - 12: $v = k \frac{\min\{v e_{\hat{\nabla}_v}^{\eta \hat{\nabla}_y}, 1\}}{\|\min\{v e_{\hat{\nabla}_v}^{\eta \hat{\nabla}_y}, 1\}\|_1}$
-

MULTIFW implements a Frank-Wolfe style algorithm to simultaneously optimize the multilinear extensions $F_1 \dots F_m$ of the discrete objectives. The algorithm proceeds over T iterations. Each iteration first identifies v^t , a good feasible point in continuous

space (Algorithm 2, line 3). Then, the current solution x^t is updated to add $\frac{1}{T}v^t$ (line 4). The final output is an approximate decomposition of x^T into integral points, produced using the algorithm of [16]. This is a technical step required for the rounding procedure.

The key challenge is to efficiently find a v^t that makes sufficient progress towards *every* objective simultaneously. We accomplish this by introducing the subroutine S-SP-MD (lines 6-12), which runs a carefully constructed version of stochastic saddle-point mirror descent [17]. The idea is to find a v for which $v \cdot \nabla_i F_i(x^{t-1})$ is large enough for all objectives i . We convert this into the saddle point problem of maximizing $\min_{i \in \mathcal{I}} v \cdot \nabla_i F_i(x^{t-1})$. \mathcal{I} denotes the set of objectives i where $W_i - F_i(x^{t-1}) \geq \epsilon$ (i.e., those where we still need to make progress). We let $\Delta(\mathcal{I})$ denote the set of all distributions over i . Our approach only requires *stochastic* gradients, a necessary feature since computing $\nabla_i F(x^{t-1})$ exactly may be intractable when the objective itself is randomized (as in influence maximization).

Specifically, we assume access to two gradient oracles. First, a stochastic gradient oracle $\mathcal{A}_{\text{grad}}^i$ for each multilinear extension F_i . Given a point x , $\mathcal{A}_{\text{grad}}^i(x)$ satisfies $\mathbb{E}[\mathcal{A}_{\text{grad}}^i(x)] = \nabla_x F_i(x)$. Second, a stochastic gradient oracle $\mathcal{A}_{\text{item}}^j$ corresponding to each item $j \in [n]$ (in influence maximization, the items are the potential seed nodes). $\mathcal{A}_{\text{item}}^j(x)$ satisfies $\mathbb{E}[\mathcal{A}_{\text{item}}^j(x)] = [\nabla_{x_j} F_1(x) \dots \nabla_{x_j} F_m(x)]$. We assume that $\|\mathcal{A}_{\text{grad}}^i(x)\|_\infty, \|\mathcal{A}_{\text{item}}^j(x)\|_\infty \leq c$ for some constant c . Linear-time oracles are available for many common submodular maximization problems (e.g., coverage functions and facility location [12]). Given such oracles, we implement a stochastic mirror descent algorithm for the maximin problem. We can interpret the algorithm as solving a game between the max player and the min player. The max player controls v , while the min player controls a variable y representing the weight put on each objective. Intuitively, the min player will put large weights where the max player is doing badly, forcing the max player to improve v . Formally, in each iteration, the players take exponentiated gradient updates (lines 8-12). The max player obtains a stochastic gradient by sampling an objective with probability proportional to the current weights y , while the min player samples an item proportional to v and uses that item's contribution to estimate the max player's current performance on each objective. We prove that these updates converge rapidly to the optimal v . With the subroutine in hand, our main algorithmic result is the following guarantee for Algorithm 1. Here, $b = \max_{i,j} f_i(\{j\})$ is the maximum value of a single item.

Theorem 2. *Given a feasible set of target values $W_1 \dots W_n$, Algorithm 1 outputs a set S such that $f_i(S) \geq (1 - \epsilon) \left(1 - \frac{m}{k(1+\epsilon')\epsilon^3}\right) \left(1 - \frac{1}{e}\right) W_i - \epsilon$ with probability at least $1 - \delta$. Asymptotically as $k \rightarrow \infty$, the approximation ratio can be set to approach $1 - 1/e$ so long as $m = o(k \log^3 k)$. The algorithm requires $O(nm)$ ϵ' -accurate value oracle calls, $O(m \frac{bk^2}{\epsilon} \log \frac{1}{\delta})$ ϵ -accurate value oracle calls, $O\left(\frac{bk^4 c^2}{\epsilon^5} \log\left(n + \frac{bk}{\delta \epsilon}\right)\right)$ calls to $\mathcal{A}_{\text{grad}}$ and $\mathcal{A}_{\text{item}}$, and $O\left(\frac{nk^2 b^2}{\epsilon^2} + \frac{mk^2 b}{\epsilon} + \frac{k^3 b^2}{\epsilon^2}\right)$ additional work.*

This says that Algorithm 1 asymptotically converges to a $(1 - \frac{1}{e})$ -approximation when the budget k is larger than the number of objectives m (i.e., the conditions under which the problem is approximable). All terms in the approximation ratio are identical

to Udvani [22], except that we improve their factor $(1 - \frac{1}{e})^2$ to $(1 - \frac{1}{e})$. The runtime is also identical apart from the time to solve the continuous problem (MULTIFW vs their corresponding subroutine). This is difficult to compare since our respective algorithms use different oracles to access the functions. However, both kinds of oracles can typically be (approximately) implemented in time $O(n)$. Udvani’s algorithm uses $O(n)$ oracle calls, while our’s requires $O(bk^4c^2 \log n)$. For large-scale problems, n typically grows much faster than k , b , and c (all of which are often constants, or near-so). Hence, trading $O(n^2)$ runtime for $O(n \log n)$ can represent a substantial improvement. We present a more detailed discussion in the appendix.

To instantiate Algorithm 1 for influence maximization, we just need to supply appropriate stochastic gradient oracles. To our knowledge, no such oracles were previously known for influence maximization, which is substantially more complicated than other submodular problems because of additional randomness in the objective; naive extensions of previous methods require $O(n^2)$ time. We provide efficient $O(kn \log n)$ time stochastic gradient oracles by introducing a randomized method to simultaneously estimate many entries of the gradient at once (details may be found in the appendix).

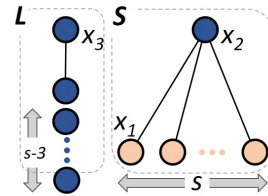
4 Price of Fairness

In this section, we show that both definitions for the Price of Fairness can be unbounded; moreover, allowing nodes to join multiple groups can, counter-intuitively, worsen the PoF. The proofs in this section use undirected graphs. As they are more restrictive, the result naturally hold for directed graphs.

Theorem 3. *As $n \rightarrow \infty$ and $p \rightarrow 0$, $PoF^{\text{Rational}} \rightarrow \infty$.*

Proof. We construct a graph G with two parts. In Part L , we have $s - 1$ vertices all disjoint except for two vertices; label one of these x_3 . In Part S , we have a star with $s + 1$ nodes. Label a leaf node x_1 and the central node x_2 . We define two groups: C_1 is comprised of the s degree-1 vertices of S , and C_2 for the remaining s vertices, which includes the vertices of L and the central vertex x_2 of the star. There are $k = 2$ seeds, and since $|C_1| = |C_2|$, they each have a fair allocation of $k_1 = k_2 = 1$ seeds. The figure below illustrates this network.

Since the subgraph induced by C_1 is comprised of isolated vertices, they have a rational allocation of $\mathcal{I}_{G[C_1]}(1) = 1$. The subgraph induced by C_2 is a collection of isolated vertices and a K_2 , its rational allocation is $\mathcal{I}_{G[C_2]}(1) = 1 + p$.



We are interested in two seeding configurations: $A = \{x_1, x_3\}$ and $B = \{x_2, x_3\}$. We can verify that configuration A is fair. The A activates $1 + p$ nodes in Part L , and $1 + p + (s - 1)p^2$ in Part S , for a total of $\mathcal{I}_G(A) = 2 + 2p + (s - 1)p^2$.

Now consider configuration B . C_1 receives ps influence, and since $p < \frac{2}{n} = \frac{1}{s}$, C_1 does not receive its group rational share of influence. However, we can verify that this seeding is optimal. Part L receives $(1 + p)$ influence, and Part S receives $1 + ps$. Therefore, $\mathcal{I}_G(B) = 2 + p + ps$.

We may then calculate our Price of Fairness:

$$PoF^{\text{Rational}} = \frac{\mathcal{I}_G^{\text{OPT}}}{\mathcal{I}_G^{\text{Rational}}} = \frac{2 + p + ps}{2 + 2p + (s - 1)p^2}$$

And if we take the limit as $n \rightarrow \infty$, $s \rightarrow \infty$, $PoF \rightarrow 1/p$. Finally, as $p \rightarrow 0$, $PoF \rightarrow \infty$.

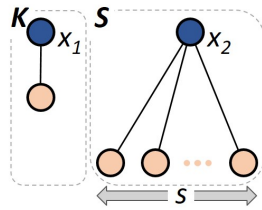
The appendix details a similar result for Maximin Fairness:

Theorem 4. PoF^{Maximin} is unbounded.

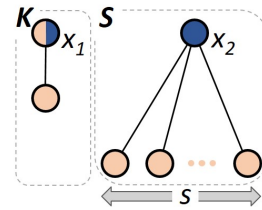
Frequently, an individual may identify with multiple groups. Intuitively, we might expect such multi-group membership to improve the influence received by different groups and make the group-fairness easier to achieve (see the appendix for an example). However, in this section, we show that this is not always true, and giving even a single node membership in a second group can cause the Price of Fairness to worsen by an arbitrarily large amount.

Theorem 5. Given graphs G with groups C_1 and C_2 , and G' with groups C'_1 and C'_2 , where $G' = G$, $C'_1 = C_1$ and C'_2 is obtained from C_2 by the addition of one vertex x_1 ($x_1 \in C_1$, $x_1 \notin C_2$). It is possible for $\lim_{n \rightarrow \infty} \frac{PoF_{G'}^{\text{Rational}}}{PoF_G^{\text{Rational}}} = \infty$.

Proof. Consider a graph G with two components: one component K contains 2 vertices joint by an edge, the other component S is a star with $s + 1$ vertices ($s \geq 1/p$). There are two groups: C_1 contains all degree-1 vertices from S and one vertex from K ; C_2 contains the other vertex x_1 from K and the central vertex x_2 from S . There is one seed ($k = 1$), and the fair allocation of seeds to each group is $k_1 = k_2 = 1$.



G with Disjoint Groups.



G' with Overlapping Groups.

Since the induced subgraphs for both groups comprise only of isolated nodes, the group rational influence for each group is $\mathcal{I}_{G[C_1]} = \mathcal{I}_{G[C_2]} = 1$. Therefore, the seed set $\{x_2\}$ is both fair and optimal, giving an expected influence of $\mathcal{I}_G(\{x_2\}) = 1 + ps$.

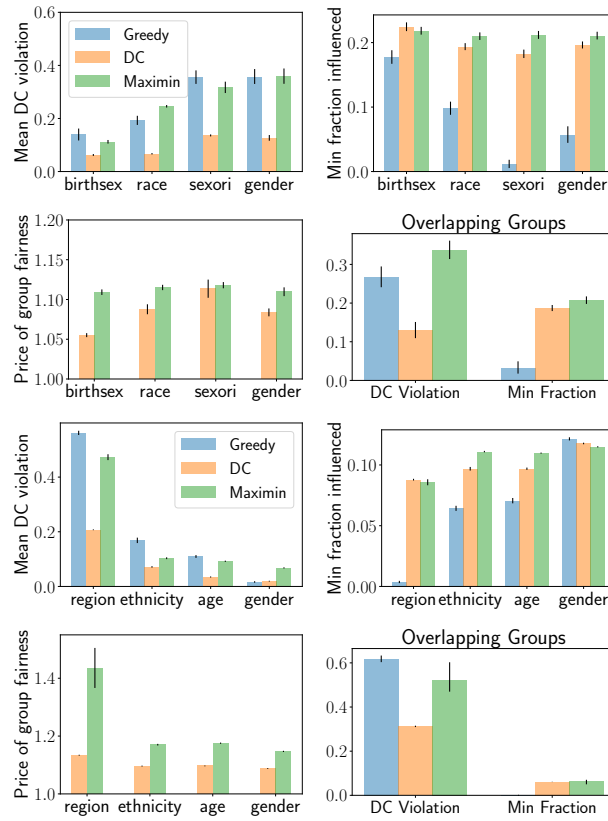


Fig. 1: Average performance on homeless youth social networks (top) and simulated Antelope Valley networks (bottom).

Now, let us modify G by letting x_1 belong to *both* communities to obtain G' , and communities C'_1 and C'_2 . The group rational influence for C'_2 remains the same (its members have not changed) but $\mathcal{I}_{G'}[C'_1]$ has increased to $1 + p$ (by seeding x_1). In fact, this forces the fair allocation to seed x_1 instead of x_2 , for a fair influence of $\mathcal{I}_{G'}(\{x_1\}) = 1 + p$.

$$\text{As } n \rightarrow \infty, \lim_{n \rightarrow \infty} \frac{PoF_{G'}^{\text{Rational}}}{PoF_G^{\text{Rational}}} = \lim_{s \rightarrow \infty} \frac{1+ps}{1+p} = \infty.$$

A more technical construction can demonstrate a similar result for Maximin Fairness, but only as $p \rightarrow \frac{1}{3}^-$; that is, $p < \frac{1}{3}$ as p approaches $\frac{1}{3}$. The proof is provided in the appendix.

Theorem 6. Given graphs G with groups C_1 and C_2 , and G' with groups C'_1 and C'_2 , where $G' = G$, $C'_1 = C_1$ and C'_2 is obtained from C_2 by the addition of one vertex x_1 ($x_1 \in C_1, x_1 \notin C_2$). It is possible for $\lim_{\substack{n \rightarrow \infty \\ p \rightarrow \frac{1}{3}^-}} \frac{PoF_{G'}^{\text{Maximin}}}{PoF_G^{\text{Maximin}}} \rightarrow \infty$.

5 Experimental results

We now investigate the empirical impact of considering fairness in influence maximization. We start with experiments on a set of four real-world social networks which have been previously used for a socially critical application: HIV prevention for homeless youth. Each network has 60-70 nodes, and represents the real-world social connections between a set of homeless youth surveyed in a major US city. Each node in the network is associated with demographic information: their birth sex, gender identity, race, and sexual orientation. Each demographic attribute gives a partition of the network into anywhere from 2 to 6 different groups. For each partition, we compare three algorithms: the standard greedy algorithm for influence maximization, which maximizes the total expected influence (Greedy), Algorithm 1 used to enforce diversity constraints (DC), and Algorithm 1 used to find a maximin fair solution (Maximin). We set the propagation probability to be $p = 0.1$ and fixed $k = 15$ seeds (varying these parameters had little impact). We average over 30 runs of the algorithms on each network (since all of the algorithms use random simulations of influence propagation), with error bars giving bootstrapped 95% confidence intervals.

Figure 1 (top) shows that the choice of solution concept has a substantial impact on the results. For the diversity constraints case, we summarize the performance of each algorithm by the mean percentage violation of the constraints over all groups. For the maximin case, we directly report the minimum fraction influenced over all groups. We see that greedy generates substantial unfairness according to either metric: it generates the highest violations of diversity constraints, and has the smallest minimum fraction influenced. Greedy actually obtains near-zero maximin value with respect to sexual orientation. This results from it assigning one seed to a minority group in a single run and zero in others.

DC performs well across the board: it reduces constraint violations by approximately 55-65% while also performing competitively with respect to the maximin metric (even without explicitly optimizing for it). As expected, the Maximin algorithm generally obtains the best maximin value. DC actually attains slightly better maximin value for one attribute (birthsex); however, the difference is within the confidence intervals and reflects slight fluctuations in the approximation quality of the algorithms. However, Maximin performs surprisingly poorly with respect to diversity constraint violations. This indicates that optimizing exclusively for equal influence spread may force the algorithm to focus on poorly connected groups which exhibit severe diminishing returns. DC is able to attain almost as much influence in such groups but is then permitted to focus its remaining budget for higher impact. Interestingly, the price of fairness is relatively small for both solution concepts, in the range 1.05-1.15 (though it is higher for maximin than for DC). This indicates that while standard influence maximization techniques can introduce substantial fairness violations, mitigating such violations may be substantially less costly in real world networks than the theoretical worst case would suggest.

Finally, the rightmost plot in the top row of Figure 1 explores an example with overlapping groups. Specifically, we consider the race and birthsex attributes so that each node belongs to two groups. Constraint violations are somewhat higher than for

either attribute individually, but the price of fairness remains small (1.07 for DC and 1.13 for Maximin).

In Figure 1 (bottom), we examine 20 synthetic networks used by Wilder et al. [26] to model an obesity prevention intervention in the Antelope Valley region of California. Each node in the network has a geographic region, ethnicity, age, and gender, and nodes are more likely to connect to those with similar attributes. Each network has 500 nodes and we set $k = 25$. Overall the results are similar to the homeless youth networks. One exception is the high price of fairness that maximin suffers with respect to the “region” attribute (over 1.4), but the other PoF values are relatively low (below 1.2). We also observe that greedy obtains the (slightly) best maximin performance for gender, likely because the network is sufficiently well-mixed across genders that fairness is not a significant concern (as confirmed by the extremely low DC violations). Absent true fairness concerns, greedy may perform slightly better since it solves a simpler optimization problem. However, in the last figure, we examine overlapping groups given by region and ethnicity and observe that greedy actually obtains zero maximin value, indicating that there is one group that it never reached across any run.

6 Conclusions

In this paper, we examine the problem of selecting key figures in a population to ensure the fair spread of vital information across all groups. This problem modifies the classic influence maximization problem with additional fairness provisions based on legal and game theoretic concepts. We examine two methods for determining these provisions, and show that the “Price of Fairness” for these provisions can be unbounded. We propose an improved algorithm for multiobjective maximization to examine this problem on real world data sets. We show that standard influence maximization techniques often neglect smaller groups, and a diversity constraint based algorithm can ensure these groups receive a fair allocation of resources at relatively little cost. As automated techniques become increasingly prevalent in society and governance, our technique will help ensure that small and marginalized groups are fairly treated.

References

1. Ahmed, F., Dickerson, J.P., Fuge, M.: Diverse weighted bipartite b -matching. In: Proc. of the 26th IJCAI. pp. 35–41 (2017)
2. Banerjee, A., Chandrasekhar, A., Duflo, E., Jackson, M.O.: The diffusion of microfinance. *Science* **341**(6144) (2013)
3. Barman, S., Biswas, A., Krishnamurthy, S.K., Narahari, Y.: Groupwise maximin fair allocation of indivisible goods. In: Proc. of the 32nd AAAI (2019)
4. Barocas, S., Selbst, A.: Big data’s disparate impact. *California Law Review* **104**, 671 (2016)
5. Benabbou, N., Chakraborty, M., Ho, V., Sliwinski, J., Zick, Y.: Diversity constraints in public housing allocation. In: Proc. of the 17th AAMAS. pp. 973–981 (2018)
6. Brederbeck, R., Faliszewski, P., Igarashi, A., Lackner, M., Skowron, P.: Multiwinner elections with diversity constraints. In: Proc. of the 32nd AAAI. pp. 933–940 (2018)
7. Chekuri, C., Vondrák, J., Zenklus, R.: Dependent randomized rounding via exchange properties of combinatorial structures. In: Proc. of the 51st FOCS. pp. 575–584 (2010)
8. Conitzer, V., Freeman, R., Shah, N., Vaughan, J.W.: Group fairness for the allocation of indivisible goods. In: Proc. of the 32nd AAAI (2019)
9. Fain, B., Munagala, K., Shah, N.: Fair allocation of indivisible public goods. In: Proc. of the 19th EC. pp. 575–592 (2018)
10. Faliszewski, P., Skowron, P., Slinko, A., Talmon, N.: Multiwinner voting: A new challenge for social choice theory. In: Endriss, U. (ed.) *Trends in Computational Social Choice*, chap. 2. AI Access (2017)
11. Hamada, N., Hsu, C., Kurata, R., Suzuki, T., Ueda, S., Yokoo, M.: Strategy-proof school choice mechanisms with minimum quotas and initial endowments. *Artificial Intelligence* **249**, 47–71 (2017)
12. Karimi, M., Lucic, M., Hassani, H., Krause, A.: Stochastic submodular maximization: The case of coverage functions. In: Proc. of the 31st NIPS. pp. 6853–6863 (2017)
13. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proc. of the 9th KDD. pp. 137–146 (2003)
14. Krause, A., McMahan, B., Guestrin, C., Gupta, A.: Robust submodular observation selection. *Jour. of ML Research* **9**(Dec), 2761–2801 (2008)
15. Li, Y., Fan, J., Wang, Y., Tan, K.L.: Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* **30**(10), 1852–1872 (2018)
16. Mirrokni, V., Leme, R.P., Vladu, A., Wong, S.C.: Tight bounds for approximate Carathéodory and beyond. In: Proc. of the 34th ICML. pp. 2440–2448 (2017)
17. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization* **19**(4), 1574–1609 (2009)
18. Schumann, C., Counts, S.N., Foster, J.S., Dickerson, J.P.: The diverse cohort selection problem: Multi-armed bandits with varied pulls. *CoRR* **abs/1709.03441** (2017)
19. Segal-Halevi, E., Suksompong, W.: Democratic fair allocation of indivisible goods. In: Proc. of the 27th IJCAI. pp. 482–488 (2018)
20. Suksompong, W.: Approximate maximin shares for groups of agents. *Mathematical Social Sciences* **92**, 40–47 (2018)
21. Todo, T., Li, R., Hu, X., Mouri, T., Iwasaki, A., Yokoo, M.: Generalizing envy-freeness toward groups of agents. In: Proc. of the 22nd IJCAI. pp. 386–392 (2011)
22. Udvani, R.: Multi-objective maximization of monotone submodular functions with cardinality constraint. In: Proc. of the 32nd NIPS. pp. 9513–9524 (2018)
23. Valente, T., Pumpuang, P.: Identifying opinion leaders to promote behavior change. *Health Education & Behavior* **34**(6), 881–896 (2007)

24. Wilder, B., Immorlica, N., Rice, E., Tambe, M.: Maximizing influence in an unknown social network. In: Proc. of the 32nd AAAI. pp. 4743–4750 (2018)
25. Wilder, B., Onasch-Vera, L., Hudson, J., Luna, J., Wilson, N., Petering, R., Woo, D., Tambe, M., Rice, E.: End-to-end influence maximization in the field. In: Proc. of the 17th AAMAS. pp. 1414–1422 (2018)
26. Wilder, B., Ou, H., de la Haye, K., Tambe, M.: Optimizing network structure for preventative health. In: Proc. of the 17th AAMAS. pp. 841–849 (2018)
27. Yadav, A., Chan, H., Jiang, A.X., Xu, H., Rice, E., Tambe, M.: Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In: Proc. of the 15th AAMAS. pp. 740–748 (2016)
28. Yadav, A., Wilder, B., Rice, E., Petering, R., Craddock, J., Yoshioka-Maxwell, A., Hemler, M., Onasch-Vera, L., Tambe, M., Woo, D.: Bridging the gap between theory and practice in influence maximization: Raising awareness about hiv among homeless youth. In: Proc. of the 27th IJCAI. pp. 5399–5403 (2018)