On the Fairness of Time-Critical Influence Maximization in Social Networks

Junaid Ali, Mahmoudreza Babaei, Abhijnan Chakraborty, Baharan Mirzasoleiman, Krishna P. Gummadi, and Adish Singla

Abstract—Influence maximization has found applications in a wide range of real-world problems, for instance, viral marketing of products in an online social network, and propagation of valuable information such as job vacancy advertisements. While existing algorithmic techniques usually aim at maximizing the total number of people influenced, the population often comprises several socially salient groups, e.g., based on gender or race. As a result, these techniques could lead to disparity across different groups in receiving important information. Furthermore, in many applications, the spread of influence is time-critical, i.e., it is only beneficial to be influenced before a deadline. As we show in this paper, such time-criticality of information could further exacerbate the disparity of influence across groups. This disparity could have far-reaching consequences, impacting people's prosperity and putting minority groups at a big disadvantage. In this work, we propose a notion of *group fairness* in *time-critical influence maximization*. We introduce surrogate objective functions to solve the influence maximization problem under fairness considerations. By exploiting the submodularity structure of our objectives, we provide computationally efficient algorithms with guarantees that are effective in enforcing fairness during the propagation process. Extensive experiments on synthetic and real-world datasets demonstrate the efficacy of our proposal.

Index Terms—Influence maximization, algorithmic fairness, social networks

1 Introduction

The problem of *Influence Maximization* has been widely studied due to its application in multiple domains such as viral marketing [1], social recommendations [2], propagation of information related to jobs, financial opportunities or public health programs [3], [4]. Over the years, extensive research efforts have focused on the cascading behavior, diffusion and spreading of ideas, or containment of diseases [1], [5], [6], [7], [8]. The idea is to identify a set of initial sources (i.e., *seed nodes*) in a social network who can influence other people (e.g., by propagating key information), and traditionally the goal has been to maximize the total number of people influenced in the process (e.g., who received the information being propagated) [6], [9], [10].

Real-world social networks, however, are often not homogeneous and comprise different groups of people. Due to the disparity in their population sizes, potentially high propensity towards creating within-group links [11], and differences in dynamics of influences among different groups [12], the structure of the social network can cause disparities in the influence maximization process. For example, selecting most of the seed nodes from the majority group might maximize the total number of influenced nodes, but very few members of the minority group may get influenced. In many application scenarios such as propagation of job or health-related information, such disparity can end up impacting people's livelihood and some groups may become impoverished in the process.

Moreover, some applications are also *time-critical* in nature [13]. For example, many job applications typically have a deadline by which one needs to apply; if information related to the application reaches someone after the deadline, it is not useful. Similarly, in viral marketing, many companies offer discount deals only for few days (hours); getting this information late does not serve the recipient(s). More worryingly, if one group of people gets influenced (i.e., they get the information) faster than other groups, it could end up exacerbating the inequality in information access. This is possible if the majority group is better connected and more central in the network than the minority group. Thus, in time-critical application scenarios, focusing on the traditional criteria of maximizing the number of influenced nodes can have a disparate impact on different groups. This disparity in time-critical applications, in turn, can put minority and under-represented groups at a big disadvantage with far-reaching consequences.

In this paper, we attempt to mitigate such unfairness in time-critical influence maximization (TCIM), and we focus on two settings: (i) where the budget (i.e., the number of seeds) is fixed and the goal is to find a seed set which maximizes the time-critical influence, we call this as TCIM-Budget

- *Junaid Ali, Krishna P. Gummadi, and Adish Singla are with the Max Planck Institute for Software Systems, 66123 Saarbrücken, Germany. E-mail: {junaid, gummadi, adishs}@mpi-sws.org.*
- *Mahmoudreza Babaei is with the Max Planck Institute for Human Development, 14195 Berlin, Germany. E-mail: babaei@mpib-berlin.mpg.de.*
- *Abhijnan Chakraborty is with the Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India. E-mail: abhijnan@iitd.ac.in.*
- *Baharan Mirzasoleiman is with the University of California, Los Angeles (UCLA), Los Angeles, CA 90095 USA. E-mail: baharan@cs.ucla.edu.*

Manuscript received 31 July 2020; revised 21 Apr. 2021; accepted 1 Oct. 2021. Date of publication 15 Oct. 2021; date of current version 3 Feb. 2023.
This work was supported in part by the European Research Council (ERC) Advanced Grant through the Foundations for Fair Social Computing Project funded under the European Union's Horizon 2020 Framework Programme under Grant 789373.
(Corresponding author: Junaid Ali.)
Recommended for acceptance by P. Bogdanov.
Digital Object Identifier no. 10.1109/TKDE.2021.3120561

problem, and (ii) where a certain quota or fraction of the population should be influenced under the prescribed time deadline, and the goal is to find such a seed set of minimal size, we call this as TCIM-COVER problem.

1.1 Our Contributions

Our first contribution is to formally introduce the notion of fairness in time-critical influence maximization, which requires that *within a prescribed time deadline, the fraction of influenced nodes should be equal across different groups*. We highlight, via experiments and an illustrative example, that the standard algorithmic techniques for solving TCIM-BUDGET and TCIM-COVER problems lead to unfair solutions, and the disparity across groups could get worse with tighter time deadline. Second, we study the effect of disparity of influence between groups: (i) by varying graph properties, such as connectivity and relative group sizes etc., and (ii) by varying TCIM algorithmic properties, such as seed budget, reach quota and time deadline etc.

We introduce two formulations of TCIM problems under fairness considerations, namely FAIRTCIM-BUDGET and FAIRTCIM-COVER. As our third contribution, we propose *monotone submodular* surrogates for solving both of these NP-Hard problems. Though the surrogate problems are still NP-Hard, we propose a greedy approximation with provable guarantees.

We evaluate our proposed solutions over several synthetic and two real-world social networks and show that they are successful in enforcing the aforementioned fairness notion. Enforcing fairness does come at the cost of a reduction in performance. However, as guaranteed by our theoretical results, our experiments indeed demonstrate that this cost of fairness, i.e., reduction in performance, is bounded for our approach.

2 RELATED WORK

In this section, we briefly review the related literature on influence maximization and algorithmic fairness.

Influence Maximization. Richardson *et al.* [1] first introduced Influence Maximization as an algorithmic problem, and proposed a heuristic approach to find a set of nodes whose initial adoption of a certain idea/product can maximize the number of further adopters. Over the years, extensive research efforts have focused on the cascading behavior, diffusion and spreading of ideas or containment of diseases, by identifying the set of influential nodes that maximizes the influence through a network (often in real-time) [1], [5], [6], [7], [8].

Typically, identifying the most influential nodes is studied in two ways: (i) using network structural properties to find the set of most central nodes [6], [14], and (ii) formulating the problem as discrete optimization [6], [9], [15]. Kempe *et al.* [6], studied influence maximization under different social contagion models and showed that submodularity of the influence function can be used to obtain provable approximation guarantees. Since then, there has been a large body of work studying various extensions [9], [10], [16], [17], [18]. However, the notion of fairness in the influence maximization problem has not been studied by this line of previous works.

Fairness in Algorithmic Decision Making. Recently a growing amount of work has focused on bias and unfairness in algorithmic decision-making systems [19], [20], [21]. The aim here is to examine and mitigate unfair decisions that may lead to discrimination. Although fairness along different dimensions of political science, moral philosophy, economics, and law has been extensively studied [22], [23], [24], [25] , only a few contemporary works have investigated fairness in influence maximization, as described next.

Contemporary Works. Very recently, Fish *et al.* [26], proposed a notion of individual fairness in information access, but did not consider the group fairness aspects. In addition, some prior works have proposed constrained optimization problems to encourage diversity in selecting the most influential nodes [27], [28], [29], [30].

A recent paper by Rahmattalabi *et al.* [31], proposes group fairness in influence maximization for robust covering problems. This method is different from ours in the following ways: i) their notion of fairness is maximizing the minimum influence for any group, while we propose parity of influence among different groups; ii) they consider a setting where seeds could be deactivated randomly while we do not have any stochasticity in seed activation; iii) they consider seed nodes to spread influence only to their immediate neighbors, while we vary the allowed time deadline and show its effect on disparity among different groups. We also demonstrate the effectiveness of our methods for different time deadlines on several datasets; iv) they propose an integer linear programming set up while we propose submodular proxies, akin to the traditional methods, which can be approximately solved using the greedy heuristic.

In concurrent works, Khajehnejad *et al.*, [32], and Tsang *et al.*, [33], proposed methods to achieve group fairness in influence maximization. However, their works are very different from our approach in three ways: i) they propose a different problem formulation with objective that does not have submodular structural properties, ii) they only study the problem under budget constraint, and iii) they do not consider the time-critical aspect of influence in their definition of fairness for influence maximization. This could result in majority groups being influenced before the minority, and can lead to disparity in applications where the timing of being influenced/informed is critical. In our work, we introduce a submodular objective that directly addresses the time-criticality in influence maximization problem under budget constraint as well as coverage constraint.

3 BACKGROUND ON TIME-CRITICAL INFLUENCE MAXIMIZATION (TCIM)

In this section, we provide the necessary background on the problem of time-critical influence maximization (henceforth, referred to as TCIM for brevity). First, we formally introduce a well-studied influence propagation model and specify the notion of time-critical influence that we consider in this paper. Then, we discuss two discrete optimization formulations to tackle the TCIM problem.

3.1 Influence Propagation in Social Network

Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of directed edges connecting these

nodes. For instance, in a social network the nodes could represent people and edges could represent friendship links between people. An undirected link between two nodes can be represented by simply considering two directed edges between these nodes.

There are two classical influence propagation models that are studied in the literature [6]: (i) Independent Cascade model (IC) and (ii) Linear Threshold (LT) model. In this paper, we will consider IC model and our results can easily be extended to the LT model.

In the IC model, there is a probability of influence associated with each edge denoted as $p_{\mathcal{E}} := \{p_e \in [0,1] : e \in \mathcal{E}\}$. Given an initial seed set $S \subseteq \mathcal{V}$, the influence propagation proceeds in discrete time steps $t = \{0, 1, 2, \ldots, \}$ as follows. At $t = 0$, the initial seed set S is "activated" (i.e., influenced). Then, at any time step $t > 0$, a node $v \in \mathcal{V}$ which was activated at time $t-1$ gets a chance to influence its neighbors (i.e., set of nodes $\{w : (v,w) \in \mathcal{E}\}$). The influence propagation process stops at time $t > 0$ if no new nodes get influenced at this time. Under the IC model, once a node is activated it stays active throughout the process and each node has only one chance to influence its neighbors.

Note that the influence propagation under IC model is a stochastic process: the stochasticity here arises because of the random outcomes of a node v influencing its neighbor w based on the Bernoulli distribution $p_{(v,w)}$. An outcome of the influence propagation process can be denoted via a set of timestamps $\{t_v \geq 0 : v \in \mathcal{V}\}$ where t_v represents the time at which a node $v \in \mathcal{V}$ was activated. We have $t_v = 0 \text{iff} v \in S$ and for convenience of notation, we define $t_v = -1$ to indicate that the node v was not activated in the process.

3.2 Utility of Time-Critical Influence

As mentioned earlier, we focus on the application settings where the spread of influence is time-critical, i.e., it is more beneficial to be influenced earlier in the process. In particular, we adopt the well-studied notion of time-critical influence as proposed by [13]. Their time-critical model is captured via a deadline τ: If a node is activated before the deadline, it receives a utility of 1, otherwise it receives no utility. This simple model captures the notion of timing in many important real-world applications such as viral marketing of an online product with limited availability, information propagation of job vacancy information, etc.

Given the influence propagation model and the notion of time-critical aspect via a deadline τ, we quantify the utility of time-critical influence for a given seed set S on a set of target nodes $Y \subseteq \mathcal{V}$ via the following:

$$f_\tau(S; Y, \mathcal{G}) = \mathbb{E}\left[\sum_{v \in Y, t_v \geq 0} \mathbb{I}(t_v \leq \tau)\right], \tag{1}$$

where the expectation is w.r.t. the randomness of the outcomes of the IC model. The function is parametrized by deadline τ, set $Y \subseteq \mathcal{V}$ representing the set of nodes over which the utility is measured (by default, one can consider $Y = \mathcal{V}$), and the underlying graph $\mathcal{G}$ along with edge activation probabilities $p_{\mathcal{E}}$. Given a fixed value of these parameters, the utility function $f_\tau : 2^{\mathcal{V}} \to \mathbb{R}_{\geq 0}$ is a set function defined over the seed set $S \subseteq \mathcal{V}$. Note that the constraint $t_v \geq 0$ represents the node was activated and the constraint $t_v \leq \tau$ represents that the activation happened before the deadline τ.

3.3 TCIM as Discrete Optimization Problem

Next, we present two settings under which we study TCIM by casting it as a discrete optimization problem.

3.3.1 Maximization Under Budget Constraint (TCIM-Budget)

In the maximization problem under budget constraint, we are given a fixed budget $B > 0$ and the goal is to find an optimal set of seed nodes that maximize the expected utility. Formally, we state the problem as

$$\max_{S \subseteq \mathcal{V}} f_\tau(S; \mathcal{V}, \mathcal{G}) \quad \text{subject to } |S| \leq B. \tag{P1}$$

3.3.2 Minimization Under Coverage Constraint (TCIM-Cover)

In the minimization problem under coverage constraint, we are given a quota $Q \in [0,1]$ representing the minimal fraction of nodes that must be activated or "covered" by the influence propagation in expectation. The goal is then to find an optimal set of seeds of minimal size that achieves the desired coverage constraint. We formally state the problem as

$$\min_{S \subseteq \mathcal{V}} |S| \quad \text{subject to } \frac{f_\tau(S; \mathcal{V}, \mathcal{G})}{|\mathcal{V}|} \geq Q. \tag{P2}$$

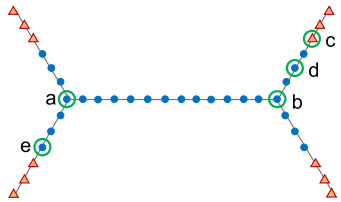
3.4 Submodularity and Approximate Solutions

Next, we present some key properties of the utility function $f_\tau(.)$ to get a better understanding of the above-mentioned optimization problems. In their seminal work, [6] showed that the utility function without time-critical deadline, i.e., $f_\infty(.) : S \to \mathbb{R}_+$, is a non-negative, monotone, submodular set function w.r.t. the optimization variable $S \subseteq \mathcal{V}$. Submodularity is an intuitive notion of diminishing returns and optimization of submodular set functions finds numerous applications in machine learning and social networks, such as influence maximization [6], sensing [34], information gathering [35], and active learning [36] (see [37] for a survey on submodular function optimization and its applications).

Chen *et al.* [13] showed that the utility function for the general time-critical setting for any τ also satisfies these properties. Submodularity is an intuitive notion of diminishing returns, stating that, for any sets $A \subseteq A' \subseteq \mathcal{V}$, and any node $a \in \mathcal{V} \backslash A'$, it holds that (omitting the parameters $\mathcal{V}$ and $\mathcal{G}$ for brevity)

$$f_\tau(A \cup \{a\}) - f_\tau(A) \geq f_\tau(A' \cup \{a\}) - f_\tau(A').$$

Existing works [37], [38], [39] have shown that (2) and (3) are NP-Hard and hence finding the optimization solution is intractable. However, on a positive note, one can exploit the submodularity property of the function to design efficient approximation algorithms with provable guarantees [37], [38]. In particular, we can run the following greedy heuristic: start from an empty set, iteratively add a new node to the set that provides the maximal marginal gain in terms of utility, and stop the algorithm when the desired constraint



	solution to TCIM-BUDGET P1				solution to FAIRTCIM-BUDGET P4			
	S	$\frac{f(S;\mathcal{V},\mathcal{G})}{\lvert\mathcal{V}\rvert}$	$\frac{f(S;\mathcal{V}_1,\mathcal{G})}{\lvert\mathcal{V}_1\rvert}$	$\frac{f(S;\mathcal{V}_2,\mathcal{G})}{\lvert\mathcal{V}_2\rvert}$	S	$\frac{f(S;\mathcal{V},\mathcal{G})}{\lvert\mathcal{V}\rvert}$	$\frac{f(S;\mathcal{V}_1,\mathcal{G})}{\lvert\mathcal{V}_1\rvert}$	$\frac{f(S;\mathcal{V}_2,\mathcal{G})}{\lvert\mathcal{V}_2\rvert}$
$\tau=\infty$	$\{a,b\}$	0.38	0.48	0.16	$\{a,c\}$	0.31	0.33	0.27
$\tau=4$	$\{a,b\}$	0.32	0.44	0.08	$\{d,e\}$	0.25	0.26	0.22
$\tau=2$	$\{a,b\}$	0.24	0.36	0.00	$\{a,c\}$	0.21	0.22	0.18

Fig. 1. An example to illustrate the disparity across groups in the standard approaches to TCIM. (Left) Graph with $|\mathcal{V}|=38$ nodes belonging to two groups shown in "blue dots" ($|\mathcal{V}_1|=26$) and "red triangles" ($|\mathcal{V}_2|=12$). (Right) We compare an optimal solution to the standard TCIM-BUDGET problem (P1) and an optimal solution to our formulation of TCIM-BUDGET with fairness considerations given by FAIRTCIM-BUDGET problem (P4). For different time critical deadlines τ, normalized utilities are reported for the whole population $\mathcal{V}$, for the "blue dots" group $\mathcal{V}_1$, and for the "red triangles" group $\mathcal{V}_2$. As τ reduces, the disparity between groups is further exacerbated in the solution to TCIM-BUDGET problem (P1). Solution to FAIRTCIM-BUDGET problem (P4) achieves high utility and low disparity for different deadlines τ.

on budget or coverage is met. This greedy algorithm provides the following guarantees for these two problems:.

- for the TCIM-BUDGET problem (P1), the greedy algorithm returns a set $\hat{S}$ that guarantees the following lower bound on the utility: $f_\tau(\hat{S};\mathcal{V},\mathcal{G}) \geq (1-\frac{1}{e})\cdot f_\tau(S^*;\mathcal{V},\mathcal{G})$ where S^* is an optimal solution to problem (P1).
- for the TCIM-COVER problem (P2), the greedy algorithm returns a set $\hat{S}$ that guarantees the following upper bound on the seed set size: $|\hat{S}| \leq \ln(1+|\mathcal{V}|)\cdot|S^*|$ where S^* is an optimal solution to problem (P2).

4 MEASURING UNFAIRNESS IN TCIM

In this section, we highlight the disparity in utility across population resulting from the solution to the standard TCIM problem formulations, and introduce a measure of unfairness in TCIM.

4.1 Socially Salient Groups and Their Utilities

The current approaches to TCIM consider all the nodes in $\mathcal{V}$ to be homogeneous. We capture the presence of different socially salient groups in the population by dividing individuals into k disjoint groups. Here, socially salient groups could be based on some sensitive attribute such as gender or race. We denote the set of nodes in each group $i \in \{1,2,\ldots,k\}$ as $\mathcal{V}_i \subseteq \mathcal{V}$, and we have $\mathcal{V} = \cup_i \mathcal{V}_i$. For any given seed set S, we define the utilities for a group i as $f_\tau(S;\mathcal{V}_i,\mathcal{G})$ by setting target nodes $Y = \mathcal{V}_i$ in Eq. (1).

4.2 Disparity in Utility Across Groups

In the standard formulations for TCIM problem, i.e., TCIM-BUDGET problem (P1) and TCIM-COVER problem (P2), the utility $f_\tau(S;\mathcal{V},\mathcal{G})$ is optimized for the whole population $\mathcal{V}$ without considering their groups. Clearly, a solution to TCIM problem can, in general, lead to high disparity in utilities of different groups.

In particular, this disparity in utility across groups arises from several factors in which two groups differ from each other. One of the factors is that the groups are of different sizes, i.e., one group is a minority. The different group sizes could, in turn, lead to selecting seed nodes from the majority group when optimizing for utility $f_\tau(S;\mathcal{V},\mathcal{G})$ in problems (2) and (3). Another factor is related to the connectivity and centrality of nodes from different groups. The solution to the optimization problems (2) and (3) tend to favor nodes which are more central and have high-connectivity. Finally, given the above two factors, we note that the disparity in influence across groups can be further exacerbated for lower values of deadline τ in the time-critical influence maximization.

In Fig. 1, we provide an example to illustrate the disparity across groups in the standard approaches to TCIM. In particular, to show this disparity, we consider the TCIM-BUDGET problem (P1), and it is easy to extend this example to show disparity in TCIM-COVER problem (P2). The graph that we consider in this example (see Fig. 1 caption for details) has the two characteristic properties that we discussed above: (i) group V_2 is in minority with less than half of the size of group V_1, (ii) group V_1 has more central nodes compared to group V_2, and (iii) nodes in group V_1 have higher connectivity than nodes in group V_2. We consider the probability of influence in the graph to be $p_e = 0.7$ for all edges, and study the optimization problem (P1) for budget $B=2$.

For different time critical deadlines τ, we report the following normalized utilities: $\frac{f(S;\mathcal{V},\mathcal{G})}{|\mathcal{V}|}$ for the whole population $\mathcal{V}$, $\frac{f(S;\mathcal{V}_1,\mathcal{G})}{|\mathcal{V}_1|}$ for the group $\mathcal{V}_1$, and $\frac{f(S;\mathcal{V}_2,\mathcal{G})}{|\mathcal{V}_2|}$ for the group $\mathcal{V}_2$. Here, normalization captures the notion of "average" utility per node in a group, and automatically allows us to account for the differences in the group sizes. As can be seen in Fig. 1, the optimal solution to the problem consistently picks set $S=\{a,b\}$ comprising of the most central and high-connectivity nodes. While these nodes maximize the total utility, they lead to a high disparity in the normalized utilities across groups. As the influence becomes more time-critical, i.e., τ is reduced, we see an increasing disparity as discussed above. For $\tau=2$, the utility of group $\mathcal{V}_2$ reduces to 0.

4.3 Measure of Unfairness

Next, in order to guide the design of fair solutions to TCIM problems, we introduce a formal notion of group unfairness in TCIM. In particular, we measure the (un-)fairness or disparity of an algorithm by the maximum *disparity in normalized utilities* across all pairs of socially salient groups, given by:

$$\max_{i,j\in\{1,2,\ldots,k\}} \left| \frac{f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}_i|} - \frac{f_\tau(S;\mathcal{V}_j,\mathcal{G})}{|\mathcal{V}_j|} \right|. \tag{2}$$

As discussed above (see Section 4.2), normalization w.r.t. group sizes captures the notion of average utility per node in a group and hence makes the measure agnostic to the group size. In the next section, we seek to design fair algorithms for TCIM problems that have low disparity (or more fairness) as measured by Eq. (2).

5 Achieving Fairness in TCIM

In this section, we seek to develop efficient algorithms for TCIM problems under fairness considerations that have low disparity measured by Eq. (2) while maintaining high performance.

5.1 Fair TCIM-Budget

5.1.1 Fairness Considerations in TCIM-Budget

A fair TCIM algorithm under budget constraint should seek to achieve the following two objectives: (i) maximizing total influence for the whole population $\mathcal{V}$ as was done in the standard TCIM-Budget problem (P1), and (ii) enforcing fairness by ensuring that disparity across different groups as per Eq. (2) is low. Clearly, enforcing fairness would lead to a reduction in total influence, and we seek to design algorithms that can achieve a good trade-off between these two objectives. We formulate the following fair variant of TCIM-Budget problem (P1) that captures this trade-off

$$\begin{aligned} &\max_{S\subseteq\mathcal{V}} \quad \underbrace{\sum_{i}^{k} f_\tau(S;\mathcal{V}_i,\mathcal{G})}_{\text{Maximize number of influenced nodes}} \\ &\text{subjectto} \quad \underbrace{|S|\leq B}_{\text{Boundseedsetsize}}, \\ &\text{and} \underbrace{\max_{i,j}\left|\frac{f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}_i|}-\frac{f_\tau(S;\mathcal{V}_j,\mathcal{G})}{|\mathcal{V}_j|}\right|\leq c}_{\text{Minimize disparity}}, \end{aligned} \tag{P3}$$

where $c\in[0,1]$ is a hyperparameter which indicates the maximum level of allowed disparity among the groups. This problem might not be feasible for all the values of c. So, one would have to tune this hyperparameter for feasibility and the desired level of disparity. Problem (P3) has two main objectives, i.e., finding B seeds which will i) *maximize the total influence*, which is exactly the same as the traditional influence maximization given in problem (P1)— here written as the sum of influences over all the groups, and, additionally, ii) *minimize the disparity of influence* between different groups up to the prescribed threshold.

We note that problem (P3) is NP-Hard and a challenging discrete optimization problem and it does not have the structural properties of submodularity as was the case for the standard TCIM-Budget problem (P1).

5.1.2 Surrogate FairTCIM-Budget With Guarantees

Instead of directly solving problem (P3), we introduce a novel surrogate problem that would allow us to indirectly trade-off the two objectives of maximizing total influence and minimizing disparity across groups, as follows:

$$\max_{S\subseteq\mathcal{V}}\sum_{i=1}^{k}\mathcal{H}(f_\tau(S;\mathcal{V}_i,\mathcal{G})) \quad \text{subject to } |S|\leq B, \tag{P4}$$

where $\mathcal{H}$ is a non-negative, monotone concave function.

Optimizing problem (P4) captures both the objectives of the original: i) *maximizing influence*: since the objective is monotonically increasing it encourages picking more influential nodes, ii) *minimizing the disparity of influence*: Passing the group influence functions through a monotone concave function $\mathcal{H}$ rewards selecting seeds that would lead to higher influence on under-represented groups early in the selection process; this in turn helps in reducing disparity across groups under the *assumption* that the under-represented groups not only have lower influence in terms of total number of nodes but also have lower influnece in terms of fraction of nodes w.r.t to their groups sizes. In other words, as we are passing the group influences through a *concave* function, the increase in the objective would be higher when under-represented groups are influenced, as demonstrated in Fig. 2.

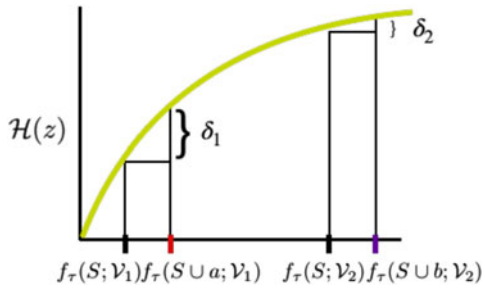


Fig. 2. Demonstration of concave function encouraging picking seeds which influence under-represented group. X-axis represents group influence and y-axis represents the value of $\mathcal{H}$ for the corresponding group influence. In this example we have two groups, $\mathcal{V}_1$ and $\mathcal{V}_2$. $\mathcal{V}_1$ is under-influenced compared to $\mathcal{V}_2$, using the seed set S. In the next iteration we have an option to either include node a or b in our seed set, both of which add the same amount of total influence. Adding node a in our seed set influences $\mathcal{V}_1$which is the under-influenced group, while adding node b influences nodes from $\mathcal{V}_2$, as demonstrated in the figure. The traditional method, given by problem (P1), would treat both of these nodes as equally good. However, since we are passing the group influences through a concave function the increase in the value of $\mathcal{H}(z)$ will be more if we pick node a, i.e., our method will pick node a because $\delta_1 > \delta_2$.

Trade-off Between Objectives. It is important to note that controlling the curvature of the concave function $\mathcal{H}$ provides an indirect way to *trade-off* between the two objectives, i.e., i) the total influence and ii) the disparity of the solution. For instance, using $\mathcal{H}(z) := \log(z)$ has higher curvature than using $\mathcal{H}(z) := \sqrt{z}$ and hence leads to lower disparity at the cost of lower total influence (this is demonstrated in the experimental results in Fig. 4a). For our illustrative example from Section 4, we report the results for an optimal solution to FairTCIM-Budget problem (P4) with $\mathcal{H}(z) := \log(z)$. As can be seen in Fig. 1, the solution leads to a drastic reduction in disparity across groups for different values of deadline τ compared to an optimal solution of the standard TCIM-Budget problem (P1) at the cost of reduction in total influence. So, if one wants to penalize disparity of influence more one can pick $\mathcal{H}$ function with higher curvature but at the expense of potentially lower total influence.

While it is intuitively clear that using the concave function $\mathcal{H}(z)$ in problem (P4) reduces disparity, we also need to ensure that the solution to this problem has high influence for the whole population $\mathcal{V}$ and that the solution can be computed efficiently. As proven in the theorem below, we can find an approximate solution to problem (P4), with guarantees on the total influence, by running the greedy heuristic (as was introduced in Section 3.4).

Theorem 1. *Let $\hat{S}$ denote the output of the greedy algorithm for problem (P4). Let S^* be an optimal solution to problem (P1).*

Then, the total influence of the greedy algorithm is guaranteed to have the following lower bound: $f_\tau(\hat{S};\mathcal{V},\mathcal{G}) \geq (1-\frac{1}{e})\cdot \mathcal{H}\big(f_\tau(S^*;\mathcal{V},\mathcal{G})\big)$.

This is equivalent to the fact that the multiplicative approximation factor of the utility of FAIRTCIM-BUDGET using greedy algorithm w.r.t. the utility of an optimal solution to TCIM-BUDGET scales as $\big((1-\frac{1}{e})\cdot\frac{\mathcal{H}(f_\tau(S^*;\mathcal{V},\mathcal{G}))}{f_\tau(S^*;\mathcal{V},\mathcal{G})}\big)$. Note that as the curvature of the concave function $\mathcal{H}$ increases, the approximation factor gets worse—this further highlights how the curvature of the function $\mathcal{H}$ provides a way to trade-off the total influence and disparity of the solution. In the case of $\mathcal{H}(z) := log(z)$, which penalizes the disparity of the solution quite severely due to high curvature, the bound on the total influence achieved by our solution is exponentially related to the optimal solution of problem (P1) which does not consider fairness. On the other hand, if $\mathcal{H}(z) := z$, i.e., $\mathcal{H}$ is an identity function, the problem reverts back to problem (P1), whose solution might have a higher total influence but could result in high disparity, as evidenced by our experimental results in Sections 6.2 and 7.2. One can pick $\mathcal{H}$ with the appropriate curvature for the desired level of penalization of the disparity of influence at the cost of total influence. *Due to lack of space, the proof of the theorem is included in the appendix, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TKDE.2021.3120561.*

5.2 Fair TCIM-Cover

5.2.1 Fairness Considerations in TCIM-COVER

A fair TCIM algorithm under coverage constraint should seek to achieve the following two objectives: (i) minimizing the size of the seed set that achieves the desired coverage constraint as was done in the standard TCIM-COVER problem (P2), and (ii) enforcing fairness by ensuring that disparity across different groups as per Eq. (2) is low. As was the case for FAIRTCIM-BUDGET problem above, enforcing fairness would lead to increasing the size of the required seed set, and we seek to design algorithms that can achieve a good trade-off between these two objectives. We formulate a fair variant of TCIM-COVER problem (P2) that captures this trade-off as follows:

$$\begin{aligned} &\min_{S\subseteq\mathcal{V}} \quad \underbrace{|S|}_{\text{Minimize seed set size}} \\ &\text{subject to} \quad \underbrace{\frac{\sum_i^k f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}|} \geq Q}_{\text{Bound fraction of influenced node}} \quad , \\ &\text{and} \underbrace{\max_{i,j}\Big|\frac{f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}_i|} - \frac{f_\tau(S;\mathcal{V}_j,\mathcal{G})}{|\mathcal{V}_j|}\Big| \leq c}_{\text{Minimize disparity}}, \end{aligned} \tag{P5}$$

where $c \in [0,1]$ is a hyperparameter, which determines the amount of disparity that is allowed. As in the case of problem (P3), it is possible that for some values of c the problem is infeasible. Problem (P5) has three objectives: i) *minimizing size of seed set* that ii) *influences a prescribed quota* of the population while ii) *minimizing disparity in the influence* among the groups.

As in Section 5.1, we note that problem (P5) is a challenging discrete optimization problem and does not have structural properties as was the case for the standard TCIM-COVER problem (P2).

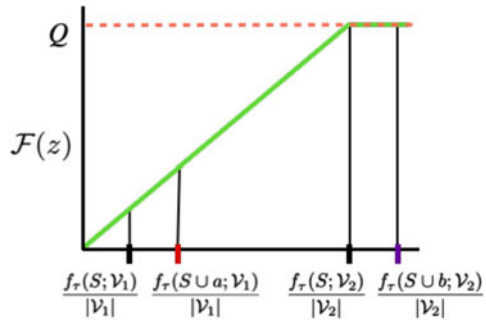


Fig. 3. where $\mathcal{F}(z) = \min\{\frac{f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}_i|}, Q\}$. Demonstration of the constraint in problem (P6). X-axis represents the fraction of group influences and y-axis represents the value of per group constraint in problem (P6) for the corresponding group influence. In this example we have two groups, $\mathcal{V}_1$ and $\mathcal{V}_2$ of roughly same size. $\mathcal{V}_1$ has not reached the prescribed quota, Q, while $\mathcal{V}_2$ has already been influenced up to the prescribed quota. In the next iteration we have an option to either include node a or node b in our seed set, both of which add the same amount of total influence. Adding node a in our seed set influences only $\mathcal{V}_1$, while adding node b influences nodes from only $\mathcal{V}_2$, as demonstrated in the figure. The traditional method, problem (P2), would treat both of these nodes as equally good candidates for including in the seed set because they add equal fraction of total influence. However, since we require all the groups to be influenced up to the required quota, selecting node a will increase our constraint value, $\mathcal{F}(z)$, while by selecting node b the constraint value would stay the same as $\mathcal{V}_2$ has already reached the required quota of influence.

5.2.2 Surrogate FAIRTCIM-COVER With Guarantees

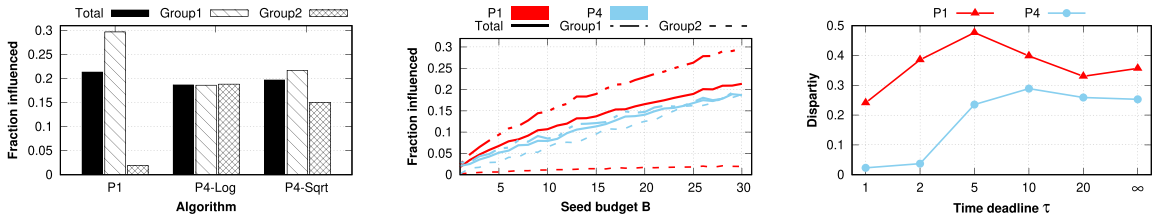
Instead of directly solving problem (P5), we introduce a novel surrogate problem that indirectly trade-offs the two objectives of minimizing the size of selected seed set and minimizing disparity, as follows:

$$\min_{S\subseteq\mathcal{V}} |S| \quad \text{subject to } \frac{f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}_i|} \geq Q \quad \forall i. \tag{P6}$$

Optimizing problem (P6) addresses all the objectives of problem (P5) by i) *minimizing the seed set size*, ii) which *influences all the groups up to the prescribed quota*, Q. iii) Thereby, *disparity* of the feasible solution is *bounded* by $(1-Q)$. The key idea of using the surrogate objective function in problem (P6) is the following: the problem has a constraint that enforces that at least Q fraction of nodes in each group are influenced by the selected seed set S; this in turn directly provides a bound on the disparity of any feasible solution to the problem as $(1-Q)$. Fig. 3 provides a demonstration of the constraints we propose.

While it is intuitively clear that the solution to problem (P6) reduces disparity, we also would like to bound the size of the final seed set and that the solution can be computed efficiently. As proven in the theorem below, we can find an approximate solution to problem (P6), with guarantees on the final seed set size, by running the greedy heuristic (as was introduced in Section 3.4).

Theorem 2. *Let us denote the output of the greedy algorithm for problem (P6) by set* $\hat{S}$. *For group* $i \in \{1,\ldots,k\}$, *let* S_i^* *denote an optimal solution to the coverage problem (P2) for the target nodes set to* $\mathcal{V}_i$, *i.e., solving problem (P2) with constraint given*



(a) Total and group influence **(b)** Varying budget B **(c)** Varying time deadline τ

Fig. 4. [Synthetic Dataset: Budget Problem] The figures show that solving TCIM-BUDGET problem (P1) can lead to disparity in number of influenced nodes belonging to different groups, while FAIRTCIM-BUDGET problem (P4) fares better in terms of achieving parity of influence, with marginally lower total influence. See Section 6.2 for further details.

by $\frac{f_\tau(S;\mathcal{V}_i,\mathcal{G})}{|\mathcal{V}_i|} \geq Q$. *Then, the size of the seed set* $\hat{S}$ *returned by the greedy algorithm is guaranteed to have the following upper bound:* $|\hat{S}| \leq \ln(1+|\mathcal{V}|)(\sum_{i=1}^{k}|S_i^*|)$.

Due to lack of space, the proof of the theorem is included in the appendix, available in the online supplemental material.

6 EVALUATION ON SYNTHETIC DATASETS

In this section, we compare the solutions of different problems on several synthetic datasets. We show that the disparity in influence is affected by varying different properties of the graphs and parameters of the algorithms.

6.1 Dataset and Experimental Setup

First we discuss how we generated the synthetic datasets and then the setup used in our experiments.

Synthetic Datasets. We consider stochastic block model to generate the synthetic datasets, particularly we consider an undirected graph with 500 nodes, where each node belongs to either group $\mathcal{V}_1$ or group $\mathcal{V}_2$. The fraction of nodes belonging to each group is determined by a parameter g (e.g., setting $g = 0.7$ results in 70% of the nodes to be randomly assigned to group $\mathcal{V}_1$). Nodes are connected based on two probabilities: (i) within-group edge probability (*Homophily*) p_{hom} and (ii) across-group edge probability (*Heterophily*) p_{het}. Placing an edge between two nodes goes as follows: given a pair of nodes (v, w), if they belong to the same group, we perform a Bernoulli trial with parameter p_{hom}; otherwise we use the parameter p_{het}. If the outcome of the trial is 1, we place an undirected edge e between these two nodes. Each edge has a probability of activation, $p_e \in [0, 1]$, with which the nodes can activate each other.

Experimental Setup. In our experiments we varied all the aforementioned properties of the graph. We vary each of these graph and algorithmic properties while rest of the properties are set to a default value. We experimented with several default values but as an illustration we include the results for the following default values: $g = 0.7$ yielding 350 nodes in $\mathcal{V}_1$ and 150 nodes in $\mathcal{V}_2$. We set $p_{hom} = 0.025$ and $p_{het} = 0.001$, which yielded 3606 total edges, out of which 2965 edges were within group $\mathcal{V}_1$, 514 within $\mathcal{V}_2$, and 127 edges connecting nodes across two groups. We used a constant activation probability on all edges given by $p_e = 0.05$. Finally, we consider the time deadline $\tau = 20$, unless explicitly stated otherwise.

Evaluating utilities, as described in Eq. (1), in closed form is intractable, so we used Monte Carlo sampling to estimate these utilities. We used 200 samples for this estimation, which yielded a stable estimation of the utility function. In all the experiments, we pick a seed set by solving the corresponding problem. Then, we use this seed set to estimate the expected number of nodes influenced in the graph using TCIM. We report the following normalized utilities: $\frac{f(S;\mathcal{V},\mathcal{G})}{|\mathcal{V}|}$ for the whole population $\mathcal{V}$, $\frac{f(S;\mathcal{V}_1,\mathcal{G})}{|\mathcal{V}_1|}$ for the group $\mathcal{V}_1$, and $\frac{f(S;\mathcal{V}_2,\mathcal{G})}{|\mathcal{V}_2|}$ for the group $\mathcal{V}_2$.

6.2 TCIM Under Budget Constraints

Next, we compare the solutions of TCIM-BUDGET problem (P1) with our solution to FAIRTCIM-BUDGET problem (P4), obtained through the greedy algorithm, i.e., by iteratively picking B seeds which yield maximum marginal gain. In all the figures discussed in this section, red color represents the results of TCIM-BUDGET problem (2), and blue color represents the results of our solution to the FAIRTCIM-BUDGET problem (P4). For the experiments in this section, we used a budget of $B = 30$ seeds.

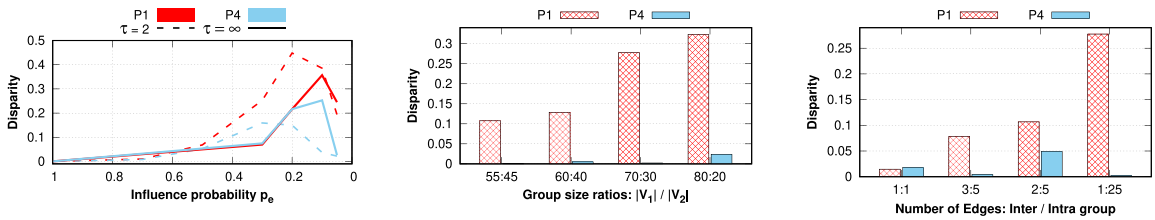
6.2.1 *Varying Algorithmic Properties*

In this section, we vary several properties of the influence maximization algorithm and answer following questions:

- *Q1*: How does the choice of $\mathcal{H}(z)$ with different curvatures affect disparity and total influence?
- *Q2*: How does varying seed budget affect disparity?
- *Q3*: How does varying time deadline affect disparity?
- *Q4*: How does varying activation probabilities on the edges affect disparity?
- *Q5*: How effective is our method in reducing disparity?
- *Q6*: How much cost does our method incur?

[Q1, Q5, Q6] Effect of Different $\mathcal{H}(z)$. Fig. 4a presents the comparison of three algorithms: one solving TCIM-BUDGET problem (P1), using the greedy heuristic; the other two solving FAIRTCIM-BUDGET problem (P4), using two realizations of the concave monotone function, $\mathcal{H}(z)$, given by: (i) $\mathcal{H}(z) := \log(z)$ and (ii) $\mathcal{H}(z) := \sqrt{z}$. Fig. 4a shows the fraction of population influenced, both overall and for every group. We can observe that solving the traditional TCIM-BUDGET problem leads to large disparity between the fraction of nodes influenced from each group: while 30% of nodes in group $\mathcal{V}_1$ are influenced, this fraction is only 2% for group $\mathcal{V}_2$.

On the other hand, our proposed solution to FAIRTCIM-BUDGET problem results in lower disparity between the groups, ensuring similar fraction of influenced nodes. We can further see that $\sqrt{z}$, with lower curvature, performs worse than $\log(z)$ in removing the disparity, however incurring



(a) Varying influence probability p_e **(b)** Varying group size proportions **(c)** Varying cliquishness

Fig. 5. [Synthetic Dataset: Budget Problem] These figures demonstrate that lower activation probabilities, uneven group sizes, and cliquishness can lead to higher disparity of influence between different groups with TCIM-BUDGET problem (P1). In comparison our proposed method, FAIRTCIM-BUDGET given by problem (P4), leads to solutions which yield lower disparity. For further details, see Section 6.2.

lower loss in total influence, as guaranteed by our theoretical results in Theorem 1. One could consider higher powers of the root to increase the curvature or increase the weights λ in problem (P4) for the under-represented group. The *key points* are: i) $\mathcal{H}(z)$ with higher curvature results in lower disparity of influence at the expense of lower total influence. ii) FAIRTCIM-BUDGET problem results in lower disparity and ii) the reduction in the total influence is only marginal as guaranteed by Theorem 1. In the subsequent figures, we only show the results of $\mathcal{H}(z) := \log(z)$ for the solution to problem (P4).

[Q2, Q5, Q6] Effect of Seed Budget. Fig. 4b shows the effect of different seed budgets on the number of influenced nodes (from different groups). Dotted and dash-dotted lines correspond to groups $\mathcal{V}_2$ and $\mathcal{V}_1$ respectively, while solid lines represent the total influence. The figure demonstrates that: (i) Disparity in the utility between both the groups increases with the increase in allowed seed budget. A reason for these differences could be the imbalances in groups sizes and average degrees, between both the groups— $\mathcal{V}_1$ and $\mathcal{V}_2$ comprise 70% and 30% of the nodes respectively. If a very big seed budget is allowed the disparity in influence might also reduce, however in many applications, due to limited resources, it is not practical to have a big budget; (ii) FAIRTCIM-BUDGET problem results in a lower disparate utility between the two groups compared to TCIM-BUDGET problem; (iii) this reduction in disparity is achieved at a very low cost to the total influence, as guaranteed by Theorem 1.

[Q3, Q5] Effect of Deadline. Fig. 4c compares disparity in the solutions of problems (P1) and (P4) as we vary the value of the deadline τ. Disparity is computed as the absolute difference between the fraction of individuals influenced in each group, given by Eq. (2). The figure demonstrates that: (i) disparity in group utilities does not have a unidirectional trend with increasing time deadline τ. One explanation for the increasing disparity— for $\tau = \{1, 2, 5\}$, could be that the seed nodes or the most influential nodes are propagating influence in *both the groups*, but as we increase the time deadline, Group V_1, with more nodes and edges, is more efficient at propagating influence compared to Group V_2, so it results in a larger disparity. But, after a threshold of increase in τ both groups are being influenced because longer cascades are allowed. Hence the disparity lowers and then plateaus, for $\tau = \{5, 10, 20, \infty\}$. One could imagine a case, as shown in the motivating example in Fig. 1, where seed nodes are surrounded by nodes of *only one group*, in this case increasing time deadline could yield a lower disparity. (ii) Our proposed method, given by problem (P4), yields solutions which result in much lower disparity.

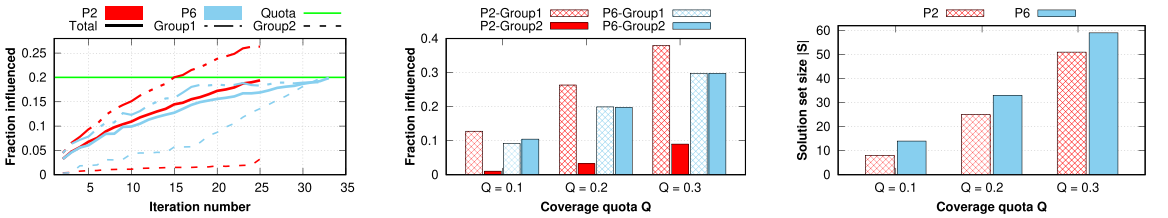
[Q4, Q5] Effect of Activation Probabilities. Fig. 5a shows the disparity in influence for different activation probabilities $p_e \in \{0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0\}$. The results show that: i) lower activation probabilities could result in larger disparity. This makes intuitive sense, since with lower activation probabilities less nodes have a chance to be influenced. We are using an imbalanced graph, both in terms of group sizes and within and across group connectivity. It is very likely that the seeds selected might belong to the majority group and will have more connections to the nodes from their own group. With low activation probabilities less number of nodes are expected to be influence and the biases in the graph structure would become more pronounced, as evidenced by the results. With the high activation probabilities more number of nodes are expected to be influenced so the disparity in the influence is lower, as demonstrated by the results. ii) Lower values of τ tend to have a higher disparity compared to the higher values of τ. The intuition presented in the previous paragraph is confirmed with this experiment. iii) Our method consistently results in a lower disparity. The difference in disparities resulting from the solution of our method compared to the solution of traditional method in more pronounced for lower activation probabilities.

6.2.2 Varying Graph Properties

In this section, we vary several graph properties and answer following evaluation questions:

- *Q1*: How does varying group sizes affect the disparity?
- *Q2*: How does varying connectivity among the groups affect the disparity?
- *Q3*: How effective is our method in reducing disparity?

[Q1, Q3] Effect of Group Sizes. Fig. 5b shows the effect of group sizes $g \in \{0.55, 0.6, 0.7, 0.8\}$. x-axis represents ratio of the nodes belonging to the two groups and y-axis represents disparity. i) The figure confirms our hypothesis that *imbalance in a graph could lead to disparate influence*, as motivated in the illustrative example given in Fig. 1. Since we are considering a $1:25$ of $p_{het} : p_{hom}$, i.e., across versus within group edge probability ratios, even slight imbalance in the group sizes could result in a high disparity. The seed nodes or influential nodes are more likely to be from the dominant group and are more likely to be connected with nodes from their own groups. ii) On the other hand our proposed method results in almost no or very little disparity of influence, as it encourages to pick seeds which influence underrepresented group.



(a) Greedy selection iterations **(b)** Varying quota Q: Influence **(c)** Varying quota Q: Solution set size $|S|$

Fig. 6. [Synthetic Dataset: Cover Problem] These figures show a comparison of TCIM-COVER problem (P2), in red, and FAIRTCIM-COVER problem (P6), in blue. They show that FAIRTCIM-COVER achieves lower disparity of influence between different groups with slightly bigger solution set sizes. See Section 6.3 for further details.

[Q2, Q3] Effect of Graph Connectivity. Fig. 5c demonstrates the importance of the graph structure, particularity connectivity between the two groups, characterized by $(p_{het}, p_{hom}) \in \{(0.025, 0.025), (0.015, 0.025), (0.01, 0.025), (0.001, 0.025)\}$. x-axis shows the ratio of across and within group edge probabilities. i) The figure validates our hypothesis that the majority group containing more influential nodes fares better in TCIM-BUDGET problem, as proposed in Fig. 1. Groups $\mathcal{V}_1$ and $\mathcal{V}_2$ comprise 70% and 30% of the nodes, respectively. As we increase the group-preferential attachment, represented by x-axis of Fig. 5c, influential nodes are more likely to have connections within the group $\mathcal{V}_1$, which in turn results in disparate influence propagation. ii) However, our proposed method performs better because it gives less weight to the nodes influenced from the majority group compared to the minority. Hence, our method encourages picking seed nodes which will influence the minority group, as explained in Fig. 2.

Takeaways. In this section we demonstrated that: (i) solving TCIM-BUDGET problem can lead to disparity of influence in different groups; (ii) the amount of disparity depends on the time deadline, activation probability, relative group sizes, budget, and connectivity of the graph; and (iii) instead, solving FAIRTCIM-BUDGET results in lower disparity of influence, with marginal reduction in overall influence, as guaranteed by Theorem 1.

6.3 TCIM Under Coverage Constraints

Next, we compare solutions of TCIM-COVER problem (3), and our solution to FAIRTCIM-COVER problem (P6). We solve both the problems using the greedy algorithm, i.e., iteratively picking seeds which maximize the constraints of problems (3) and (P6) until the required quota is reached. The goal is to reach the prescribed quota Q, with minimum number of seeds. In all the figures discussed in this section, red color represents the results of TCIM-COVER problem (3), and blue color represents the results of our solution to FAIRTCIM-COVER problem (P6). We answer the following question in this section:

- *Q1*: How does our method fare compared to the traditional method over the *iterations of the algorithm*?
- *Q2*: How effective is our method in reducing disparity for different *reach quotas*?
- *Q3*: How much cost does our method incur?

[Q1] Effect of Iterations. Fig. 6a shows how the fraction of population influenced changes with seed selection at each iteration. Solid lines represent total influence while dash-dotted lines and dotted lines represent groups $\mathcal{V}_1$ and $\mathcal{V}_2$, respectively. In this experiment, Q was set to 0.2 which is represented by the horizontal green line. The figure demonstrates that: (i) both methods reach the required quota of the population; (ii) however, only the solution set of FAIRTCIM-COVER problem (P6) reaches the required quota in both the groups; (iii) while maintaining roughly similar utility for both the groups throughout the iterations; (iv) and it does so *at a small expense of additional seeds*, as guaranteed in Theorem 2.

[Q2, Q3] Effect of Quota Q. Fig. 6b shows fractions of individuals that are influenced for different quota Q: (i) for different values of the required quota, traditional method given by problem (P2) results in disparate utility between both the groups which is most likely due imbalance in group sizes and connectivity. (ii) Seeds selected by solving problem (P6) result in a more equal utility because our method explicitly requires every group to be influence up to quota Q. Depending on the graph structure, our method could result in a disparity up to $1 - Q$. The objective in the constraint given in problem (P6) *only* increases if nodes belonging to the groups are influenced which have not reached the required quota, as demonstrated in Fig. 3. A higher disparity between groups could occur when it is not possible to influence the under-influenced group without influencing the already over-influenced group. In practice a higher disparity could occur, e.g., if one of the groups is very small and very sparsely connected within the group, which is unlikely to occur in practice. (iii) FAIRTCIM-COVER problem (P6) uses only a small number of additional seeds, as guaranteed by Theorem 2.

Takeaways. We compared the result of TCIM-COVER problem (P2) and our solution to FAIRTCIM-COVER problem (P6). The results show that: (i) both methods reach the same fraction of the population; (ii) however, only FAIRTCIM-COVER problem results in seed sets influencing the required quota in *all the groups* and results in *a very low disparity* between groups; and (iii) lastly, FAIRTCIM-COVER yields *only* slightly larger solution sets as guaranteed by Theorem 2.

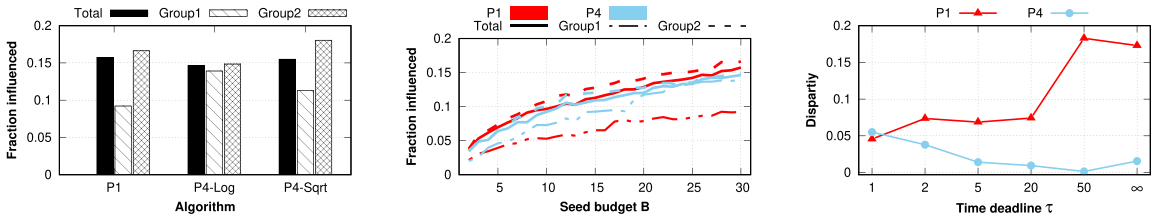
7 EXPERIMENTS ON REAL-WORLD DATASETS

In this section, we evaluate our proposed solutions using two real-world datasets. We describe the datasets and the details of the experiments, and then present our findings.

7.1 Dataset and Experimental Setup

Next, we describe the datasets we used to evaluate our proposed methods, followed by the experimental setup.

Rice-Facebook Dataset. To evaluate our proposed methods, we used *Rice-Facebook* dataset collected by [40], where they



(a) Total and Group Influence **(b)** Varying budget B **(c)** Varying time deadline τ

Fig. 7. [Rice-Facebook Dataset: Budget Problem] Comparison of results solving TCIM-BUDGET problem (P1) and FAIRTCIM-BUDGET (P4). We experimented with 4 groups and total influence includes all the groups, but we show group influences and disparity for only two groups which showed the maximum disparity. The results demonstrate that our method, given by problem (P4), yields seed set which propagate influence in a more fair manner, at the cost of a marginally lower total influence. See Section 7.2 for further details.

capture the connections between students at the Rice University. The resulting network consists of 1205 nodes and 42443 undirected edges. Each node has 3 attributes: (i) the residential college id (a number between $[1-9]$), (ii) age (a number between $[18-22]$), and (iii) a major ID (which is in the range $[1-60]$).

We grouped the nodes (students) into four groups based on their age attributes. We experimented with all four groups while running our algorithms but present the results using only 2 groups which showed the *highest disparity*. We considered nodes with ages 18 and 19 as group $\mathcal{V}_1$ and age 20 as group $\mathcal{V}_2$. Group $\mathcal{V}_1$ has 97 nodes and 513 within-group edges. Whereas, group $\mathcal{V}_2$ has 344 nodes and 7441 within-group edges. Overall, there are 3350 across-group edges going between nodes in $\mathcal{V}_1$ and $\mathcal{V}_2$.

Instagram-Activities Dataset. This dataset was gathered by [41]. It comprises 553628 nodes and 652830 undirected edges. The nodes represent a subset of Instagram users. There exists an edge between two nodes if either of them have liked or commented on each other's photos. Each node has a binary-valued gender attribute, i.e., male or female. 45.5% of the nodes belong to the male group. There are 179668 within-group edge among males and 201083 within-group edges among females, while there are 136039 across-group edges.

Experimental Setup. In all the experiments using *Rice-Facebook* dataset, we show the results for activation probability $p_e = 0.01$. All the other parameter were the same as described in Section 6.1. For experiments using *Instagram-Activities* dataset we show the results with activation probability $p_e = 0.06$, time deadline $\tau = 2$, reach quota $Q = \{0.0015, 0.002\}$ and seed budget $B = 30$. We also experimented with other values of these parameters and get similar results. For *Instagram-Activities* we restrict the seeds to be picked from 5000 randomly selected nodes from the graph. However the influence was evaluated and propagated on the *entire* network. We used 500 sample for *Facebook-Rice* dataset and 10000 samples for *Instagram-Activities* dataset for Monte Carlo estimation of the influence of a node, which yielded very low-variance influence estimates.

7.2 TCIM Under Budget Constraint

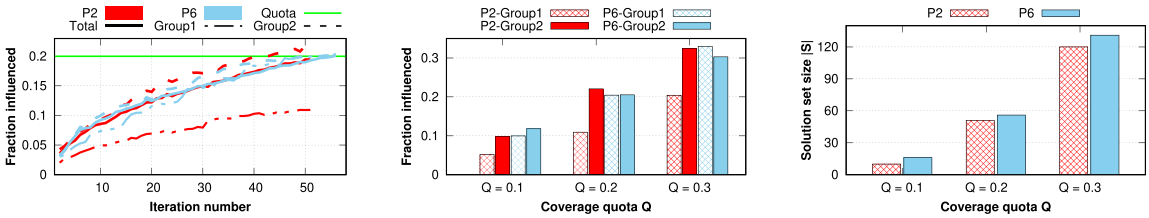
In this section, we compare the results of TCIM-BUDGET problem (P1) and our solution to FAIRTCIM-BUDGET problem (P4). Red color in all the figures discussed in this section corresponds to the solution of TCIM-BUDGET problem (P1) and the blue color corresponds to our solution of FAIRTCIM-BUDGET problem (P4). In all the experiments in this section we used a seed budget $B = 30$. We answer the following evaluation questions using two *real-world datasets* in this section:

- *Q1*: How does the choice of $\mathcal{H}(z)$ with different curvatures affect disparity?
- *Q2*: How does varying seed budget affect disparity?
- *Q3*: How does varying time deadline affect disparity?
- *Q4*: How effective is our method in reducing disparity?
- *Q5*: How much cost does our method incur?

[Q1, Q4, Q5] Effect of Different $\mathcal{H}(z)$. In Figs. 7 a and 9 a, we compare the results of TCIM-BUDGET problem (P1) and FAIRTCIM-BUDGET problem (P4) using two realizations of $\mathcal{H}(z)$, given by: (i) $\mathcal{H}(z) := \log(z)$ and (ii) $\mathcal{H}(z) := \sqrt{z}$. In Figs. 7a the total influence are shown for all the 4 groups while the group influences are shown for 2 out of the 4 groups which showed the maximum disparity. The results demonstrates that: (i) At a marginal reduction of total influence, as guaranteed by Theorem 1, our proposed method significantly reduces disparity in influence in case of Rice-Facebook dataset. However, in the *Instagram-Activities* dataset solving FAIRTCIM-BUDGET problem results in a *higher* total influence while achieving same or lower disparity for both the groups. This is in line with the finding by [42], which, using this dataset, shows that picking more diverse seeds could increase the total influence compared to greedy degree based seeding strategy. Greedy heuristic is just an approximation of the optimal solution. The optimal solution of the unfair problem cannot yield a lower influence compared to the optimal solution of the fair problem, as it adds additional constraints; (ii) as hypothesized in Section 5.1, a higher curvature function, $\mathcal{H}(z) := \log(z)$, leads to a bigger reduction in disparity compared to $\mathcal{H}(z) := \sqrt{z}$. In *Instagram-Activities* dataset $\mathcal{H}(z) := \sqrt{z}$ does not reduce disparity, however it does result in a higher fraction of influence in under-influenced group.

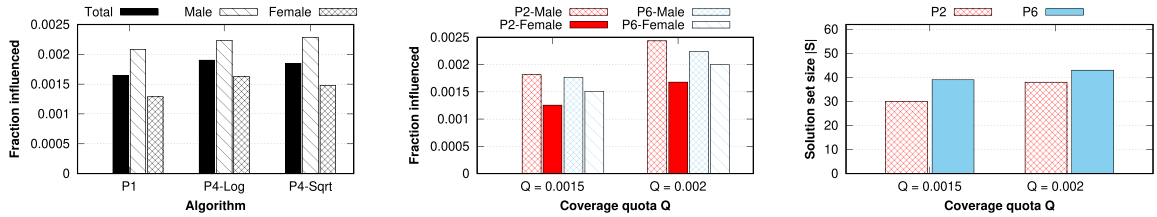
[Q2, Q4, Q5] Effect of Seed Budget. Fig. 7b demonstrates the effect of allowed seed budget on the group and total influences. Groups $\mathcal{V}_1$ and $\mathcal{V}_2$ are represented by dash-dotted lines and dotted lines respectively and solid lines correspond to total influence. Similar to the results on synthetic dataset presented in Section 6.2, i) the disparity between the groups seems to increase with increasing budget and ii) our method consistently results in lower disparity for different seed budgets, iii) while incurring a very small cost of total influence.

[Q3, Q4] Effect of Time Deadline. Fig. 7c shows the effect of different time deadlines on the disparity between group influences, as calculated by Eq. (2). It demonstrates that: (i)



(a) Greedy selection iterations **(b)** Varying quota Q: Influence **(c)** Varying Q: Solution set size $|S|$

Fig. 8. [Rice-Facebook Dataset: Cover Problem] These figures demonstrate the results of TCIM-COVER problem (P2), in red, and FAIRTCIM-COVER problem (P6), in blue. We experimented with 4 groups and total influence includes all the groups but we show group influences for the two groups which had maximum disparity. The results show that our method achieves a more equal coverage for all the groups at the expense of only slightly larger seed sets. See Section 7.3 for further details.



(a) Budget Problem: Influence and Cost **(b)** Cover Problem: Influence **(c)** Cover Problem: Cost

Fig. 9. [Instagram-Activities Dataset] These figures demonstrate a comparison of TCIM-BUDGET versus FAIRTCIM-BUDGET and TCIM-COVER versus FAIRTCIM-COVER problems. The results show that our methods fare better compared to the traditional methods. Even though the fraction of influence seems small, since the graph comprises 0.5m nodes, the differences in fractions are significant in total numbers.

the disparity of influence among groups increases as the value of τ increase, refer to Section 6.2 for an intuitive explanation and, ii) our method is very effective in reducing disparity for different values of τ.

Takeaways. We demonstrated that: (i) FAIRTCIM-BUDGET, our proposed method, yields more fair solutions; (ii) this fairness is achieved at a very small reduction of the total influence compared to TCIM-BUDGET problem, as guaranteed by Theorem 1.

7.3 TCIM Under Coverage Constraint

Next, we compare TCIM-COVER problem (P2) and our solution to FAIRTCIM-BUDGET problem (P6). Red color in all the figures discussed in this section corresponds to the solution of TCIM-COVER problem (P2) and the blue color corresponds to our solution of FAIRTCIM-COVER problem (P6). We answer the following evaluation question using a *real-world* dataset.

- *Q1*: How does our method fare compared to the traditional method over the *iterations of the algorithm*?
- *Q2*: How effective is our method in reducing disparity for different *reach quotas*?
- *Q3*: How much cost does our method incur?

[Q1] Effect of Iterations. In Fig. 8a we compare iterations of problem (P2) and problem (P6), realized with the log function. In each iteration, one seed is selected. Green line represents the required quota of coverage. Dashed-dotted lines, dotted lines and solid lines represent group $\mathcal{V}_1$, group $\mathcal{V}_2$ and total population, respectively. Similar to the results on Synthetic dataset, i) our method consistently results in lower disparity between the two groups, which showed the highest disparity, throughout the iteration of the seed selection algorithm; ii) our method influences all the groups up to prescribed quota; iii) by using small number of additional seeds.

[Q2, Q3] Effect of Quota. Figs. 8b, 8c, 9a and 9c demonstrate similar results to the synthetic dataset described in Section 6.3. The *keypoint* is that all the groups are covered up to the required quotas with the solution set of FAIRTCIM-COVER problem by using only a small number of additional seeds.

Takeaways. We compared the TCIM-COVER and FAIRTCIM-COVER problems in this section using a real world dataset. The results demonstrate that our method is i) effective in reducing disparity ii) by using a small additional number of seeds.

8 CONCLUSION

In this paper, we considered the important problem of time-critical influence maximization (TCIM) under (i) budget constraint (TCIM-BUDGET) and (ii) coverage constraint (TCIM-COVER). We showed that the existing algorithmic techniques aimed at maximizing total influence in the population could lead to a huge disparity in utility across the underlying groups. This can put minority groups at a big disadvantage with far-reaching consequences.

To ensure that different groups are fairly treated, we proposed a notion of fairness and formulated two novel problems to solve TCIM under fairness considerations, namely, FAIRTCIM-BUDGET and FAIRTCIM-COVER. By introducing surrogate objective functions with submodular structural properties, we provided computationally efficient algorithms with desirable guarantees. Experiments over synthetic and real-world datasets demonstrated that our algorithms lead to low disparity in the time-critical influence propagation. This work opens up a variety of new research problems, including extensions to different notions of fairness, considering more complex models of time-criticality in information propagation (such as discounting with time), and developing new optimization methods for solving the fair TCIM problem formulations.

REFERENCES

[1] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2002, pp. 61–70.

[2] M. Ye, X. Liu, and W.-C. Lee, "Exploring social influence for recommendation: A generative model approach," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2012, pp. 671–680.

[3] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, "The diffusion of Microfinance," *Science*, vol. 341, no. 6144, 2013, Art. no. 1236498.

[4] A. Yadav, H. Chan, A. Xin Jiang, H. Xu, E. Rice, and M. Tambe, "Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2016, pp. 740–748.

[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2007, pp. 420–429.

[6] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2003, pp. 137–146.

[7] J. Wallinga and P. Teunis, "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures," *Amer. J. Epidemiol.*, vol. 160, no. 6, pp. 509–516, 2004.

[8] M. Gomez-Rodriguez , J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 4, 2010, Art. no. 21.

[9] A. Goyal, F. Bonchi, L. V. Lakshmanan, and S. Venkatasubramanian, "On minimizing budget and time in influence propagation over social networks," *Social Netw. Anal. Mining*, vol. 3, no. 2, pp. 179–192, 2013.

[10] T. Carnes, C. Nagarajan, S. M. Wild, and A. Van Zuylen , "Maximizing influence in a competitive social network: A follower's perspective," in *Proc. 9th Int. Conf. Electron. Commerce*, 2007, pp. 351–360.

[11] M. McPherson, L. Smith-Lovin , and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.

[12] A. Singla and I. Weber, "Camera brand congruence in the flickr social graph," in *Proc. 2nd ACM Int. Conf. Web Search Data Mining*, 2009, pp. 252–261.

[13] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 592–598.

[14] N. Kourtellis, T. Alahakoon, R. Simha, A. Iamnitchi, and R. Tripathi, "Identifying high betweenness centrality nodes in large social networks," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 899–914, 2013.

[15] M. Babaei, B. Mirzasoleiman, M. Jalili, and M. A. Safari, "Revenue maximization in social networks through discounting," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 1249–1262, 2013.

[16] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Proc. Int. Workshop Web Internet Econ.*, 2007, pp. 306–311.

[17] C. Budak, D. Agrawal, and A. El Abbadi , "Limiting the spread of misinformation in social networks," in*Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 665–674.

[18] K. Huang, S. Wang, G. Bevilacqua, X. Xiao, and L. V. Lakshmanan, "Revisiting the stop-and-stare algorithms for influence maximization," *Proc. VLDB Endowment*, vol. 10, no. 9, pp. 913–924, 2017.

[19] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proc. 3rd Innov. Theor. Comput. Sci. Conf.*, 2012, pp. 214–226.

[20] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian, "Certifying and removing disparate impact," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2015, pp. 259–268.

[21] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Proc. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3315–3323.

[22] V. Conitzer, R. Freeman, N. Shah, and J. W. Vaughan, "Group fairness for the allocation of indivisible goods," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 1853–1860.

[23] B. Fain, K. Munagala, and N. Shah, "Fair allocation of indivisible public goods," in *Proc. ACM Conf. Econ. Comput*, 2018, pp. 575–592.

[24] E. Segal-Halevi and W. Suksompong, "Democratic fair allocation of indivisible goods," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 482–488.

[25] W. Suksompong, "Approximate maximin shares for groups of agents," *Math. Social Sci.*, vol. 92, pp. 40–47, 2018.

[26] B. Fish, A. Bashardoust, Danah Boyd, S. A. Friedler, C. Scheidegger, and S. Venkatasubramanian, "Gaps in information access in social networks," in *Proc. World Wide Web Conf.*, 2019, pp. 480–490.

[27] R. Bredereck, P. Faliszewski, A. Igarashi, M. Lackner, and P. Skowron, "Multiwinner elections with diversity constraints," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 933–940.

[28] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon, "Multiwinner voting: A new challenge for social choice theory," *Trends Comput. Soc. Choice*, vol. 74, pp. 27–43, 2017.

[29] N. Benabbou, M. Chakraborty, X.-V. Ho, J. Sliwinski, and Y. Zick, "Diversity constraints in public housing allocation," in *Proc. 17th Int. Conf. Auton. Agents MultiAgent Syst.*, 2018, pp. 973–981.

[30] S. Aghaei, M. J. Azizi, and P. Vayanos, "Learning optimal and fair decision trees for non-discriminative decision-making," 2019, *arXiv: 1903.10598*.

[31] A. Rahmattalabi *et al.*, "Exploring algorithmic fairness in robust graph covering problems," in *Proc. Neural Inf. Process. Syst.*, 2019, pp. 15750–15761.

[32] M. Khajehnejad, A. A. Rezaei, M. Babaei, J. Hoffmann, M. Jalili, and A. Weller, "Adversarial graph embeddings for fair influence maximization over social networks," 2020, *arXiv:2005.04074*.

[33] A. Tsang, B. Wilder, E. Rice, M. Tambe, and Y. Zick, "Group-Fairness in influence maximization," 2019, *arXiv:1903.00967*.

[34] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proc. 22nd Nat. Conf. Artif. Intell.*, 2007, pp. 1650–1654.

[35] A. Singla, E. Horvitz, P. Kohli, R. White, and A. Krause, "Information gathering in networks via active exploration," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 891–988.

[36] A. Guillory and J. A. Bilmes, "Active semi-supervised learning using submodular functions," in *Proc. 27th Conf. Uncertainty Artif. Intell.*, 2011, pp. 274–282.

[37] A. Krause and D. Golovin, "Submodular function maximization," in *Tractability: Practical Approaches to Hard Problems*, Cambridge, U.K.: Cambridge Univ. Press, 2014, pp. 71–104.

[38] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of the approximations for maximizing submodular set functions," *Math. Prog.*, vol. 14, pp. 265–294, 1978.

[39] U. Feige, "A threshold of ln n for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, 1998.

[40] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 2010, pp. 251–260.

[41] A.-A. Stoica, C. Riederer, and A. Chaintreau, "Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity," in *Proc. World Wide Web Conf.*, 2018, pp. 923–932.

[42] A.-A. Stoica and A. Chaintreau, "Fairness in social influence maximization," in *Proc. World Wide Web Conf.*, 2019, pp. 569–574.