On the Fairness of Time-Critical Influence Maximization in Social Networks (Extended Abstract)

Junaid Ali†, Mahmoudreza Babaei‡, Abhijnan Chakraborty*, Baharan Mirzasoleiman**, Krishna P. Gummadi†, Adish Singla†
†Max Planck Institute for Software Systems,
‡Max Planck Institute for Human Development,
*Indian Institute of Technology Delhi, **UCLA

2022 IEEE 38th International Conference on Data Engineering (ICDE) | 978-1-6654-0883-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICDE53745.2022.00147

***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.**

I. Introduction

The problem of *Influence Maximization* has been widely studied due to its application in multiple domains such as viral marketing, social recommendations, propagation of information related to jobs, financial opportunities or public health programs. 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) [1]

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 , and differences in dynamics of influences among different groups, the structure of the social network can cause disparities in the influence maximization process.

Moreover, some applications are also *time-critical* in nature [2]. 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. 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. 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.

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.

II. Background

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges connecting these nodes. 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 $\mathcal{V} = \cup_i \mathcal{V}_i$. In this paper, we assume the independent cascade model, i.e., each edge is associated with a probability with which it can influence its neighbor. Our results also easily extend to the other diffusion models, e.g., linear threshold model.

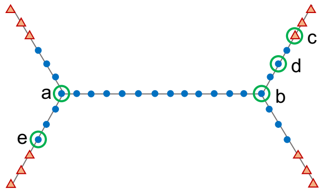
We adopt the well-studied notion of time-critical influence as proposed by [2]. Their time-critical model is captured via a deadline τ: If a node is activated before the deadline, it receives a utility of 1, otherwise it receives no utility.

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

Given a budget B one can define TCIM-BUDGET:

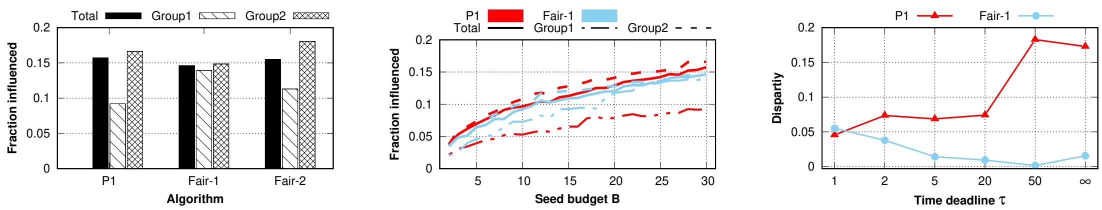
$$\max_{S \subseteq \mathcal{V}} f_\tau(S; \mathcal{V}, \mathcal{G}) \quad \text{subject to } |S| \leq B. \tag{P1}$$

2375-026X/22/$31.00 ©2022 IEEE
DOI 10.1109/ICDE53745.2022.00147



	solution to TCIM-BUDGET				solution to FAIRTCIM-BUDGET			
	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 proposal FAIRTCIM-BUDGET. 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.



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

Fig. 2: [Rice-Facebook Dataset: Budget Problem] Comparison of results solving TCIM-BUDGET problem (P1) and FAIRTCIM-BUDGET (Fair-1 and Fair-2, two versions). 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 yields seed set which propagate influence in a more fair manner, at the cost of a marginally lower total influence. See the full paper for further details.

Given the quota Q one can define TCIM-COVER:

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

III. TECHNICAL CONTRIBUTIONS

In this section, we provide a brief overview of the technical contributions of our work.

Notion of disparity. 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:

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

Achieving Fairness in TCIM. In this section, we provide an overview of our proposals to design fair versions of both the TCIM problems P1 and P2.

In particular, we have **two objectives**: i) maximize total influence for the whole population $\mathcal{V}$ as was done in the standard TCIM-BUDGET problem P1 or find the smallest seed set that can influence a prescribed fraction of the population as was done in the standard TCIM-COVER problem P2, and (ii) enforce 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 or larger seed set sizes, and we seek to design algorithms that can achieve a good trade-off between these two objectives. We call the TCIM-BUDGET problem under the fairness constraint FAIRTCIM-BUDGET and TCIM-COVER under the fairness constraint FAIRTCIM-COVER problem.

Unfortunately, solving FAIRTCIM-BUDGET and FAIRTCIM-COVER is intractable. Additionally, both the formulations are also not submodular unlike their unfair counterparts. This means that we cannot get the approximate solutions using the greedy algorithm. In order to solve this problem, we provide surrogate submodular functions which capture the aforementioned *two objectives*, for both FAIRTCIM-BUDGET and FAIRTCIM-COVER. Since, our surrogate functions are submodular we can use the greedy algorithm to find the approximate solutions. Additionally, we provide bounds on the reduction of influence, for FAIRTCIM-BUDGET problem, and bounds on the increase in the seed set size for FAIRTCIM-COVER problem.

IV. EXPERIMENTAL CONTRIBUTIONS

We evaluate our proposed methods for FAIRTCIM-BUDGET and FAIRTCIM-COVER problems on 3 real world and several synthetic datasets. 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.

REFERENCES

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *KDD*, 2003.

[2] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," in *AAAI*, 2012.