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

Junaid Ali[†] Mahmoudreza Babaei[†] Abhijnan Chakraborty[†]

Baharan Mirzasoleiman[‡] Krishna P. Gummadi[†] 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 information propagation of valuable information such as job vacancy advertisements and health-related information. 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 of these applications, the spread of influence is time-critical, i.e., it is only beneficial to be influenced before a time deadline. As we show in this paper, the time-criticality of the information could further exacerbate the disparity of influence across groups. This disparity, introduced by algorithms aimed at maximizing total influence, 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.

1 Introduction

The problem of *Influence Maximization* has been widely studied due to its application in multiple domains such as viral marketing [31], social recommendations [38], propagation of information related to jobs, financial opportunities or public health programs [3, 37]. Over the years, extensive research efforts have focused on the cascading behavior, diffusion and spreading of ideas, or containment of diseases, [25]; [21]; [31]; [36]; and [16]. 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) [26]; [22]; [21]; [17]; [2]; [5]; [7]; [8].

Real-world social networks, however, are often not homogeneous and comprise different groups of people. Due to the disparity in group sizes, the potentially high propensity towards creating within-group links [28], and differences in dynamics of influences among different groups [32], the structure of the social network can cause a disparity 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 [10]. 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 doesn't 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 than the minority group, and they are more central in the network. 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 present in the network. This disparity in time-critical applications, in turn, can bring 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 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.

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 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.

We introduce two formulations of TCIM problems under fairness considerations, namely FAIRTCIM-BUDGET and FAIRTCIM-COVER. However, directly optimizing FAIRTCIM-BUDGET problem and FAIRTCIM-COVER problem are computationally challenging, and comes without any structural properties. As our second contribution, we propose *monotone submodular* surrogates for solving both of these problems that capture the tradeoff between maximizing total influence and minimizing unfairness (i.e., disparity across groups). Although the surrogate problems are still NP-Hard, we propose a greedy approximation with provable guarantees.

We evaluate our proposed solutions over a real-world social network and show that they are successful in enforcing the aforementioned fairness notion. As expected, enforcing this fairness comes at the cost of a reduction in performance, i.e., lowering the total influence, in case of FAIRTCIM-BUDGET problem, and solutions with larger seed set sizes, in case of FAIRTCIM-COVER problem. 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.

Contemporary Works. Fairness has mostly been studied in algorithmic decision-making system, e.g., [34]; [9]; [11]; [19]; [14]; and [20], however, only a few contemporary works have investigated fairness in influence maximization, as described next.

Very recently, [15] 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 [6, 12, 4, 1]. [39] proposed a greedy heuristic to enforce group fairness where multiple sources of influence are in competition.

In a concurrent work, [35] propose a method to achieve group fairness in influence maximization. However, their work is 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.



	solution to TCIM-BUDGET					solution to FAIRTCIM-BUDGET			
	S	$\frac{f(S;V,G)}{ V }$	$\frac{f(S; \mathcal{V}_1, \mathcal{G})}{ \mathcal{V}_1 }$	$\frac{f(S; V_2, G)}{ V_2 }$		S	$\frac{f(S; \mathcal{V}, \mathcal{G})}{ \mathcal{V} }$	$\frac{f(S; \mathcal{V}_1, \mathcal{G})}{ \mathcal{V}_1 }$	$\frac{f(S; V_2, G)}{ V_2 }$
$\tau = \infty$	$\{a,b\}$	0.38	0.48	0.16		$\{a,c\}$	0.31	0.33	0.27
$ au = \infty$ $ au = 4$ $ au = 2$	$\{a,b\}$	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

Figure 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 and an optimal solution to our formulation of TCIM-BUDGET with fairness considerations. 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. Solution to FAIRTCIM-BUDGET problem achieves high utility and low disparity for different deadlines τ .

2 Disparity in TCIM and Introducing Notions of Fairness

In this section, we first introduce the utility funciton of time-critical influence function, then, we highlight the disparity in utility across population resulting from the solution to the standard TCIM problem formulations, and introduce a notion of fairness in TCIM.

2.1 Utility of Time-Critical Influence

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:

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

where the expectation is w.r.t. the activation probabilities associated to each edge under independant cascade 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 .

2.2 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. 3.

2.3 Disparity in Utility Across Groups

In the standard formulations for TCIM problem, i.e., TCIM-BUDGET problem and TCIM-COVER problem, 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 standard TCIM-BUDGET problem and TCIM-COVER problem. Another factor is related to the connectivity and centrality of nodes from different groups. The solution to these standard optimization problems 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 Figure 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, and it is easy to extend this example to show disparity in TCIM-COVER problem. The graph that we consider in this example (see Figure 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 TCIM-BUDGET problem 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 Figure 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.

2.4 Notion of Fairness

Next, in order to guide the design of fair solutions to TCIM problems, we introduce a formal notion of group fairness 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:

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

As discussed above (see Section 2.3), 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.

3 Solutions and Results

Due to lack of space, we only provide a sketch of our proposed solution. We solve the TCIM-BUDGET problem and the TCIM-COVER problem with additionally considering the fairness notion introduced in section 2.4, which we call FAIRTCIM-BUDGET problem and FAIRTCIM-COVER problem, respectively. However, adding the fairness measure yields objectives which remain NP-Hard as the original problem, but, additionally, they are no longer submodular—a property used to approximately solve the traditional problem with guarantees using the greedy heuristic. We propose submodular proxies for both FAIRTCIM-BUDGET and FAIRTCIM-COVER problems. In the case of FAIRTCIM-BUDGET, we propose to maximize sum of concave functions whose arguments are group influences. The resultant function encourages selecting seeds which influence underrepresented groups. The penalty of disparity in the group influences would be dependent on the curvature of the concave function. For FAIRTCIM-COVER problem, we propose that every group should reach the required quota of influence instead of just the entire population. This way every group is ensured a minimum reach. Since our proposed proxies are submodular and increasing functions of the seed set, we can approximate the solutions of our proposed proxy objectives using a greedy heuristic. We, additionally, provide bounds: (i) on the minimum expected influence achieved by our solution, in the case of FAIRTCIM-BUDGET problem, and (ii) on the maximum length of the solution set size, in the case of FAIRTCIM-COVER problem. We evaluate our proposed solutions using a real and a synthetic dataset and show that: (i) our methods are effective in enforcing the proposed fairness notion, and (ii) the cost of achieving this fairness is bounded as guaranteed by our theoretical results. We also 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. The proposed solutions, the theoretical guarantees and the corresponding empirical evaluations can be found in detail in the supplementary material.

References

- [1] Sina Aghaei, Mohammad Javad Azizi, and Phebe Vayanos. Learning optimal and fair decision trees for non-discriminative decision-making. *arXiv preprint arXiv:1903.10598*, 2019.
- [2] Mahmoudreza Babaei, Baharan Mirzasoleiman, Mahdi Jalili, and Mohammad Ali Safari. Revenue maximization in social networks through discounting. *Social Network Analysis and Mining*, 3(4):1249–1262, 2013.
- [3] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013.
- [4] Nawal Benabbou, Mithun Chakraborty, Xuan-Vinh Ho, Jakub Sliwinski, and Yair Zick. Diversity constraints in public housing allocation. In *AAMAS*, pages 973–981, 2018.
- [5] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *International workshop on web and internet economics*, pages 306–311. Springer, 2007.
- [6] Robert Bredereck, Piotr Faliszewski, Ayumi Igarashi, Martin Lackner, and Piotr Skowron. Multiwinner elections with diversity constraints. In *AAAI*, 2018.
- [7] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, 2011.
- [8] Tim Carnes, Chandrashekhar Nagarajan, Stefan M Wild, and Anke Van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In EC, pages 351–360. ACM, 2007.
- [9] Abhijnan Chakraborty, Gourab K Patro, Niloy Ganguly, Krishna P Gummadi, and Patrick Loiseau. Equality of voice: Towards fair representation in crowdsourced top-k recommendations. In *FAT**, pages 129–138. ACM, 2019.
- [10] Wei Chen, Wei Lu, and Ning Zhang. Time-critical influence maximization in social networks with time-delayed diffusion process. In *AAAI*, 2012.
- [11] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *ITCS*, pages 214–226. ACM, 2012.
- [12] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice*, 74, 2017.
- [13] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45:314–318, 1998.
- [14] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In KDD, pages 259–268. ACM, 2015.
- [15] Benjamin Fish, Ashkan Bashardoust, danah boyd, Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. Gaps in Information Access in Social Networks. In *WWW*, 2019.
- [16] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring Networks of Diffusion and Influence. In KDD, 2010.
- [17] Amit Goyal, Francesco Bonchi, Laks VS Lakshmanan, and Suresh Venkatasubramanian. On minimizing budget and time in influence propagation over social networks. *Social network* analysis and mining, 3(2):179–192, 2013.
- [18] Andrew Guillory and Jeff A. Bilmes. Active semi-supervised learning using submodular functions. In *UAI*, pages 274–282, 2011.
- [19] Sara Hajian, Francesco Bonchi, and Carlos Castillo. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *KDD*, pages 2125–2126. ACM, 2016.

- [20] Moritz Hardt, Eric Price, Nati Srebro, et al. Equality of opportunity in supervised learning. In NIPS, pages 3315–3323, 2016.
- [21] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.
- [22] Nicolas Kourtellis, Tharaka Alahakoon, Ramanuja Simha, Adriana Iamnitchi, and Rahul Tripathi. Identifying high betweenness centrality nodes in large social networks. *Social Network Analysis and Mining*, 3(4):899–914, 2013.
- [23] Andreas Krause and Daniel Golovin. Submodular function maximization., 2014.
- [24] Andreas Krause and Carlos Guestrin. Near-optimal observation selection using submodular functions. In *AAAI*, pages 1650–1654, 2007.
- [25] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In KDD, pages 420–429. ACM, 2007.
- [26] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence maximization on social graphs: A survey. *TKDE*, 30(10):1852–1872, 2018.
- [27] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *Proc. ACL: HLT Vol. 1*, pages 510–520, 2011.
- [28] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.
- [29] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are who you know: inferring user profiles in online social networks. In WSDM, pages 251–260. ACM, 2010.
- [30] G.L. Nemhauser, L.A. Wolsey, and M. Fisher. An analysis of the approximations for maximizing submodular set functions. *Math. Prog.*, 14:265–294, 1978.
- [31] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In KDD, 2002.
- [32] Adish Singla and Ingmar Weber. Camera brand congruence in the flickr social graph. In *WSDM*, pages 252–261, 2009.
- [33] Adish Singla, Eric Horvitz, Pushmeet Kohli, Ryen White, and Andreas Krause. Information gathering in networks via active exploration. In *IJCAI*, pages 891–988, 2015.
- [34] Till Speicher, Hoda Heidari, Nina Grgic-Hlaca, Krishna P Gummadi, Adish Singla, Adrian Weller, and Muhammad Bilal Zafar. A unified approach to quantifying algorithmic unfairness: Measuring individual &group unfairness via inequality indices. In *KDD*, pages 2239–2248. ACM, 2018.
- [35] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. Group-Fairness in Influence Maximization. *arXiv preprint arXiv:1903.00967*, 2019.
- [36] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160 (6):509–516, 2004.
- [37] Amulya Yadav, Hau Chan, Albert Xin Jiang, Haifeng Xu, Eric Rice, and Milind Tambe. Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In *AAMAS*, pages 740–748, 2016.
- [38] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation: a generative model approach. In *SIGIR*, pages 671–680. ACM, 2012.
- [39] Ying Yu, Jinglan Jia, Deying Li, and Yuqing Zhu. Fair multi-influence maximization in competitive social networks. In *International Conference on Wireless Algorithms, Systems, and Applications*, pages 253–265. Springer, 2017.

A 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.

A.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 [21]: (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 $t_v = 0$ 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.

A.2 Utility of Time-Critical Influence

As motivated in the introduction, 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 [10]. 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:

$$f_{\tau}(S; Y, \mathcal{G}) = \mathbb{E}\left[\sum_{v \in Y, t_v \ge 0} \mathbb{I}(t_v \le \tau)\right],$$
 (3)

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 .

A.3 TCIM as Discrete Optimization Problem

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

A.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

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

A.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

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

A.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 optmization problems. In their seminal work, [21] 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 [21], sensing [24], information gathering [33], and active learning [18] (see [23] for a survey on submodular function optimization and its applications).

Chen et al. [10] 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 $a \in \mathcal{V} \setminus A'$, it holds that (omitting the parameters \mathcal{V} and \mathcal{G} for brevity):

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

The fact that the utility function is submodular in turn implies that these two problems P1 and P2 are NP-Hard and hence finding the optimization solution is intractable [30, 13, 23]. However, on a positive note, one can exploit the submodularity property of the function to design efficient approximation algorithms with provable guarantees [30, 23]. 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 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.

B 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.

B.1 Fair Maximization under Budget Constraint (FAIRTCIM-BUDGET)

B.1.1 Introducing 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:

$$\max_{S \subseteq \mathcal{V}} \left(f_{\tau}(S; \mathcal{V}, \mathcal{G}) - \gamma \cdot \max_{i,j} \left| \frac{f_{\tau}(S; \mathcal{V}_{i}, \mathcal{G})}{|\mathcal{V}_{i}|} - \frac{f_{\tau}(S; \mathcal{V}_{j}, \mathcal{G})}{|\mathcal{V}_{j}|} \right| \right)$$
subject to $|S| < B$, (P3)

where $\gamma \geq 0$ is a fixed regularization factor to trade-off the total influence and disparity of the solution.

B.1.2 Surrogate problem for FAIRTCIM-BUDGET with guarantees

We note that problem P3 is a challenging discrete optimization problem and the objective function does not have structural properties of submodularity as was the case for the standard TCIM-BUDGET problem P1. 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:

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

where \mathcal{H} is a non-negative, monotone concave function and $\lambda_i \geq 0$ are fixed scalars.

The key idea of using the surrogate objective function in problem P4 is the following: 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. It is important to note that controlling the curvature of the concave function $\mathcal H$ provides an indirect way to trade-off the total influence and 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 (this is demonstrated in the experimental results in Figure 3a). For our illustrative example from Section 2, we report the results for an optimal solution to FAIRTCIM-BUDGET problem P4 with $\mathcal H(z) := \log(z)$. As can be seen in Figure 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.

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 A.4).

Theorem 1 Let \hat{S} denote the output of the greedy algorithm for problem P4 with $\lambda_i = 1 \ \forall i \in [k]$. 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: $f_{\tau}(\hat{S}; \mathcal{V}, \mathcal{G}) \geq (1 - \frac{1}{e}) \cdot \mathcal{H}(f_{\tau}(S^*; \mathcal{V}, \mathcal{G}))$.

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 $\left((1-\frac{1}{e})\cdot\frac{\mathcal{H}(f_{\tau}(S^*;\mathcal{V},\mathcal{G}))}{f_{\tau}(S^*;\mathcal{V},\mathcal{G})}\right)$. 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. We provide sketches of proofs below.

Proof [Proof sketch] There are two key steps in proving the above theorem. The first step is showing that the objective function in problem P4 is monotone submodular function, where we use a result

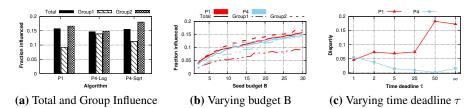


Figure 2: Comparison of results solving TCIM-BUDGET problem P1 and FAIRTCIM-BUDGET P4. The results demonstrate that our method, given by problem P4, yields seed set which propagate influence in a more fair manner, at the cost of a marginally lower total influence.

from [27] that composition of a non-decreasing concave and a non-decreasing submodular function is submodular. The second step is to bound the utility of the optimal solution of problem P4 w.r.t. the optimal solution of problem P1. We prove that this can be bounded by a multiplicative factor of $\frac{\mathcal{H}(f_{\tau}(S^*; \mathcal{V}, \mathcal{G}))}{f_{\tau}(S^*; \mathcal{V}, \mathcal{G})}$. Combining bounds from individual steps and by applying inequalities resulting from the concavity of \mathcal{H} gives us the final result.

B.2 Fair Minimization under Coverage Constraint (FAIRTCIM-COVER)

B.2.1 Introducing 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:

$$\min_{S \subseteq \mathcal{V}} \left(|S| + \gamma \cdot \max_{i,j} \left| \frac{f_{\tau}(S; \mathcal{V}_{i}, \mathcal{G})}{|\mathcal{V}_{i}|} - \frac{f_{\tau}(S; \mathcal{V}_{j}, \mathcal{G})}{|\mathcal{V}_{j}|} \right| \right)$$
subject to
$$\frac{f_{\tau}(S; \mathcal{V}, \mathcal{G})}{|\mathcal{V}|} \ge Q,$$
(P5)

where $\gamma \geq 0$ is a fixed regularization factor to trade-off the size of seed set and disparity of the solution.

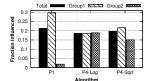
B.2.2 Surrogate problem for FAIRTCIM-COVER with guarantees

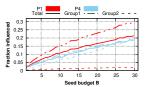
As in Section B.1, we note that problem P5 is a challenging discrete optimization problem and does not have structural properties as was the case for the standard TCIM-COVER problem P2. Instead of directly solving problem P5, we introduce a novel surrogate problem that indirectly trade-offs the two objectives of minimizing the size of selected seed set and minimizing disparity, as follows:

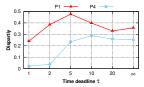
$$\min_{S \subseteq \mathcal{V}} |S| \quad \text{ subject to } \qquad \qquad \sum_{i=1}^k \min \left\{ \frac{f_\tau(S; \mathcal{V}_i, \mathcal{G})}{|\mathcal{V}_i|}, Q \right\} \geq k \cdot Q. \tag{P6}$$

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).

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

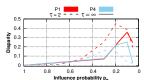


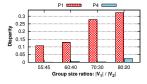


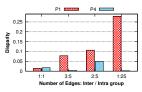


- (a) Total and group influence
- (b) Varying budget B
- (c) Varying time deadline τ

Figure 3: [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 C.2 for further details.







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

Figure 4: [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 C.2.

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 A.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 V_i , i.e., solving problem P2 with constraint given 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: $|\hat{S}| \leq \ln(1+|\mathcal{V}|) \left(\sum_{i=1}^k |S_i^*|\right)$.

Proof [Proof sketch] There are two key steps in proving the above theorem. The first step is showing that the objective function in the constraint of problem P6 is monotone submodular function by using several composition properties of submodular functions [23]. The second step is to provide an upper bound on the optimal solution of problem P6 as $\left(\sum_{i=1}^{k} |S_i^*|\right)$.

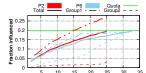
C Evaluation on Synthetic Dataset

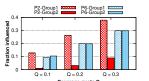
In this section, we compare the solutions of different problems on a synthetic dataset. We show the effects of different properties of the synthetic graph and parameters of the algorithms.

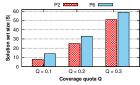
C.1 Dataset and Experimental Setup

First, we discuss the synthetic dataset generation process and the setup used in our experiments.

Synthetic dataset. We consider an undirected graph with 500 nodes, where each node belongs to either group V_1 or group 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 V_1). Nodes are connected based on two probabilities: (i) within-group edge probability (Homophily) P_{hom} and (ii) across-group edge probability (Homophily) P_{hom} . 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







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

Figure 5: [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 C.3 for further details.

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 used 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. 3, 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: $\frac{f(S;\mathcal{V}_1,\mathcal{G})}{|\mathcal{V}|}$ 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 .

C.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 greedily picking B seeds which maximize the objective functions. In all the figures discussed in this section, red color represents the results of TCIM-BUDGET problem P1, 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.

Effect of different $\mathcal{H}(z)$. Figure 3a presents the comparison of three algorithms: one solving TCIM-BUDGET problem P1, using the greedy heuristic; the other two solving FAIRTCIM-BUDGET problem P4, where we use two realizations of the concave monotone function, $\mathcal{H}(z)$, given by: (i) $\mathcal{H}(z) := \log(z)$ and (ii) $\mathcal{H}(z) := \sqrt{z}$. Figure 3a 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} performs worse than $\log(z)$ in removing the disparity, however incurring lower loss in total influence, which is expected as it has lower curvature than $\log(z)$. 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 point* to notice is that 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. **Effect of seed budget.** Figure 3b 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. We can see that

problem P4, with a small reduction in total influence, reduces a lot of disparity between the groups. We also observe that higher budget could lead to more disparity in an imbalanced graph.

Effect of deadline. Figure 3c compares disparity as we vary the value of the deadline τ , in problems P1 and P4. Disparity is computed as the absolute difference between the fraction of individuals influenced in each group, given by Eq. 2. The figure shows that the disparity depends on the value of τ and, in this case, our method is more effective for lower values of τ .

Effect of activation probabilities. Figure 4a 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 lower values of τ and lower activation probabilities could result in larger disparity and our method consistently performs better.

Effect of group sizes. Figure 4b 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 different groups and y-axis represents disparity. The figure confirms our hypothesis that *imbalance in a graph could lead to disparate influence*, as motivated in the illustrative example given in Figure 1.

Effect of graph structure. Figure 4c demonstrates the importance of the graph structure $(p_{het}, p_{hom}) \in \{(0.025, 0.025), (0.015, 0.025), (0.01, 0.025), (0.001, 0.025)\}$. x-axis shows the ratio of across and within group edge probabilities. The figure validates our hypothesis that the majority group containing more influential nodes fares better in TCIM-BUDGET problem, as proposed in Figure 1.

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 limit, 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.

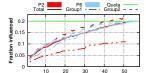
C.3 TCIM under Coverage Constraints

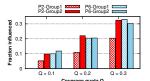
Next, we compare solutions of TCIM-COVER problem P2, and our solution to FAIRTCIM-COVER problem P6. The goal is to reach the prescribed quota Q, while trying to minimize the number of seeds. In all the figures discussed in this section, red color represents the results of TCIM-COVER problem P2, and blue color represents the results of our solution to FAIRTCIM-COVER problem P6.

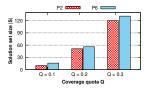
Effect of iterations. Figure 5a shows how the fraction of population influenced changes with seed selection at each iteration. Solid lines represent total influence while dash-dotted lines and dotted lines represent groups \mathcal{V}_1 and \mathcal{V}_2 , respectively. In this experiment, Q was set to 0.2 which is represented by the horizontal green line. The figure demonstrates that in FAIRTCIM-COVER, the fraction of influenced individuals are more than Q in both groups. On the other hand, in the TCIM-COVER, the fraction of influenced individuals in group \mathcal{V}_1 is larger than Q; however, the fraction of influenced individuals in group \mathcal{V}_2 is much smaller than Q. The figure also shows that this fairness is achieved with only marginal increase in the solution set sizes, as guaranteed in Theorem 2.

Effect of quota Q. Figure 5b shows fractions of individuals that are influenced for different quota Q: (i) for the TCIM-COVER problem, the disparity in the fraction of influenced individuals in different groups persists with increased quota; (ii) however, our proposed method results in both groups reaching the specified coverage Q. Figure 5c shows the coverage Q on the x-axis and number of seeds used on the y-axis. The figure shows that our method results in only slightly larger seed sets compared to the TCIM-COVER problem, as guaranteed by Theorem 2.

Takeaways. We compared the results of TCIM-COVER problem, P2, and our solution to FAIRTCIM-COVER problem, P6. The results demonstrate that: (i) both methods influence Q fraction of the total population; (ii) however, the TCIM-COVER approach can result in disparate coverage, i.e., the coverage in some groups could be much lower than the prescribed fraction Q; (iii) FAIRTCIM-COVER, on the other hand, ensures coverage of at least Q fraction in both groups; and (iv) additionally, FAIRTCIM-COVER only yields slightly larger solution set sizes, as guaranteed by Theorem 2.







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

Figure 6: The figure demonstrates the results of TCIM-COVER problem P2, in red, and FAIRTCIM-COVER problem P6, in blue. The results show that our method achieves a more equal coverage for all the groups at the expense of only slightly larger seed sets.

Evaluation on Real Dataset

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

D.1 Dataset and Experimental Setup

Next, we describe the dataset we used to evaluate our proposed methods, followed by the experimental

Rice-Facebook dataset. To evaluate our proposed methods, we used Rice-Facebook dataset collected by Mislove et al. [29], where they 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 two 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 V_1 and age 20 as group V_2 . Group V_1 has 97 nodes and 513 within-group edges. Whereas, group 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 .

Experimental Setup. In all the experiments in this section, we used activation probability $p_e = 0.01$. All the other parameter were the same as described in Section C.1.

D.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.

Effect of different $\mathcal{H}(z)$. In Figure 2a, 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) $\mathcal{H}(z) := \sqrt{z}$. The figure demonstrates that: (i) At a marginal reduction of total influence, as guaranteed by Theorem 1, our proposed method reduces disparity in influence; (ii) as hypothesized in Section B.1, a higher curvature function, $\mathcal{H}(z) := \log(z)$, leads to a bigger reduction in disparity compared to $\mathcal{H}(z) := \sqrt{z}$. Effect of seed budget. Figure 2b demonstrates the effect of allowed seed budget on the group and total influences. Groups V_1 and V_2 are represented by dash-dotted lines and dotted lines respectively and solid lines correspond to total influence. The figure demonstrates that: (i) Disparity in the utility between both the groups increases with increase in allowed seed budget. A reason for these differences could be the imbalances in groups sizes and average degrees, between both the groups. 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 allow 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.

Effect of time deadline. Figure 2c shows the effect of different time deadlines on the disparity between group influences, as calculated by Eq. 2. It demonstrates that: (i) our proposed method, given by problem P4, yields solutions which result in lower disparity; (ii) in this case disparity in group utilities increases as the time deadline is increased. One explanation could be that the seed nodes are propagating influence in both the groups, but as we increase the time deadline Group V_2 , with more nodes and edges, is more efficient at propagating influence compared to Group V_1 , so it results in a larger disparity. One could imagine a case, as shown in the motivating example in Figure 1, where seed nodes are surrounded by nodes of only one group, in this case a longer time deadline could yield a lower disparity.

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.

D.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.

Effect of iterations. In Figure 6a 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. 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.

Effect of quota. Figure 6b demonstrates that: (i) for different values of the required quota, problem P2 results in disparate utility between both the groups; (ii) while seeds selected by solving problem P6 result in a more equal utility; (iii) FAIRTCIM-COVER problem P6 uses only a small number of additional seeds, as shown by Figure 6c.

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 (iii) lastly, FAIRTCIM-COVER yields only slightly larger solution sets as guaranteed by Theorem 2.