Titre: Title:	A unifying framework for fairness-aware influence maximization
Auteurs: Authors:	Golnoosh Farnadi, Behrouz Babaki et Michel Gendreau
Date:	2020
Type:	Communication de conférence / Conference or workshop item
Référence: Citation:	Farnadi, G., Babaki, B. & Gendreau, M. (2020, avril). *A unifying framework for fairness-aware influence maximization.* Communication écrite présentée à WWW '20: The Web Conference 2020, Taipei, Taiwan. doi:10.1145/3366424.3383555

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/9245/
Version:	Version finale avant publication / Accepted version Révisé par les pairs / Refereed
Conditions d'utilisation: Terms of Use:	CC BY

Nom de la conférence: Conference Name:	WWW '20: The Web Conference 2020
Date et lieu: Date and Location:	Taipei, Taiwan
Maison d'édition: Publisher:	Association for Computing Machinery
URL officiel: Official URL:	https://doi.org/10.1145/3366424.3383555
Mention légale: Legal notice:	

A Unifying Framework for Fairness-Aware Influence Maximization

Golnoosh Farnadi
Université de Montréal, Mila
Montréal, Canada

Behrouz Babaki
Polytechnique Montréal
Montréal, Canada

Michel Gendreau
Polytechnique Montréal
Montréal, Canada

ABSTRACT

The problem of selecting a subset of nodes with greatest influence in a graph, commonly known as influence maximization, has been well studied over the past decade. This problem has real world applications which can potentially affect lives of individuals. Algorithmic decision making in such domains raises concerns about their societal implications. One of these concerns, which surprisingly has only received limited attention so far, is algorithmic bias and fairness. We propose a flexible framework that extends and unifies the existing works in fairness-aware influence maximization. This framework is based on an integer programming formulation of the influence maximization problem. The fairness requirements are enforced by adding linear constraints or modifying the objective function. Contrary to the previous work which designs specific algorithms for each variant, we develop a formalism which is general enough for specifying different notions of fairness. A problem defined in this formalism can be then solved using efficient mixed integer programming solvers. The experimental evaluation indicates that our framework not only is general but also is competitive with existing algorithms.

CCS CONCEPTS

• **Social and professional topics** → *User characteristics*; • **Theory of computation** → *Mathematical optimization*; • **Networks** → Online social networks.

KEYWORDS

Group Fairness, Influence Maximization, Mixed Integer Programming

1 INTRODUCTION

The problem of finding the set of nodes which have the maximum influence in a graph is known as the *Influence Maximization* (IM) problem. This problem has been well studied over the past decade

and has various applications in domains such as marketing, healthcare, communication, education, agriculture, and epidemiology, among many others [14, 17, 22, 25].

In recent years, with the increasing use of automatic decision making, the concerns about their impact on people's lives has increased. The majority of the existing works on fairness in automated decision making focuses on machine learning approaches, and in particular classification tasks. There exists a vast literature on various techniques to address algorithmic bias in machine learning. These techniques range from processing the data to removing historical discrimination, limiting the unjust treatment of similar individuals, and establishing statistical guarantees to ensure that groups of individuals are not classified based on their sensitive attributes such as race, age, and gender.

In this paper we focus on the less-studied problem of fairness in influence maximization. In many countries, controlling the discrimination in certain domains such as education and hiring is regulated by law. Recently these regulations have been extend to online advertisement (for example, see [26]). An important application of the IM problem is advertising. In fact, it was first introduced under the name of *viral marketing* and aimed at maximizing the profit of an advertiser who targets individuals in a social network [24]. In addition, IM has been used in various applications where the main focus is social good, such as HIV prevention for homeless youth [32] and financial inclusion [1]. Hence it is crucial to ensure that maximizing influence in a network is performed in a way that ensures a diverse spread of influence among various groups and communities. For instance, a health awareness campaign needs to ensure that gender, race, and sexual orientation of the people in the community has no influence on their access to the information.

Fairness is a subjective matter and depending on a situation, different definitions of fairness may be employed. In machine learning there are more than twenty different definitions of fairness [30] and there is some debate on which definition is most suitable in each situation. Similar to any automated decision making framework, it is crucial to develop different variants of fairness-aware IM for different situations. One possible approach for doing so is to design a specific algorithm for each definition. In contrast, we develop a formalism backed by a generic solver to handle all problem variants. As a result, the task of developing a fairness-aware IM algorithm is reduced to specifying the fairness measure in this formalism. Our contributions are:

(1) We review and introduce a number of fairness measures for modeling fairness-aware IM problems. These measures serve as building blocks in our framework. The group-level measures, namely *equity*, *equality*, *maximin* and *diversity* focus on the fair distribution of influence across different groups.

(2) We propose a flexible framework for modeling and solving fairness-aware IM problems. Unlike the majority of methods in IM which solve the problems sub-optimally (possibly with guarantees), our framework obtains optimal solutions.

(3) We conduct extensive experiments to evaluate IM methods subject to various fairness measures. We demonstrate the trade-offs between different fairness measures and the influence.

(4) We compare our proposed approach with existing IM methods on various metrics and show that our proposed approach not only is general but also is competitive with existing algorithms.

The rest of the paper is structured as follows: in Section 2 we review the IM problem and a few IM methods. Next, in Section 3, we address fairness in the IM problem by proposing four notions of fairness. We then convert IM and fairness-aware IM problems into Mixed Integer Programming problems in Section 4. We then show how to encode fairness notions into linear expressions. We evaluate our framework in Section 5, and compare our framework to the related work in Section 6. Finally, we provide numbers of future directions in Section 7 and conclude the paper in Section 8.

2 BACKGROUND

In this section, we review several key topics on which our proposed approach for fairness-aware influence maximization (IM) relies. We begin by reviewing the influence maximization problem and then present a brief overview of a few popular methods for solving this problem.

2.1 The Influence Maximization problem

The IM problem was first studied by Richardson and Domingos [24] as the problem of maximizing the profit of an advertiser in a social network. However, this problem was first presented under the title of influence maximization by Kempe et al. [12].

DEFINITION 1 (THE INFLUENCE MAXIMIZATION (IM) PROBLEM). *Assume a graph $G = (V, E)$ and a diffusion model M that captures the stochastic process of spreading information on G. The* influence *function $\sigma_{G,M} : 2^V \to \mathbb{R}_{\geq 0}$ is a set function defined on any subset of nodes. This function captures the expected number of influenced nodes according to M. The goal of the IM problem is to find a seed set $S \subset V$ with $|S| = k$ such that $\sigma_{G,M}(S)$ is maximized. For the rest of the paper, when the context is clear, we omit the subscript of $\sigma_{G,M}(S)$.*

The influence of each user is defined based on a *diffusion model*. In a generic diffusion model, each user $u \in V$ is either *active* or *inactive*. First, we assume all users in G are inactive. Next, a set of k users called the *seed set S* is selected from G, and all the users in S are categorized as active users. Then a stochastic activation process begins and continues until no further activation is possible. There exists an extensive amount of literature on various diffusion models such as linear threshold model, triggering model, and time aware model. In this paper, we focus on the most commonly used model, called the *Independent Cascade (IC) model*.

DEFINITION 2 (THE INDEPENDENT CASCADE (IC) MODEL). *Given $G = (V, E)$, where each edge $e \in E$ has a propagation probability $p(e) \in [0, 1]$, the influence propagation process defined by IC is as*

follows: At the first step, only nodes in the seed set S are active. If node u is activated at step i and it is connected to an inactive node v by a directed edge e, there is a probability $p(e)$ that u activates v at step $i + 1$. Each node only has one chance to influence its neighbors. An activated node always remains activated. This process continues until it is not possible to activate further nodes.

The quality of a solution of an IM problem is determined by its *spread function*.

DEFINITION 3 (THE SPREAD FUNCTION). *Given $G = (V, E)$ and $S, A \subset V$, the spread function $I(S, A)$ is the number of nodes in A activated according to a realization of the IC model starting from the seed set S. Note that $\mathbb{E}[I(S, V)] = \sigma(S)$.*

Computation of $\mathbb{E}[I(S, V)]$ is #$\mathcal{P}$-hard [12]. However, Kempe et al. propose a Monte Carlo approach for estimating $\mathbb{E}[I(S, V)]$ from samples of G. Assume a distribution over induced subgraphs of G, such that the probability of subgraph $g = (V_g, E_g)$ is as follows:

$$p(g) = \prod_{e \in E_g} p_e \prod_{e \in E \setminus E_g} (1 - p_e) \tag{1}$$

Moreover, for a given subgraph g, let $R_g(S)$ denote the set of nodes in V_g which are reachable from S in g. Kempe et al. show that under the IC model the influence function $\sigma(S)$ is equal to the expected value of $R_g(S)$, that is:

$$\mathbb{E}[I(S, V)] = \sum_g p(g) R_g(S) \tag{2}$$

This suggests a method for estimating $\mathbb{E}[I(S, V)]$: Given p instances of G, $\mathbb{E}[I(S, V)] \approx \sum_{i=1}^{P} R_{g_i}(S)/p$. To sample an instance g according to the probability distribution of Equation 1, we remove each edge $e \in E$ with probability $1 - p(e)$.

2.2 Solving the IM problem

It has been shown that the IM problem is generally NP-hard and hence intractable unless $P = NP$. However, when the influence function has certain properties, the optimal solution can be approximated by an efficient algorithm. These properties are *monotonicity* and *submodularity*.

DEFINITION 4 (MONOTONICITY). *An influence function $\sigma(.)$ is monotone iff $\sigma(S') \leq \sigma(S)$ for any $S' \subset S \subseteq V$.*

The monotonicity means that adding more nodes to a seed set S increases the influence spread.

DEFINITION 5 (SUBMODULARITY). *An influence function $\sigma(.)$ is submodular iff $\sigma(\{v\} \cup S') - \sigma(S') \geq \sigma(\{v\} \cup S) - \sigma(S)$ for any $S' \subset S \subseteq V$ and $v \in V \setminus S$.*

Kempe et al. [12] introduced a a greedy algorithm which obtains an $1 - \frac{1}{e}$ approximation for the IM problem, assuming that the influence function is monotone and submodular. They also demonstrate that under the IC diffusion model, the influence function has these properties.

DEFINITION 6 (THE GREEDY ALGORITHM FOR IM). *The greedy algorithm starts with an empty set S and iteratively selects a user which provides the maximum marginal gain to the influence function $\sigma(.)$ and adds it to the set S.*

Despite the efficiency of the greedy method, IM is still a very challenging problem to solve because evaluating the influence function $\sigma(.)$ is #$\mathcal{P}$-hard (see Li et al. [20] for details). Existing algorithms for solving IM are categorized into three main groups according to the way that they approximate the influence function. The *simulation based* methods such as *celf* [17] perform Monte-Carlo sampling for evaluating influence function. The *proxy based* algorithms such as *simpath* [9] replace the influence functions with proxy functions. The main focus of this type of algorithms is efficiency. Finally, the *sketch based* algorithms such as *tim* [28] and *imm* [27] speed up the evaluation of the influence function by first constructing sketches, which are subgraphs induced by an instance of the influence process. There are two types of sketches, namely forward influence (FI) and reverse reachable (RR). We refer the reader to a recent survey for a detailed discussion of variants of IM and existing methods for solving them [20].

3 FAIRNESS-AWARE INFLUENCE MAXIMIZATION

To address fairness in an IM problem, we first need to define what we mean by fairness-aware IM. There are mainly two categories of fairness metrics in machine learning: *group fairness* and *individual fairness*. Group fairness metrics are statistical measures which are defined based on an attribute or set of attributes that defines two groups of individuals. Measure such as *equal opportunity*, *equalized odds*, and *demographic parity* [10] among many others attempt to define a metric that satisfies a fairness property among two groups of individuals. While group fairness approaches focus on group-level fairness, individual fairness [5] considers similar treatment among two similar individuals. In this paper our main focus is on defining various group fairness notions in IM.

To represent different populations in G, assume we have m groups, i.e., $C = \{C_1, C_2, \ldots, C_m\}$ where each node $v \in V$ belongs to at least one group, i.e., $V_{C_1} \cup V_{C_2} \cup \ldots \cup V_{C_m} = V$.

The first notion of fairness that we introduce is *equality* which focuses on the fair allocation of seeds to the groups proportional to the size of the group within the population. This notion of fairness satisfies the fair distribution of resources among groups. Assuming an advertising campaign that aims to spend its budget fairly among various demographic groups, which means to have fair share of leaders among different groups.

DEFINITION 7 (EQUALITY). *Given a a set of groups C, the equality constraint requires that the number of seed nodes in each group $C \in C$ is proportional to the population ratio of that group, that is, $|S \cap V_C|/k \approx |C|/|V|$.*

The second notions is *equity*. While equality focuses on the fair distribution among the seed nodes, equity aims to minimize the unfair outcome by fair treatment among different groups proportional to their size in the network. Therefore, we may start with disproportionate share of various groups that reaches a more diverse population at the end.

DEFINITION 8 (EQUITY). *Given a set of groups C, the equity constraint requires that the expected number of influenced nodes in a group $C \in C$ is proportional to the its population ratio, i.e. $\mathbb{E}[I(S, V_C)]/\mathbb{E}[I(S, V)] \approx |C|/|V|$.*

The two measures that we visit next, namely *maximin* and *diversity* were first introduced by [29]. The maximin measure is closely related to equity. This measure is closely related to the legal notion of *disparate impact* where the goal is to minimize the gap between different groups in terms of received influence relative to their size. We want to maximize the minimum relative influence received by any group.

DEFINITION 9 (MAXIMIN). *Given a set of groups C, the maximin criterion aims to maximize $\min_{C \in C} \mathbb{E}[I(S, V_C)]/|C|$, which is the minimum influence among all groups relative to their populations.*

If members of one community are not well-connected in a network, enforcing maximin fairness becomes very costly as assigning seeds to the members of that community leads to limited impact on the overall influence. The diversity measure overcomes this problem by allocating resources according to the internal topology of each community. It is done in two stages. In the first stage, a fraction of seeds proportionate to the size of each community is used for influencing the induced sub-graph of that community. In the second stage, the internal spread of influence among each community is used as a lower bound on the influence received by the nodes of that group. Hence, diversity is not giving each group fair allocation of resources like equality nor fair share of influence similar to equity and maximin. Instead, it guarantees that each group receives influence at least equal to their internal spread of influence.

DEFINITION 10 (DIVERSITY). *For a group C, let G_c denote the graph induced from G by the nodes in C, $k_c = \lceil k \cdot |C|/|V| \rceil$, and $OPT_C = \max_{S \subset C:|S|=k_C} \sigma_{G_C}(S)$. Given a set of groups C, the diversity constraint requires that for each $C \in C$ it holds that $\mathbb{E}[I(S, V_C)] \geq OPT_C$.*

3.1 Hardness of fairness-aware IM

As we discussed in Section 2, the IM problems are computationally challenging to solve. However since the influence function is monotone and submodular, the IM problems can be solved with a greedy algorithm with a $1 - \frac{1}{e}$ approximate ratio. However, the influence function of the fairness-aware IM problems does not have these properties.

THEOREM 1. *The influence function of the fairness-aware IM problems with equity, equality, maximin, and diversity notions of fairness is neither monotone nor submodular.*

This theorem is proved for maximin and diversity measures in [29]. This proof can be easily extended to the other two variants.

4 FORMULATING FAIRNESS-AWARE IM IN MIXED INTEGER PROGRAMMING

We will now show how the different variations of fairness-aware IM which we presented earlier can be modeled and solved as instances of a general framework. This framework is based on converting the IM problems into Mixed Integer linear Programming (MIP) problems. This framework is declarative, meaning that the user only needs to specify the problem, and the solving is done by a MIP solver. This eliminates the need for developing specific algorithms for different problem settings.

Table 1: Definitions and MIP encodings of fairness measures in influence maximization ($0 \leq \epsilon < 1$).

name	mathematical notation	mixed integer programming encoding
equality	$\|S \cap V_C\|/k \approx \|C\|/\|V\|$	$\sum_{j \in C} s_j \leq k \cdot \|C\|/\|V\| + \epsilon$ $\sum_{j \in C} s_j \geq k \cdot \|C\|/\|V\| - \epsilon$
equity	$\mathbb{E}[I(S, V_C)]/\mathbb{E}[I(S, V)] \approx \|C\|/\|V\|$	$\sum_{p=1}^{m} \sum_{j=1}^{n} a_i^p - \|C\|/\|V\| \cdot \sum_{p=1}^{m} \sum_{j=1}^{n} a_j^p \leq \epsilon$ $\sum_{p=1}^{m} \sum_{j=1}^{n} a_i^p - \|C\|/\|V\| \cdot \sum_{p=1}^{m} \sum_{j=1}^{n} a_j^p \geq -\epsilon$
maximin	maximize $\min_{C \in \mathcal{C}} \mathbb{E}[I(S, V_C)]/\|C\|$	maximize $x_{\min}$ $x_{\min} \cdot \|C\| \leq \sum_{p=1}^{m} \sum_{j \in C} a_j^p \quad \forall C \in \mathcal{C}$
diversity	$\mathbb{E}[I(S, V_C)] \geq \mathrm{OPT}_C$	$\sum_{p=1}^{m} \sum_{j \in C} a_j^p \geq \mathrm{OPT}_C$

4.1 Encoding Influence Maximization in MIP

We start with presenting the MIP encoding for the basic influence maximization problem. This MIP encoding is taken from [19]. We introduce the binary variables $s_i, i \in \{1, \ldots, n\}$ as indicators of whether node i is selected as a seed. Recall that the influence function can be approximated by the average count of *influenced* nodes in a number of samples generated from the graph. Assume that we have m such samples. Let the variable a_i^p denote whether node v_i is influenced in sample p. Finally, given a graph $G = (V, E)$, let us define $H(G, i)$ as the set of indices of nodes that can influence v_i, that is, $H(G, i) = \{i\} \cup \{j : v_j \in V$ and there is a path from v_j to v_i in $G\}$. Then the IM problem can be expressed as:

$$\text{maximize} \sum_{p=1}^{m} \sum_{i=1}^{n} a_i^p$$

s.t.

$$a_i^p \leq \sum_{j \in H(g^p, i)} s_j \qquad 1 \leq i \leq n, \, 1 \leq p \leq m \qquad (3)$$

$$\sum_{i=1}^{n} s_i \leq k \qquad\qquad\qquad\qquad\qquad (4)$$

$$s_i \in \{0, 1\} \qquad\qquad\qquad 1 \leq i \leq n \qquad (5)$$

$$a_i^p \in [0, 1] \qquad\qquad 1 \leq i \leq n, \, 1 \leq p \leq m \qquad (6)$$

When the objective is maximizing the influence spread, constraints of Equation 3 are sufficient to ensure that a_i^p is one if and only if node v_i is influenced in sample p. We can make this condition hold regardless of the objective function, too. For this to hold, we should enforce that $a_i^p \in \{0, 1\}$ and $a_i^p \geq |H(g^p, i)| \sum_{j \in H(g^p, i)} s_j$.

4.2 Encoding Fairness Measures in MIP

We will now show that the fairness measures introduced in Section 3 can be encoded as linear expressions. This implies that one can turn the MIP formulation of IM into a fairness-aware model by including these expressions as constraints or objective function.

4.2.1 equality. The equality measure concerns the number of seeds dedicated to each group. For group C, this value is captured by the expression $\sum_{j \in C} s_j$.

4.2.2 equity. The core constructs of the equity constraint are the total influence $\mathbb{E}[I(S, V)]$ and the group influence $\mathbb{E}[I(S, V_C)]$. Since we

are approximating the influence by sum of spread in samples, these quantities can be expressed by $\sum_{p=1}^{m} \sum_{j=1}^{n} a_j^p$ and $\sum_{p=1}^{m} \sum_{j \in C} a_j^p$, respectively.

4.2.3 maximin. This measure concerns the *smallest* group influence. To encode this quantity, we introduce an auxiliary variable x_{min} and constrain it to be smaller or equal to the group influence of all groups. When the objective is maximizing x_{min}, it will be equal to the smallest group influence.

4.2.4 diversity. The diversity constraints are inequality constraints over the influence in a group and the whole graph. These inequalities can be simply formulated using the encodings of $\mathbb{E}[I(S, V_C)]$ and $\mathbb{E}[I(S, V)]$ which were introduced before. Table 1 summarizes the MIP encodings of all fairness measures.

5 EXPERIMENTAL EVALUATION

We ran experiments on machines with 40 Intel Skylake processors (2.4GHz) and 202GB of RAM running Linux Centos 7. The timeout was 7200 seconds. The MIP solver is Gurobi-8.1.1[1]. The code and data will be available upon publication of this paper.

We examined our framework using the 20 synthetic networks that model an obesity prevention intervention in the Antelope Valley region of California [31]. Each network in the Antelope Valley dataset has 500 nodes, and 1576–1697 edges. Nodes in this dataset have labels from which we select the sensitive attributes *gender*, *age* and *ethnicity* to determine the groups. In this dataset gender includes 2 classes, age has 7 classes and ethnicity contains 5 classes. In all the experiments, similar to [29], we set $k = 25$. For the equality and equity, we set $\epsilon = 0.1$. All evaluations are under the IC model with the commonly used edge probabilities $p(e_{u,v}) = 1/\text{IN-DEGREE}(v)$. The number of samples in the MIP models is 100 unless otherwise stated, and the solution quality is evaluated by simulation over $10,000$ samples.

5.1 Evaluation of Fairness-aware IM

Our fairness-aware models are all based on a MIP model for the IM problem. To evaluate various notions of fairness, we first compare the quality of solutions of the MIP model. We aim to determine how the number of samples affects the performance of the fairness-blind MIP model. We evaluate our MIP model with various number of

---

[1] www.gurobi.com

samples in {10, 100, 200, 500, 1000}. Figure 1 shows that the average quality of the solutions obtained by the MIP models increases with the number of samples.
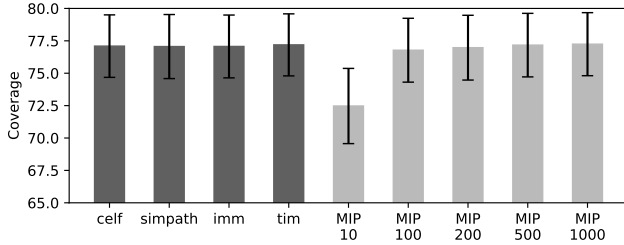
Figure 1: Coverage obtained by MIP models and existing IM algorithms. The numbers below the MIP models show the number of samples.

To determine the quality of the solutions obtained by the MIP model, we also compare the performance of the MIP models with the state-of-the-art algorithms for influence maximization, namely celf, simpath, imm and tim (a brief explanation of each algorithm is provided in Section 2). All competing methods use 10, 000 samples. The results presented in Figure 1 show that despite using fewer samples, the MIP models provides comparable or even better solutions than the competing methods. Note that increasing the size of the samples in MIP significantly increase the run time of the models. The results show that models with only 100 samples are solved within reasonable time and are competitive in terms of solution quality. Using less than 100 samples is not practical as the quality of solutions drops drastically. Using 200 samples, models perform very similar to the celf, simpath and imm models. The best performing state-of-the-art IM approach is tim and the MIP models with 1000 samples perform similarly to this approach in terms of overall coverage. Considering the trade-off between the runtime an solution quality, we choose to run the fairness-aware MIP models with 100 samples.

Next, we determine the quality of our fairness-aware models by comparing our MIP models for fairness-aware IM according to their overall coverage. Figure 2 present the results for all labels. We compare the performance of our MIP models for fairness-aware IM with the best performing state-of-the-art fairness-blind IM, i.e., tim.

Tsang et. al [29] propose a metric called *price of fairness* (POF) to determine the cost of fair allocation among diverse groups. Price of fairness is defined as the ratio of optimal unconstrained total influence to the total influence subject to fairness constraints. Figure 2 presents the price of fairness for our MIP models.

An interesting observation is that the price of fairness is higher when the number of communities are larger. Hence, POF is higher when we consider age which has 7 categories compared with ethnicity with 5 categories and gender with only two labels. Moreover, the POF is less for those methods which focus on providing equality, i.e., equal number of influenced nodes in each community. Hence, for the case of equality, in which we assign fair seeds to each group, is really high. Equality and equity have higher POF and less overall coverage compared to the maximin and diversity methods. And

the maximin approach has the overall best coverage among the fairness notions across different attributes.

The choice of fairness measure has a substantial effect on the result. To better understand the differences among different notions of fairness, we perform a deeper analysis on their differences. In Figure 4, we evaluate all fairness-aware methods using *L1 norm*, which calculates the sum of absolute differences of coverage in various communities from the average coverage. This measure quantifies the degree of balance in terms of coverage among various communities. From Figure 3 and Figure 4, we observe that both maximin and equity are balanced wrt the number of influenced nodes in each community whereas the diversity approach performs poorly. Interestingly, even the blind approach, i.e., tim, performs better than the diversity approach. As expected, the performance of equality is not better than the equity or the maximin approaches. However it still performs better than the blind model as it at-least guarantees to provide similar opportunities for each community.

We also evaluate the notion of fairness based on their coverage for the most disadvantaged group. Both equity and maximin methods perform well to give opportunity to the least advantaged group while the diversity and equality methods perform poorly in this metric. Equality still performs better than the blind method while diversity performs even worst than TIM when the attributes are not balanced, i.e., for the case of age and ethnicity.

Among the four notions of fairness that we study in this section, the maximin approach not only has higher coverage and lower POF, but it better distributes the opportunity among various communities and takes care of the most disadvantaged group. Note that in all these experiments we set $\epsilon = 0.1$ meaning that the differences between the groups using the equity approach is only 10%. However, it is possible to use higher/lower ϵ to control the distribution of opportunity among various groups. Moreover, the maximin approach can potentially perform very costly if our disadvantaged community is poorly connected. While the structure of the graph and how the members of each community are connected to each other has a huge impact on the performance of maximin, other approaches are less affected by the topology of the graph. Finally, the stochastic nature of propagation can affect the performance and one can measure whether different communities propagate similarly or not. To this end, the equality measure can provide a guarantee that all communities are treated similarly and resources are distributed evenly among them and it is due to them to propagate it accordingly.

5.2 Comparison with existing methods

We will now compare the performance of existing methods for fairness-aware influence maximization with their counterpart MIP models. The group maximin and diversity problems have been introduced by Tsang et al. [29] and solved using multi-objective submodular optimization. We compare the quality of solutions obtained by this approach (which we will refer to as *fairIM*) with the MIP models of maximin and diversity.

5.2.1 Maximin. The objective in the maximin problem is to maximize the influence received by the least well-off group. We compare the two approaches in terms of this quantity. Figure 5 shows the average value of this quantity over all networks for each attribute.
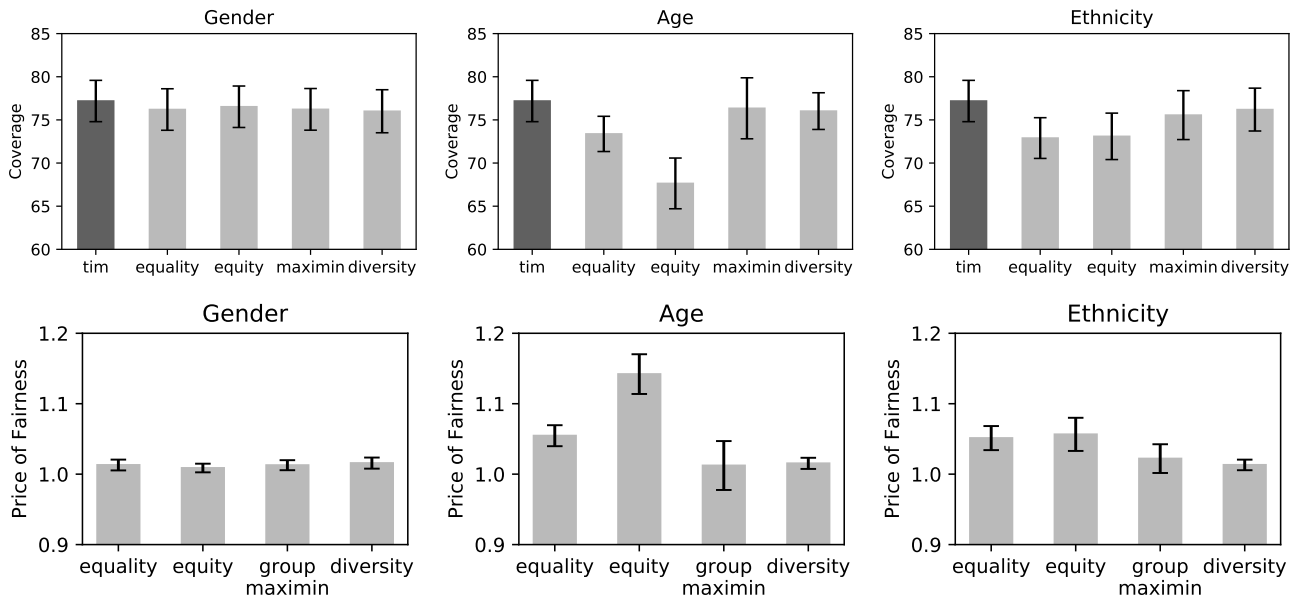
Figure 2: Coverage (top row) and price of fairness (bottom row) obtained by fairness-aware MIP models for attributes gender, age, and ethnicity.
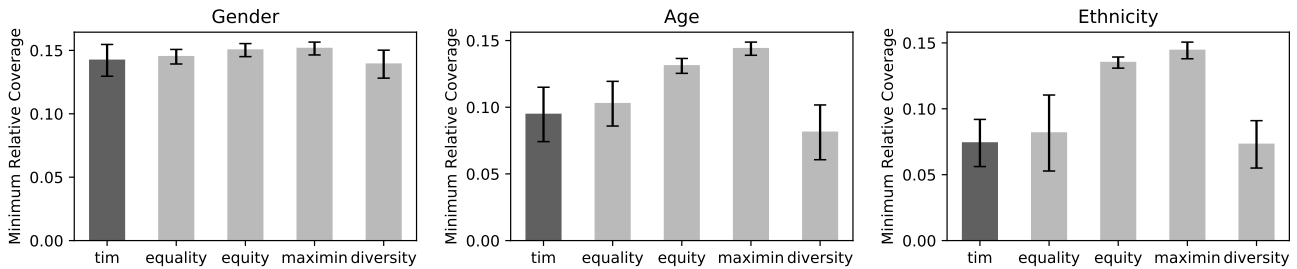

Figure 3: Comparing various notion of fairness in IM in terms of the coverage received by the most disadvantaged group for attributes gender, age, and ethnicity.
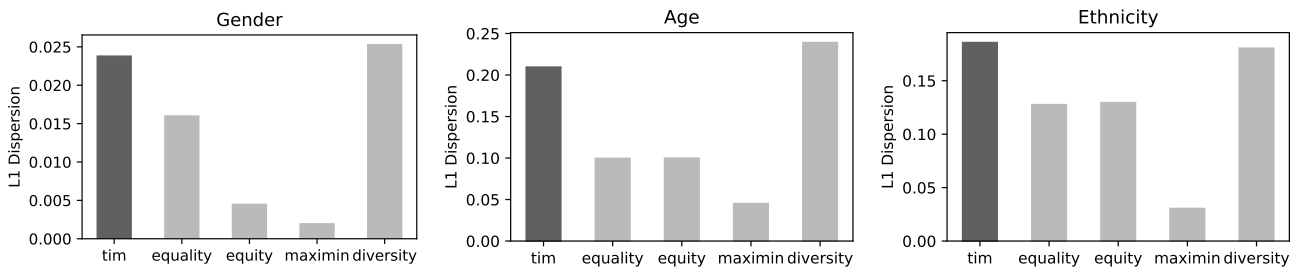

Figure 4: L1 Dispersion obtained by fairness-aware MIP models for attributes gender, age, and ethnicity.

The results indicate that the MIP model is competitive with the existing method. The advantage of our approach over fairIM is more visible for the *gender* attribute. Interestingly, this improvement in fairness does not lead to a drop in overall influence. This is reflected in Figure 6 which shows the price of fairness for both methods.

5.2.2 Diversity. We compare our MIP model and fairIM for solving the problem of influence maximization subject to diversity constraints [29]. The objective function in these problems is the total coverage. Figure 7 compare the solutions obtained by these methods in terms of their objective function. The results indicate that the
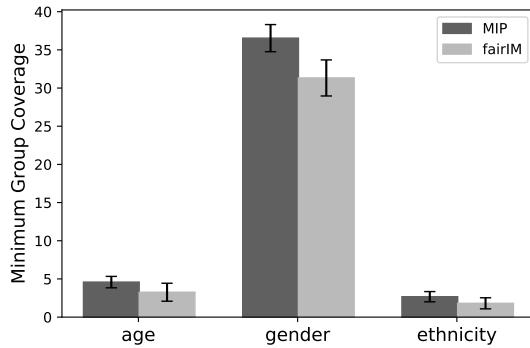
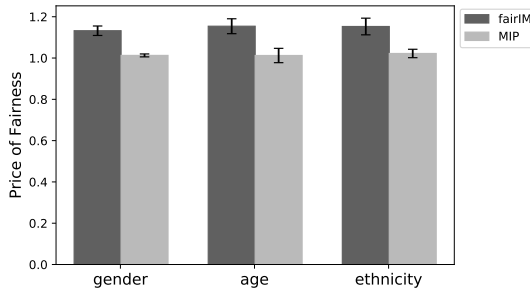**Figure 5: Comparing the objective values obtained by MIP and fairIM methods for the maximin problem.**

Figure 6: Comparing the price of fairness in solutions obtained by MIP and fairIM methods for the maximin problem.

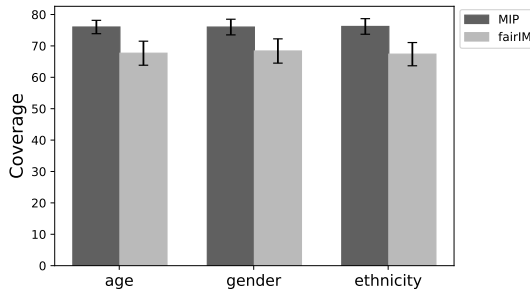MIP model consistently outperforms fairIM in terms of objective value.

Figure 7: Comparing the objective values obtained by MIP and fairIM methods for the diversity problem.

Next we compare the two methods in terms of violation of diversity constraints. Although these constraints are enforced as hard constraints in the MIP model, they can be still violated when evaluated on a set of samples different from the ones included in the MIP model. Figure 8 shows the average violations on all graphs for different attributes. The results of Figures 7 and 8 indicate that the MIP model dominates fairIM both in terms of objective value and

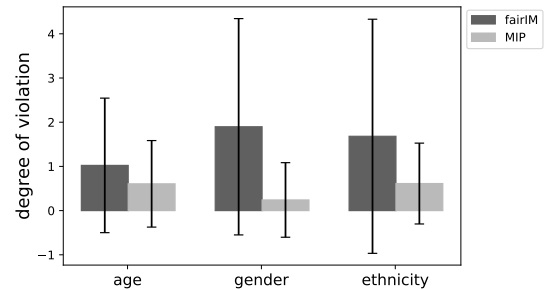constraint violation, while obtaining solutions within a reasonable runtime.

Figure 8: Comparing the violation of diversity constraints in solutions obtained by MIP and fairIM methods.

6 RELATED WORK

In this section we position our work in relation to the literature of influence maximization and fairness in AI.

Influence Maximization. Although IM problem has been introduced in 2003 by Kempe et al. and despite the fact that it has been proved that a greedy algorithm can solve it with a $1 - 1/e$-optimally, a large literature exists in AI community on various approaches on making it more scalable [21], time-aware [13], cost-aware [23], topic-aware [2] and location-aware [18]. Also, besides the technical aspects of the IM problem, there have been efforts to address the challenges of deploying IM in real-world applications, e.g., healthcare [32]. However, there is a lack in the literature in addressing fairness-aware IM problems. There are only a few attempts in recent years to introduce definitions of fairness in IM. Group fairness has been first studied by Tsang et al. [29]. Similarly, the definition of diversity for access gap is first introduced by Fish et al. [7] as an individual-level definition of fairness in IM. Balance constraint as two-stage problem is introduced in Gershtein et al. [8]. Also the topology of the network for various labels to increase diversity among the influenced nodes is studied in [3]. Our work differs from these studies because first, we look at the IM problem from the generic perspective. Our approach not only allows us to define all the current definitions of fairness in our frameworks, but also gives the possibility to define new fairness concepts. For example, any definition consisting of a combination of existing fairness measures can be easily encoded in our framework. In addition, in the experimental evaluation, our approach has proved to be more efficient than the existing methods.

Fairness in AI. The most notable approaches include fairness through awareness [5], individual fairness [34], statistical parity, disparate impact, and group fairness [4, 6, 11], counterfactual fairness [16], preference-based fairness [33], and equality of opportunity [10]. The goal of the above mentioned works is to assure the fair treatment of individuals or groups that are identified by sensitive attributes. The main goal of individual fairness measures is that similar individuals are treated similarly. The most well-known individual fairness notion is the Lipschitz condition [5] which bounds

the distance between the outcome of the algorithm for two individuals by a function of their similarity. In this paper, we focus on group fairness. A notable example of group fairness is demographic parity, which expresses that a system is fair if the probability of getting a favorable decision is equal between two groups. Assuming that the favorable decision in the IM problem is receiving influence, then *equality* and *maxmin* are closely related to the definition of demographic parity. Our notion of group allow us to satisfy demographic parity for any arbitrary subgroups defined over the set of sensitive attributes. Note that subgroup can be expressed as a union of joint assignments to the sensitive features.

7 DISCUSSION AND FUTURE DIRECTIONS

Enforcing fairness in the IM problem is a challenging task and only recently received some attention. Most of the existing literature on the IM problem revolves around the submodularity of the influence function. This property plays an important role in designing effective and efficient IM solutions. However, fairness-aware IM has to deal with non-submodular influence functions, and therefore most of the existing attempts for the IM problem are no longer effective and need to be re-visited in this context.

Scalability is also a major concern for any IM algorithm to make it feasible to be used in real life applications. For the fairness-aware IM the issue is even harder to tackle due to its complex nature. In this paper, as in similar existing work, we use MC sampling to estimate the influence function, which increases the size of the MIP problem relative to the number of the samples. An interesting path to explore in the future is to design more effective and efficient sampling approaches to reduce the size of the optimization problem.

Besides scalability, an important issue with the current sampling techniques is that they cannot provide any robustness guarantee. It has been shown that a small change in the graph structure, which is very common in real world networks, can change the solution drastically. Providing a robust fairness-aware IM solution may require new diffusion models to encode the dynamic nature of the network.

In machine learning, many of the fairness metrics are defined in terms of the confusion matrix, which consists of counts of true positives, false positives, true negatives and false negatives. As Kleinberg et al. [15] noted by the impossibility theorem, it is often impractical to satisfy all fairness definitions simultaneously. Although it is feasible to express all fairness measures in IM as linear constraints in our framework, it is important to note that such limitation exists and we are not able to satisfy all definitions of fairness e.g., equality, equity, maximin and diversity at the same time.

We show that addressing fairness to ensure diverse allocation of resources in graphs is expensive. An interesting path to explore is to study the trade-offs between the price of fairness and various fairness metrics in IM. Since our framework is highly expressive, we can impose such control on the price of fairness through addition of suitable linear constraints.

Finally, in this work we mainly focus on group-level definitions of fairness in IM. However, our framework is easily extensible to individual fairness measures in which the main focus is the welfare of all individuals despite the communities to which they belong.

Addressing individual fairness in our framework is thus a promising future direction.

8 CONCLUSION

In this paper, we demonstrate a unifying framework for the fairness-aware IM problems from an optimization perspective. Since fairness is subjective and IM problem are applied in different real-world domains, a single fairness definition cannot be used in all situations. We present multiple definitions of fairness and demonstrate their relative advantages along with their differences depending on the context. We also study the trade-offs between enforcing fairness and the loss of total influence. We propose a unifying framework which allows the user to combine various definitions of fairness (if desired) and potentially encode new definitions.

We give new perspectives to consider diversity in IM problem at the group level given node labels. Our extensive evaluation indicate that addressing discrimination is expensive no matter which group-level notion of fairness we choose. We demonstrate that each notion of fairness has certain properties and we may not be able to satisfy all these properties with a single fairness metric. Finally, we discuss the limitations and opportunities of fairness-aware IM and mention a number of open questions to explore in the future.

Acknowledgments

REFERENCES

[1] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2013. The diffusion of microfinance. *Science* 341, 6144 (2013), 1236498.
[2] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2013. Topic-aware social influence propagation models. *Knowledge and information systems* 37, 3 (2013), 555–584.
[3] Antonio Caliò, Roberto Interdonato, Chiara Pulice, and Andrea Tagarelli. 2018. Topology-driven diversity for targeted influence maximization with application to user engagement in social networks. *IEEE Transactions on Knowledge and Data Engineering* 30, 12 (2018), 2421–2434.
[4] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *CoRR* abs/1703.00056 (2017).
[5] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. ACM, 214–226.
[6] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 259–268.
[7] Benjamin Fish, Ashkan Bashardoust, danah boyd, Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2019. Gaps in Information Access in Social Networks. *The Web Conference (WWWW)* (2019).
[8] Shay Gershtein, Tova Milo, Brit Youngmann, and Gal Zeevi. 2018. IM Balanced: Influence Maximization Under Balance Constraints. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1919–1922.
[9] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. 2011. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *2011 IEEE 11th international conference on data mining*. IEEE, 211–220.
[10] Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*. 3315–3323.
[11] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2012. Fairness-aware classifier with prejudice remover regularizer. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 35–50.

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

[13] Jinha Kim, Wonyeol Lee, and Hwanjo Yu. 2014. CT-IC: Continuously activated and time-restricted independent cascade model for viral marketing. *Knowledge-Based Systems* 62 (2014), 57–68.

[14] Masahiro Kimura, Kazumi Saito, and Hiroshi Motoda. 2008. Minimizing the Spread of Contamination by Blocking Links in a Network.. In *AAAI*, Vol. 8. 1175–1180.

[15] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2016. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807* (2016).

[16] Matt J. Kusner, Joshua R. Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual Fairness. In *NIPS*. 4069–4079.

[17] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 420–429.

[18] Guoliang Li, Shuo Chen, Jianhua Feng, Kian-lee Tan, and Wen-syan Li. 2014. Efficient location-aware influence maximization. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. ACM, 87–98.

[19] Xiang Li, J David Smith, Thang N Dinh, and My T Thai. 2019. TipTop: Almost Exact Solutions for Influence Maximization in Billion-Scale Networks. *IEEE/ACM Transactions on Networking (TON)* 27, 2 (2019), 649–661.

[20] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. 2018. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* 30, 10 (2018), 1852–1872.

[21] Feng Lu, Weikang Zhang, Liwen Shao, Xunfei Jiang, Peng Xu, and Hai Jin. 2017. Scalable influence maximization under independent cascade model. *Journal of Network and Computer Applications* 86 (2017), 15–23.

[22] Flaviano Morone and Hernán A Makse. 2015. Influence maximization in complex networks through optimal percolation. *Nature* 524, 7563 (2015), 65.

[23] Hung T Nguyen, Thang N Dinh, and My T Thai. 2016. Cost-aware targeted viral marketing in billion-scale networks. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

[24] Matthew Richardson and Pedro Domingos. 2002. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 61–70.

[25] Yaron Singer. 2012. How to win friends and influence people, truthfully: influence maximization mechanisms for social networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 733–742.

[26] Richard Steppe. 2017. Online price discrimination and personal data: A General Data Protection Regulation perspective. *Computer law & security review* 33, 6 (2017), 768–785.

[27] Youze Tang, Yanchen Shi, and Xiaokui Xiao. 2015. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 1539–1554.

[28] Youze Tang, Xiaokui Xiao, and Yanchen Shi. 2014. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. ACM, 75–86.

[29] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. 2019. Group-fairness in influence maximization. *arXiv preprint arXiv:1903.00967* (2019).

[30] Sahil Verma and Julia Rubin. 2018. Fairness definitions explained. In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*. IEEE, 1–7.

[31] Bryan Wilder, Han-Ching Ou, Kayla de la Haye, and Milind Tambe. 2018. Optimizing Network Structure for Preventative Health. In *AAMAS*. International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, USA / ACM, 841–849.

[32] Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, and Darlene Woo. 2018. Bridging the Gap Between Theory and Practice in Influence Maximization: Raising Awareness about HIV among Homeless Youth.. In *IJCAI*. 5399–5403.

[33] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, Krishna P. Gummadi, and Adrian Weller. 2017. From Parity to Preference-based Notions of Fairness in Classification. In *NIPS*. 228–238.

[34] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning fair representations. In *International Conference on Machine Learning*. 325–333.